Biomedical Informatics

Computer Applications in Health Care and Biomedicine

Edward H. Shortliffe
James J. Cimino
Editors
Michael F. Chiang
Co-Editor

Fifth Edition

Springer

Biomedical Informatics

Edward H. Shortliffe • James J. Cimino
Editors
Michael F. Chiang
Co-Editor

Biomedical Informatics

Computer Applications in Health Care and Biomedicine

5th Edition

Springer

Editors
Edward H. Shortliffe
Biomedical Informatics
Columbia University
New York, NY, USA

James J. Cimino
Informatics Institute
University of Alabama at Birmingham
Birmingham, AL, USA

Co-Editor
Michael F. Chiang
National Eye Institute
National Institutes of Health
Bethesda, MD, USA

ISBN 978-3-030-58720-8 ISBN 978-3-030-58721-5 (eBook)
https://doi.org/10.1007/978-3-030-58721-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This volume is dedicated to AMIA, the principal professional association for the editors. Born as the American Medical Informatics Association in 1990, AMIA is now preferentially known simply by its acronym and has grown to include some 5500 members who are dedicated to all aspects of biomedical informatics. AMIA and this textbook have evolved in parallel for four decades, and we thank the organization and its members for all they have done for the field and for health care and biomedicine. May both AMIA and this volume evolve and prosper in parallel for years to come.

Foreword

Health and biomedicine are in the midst of revolutionary change. Health care, mental health, and public health are converging as discovery science reveals these traditional "silos" share biologic pathways and collaborative management demonstrates better outcomes. Health care reimbursement is increasingly framed in terms of paying for outcomes achieved through value-based purchasing and population health management. Individuals are more engaged in their health and wellness decisions, using personal biomedical monitoring devices and testing services and engaging in citizen science. Systems biology is revealing the complex interactions among a person's genome, microbiome, immune system, neurologic system, social factors, and environment. Novel biomarkers and therapeutics exploit these interactions.

These advances are fueled by digitization and generation of data at an unprecedented scale. The volume of health care data has multiplied 8 times since 2013 and is projected to grow at a compound annual rate of 36% between 2018 and 2025[1]. The rate of growth of biomedical research data is comparable[2]. When you consider recent estimates that socioeconomics, health behaviors, and environment—factors outside of the domain of health care and biomedicine—contribute as much as 80% to health outcomes[3], the variety and scale of health-related data are breathtaking.

Biomedical informatics provides the scientific basis for making sense of these data—methods and tools to structure, mine, visualize, and reason with data and information. Biomedical informatics also provides the scientific basis for incorporating data and information into effective workflows—techniques to link people, process, and technology into systems; methods to evaluate systems and technology components; and methods to facilitate system-level change.

Biomedical informatics grew out of efforts to understand biomedical reasoning[4], such as artificial intelligence; to develop medical systems, such as multiphasic screening[5]; and to write computer programs to solve clinical problems, such as diagnosis and treatment of acid-base disorders[6]. By the late 1970s, "medical informatics" was used interchangeably with "computer applications in medical care". As computer programs were written for various allied health disciplines, nursing informatics, dental informatics, and public health informatics emerged. The 1980s saw the emergence of computational biology for applications such as scientific visualization and bioinformatics to support tasks such as DNA sequence analysis.

Biomedical Informatics: Computer Applications in Health Care and Biomedicine provided the first comprehensive guide to the field with its first edition in 1990. That edition and the subsequent three have served as the core syllabus for introductory courses in informatics and as a reference source for those seeking advanced training or working in the field. The fifth edition carries on the tradition with new topics, comprehensive glossary, reading lists, and citations.

I encourage people who are considering formal education in biomedical informatics to use this book to sample the field. The book's framework provides a guide for educators from junior high to graduate school as they design introductory courses in biomedical informatics. It is the basic text for students entering the field.

With digitization and data driving change across the health and biomedicine ecosystem, everyone in the ecosystem will benefit from reading *Biomedical Informatics* and using it as a handbook to guide their work. The following is a sample of questions readers can turn to the book to explore:

- **Practicing health professionals**—How do I recognize an information need? How do I quickly scan and filter information to answer a question? How do I sense the fitness of the information to answer my question? How do I configure my electronic health record to focus my attention and save time? How do I recognize when to override decision support? How do I analyze data from my practice to identify learning and improvement opportunities? How do I engage with patients outside of face-to-face encounters?
- **Quality improvement teams**—How might we detect if the outcome we are trying to improve is changing in the desired direction? Are data available in our operational systems that are fit for that purpose? What combination of pattern detection algorithm, workflow process, decision support, and training might work together to change the outcome? How can we adapt operational processes and systems to test the change and to scale if it proves effective?
- **Discovery science teams**—How do data about biological systems differ from data about physical systems? How do we decide when to use integrative analytic approaches and when to use reductionist approaches? How much context do we need to keep about data we create and how do we structure the metadata? How do we optimize compute and storage platforms? How might we leverage electronic health record-derived phenotype to generate hypotheses?
- **Artificial intelligence researchers or health "app" developers**—What health outcome am I trying to change? Do I need a detection, prediction, or classification algorithm? What sources of data might be fit for that purpose? What type of intervention might change the outcome? Who would be the best target for the intervention? What is the best place in their workflow to incorporate the intervention?
- **Health system leaders**—How do we restructure team roles and electronic health record workflows to reduce clinician burnout and improve care quality? How do we take advantage of technology-enabled self-management and virtual visits to increase adherence and close gaps in care? How do we continuously evaluate evidence and implement or de-implement guidelines and decision support across our system? How do we leverage technology to deploy context-sensitive just-in-time learning across our system?
- **Health policy makers**—How might we enhance health information privacy and security and reduce barriers to using data for population

health, health care quality improvement, and discovery? To what degree is de-identification a safeguard? What combination of legislative mandate, executive action, and industry-driven innovation will accelerate health data interoperability and business agility? How might federal and state governments enable communities to access small area data to inform their collective action to improve community health and well-being?

You have taken the first step in exploring these frontiers by picking up this book. Enjoy!

References

1. IDC White Paper, The Digitization of the World from Edge to Core, #US44413318 November 2018 ► https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf
2. Vamathevan, J. V., Apweiler, R., & Birney, W. (2019). Biomedical data resources: Bioinformatics infrastructure for biomedical data science. *Annual Review of Biomedical Data Science*, *2*:199–222.
3. County Health Rankings Program and Roadmap ► https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model
4. Ledley, R., & Lusted, L. (1959). Reasoning foundations of medical diagnosis. *Science*, *130*, 9–21.
5. Collen, M. F. (1966). Periodic health examinations using an automated multitest laboratory. *JAMA*, *195*, 830–3.
6. Bleich, H. L. (1971). The computer as a consultant. *New England Journal of Medicine*, *284*, 141–7.

William W. Stead, MD, FACMI
Chief Strategy Officer, Vanderbilt University Medical Center, McKesson Foundation Professor, Department of Biomedical Informatics, Professor, Department of Medicine, Vanderbilt University
Nashville, TN, USA
September 2019

Preface to the Fifth Edition

The world of biomedical research and health care has changed remarkably in the 30 years since the first edition of this book was published. So too has the world of computing and communications and thus the underlying scientific issues that sit at the intersections among biomedical science, patient care, public health, and information technology. It is no longer necessary to argue that it has become impossible to practice modern medicine, or to conduct modern biological research, without information technologies. Since the initiation of the Human Genome Project three decades ago, life scientists have been generating data at a rate that defies traditional methods for information management and data analysis.

Health professionals also are constantly reminded that a large percentage of their activities relates to information management—for example, obtaining and recording information about patients, consulting colleagues, reading and assessing the scientific literature, planning diagnostic procedures, devising strategies for patient care, interpreting results of laboratory and radiologic studies, or conducting case-based and population-based research. Artificial intelligence, "big data," and data science are having unprecedented impact on the world, with the biomedical field a particularly active and visible component of such activity.

It is complexity and uncertainty, plus society's overriding concern for patient well-being, and the resulting need for optimal decision making, that set medicine and health apart from many other information-intensive fields. Our desire to provide the best possible health and health care for our society gives a special significance to the effective organization and management of the huge bodies of data with which health professionals and biomedical researchers must deal. It also suggests the need for specialized approaches and for skilled scientists who are knowledgeable about human biology, clinical care, information technologies, and the scientific issues that drive the effective use of such technologies in the biomedical context.

Information Management in Biomedicine

The clinical and research influence of biomedical-computing systems is remarkably broad. Clinical information systems, which provide communication and information-management functions, are now installed in essentially all health care institutions. Physicians can search entire drug indexes in a few seconds, using the information provided by a computer program to anticipate harmful side effects or drug interactions. Electrocardiograms (ECGs) are typically analyzed initially by computer programs, and similar techniques are being applied for interpretation of pulmonary-function tests and a variety of laboratory and radiologic abnormalities. Devices with embedded processors routinely monitor patients and provide warnings in critical-care settings, such as the

intensive-care unit (ICU) or the operating room. Both biomedical researchers and clinicians regularly use computer programs to search the medical literature, and modern clinical research would be severely hampered without computer-based data-storage techniques and statistical analysis systems. Machine learning methods and artificial intelligence are generating remarkable results in medical settings. These have attracted attention not only from the news media, patients, and clinicians but also from health system leaders and from major corporations and startup companies that are offering new approaches to patient care and health information management. Advanced decision-support tools also are emerging from research laboratories, are being integrated with patient-care systems, and are beginning to have a profound effect on the way medicine is practiced.

Despite this extensive use of computers in health care settings and biomedical research, and a resulting expansion of interest in learning more about biomedical computing, many life scientists, health-science students, and professionals have found it difficult to obtain a comprehensive and rigorous, but nontechnical, overview of the field. Both practitioners and basic scientists are recognizing that thorough preparation for their professional futures requires that they gain an understanding of the state of the art in biomedical computing, of the current and future capabilities *and* limitations of the technology, and of the way in which such developments fit within the scientific, social, and financial context of biomedicine and our health care system. In turn, the future of the biomedical-computing field will be largely determined by how well health professionals and biomedical scientists are prepared to guide and to capitalize upon the discipline's development.

This book is intended to meet this growing need for such well-equipped professionals. The first edition appeared in 1990 (published by Addison-Wesley) and was used extensively in courses on medical informatics throughout the world (in some cases with translations to other languages). It was updated with a second edition (published by Springer) in 2000, responding to the remarkable changes that occurred during the 1990s, most notably the Human Genome Project and the introduction of the World Wide Web with its impact on adoption and acceptance of the Internet. The third edition (again published by Springer) appeared in 2006, reflecting the ongoing rapid evolution of both technology and health- and biomedically related applications, plus the emerging government recognition of the key role that health information technology would need to play in promoting quality, safety, and efficiency in patient care. With that edition the title of the book was changed from *Medical Informatics* to *Biomedical Informatics*, reflecting (as is discussed in ▶ Chap. 1) both the increasing breadth of the basic discipline and the evolving new name for academic units, societies, research programs, and publications in the field. The fourth edition (published by Springer in 2014) followed the same conceptual framework for learning about the science that underlies applications of computing and communications technology in biomedicine and health care, for understanding the state of the art in computer applications in clinical care and biology, for critiquing existing systems, and for anticipating future directions that the field may take.

In many respects, the fourth edition was very different from its predecessors, however. Most importantly, it reflected the remarkable changes in computing and communications that continued to occur, most notably in communications, networking, and health information technology policy, and the exploding interest in the role that information technology must play in systems integration and the melding of genomics with innovations in clinical practice and treatment. Several new chapters were introduced and most of the remaining ones underwent extensive revision.

In this fifth edition, we have found that two previous single-chapter topics have expanded to warrant two complementary chapters, specifically Cognitive Science (split into *Cognitive Informatics* and *Human-Computer Interaction, Usability, and Workflow*) and Consumer Health Informatics and Personal Health Records (split into *Personal Health Informatics* and *mHealth and Applications*). There is a new chapter on precision medicine, which has emerged in the past 6 years as a unique area of special interest. Those readers who are familiar with the first four editions will find that the organization and philosophy are essentially unchanged (although bioinformatics, as a set of methodologies, is now considered a "recurrent theme" rather than an "application"), but the content is either new or extensively updated.[1]

This book differs from other introductions to the field in its broad coverage and in its emphasis on the field's conceptual underpinnings rather than on technical details. Our book presumes no health- or computer-science background, but it does assume that you are interested in a comprehensive domain summary that stresses the underlying concepts and that introduces technical details only to the extent that they are necessary to meet the principal goal. Recent specialized texts are available to cover the technical underpinnings of many topics in this book; many are cited as suggested readings throughout the book, or are cited in the text for those who wish to pursue a more technical exposure to a topic.

Overview and Guide to Use of This Book

This book is written as a text so that it can be used in formal courses, but we have adopted a broad view of the population for whom it is intended. Thus, it may be used not only by students of medicine and of the other health professions but also as an introductory text by future biomedical informatics professionals, as well as for self-study and for reference by practitioners, including those who are pursuing formal board certification in clinical informatics (as is discussed in more detail later in this

1 As with the first four editions, this book has tended to draw both its examples and its contributors from North America. There is excellent work in other parts of the world as well, although variations in health care systems, and especially financing, do tend to change the way in which systems evolve from one country to the next. The basic concepts are identical, however, so the book is intended to be useful in educational programs in other parts of the world as well.

"Preface"). The book is probably too detailed for use in a 2- or 3-day continuing-education course, although it could be introduced as a reference for further independent study.

Our principal goal in writing this text is to teach *concepts* in biomedical informatics—the study of biomedical information and its use in decision making—and to illustrate them in the context of descriptions of representative systems that are in use today or that taught us lessons in the past. As you will see, biomedical informatics is more than the study of computers in biomedicine, and we have organized the book to emphasize that point. ▶ Chapter 1 first sets the stage for the rest of the book by providing a glimpse of the future, defining important terms and concepts, describing the content of the field, explaining the connections between biomedical informatics and related disciplines, and discussing the forces that have influenced research in biomedical informatics and its integration into clinical practice and biological research.

Broad issues regarding the nature of data, information, and knowledge pervade all areas of application, as do concepts related to optimal decision making. ▶ Chapters 2 and 3 focus on these topics but mention computers only in passing. They serve as the foundation for all that follows. ▶ Chapters 4 and 5 on cognitive science issues enhance the discussions in ▶ Chaps. 2 and 3, pointing out that decision making and behavior are deeply rooted in the ways in which information is processed by the human mind. Key concepts underlying system design, human-computer interaction, patient safety, educational technology, and decision making are introduced in these chapters.

▶ Chapter 6 introduces the central notions of software engineering that are important for understanding the applications described later. We have dropped a chapter from previous editions that dealt broadly with system architectures, networking, and computer-system design. This topic is more about engineering than informatics, it changes rapidly, and there are excellent books on this subject to which students can turn if they need more information on these topics.

▶ Chapter 7 summarizes the issues of standards development, focusing in particular on data exchange and issues related to sharing of clinical data. This important and rapidly evolving topic warrants inclusion given the evolution of the health information exchange, institutional system integration challenges, federal government directives, and the increasingly central role of standards in enabling clinical systems to have their desired influence on health care practices.

▶ Chapter 8 addresses a topic of increasing practical relevance in both the clinical and biological worlds: natural language understanding and the processing of biomedical texts. The importance of these methods is clear when one considers the amount of information contained in free-text notes or reports (either dictated and transcribed or increasingly created using speech-understanding systems) or in the published biomedical literature. Even with efforts to encourage structured data entry in clinical systems, there will likely always be an important role for techniques that allow computer systems to extract meaning from natural language documents.

▶ Chapter 9 recognizes that bioinformatics is not just an application area but rather a fundamental area of study. The chapter introduces

many of the concepts and analytical tools that underlie modern computational approaches to the management of human biological data, especially in areas such as genomics and proteomics. Applications of bioinformatics related to human health and disease later appear in a chapter on "Translational Bioinformatics" (▶ Chap. 26).

▶ Chapter 10 is a comprehensive introduction to the conceptual underpinnings of biomedical and clinical image capture, analysis, interpretation, and use. This overview of the basic issues and imaging modalities serves as background for ▶ Chap. 22, which deals with imaging applications issues, highlighted in the world of radiological imaging and image management (e.g., in picture archiving and communication systems).

▶ Chapter 11 considers personal health informatics not as a set of applications (which are covered in ▶ Chap. 19), but as introductory concepts that relate to this topic, such as notions of the digital self and the digital divide, patient-generated health data, and how a focus on the patient (or on healthy individuals) affects both the person and the field of biomedical informatics.

▶ Chapter 12 addresses the key legal and ethical issues that have arisen when health information systems are considered. Then, in ▶ Chap. 13, the challenges associated with technology assessment and with the evaluation of clinical information systems are introduced.

▶ Chapters 14–28 (which include two new chapters in this edition, including one on mHealth and another on precision medicine) survey many of the key biomedical areas in which informatics methods are being used. Each chapter explains the conceptual and organizational issues in building that type of system, reviews the pertinent history, and examines the barriers to successful implementations.

▶ Chapter 29 reprises and updates a chapter that was new in the fourth edition, providing a summary of the rapidly evolving policy issues related to health information technology. Although the emphasis is on US government policy, there is some discussion of issues that clearly generalize both to states (in the USA) and to other countries.

The book concludes in ▶ Chap. 30 with a look to the future—a vision of how informatics concepts, computers, and advanced communication devices one day may pervade every aspect of biomedical research and clinical practice. Rather than offering a single point of view developed by a group of forward thinkers, as was offered in the fourth edition, we have invited seven prominent and innovative thinkers to contribute their own views. We integrate these seven future perspectives (representing clinical medicine, nursing, health policy, translational bioinformatics, academic informatics, the information technology industry, and the federal government) into a chapter where the editors have synthesized the seven perspectives after building on how an analysis of the past helps to inform the future of this dynamic field.

The Study of Computer Applications in Biomedicine

The actual and potential uses of computers in health care and biomedicine form a remarkably broad and complex topic. However, just as you do not need to understand how a telephone or an ATM machine works

to make good use of it and to tell when it is functioning poorly, we believe that technical biomedical-computing skills are not needed by health workers and life scientists who wish simply to become effective users of evolving information technologies. On the other hand, such technical skills are of course necessary for individuals with career commitment to developing information systems for biomedical and health environments. Thus, this book will neither teach you to be a programmer nor show you how to fix a broken computer (although it might motivate you to learn how to do both). It also will not tell you about every important biomedical-computing system or application; we shall use an extensive bibliography included with each chapter to direct you to a wealth of literature where review articles and individual project reports can be found. We describe specific systems only as examples that can provide you with an understanding of the conceptual and organizational issues to be addressed in building systems for such uses. Examples also help to reveal the remaining barriers to successful implementations. Some of the application systems described in the book are well established, even in the commercial marketplace. Others are just beginning to be used broadly in biomedical settings. Several are still largely confined to the research laboratory.

Because we wish to emphasize the concepts underlying this field, we generally limit the discussion of technical implementation details. The computer-science issues can be learned from other courses and other textbooks. One exception, however, is our emphasis on the details of decision science as they relate to biomedical problem solving (▶ Chaps. 3 and 24). These topics generally are not presented in computer-science courses, yet they play a central role in the intelligent use of biomedical data and knowledge. Sections on medical decision making and computer-assisted decision support accordingly include more technical detail than you will find in other chapters.

All chapters include an annotated list of "Suggested Readings" to which you can turn if you have a particular interest in a topic, and there is a comprehensive set of references with each chapter. We use **boldface** print to indicate the key terms of each chapter; the definitions of these terms are included in the "Glossary" at the end of the book. Because many of the issues in biomedical informatics are conceptual, we have included "Questions for Discussion" at the end of each chapter. You will quickly discover that most of these questions do not have "right" answers. They are intended to illuminate key issues in the field and to motivate you to examine additional readings and new areas of research.

It is inherently limiting to learn about computer applications solely by reading about them. We accordingly encourage you to complement your studies by seeing real systems in use—ideally by using them yourself. Your understanding of system limitations and of what *you* would do to improve a biomedical-computing system will be greatly enhanced if you have had personal experience with representative applications. Be aggressive in seeking opportunities to observe and use working systems.

In a field that is changing as rapidly as biomedical informatics is, it is difficult ever to feel that you have knowledge that is completely current.

However, the conceptual basis for study changes much more slowly than do the detailed technological issues. Thus, the lessons you learn from this volume will provide you with a foundation on which you can continue to build in the years ahead.

The Need for a Course in Biomedical Informatics

A suggestion that new courses are needed in the curricula for students of the health professions is generally not met with enthusiasm. If anything, educators and students have been clamoring for *reduced* lecture time, for more emphasis on small group sessions, and for more free time for problem solving and reflection. Yet, in recent decades, many studies and reports have specifically identified biomedical informatics, including computer applications, as an area in which new educational opportunities need to be developed so that physicians and other health professionals will be better prepared for clinical practice. As early as 1984, the Association of American Medical Colleges (AAMC) recommended the formation of new academic units in biomedical informatics in our medical schools, and subsequent studies and reports have continued to stress the importance of the field and the need for its inclusion in the educational environments of health professionals.

The reason for this strong recommendation is clear: *The practice of medicine is inextricably entwined with the management of information*. In the past, practitioners handled medical information through resources such as the nearest hospital or medical-school library; personal collections of books, journals, and reprints; files of patient records; consultation with colleagues; manual office bookkeeping; and (all-too-often flawed) memorization. Although these techniques continue to be variably valuable, information technology is offering new methods for finding, filing, and sorting information: online bibliographic retrieval systems, including full-text publications; personal computers, laptops, tablets, and smart phones, with database software to maintain personal information and commonly used references; office-practice and clinical information systems and EHRs to capture, communicate, and preserve key elements of the health record; information retrieval and consultation systems to provide assistance when an answer to a question is needed rapidly; practice-management systems to integrate billing and receivable functions with other aspects of office or clinic organization; and other online information resources that help to reduce the pressure to memorize in a field that defies total mastery of all but its narrowest aspects. With such a pervasive and inevitable role for computers in clinical practice, and with a growing failure of traditional techniques to deal with the rapidly increasing information-management needs of practitioners, it has become obvious to many people that an essential topic has emerged for study in schools and clinical training programs (such as residencies) that train medical and other health professionals.

What is less clear is how the subject should be taught in medical schools or other health professional degree programs, and to what extent it should be left for postgraduate education. We believe that top-

ics in biomedical informatics are best taught and learned in the context of health-science training, which allows concepts from both the health sciences and informatics science to be integrated. Biomedical-computing novices are likely to have only limited opportunities for intensive study of the material once their health-professional training has been completed, although elective opportunities for informatics rotations are now offered to residents in many academic medical centers.

The format of biomedical informatics education has evolved as faculty members have been hired to carry out informatics research and to develop courses at more health-science schools, and as the emphasis on lectures as the primary teaching method continues to diminish. Computers will be used increasingly as teaching tools and as devices for communication, problem solving, and data sharing among students and faculty. Indeed, the recent COVID-19 pandemic has moved many traditional medical teaching experiences from the classroom to online teaching environments using video conferencing and on-demand access to course materials. Such experiences do not teach informatics (unless that is the topic of the course), but they have rapidly engaged both faculty and students in technology-intensive teaching and learning experiences. The acceptance of computing, and dependence upon it, has already influenced faculty, trainees, and curriculum committees. This book is designed to be used in a traditional introductory course, whether taught online or in a classroom, although the "Questions for Discussion" also could be used to focus conversation in small seminars and working groups. Integration of biomedical informatics topics into clinical experiences has also become more common. The goal is increasingly to provide instruction in biomedical informatics whenever this field is most relevant to the topic the student is studying. This aim requires educational opportunities throughout the years of formal training, supplemented by continuing-education programs after graduation.

The goal of integrating biomedicine and biomedical informatics is to provide a mechanism for increasing the sophistication of health professionals, so that they know and understand the available resources. They also should be familiar with biomedical computing's successes and failures, its research frontiers, and its limitations, so that they can avoid repeating the mistakes of the past. Study of biomedical informatics also should improve their skills in information management and problem solving. With a suitable integration of hands-on computer experience, computer-mediated learning, courses in clinical problem solving, and study of the material in this volume, health-science students will be well prepared to make effective use of computational tools and information management in health care delivery.

The Need for Specialists in Biomedical Informatics

As mentioned, this book also is intended to be used as an introductory text in programs of study for people who intend to make their professional careers in biomedical informatics. If we have persuaded you that a course in biomedical informatics is needed, then the requirement for

trained faculty to teach the courses will be obvious. Some people might argue, however, that a course on this subject could be taught by a computer scientist who had an interest in biomedical computing, or by a physician or biologist who had taken a few computing courses. Indeed, in the past, most teaching—and research—has been undertaken by faculty trained primarily in one of the fields and later drawn to the other. Today, however, schools have come to realize the need for professionals trained specifically at the interfaces among biomedicine, biomedical informatics, and related disciplines such as computer science, statistics, cognitive science, health economics, and medical ethics.

This book outlines a first course for students training for careers in the biomedical informatics field. We specifically address the need for an educational experience in which computing and information-science concepts are synthesized with biomedical issues regarding research, training, and clinical practice. It is the *integration* of the related disciplines that originally was lacking in the educational opportunities available to students with career interests in biomedical informatics. Schools are establishing such courses and training programs in growing numbers, but their efforts have been constrained by a lack of faculty who have a broad familiarity with the field and who can develop curricula for students of the health professions as well as of informatics itself.

The increasing introduction of computing techniques into biomedical environments requires that well-trained individuals be available not only to teach students but also to design, develop, select, and manage the biomedical-computing systems of tomorrow. There is a wide range of context-dependent computing issues that people can appreciate only by working on problems defined by the health care setting and its constraints. The field's development has been hampered because there are relatively few trained personnel to design research programs, to carry out the experimental and developmental activities, and to provide academic leadership in biomedical informatics. A frequently cited problem is the difficulty a health professional (or a biologist) and a technically trained computer scientist experience when they try to communicate with one another. The vocabularies of the two fields are complex and have little overlap, and there is a process of acculturation to biomedicine that is difficult for computer scientists to appreciate through distant observation. Thus, interdisciplinary research and development projects are more likely to be successful when they are led by people who can effectively bridge the biomedical and computing fields. Such professionals often can facilitate sensitive communication among program personnel whose backgrounds and training differ substantially.

Hospitals and health systems have begun to learn that they need such individuals, especially with the increasing implementation of, and dependence upon, EHRs and related clinical systems. The creation of a *Chief Medical Information Officer* (CMIO) has now become a common innovation. As the concept became popular, however, questions arose about how to identify and evaluate candidates for such key institutional roles. The need for some kind of suitable certification process became clear—one that would require individuals to demonstrate both formal training and the broad skills and knowledge that were required. Thus,

the American Medical Informatics Association (AMIA) and its members began to develop plans for a formal certification program. For physicians, the most meaningful approach was to create a formal medical subspecialty in clinical informatics. Working with the American Board of Preventive Medicine and the parent organization, the American Board of Medical Specialties (ABMS), AMIA helped to obtain approval for a subspecialty board that would allow medical specialists, with board certification in any ABMS specialty (such as pediatrics, internal medicine, radiology, pathology, preventive medicine) to pursue subspecialty board certification in clinical informatics. This proposal was ultimately approved by the ABMS in 2011, and the board examination was first administered in 2013[2]. After a period during which currently active clinical informatics physician experts could sit for their clinical informatics boards, board eligibility now requires a formal fellowship in clinical informatics. This is similar to the fellowship requirement for other subspecialties such as cardiology, nephrology, and the like. Many health care institutions now offer formal clinical informatics fellowships for physicians who have completed a residency in one of the almost 30 ABMS specialties. These individuals are now often turning to this volume as a resource to help them to prepare for their board examinations.

It is exciting to be working in a field that is maturing and that is having a beneficial effect on society. There is ample opportunity remaining for innovation as new technologies evolve and fundamental computing problems succumb to the creativity and hard work of our colleagues. In light of the increasing sophistication and specialization required in computer science in general, it is hardly surprising that a new discipline should arise at that field's interface with biomedicine. This book is dedicated to clarifying the definition and to nurturing the effectiveness of that discipline: biomedical informatics.

Edward H. Shortliffe
New York, NY, USA

James J. Cimino
Birmingham, AL, USA

Michael F. Chiang
Bethesda, MD, USA
June 2020

2 AMIA is currently developing a Health Informatics Certification program (AHIC) for individuals who seek professional certification in health-related informatics but are not physicians or are otherwise not eligible to take the ABMS board certification exam. ► https://www.amia.org/ahic (Accessed June 10, 2020).

Acknowledgments

In the 1980s, when I was based at Stanford University, I conferred with colleagues Larry Fagan and Gio Wiederhold, and we decided to compile the first comprehensive textbook on what was then called medical informatics. As it turned out, none of us predicted the enormity of the task we were about to undertake. Our challenge was to create a multiauthored textbook that captured the collective expertise of leaders in the field yet was cohesive in content and style. The concept for the book was first developed in 1982. We had begun to teach a course on computer applications in health care at Stanford's School of Medicine and had quickly determined that there was no comprehensive introductory text on the subject. Despite several published collections of research descriptions and subject reviews, none had been developed to meet the needs of a rigorous introductory course. The thought of writing a textbook was daunting due to the diversity of topics. None of us felt that he was sufficiently expert in the full range of important subjects for us to write the book ourselves. Yet we wanted to avoid putting together a collection of disconnected chapters containing assorted subject reviews. Thus, we decided to solicit contributions from leaders in the pertinent fields but to provide organizational guidelines in advance for each chapter. We also urged contributors to avoid writing subject reviews but, instead, to focus on the key conceptual topics in their field and to pick a handful of examples to illustrate their didactic points.

As the draft chapters began to come in, we realized that major editing would be required if we were to achieve our goals of cohesiveness and a uniform orientation across all the chapters. We were thus delighted when, in 1987, Leslie Perreault, a graduate of our informatics training program, assumed responsibility for reworking the individual chapters to make an integral whole and for bringing the project to completion. The final product, published in 1990, was the result of many compromises, heavy editing, detailed rewriting, and numerous iterations. We were gratified by the positive response to the book when it finally appeared, and especially by the students of biomedical informatics who have often come to us at scientific meetings and told us about their appreciation of the book.

As the 1990s progressed, however, we began to realize that, despite our emphasis on basic concepts in the field (rather than a survey of existing systems), the volume was beginning to show its age. A great deal had changed since the initial chapters were written, and it became clear that a new edition would be required. The original editors discussed the project and decided that we should redesign the book, solicit updated chapters, and publish a new edition. Leslie Perreault by this time was a busy Director at First Consulting Group in New York City and would not have as much time to devote to the project as she had when we did the first edition. With trepidation, in light of our knowledge of the work that would be involved, we embarked on the new project.

As before, the chapter authors did a marvelous job, trying to meet our deadlines, putting up with editing changes that were designed to

bring a uniform style to the book, and contributing excellent chapters that nicely reflected the changes in the field during the preceding decade.

No sooner had the second edition appeared in print in 2000 than we started to get inquiries about when the next update would appear. We began to realize that the maintenance of a textbook in a field such as biomedical informatics was nearly a constant, ongoing process. By this time I had moved to Columbia University and the initial group of editors had largely disbanded to take on other responsibilities, with Leslie Perreault no longer available. Accordingly, as plans for a third edition began to take shape, my Columbia colleague Jim Cimino joined me as the new associate editor, whereas Drs. Fagan, Wiederhold, and Perreault continued to be involved as chapter authors. Once again the authors did their best to try to meet our deadlines as the third edition took shape. This time we added several chapters, attempting to cover additional key topics that readers and authors had identified as being necessary enhancements to the earlier editions. We were once again extremely appreciative of all the authors' commitment and for the excellence of their work on behalf of the book and the field.

Predictably, it was only a short time after the publication of the third edition in 2006 that we began to get queries about a fourth edition. We resisted for a year or two, but it became clear that the third edition was becoming rapidly stale in some key areas and that there were new topics that were not in the book and needed to be added. With that in mind we, in consultation with Grant Weston from Springer's offices in London, agreed to embark on a fourth edition. Progress was slowed by my professional moves (to Phoenix, Arizona, then Houston, Texas, and then back to New York) with a very busy 3-year stint as President and CEO of the American Medical Informatics Association. Similarly, Jim Cimino left Columbia to assume new responsibilities at the NIH Clinical Center in Bethesda, MD. With several new chapters in mind, and the need to change authors of some of the existing chapters due to retirements (this too will happen, even in a young field like informatics), we began working on the fourth edition, finally completing the effort with publication in early 2014.

Now, seven years later, we are completing the fifth edition of the volume. It was not long after the publication of the fourth edition that we began to get requests for a new edition that would include many of the new and emerging topics that had not made it into the 2014 publication. With the introduction of new chapters, major revisions to previous chapters, and some reordering of authors or introduction of new ones, we have attempted to assure that this new edition will fill the necessary gaps and engage our readers with its currency and relevance. As Jim Cimino (now directing the Informatics Institute at the University of Alabama in Birmingham) and I considered the development of this edition, we realized that we were not getting any younger and it would be wise to craft a succession plan so that others could handle the inevitable requests for a sixth and subsequent editions. We were delighted when Michael Chiang agreed to join us as an associate editor, coauthoring three chapters and becoming fully involved in the book's philosophy and the editing tasks involved. Michael was a postdoctoral informatics

trainee at Columbia when we were both there on the faculty. A well-known pediatric ophthalmologist, he is now balancing his clinical career with an active set of research and academic activities in biomedical informatics. We believe that Michael will be a perfect person to carry the book into the future as Jim and I (both of whom view the book as a significant component of our professional life's work) phase out our own involvement after this edition. I should add that, in mid-2020, Michael was named director of the National Eye Institute at NIH, which offers further evidence of his accomplishments as a ophthalmologist, researcher, and informatician.

For this edition we owe particular gratitude to Elektra McDermott, our developmental editor, whose rigorous attention to detail has been crucial given the size and the complexity of the undertaking. At Springer we have been delighted to work once again with Grant Weston, Executive Editor in their Medicine and Life Sciences division, who has been extremely supportive despite our missed deadlines. And I want to offer my sincere personal thanks to Jim Cimino, who has been a superb and talented collaborator in this effort for the last three editions. Without his hard work and expertise, we would still be struggling to complete the massive editing job associated with this now very long manuscript.

Edward H. Shortliffe
New York, NY, USA
December 2020

Contents

Editors

Edward H. Shortliffe, MD, PhD, MACP, FACMI Biomedical Informatics, Columbia University, Arizona State University, and Weill Cornell Medical College, New York, NY, USA

Columbia University, New York, NY, USA

Arizona State University, Phoenix, AZ, USA

Weill Cornell Medical College, New York, NY, USA
ted@shortliffe.net

James J. Cimino, MD, FACP, FACMI Informatics Institute, University of Alabama at Birmingham, Birmingham, AL, USA
jamescimino@uabmc.edu

Associate Editor

Michael F. Chiang, MD, MA, FACMI National Eye Institute, National Institutes of Health, Bethesda, MD, USA
michael.chiang@nih.gov

Contributors

Russ B. Altman, MD, PhD, FACMI Departments of Bioengineering, Genetics and Medicine, Stanford University, Stanford, CA, USA
russ.altman@stanford.edu

Suzanne Bakken, RN, PhD, FAAN, FACMI, FIAHSI Department of Biomedical Informatics, Vagelos College of Physicians and Surgeons, School of Nursing, and Data Science Institute, Columbia University, New York, NY, USA
sbh22@cumc.columbia.edu

David W. Bates, MD, MSc, FACMI Division of General Internal, Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA
dbates@partners.org

David Blumenthal, MD, MPP Commonwealth Fund, New York, NY, USA
db@cmwf.org

Patricia Flatley Brennan, RN, PhD, FACMI National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
patti.brennan@nih.gov

Eun Kyoung Choe, PhD College of Information Studies, University of Maryland, College Park, College Park, MD, USA
choe@umd.edu

Terry P. Clemmer, MD Pulmonary – Critical Care Medicine, Intermountain Healthcare (Retired), Salt Lake City, UT, USA
terry.clemmer@imail.org

Sarah Collins Rossetti, PhD, RN, FACMI Department of Biomedical Informatics, School of Nursing, Columbia University Medical Center, New York, NY, USA
sac2125@cumc.columbia.edu

Todd Cooper Trusted Solutions Foundry, Inc., San Diego, CA, USA

Robert M. Cronin, MD, MS, MEng Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA

The Department of Internal Medicine, The Ohio State University, Columbus, OH, USA
robert.cronin@vumc.org

George Demiris, PhD, FACMI Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
gdemiris@nursing.upenn.edu

Dina Demner-Fushman, MD, PhD, FACMI National Library of Medicine, Lister Hill National Center for Biomedical Communications, Bethesda, MD, USA
ddemner@mail.nih.gov

Joshua C. Denny, MD, MS, FACMI All of Us Research Program, National Institutes of Health, Bethesda, MD, USA
josh.denny@vumc.org

Parvati Dev, PhD, FACMI SimTabs, Los Altos Hills, CA, USA
parvati@parvatidev.org

Patricia C. Dykes, PhD, MA, RN, FAAN, FACMI Center for Patient Safety Research & Practice, Brigham and Women's Hospital, Boston, MA, USA
pdykes@bwh.harvard.edu

Noémie Elhadad, PhD, FACMI Department of Biomedical Informatics, Columbia University, New York, NY, USA
noemie.elhadad@columbia.edu

Peter J. Embi, MD, MS, FACMI Regenstrief Institute and Indiana University School of Medicine, Indianapolis, IN, USA

Bradley J. Erickson, MD, PhD Department of Radiology, Mayo Clinic, Rochester, MN, USA
bje@mayo.edu

Carol Friedman, PhD, FACMI Department of Biomedical Informatics, Columbia University, New York, NY, USA

Charles P. Friedman, PhD, FACMI, FIAHSI Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, USA
cpfried@umich.edu

Mark Frisse, MD, MS, MBA, FACMI Biomedical Informatics, Vanderbilt University, Nashville, TN, USA
mark.frisse@vanderbilt.edu

Jeremy D. Goldhaber-Fiebert, PhD Center for Primary Care and Outcomes Research/Center for Health Policy, Stanford University, Stanford, CA, USA
jeremygf@stanford.edu

Kenneth W. Goodman, PhD, FACMI Institute for Bioethics and Health Policy, University of Miami Miller School of Medicine, Miami, FL, USA
kgoodman@med.miami.edu

Robert A. Greenes, MD, PhD, FACMI Department of Biomedical Informatics, Arizona State University, Mayo Clinic, Scottsdale, AZ, USA

Hayit Greenspan, PhD Tel Aviv University, Tel Aviv, Israel

W. Edward Hammond, PhD, FACMI Duke Center for Health Informatics, Duke University Medical Center, Durham, NC, USA

Vitaly Herasevich, MD, PhD Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN, USA
vitaly@mayo.edu

William Hersh, MD, FACP, FACMI, FAMIA, FIAHSI Department of Medical Informatics & Clinical Epidemiology, School of Medicine, Oregon Health & Science University, Portland, OR, USA
hersh@ohsu.edu

Assaf Hoogi, PhD Computer Science Department, Ariel University, Israel, Ariel, Israel

Eric Horvitz, MD, PhD, FACMI Microsoft, Redmond, WA, USA
horvitz@microsoft.com

George Hripcsak, MD, MS, FACMI Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, USA
hripcsak@columbia.edu

Charles Jaffe, MD, PhD, FACMI Health Level Seven International, Ann Arbor, MI, USA
cjaffe@hl7.org

Holly Jimison, PhD, FACMI Khoury College of Computer Sciences & Bouve College of Health Sciences, Northeastern University, Boston, MA, USA
h.jimison@northeastern.edu

Kevin B. Johnson, MD, MS, FACMI Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA
kevin.johnson@vumc.org

Thomas Kannampallil, PhD Institute for Informatics, Washington University School of Medicine, St Louis, MO, USA
thomas.k@wustl.edu

David R. Kaufman, PhD, FACMI Medical Informatics, SUNY Downstate Medical Center, Brooklyn, NY, USA
david.kaufman@downstate.edu

Kensaku Kawamoto, MD, PhD, MHS, FACMI University of Utah, Salt Lake City, UT, USA

Predrag Klasnja, PhD School of Information, University of Michigan, Ann Arbor, MI, USA
klasnja@umich.edu

Wayne R. Kubick, BA, MBA Health Level 7 International, Ann Arbor, MI, USA

Martin LaVenture, MPH, PhD, FACMI Informatics Savvy Advisors, Edina, MN, USA

Institute for Health Informatics (IHI), University of Minnesota, Minneapolis, MN, USA
marty@umn.edu

Russell B. Leftwich, MD InterSystems, Cambridge, MA, USA

Vanderbilt University Medical Center, Department of Biomedical Informatics, Nashville, TN, USA

Roger G. Mark, MD, PhD Department of Electrical Engineering and Computer Science (EECS), Institute of Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA
rgmark@mit.edu

Clement J. McDonald, MD, MS, FACMI Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
clemmcdonald@mail.nih.gov; CJMcDona@Regenstrief.org

Genevieve B. Melton, MD, PhD, FACMI Department of Surgery and Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA
gmelton@umn.edu

Blackford Middleton, MD, MPH, MSc, FACMI Apervita, Inc, Chicago, IL, USA

Matt Might, PhD UAB Hugh Kaul Precision Medicine Institute, University of Alabama at Birmingham, Birmingham, AL, USA
might@uab.edu

Randolph A. Miller, MD, FACMI Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
randolph.a.miller@vanderbilt.edu

Sean D. Mooney, PhD, FACMI Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA
sdmooney@uw.edu

Judy Murphy, RN, FACMI, LFHIMSS, FAAN IBM Global Healthcare, Armonk, NY, USA
judy.murphy29@gmail.com

Mark A. Musen, MD, PhD, FACMI Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA
musen@stanford.edu

Viet Nguyen, MD Stratametrics LLC, Salt Lake City, UT, USA

Douglas K. Owens, MD, MS Center for Primary Care and Outcomes Research/ Center for Health Policy, Stanford University, Stanford, CA, USA
owens@stanford.edu

Judy G. Ozbolt, PhD, RN(Ret), FAAN, FACMI Department of Organizational Systems and Adult Health, University of Maryland School of Nursing, Baltimore, MD, USA
judy.ozbolt@gmail.com

Vimla L. Patel, PhD, DSc, FACMI Center for Cognitive Studies in Medicine and Public Health, The New York Academy of Medicine, New York, NY, USA
vpatel@nyam.org

Philip R. O. Payne, PhD, FACMI Institute for Informatics, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
prpayne@wustl.edu

Brian W. Pickering, MB, BCh Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN, USA
pickering.brian@mayo.edu

Wanda Pratt, PhD, FACMI Information School, University of Washington, Seattle, WA, USA
wpratt@uw.edu

William C. Reed, MS, PhD Huntzinger Management Group, Inc, Moosic, PA, USA
wcreed@huntzingergroup.com

David A. Ross, DSc Task Force for Global Health, Decatur, GA, USA
dross@taskforce.org

Daniel L. Rubin, MD, MS, FSIIM, FACMI Departments of Biomedical Data Science, Radiology, and Medicine, Stanford University, Stanford, CA, USA
dlrubin@stanford.edu

Robert S. Rudin, PhD RAND Healthcare, RAND Corporation, Boston, MA, USA
rrudin@rand.org

Titus Schleyer, DMD, PhD, FAMIA, FACMI Department of Medicine, Division of General Internal Medicine and Geriatrics, Indiana University School of Medicine, Indianapolis, IN, USA
schleyer@regenstrief.org

Nigam H. Shah, MBBS, PhD, FACMI Department of Medicine, Stanford University, Stanford, CA, USA
nigam@stanford.edu

Harold C. Sox, MD, MACP Patient-Centered Outcomes Research Institute, Washington, DC, USA

Geisel School of Medicine at Dartmouth, Dartmouth College, Hanover, NH, USA
hsox@comcast.net

Catherine Staes, PhD, MPH, RN, FACMI Nursing Informatics Program, College of Nursing, University of Utah, Lake City, UT, USA
Catherine.Staes@hsc.utah.edu

Justin B. Starren, MD, PhD Division of Health and Biomedical Informatics, Departments of Preventive Medicine and Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
Justin.Starren@northwestern.edu

Paul C. Tang, MD, MS, FACMI Clinical Excellence Research Center, Stanford University, Stanford, CA, USA
paultang@stanford.edu

Peter Tarczy-Horonch, MD, FACMI Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA
pth@uw.edu

Jessica D. Tenenbaum, PhD, FACMI Department of Biostatistics & Bioinformatics, Duke University, Durham, NC, USA
jessie.tenenbaum@duke.edu

David K. Vawdrey, PhD, FACMI Geisinger, Danville, PA, USA

Lynn Harold Vogel, AB, AM, PhD LH Vogel Consulting, LLC, Ridgewood, NJ, USA
lynn@lhvogelconsulting.com

Robert M. Wachter, MD Department of Medicine, University of California, San Francisco, San Francisco, CA, USA
Robert.Wachter@ucsf.edu

Adam B. Wilcox, MD, PhD, FACMI University of Washington, Seattle, WA, USA
abwilcox@uw.edu

Jeremy C. Wyatt, MBBS, FRCP, FACMI, FFCI Wessex Institute of Health Research, Faculty of Medicine, University of Southampton, Southampton, UK
j.c.wyatt@soton.ac.uk

William A. Yasnoff, MD, PhD, FACMI National Health Information Infrastructure Advisors, Portland, OR, USA

Division of Health Sciences Informatics, Johns Hopkins University, Portland, OR, USA

National Health Information Infrastructure (NHII) Advisors and Johns Hopkins University, Portland, OR, USA
william.yasnoff@nhiiadvisors.com

About the Future Perspective Authors (Chapter 30)

David Blumenthal, MD, MPP
became President and CEO of the Commonwealth Fund, a national health care philanthropy based in New York City, in January, 2013. Previously, he served as Chief Health Information and Innovation Officer at Partners Health System in Boston, MA, and was Samuel O. Thier Professor of Medicine and Professor of Health Care Policy at Massachusetts General Hospital/ Harvard Medical School. From 2009 to 2011, Dr. Blumenthal was the National Coordinator for Health Information Technology under President Barack Obama. In this role he was charged with building an interoperable, private and secure nationwide health information system and supporting the widespread, meaningful use of health IT. Prior to that, Dr. Blumenthal was a practicing primary care physician, director of the Institute for Health Policy, and professor of medicine and health policy at Massachusetts General Hospital/Partners Healthcare System and Harvard Medical School. As a renowned health services researcher and national authority on health IT adoption, Dr. Blumenthal has authored over 300 scholarly publications, including the seminal studies on the adoption and use of health information technology in the United States. Dr. Blumenthal received his undergraduate, medical, and public policy degrees from Harvard University and completed his residency in internal medicine at Massachusetts General Hospital.

Patricia Flatley Brennan, RN, PhD, FAAN, FACMI
is the Director of the National Library of Medicine (NLM) and Adjunct Investigator in the National Institute of Nursing Research's Advanced Visualization Branch at the National Institutes of Health (NIH). As the world's largest biomedical library, NLM produces digital information resources used by scientists, health professionals, and members of the public. A leader in research in computational health informatics, NLM plays a pivotal role in translating research into practice. NLM's research and information services support scientific discovery, health care, and public health. Prior to joining

NLM, Dr. Brennan was the Lillian L. Moehlman Bascom Professor at the School of Nursing and College of Engineering at the University of Wisconsin–Madison. Dr. Brennan is a pioneer in the development of innovative information systems and services such as ComputerLink, an electronic network designed to reduce isolation and improve self-care among home care patients. She directed HeartCare, a web-based information and communication service that helps home-dwelling cardiac patients recover faster and with fewer symptoms, and also directed Project HealthDesign, an initiative designed to stimulate the next generation of personal health records. Her professional accomplishments reflect her background, which unites engineering, information technology, and clinical care to improve public health and ensure the best possible experience in patient care. A past president of the American Medical Informatics Association, Dr. Brennan was elected to the National Academy of Medicine in 2001. She is a fellow of the American Academy of Nursing, the American College of Medical Informatics, and the New York Academy of Medicine. In 2020, Dr. Brennan was inducted into the American Institute for Medical and Biological Engineering (AIMBE). The AIMBE College of Fellows is among the highest professional distinctions accorded to a medical and biological engineer.

Mark Frisse, MD, MS, MBA

is an Emeritus Professor in the Department of Biomedical Informatics at Vanderbilt University Medical Center. Until July 2020, he held the Accenture Endowed Chair within the Department. His interests are directed toward the potential contribution to economically sustainable health care through more effective use of informatics and information technology. In addition to his teaching and research responsibilities at the Vanderbilt Medical School, he also directed the Masters of Management in Health Care graduate program at Vanderbilt's Owen Graduate School of Management and led a health information technology course for the executive Masters of Health Care Delivery Science program at Tuck School of Business, Dartmouth College. Previously, Dr. Frisse has held leadership positions at Washington University, Express Scripts, and the First Consulting Group.

At Express Scripts, he was Chief Medical Officer and was responsible for their Practice Patterns Science division and helped found RxHub. At First Consulting, he led engagements in vendor selection, quality governance, physician information technology leadership development, and clinician governance. A board-certified internist with fellowship training in medical oncology, he obtained his bachelor's degree from the University of Notre Dame in 1974 and his MD from Washington University in 1978. He received a master's degree in medical information science from Stanford University in 1987 and an MBA from Washington University in 1997. He is a fellow of the American College of Physicians, the American College of Medical Informatics, and the New York Academy of Medicine. He is an elected member of the National Academy of Medicine.

Eric Horvitz

is a technical fellow at Microsoft, where he serves as the company's first Chief Scientific Officer. He previously served as director of Microsoft Research Labs. He has pursued principles and applications of AI with contributions in machine learning, perception, natural language understanding, and decision making. His research centers on challenges with uses of AI amidst the complexities of the open world, including uses of probabilistic and decision-theoretic representations for reasoning and action, models of bounded rationality, and human-AI complementarity and coordination. His efforts and collaborations have led to fielded systems in healthcare, transportation, ecommerce, operating systems, and aerospace. He received the Feigenbaum Prize and the Allen Newell Prize for contributions to AI. He was inducted into the CHI Academy for contributions at the intersection of AI and human-computer interaction. He has been elected fellow of the American College of Medical Informatics (ACMI), National Academy of Engineering (NAE), Association of Computing Machinery (ACM), Association for the Advancement of AI (AAAI), the American Association for the Advancement of Science (AAAS), the American Academy of Arts and Sciences, and the American Philosophical Society. He has served as president of the AAAI, and on advisory commit-

tees for the NSF, NIH, and the U.S. Department of Defense. He currently serves on the Board of Regents of the National Library of Medicine and on the scientific advisory board of the Allen Institute for AI. He earned his Bachelors in biophysics at Binghamton University and a PhD in biomedical informatics and his MD at Stanford University.

Judy Murphy, RN, FACMI, FHIMSS, FAAN
is a nursing executive with a long history in health informatics. She was Chief Nursing Officer for IBM Global Healthcare, where she was strategic advisor regarding provider IT solutions to achieve the quadruple aim. Prior to working at IBM, she was Deputy National Coordinator for Programs and Policy at the ONC in Washington D.C. where she led federal efforts to assist healthcare providers in adopting health IT to improve care and to promote consumers' understanding and use of health IT for their own health. She came to ONC with 25 years of health informatics experience at Aurora Health Care in Wisconsin, a large integrated delivery network. As Vice President-EHR Applications, she led their EHR program since 1995, when Aurora was an early adopter of technology. She publishes and lectures nationally and internationally. She served on the AMIA and HIMSS Board of Directors and is a Fellow in the American Academy of Nursing, the American College of Medical Informatics and HIMSS. She has received numerous awards, including the HIMSS 2018 Most Influential Women in Health IT, the AMIA 2014 Don Eugene Detmer Award for Health Policy Contributions in Informatics, the HIMSS 2014 Federal Health IT Leadership Award, and the HIMSS 2006 Nursing Informatics Leadership Award.

Peter Tarczy-Hornoch, MD
is Professor and Chair of the Department of Biomedical Informatics and Medical Education at the University of Washington (UW). His background includes computer science, bioengineering, biomedical informatics, medicine (Pediatrics and Neonatology) with undergraduate and medical degrees from Stanford, Residency at the University of Minnesota, Fellowship at the University of Washington. He has played a key role in establishing and growing the biomedical informatics research, teaching and practice activities at UW

for over two decades. The unifying theme of his research has been data integration of electronic biomedical data (clinical and genomic) both for a) knowledge discovery and b) in order to integrate this knowledge with clinical data at the point of care for decision support. His current research focuses on a) secondary use of electronic medical record (EMR) for translational research including outcomes research, learning healthcare systems, patient accrual and biospecimen acquisition based on complex phenotypic eligibility criteria, b) the use of EMR systems for cross institutional comparative effectiveness research, and c) integration of genomic data into the EMR for clinical decision support. Key past research has included data integration and knowledge base creation for genomic testing. He has served in a number of national leadership roles including founding chair of the AMIA Genomics (now Genomics and Translational Bioinformatics) working group.

Robert M. Wachter, MD

is Professor and Chair of the Department of Medicine at the University of California, San Francisco. He is author of 300 articles and 6 books and is a frequent contributor to the *New York Times* and *Wall Street Journal*. He coined the term "hospitalist" in 1996 and is often considered the "father" of the hospitalist field, the fastest growing specialty in the history of modern medicine. He is past president of the Society of Hospital Medicine and past chair of the American Board of Internal Medicine. In the safety and quality arenas, his book, *Understanding Patient Safety*, is the world's top selling safety primer. In 2004, he received the John M. Eisenberg Award, the nation's top honor in patient safety. Twelve times, *Modern Healthcare* magazine has ranked him as one of the 50 most influential physician-executives in the U.S.; he was #1 on the list in 2015. His 2015 book, *The Digital Doctor: Hope, Hype and Harm at the Dawn of Medicine's Computer Age*, was a *New York Times* science best-seller. In 2016, he chaired a blue-ribbon commission advising England's National Health Service on its digital strategy. He is an elected member of the National Academy of Medicine.

About the Editors

Edward H. Shortliffe

is Chair Emeritus and Adjunct Professor in the Department of Biomedical Informatics at Columbia University's Vagelos College of Physicians and Surgeons. Previously he served as President and CEO of the American Medical Informatics Association. He was Professor of Biomedical Informatics at the University of Texas Health Science Center in Houston and at Arizona State University. A board-certified internist, he was Founding Dean of the University of Arizona College of Medicine – Phoenix and served as Professor of Biomedical Informatics and of Medicine at Columbia University. Before that he was Professor of Medicine and of Computer Science at Stanford University. Honors include his election to membership in the National Academy of Medicine (where he served on the executive council for 6 years and has chaired the membership committee) and in the American Society for Clinical Investigation. He has also been elected to fellowship in the American College of Medical Informatics and the American Association for Artificial Intelligence. A Master of the American College of Physicians (ACP), he held a position for 6 years on that organization's Board of Regents. He is Editor Emeritus of the *Journal of Biomedical Informatics* and has served on the editorial boards for several other biomedical informatics publications. In the early 1980s, he was recipient of a research career development award from the National Library of Medicine. In addition, he received the Grace Murray Hopper Award of the Association for Computing Machinery in 1976, the Morris F. Collen Award of the American College of Medical Informatics in 2006, and was a Henry J. Kaiser Family Foundation Faculty Scholar in General Internal Medicine. He has served on the oversight committee for the Division on Engineering and Physical Sciences (National Academy of Sciences), the National Committee on Vital and Health Statistics (NCVHS), and on the President's Information Technology Advisory Committee (PITAC). Dr. Shortliffe has authored over 350 articles and books in the fields of biomedical computing and artificial intelligence.

James J. Cimino

is a board-certified internist who completed a National Library of Medicine informatics fellowship at the Massachusetts General Hospital and Harvard University and then went on to an academic position at Columbia University College of Physicians and Surgeons and the Presbyterian Hospital in New York. He spent 20 years at Columbia, carrying out clinical informatics research, building clinical information systems, teaching medical informatics and medicine, and caring for patients, rising to the rank of full professor in both Biomedical Informatics and Medicine. His principal research areas there included desiderata for controlled terminologies, mobile and Web-based clinical information systems for clinicians and patients, and a context-aware form of clinical decision support called "infobuttons." In 2008, he moved to the National Institutes of Health, where he was the Chief of the Laboratory for Informatics Development and a Tenured Investigator at the NIH Clinical Center and the National Library of Medicine. His principal project involved the development of the Biomedical Translational Research Information System (BTRIS), an NIH-wide clinical research data resource. In 2015, he left the NIH to be the inaugural Director of the Informatics Institute at the University of Alabama at Birmingham. The Institute is charged with improving informatics research, education, and service across the University, supporting the Personalized Medicine Institute, the Center for Genomic Medicine, and the University Health System Foundation, including improvement of and access to electronic health records. He holds the rank of Tenured Professor in Medicine and is the Chief for the Informatics Section in the Division of General Internal Medicine. He continues to conduct research in clinical informatics and clinical research informatics, he was Director of the NLM's weeklong Biomedical Informatics course for 16 years, and teaches at Columbia University and Georgetown University as an Adjunct Professor. He is an Associate Editor of the *Journal of Biomedical Informatics*. His honors include Fellowships of the American College of Physicians, the New York Academy of Medicine and the American College of Medical Informatics (Past President), the Priscilla Mayden Award from the University of Utah, the Donald A.B. Lindberg Award for Innovation in Informatics and the President's Award, both from the American Medical

Informatics Association, the Morris F. Collen Award of the American College of Medical Informatics, the Medal of Honor from New York Medical College, the NIH Clinical Center Director's Award (twice), and induction into the National Academy of Medicine (formerly the Institute of Medicine).

Michael F. Chiang

is Director of the National Eye Institute, at the National Institutes of Health in Bethesda, Maryland. His clinical practice focuses on pediatric ophthalmology, and he is board-certified in clinical informatics. His research develops and applies biomedical informatics methods to clinical ophthalmology in areas such as retinopathy of prematurity (ROP), telehealth, artificial intelligence, clinical information systems, genotype-phenotype correlation, and data analytics. His group has published over 200 peer-reviewed papers, and has developed an assistive artificial intelligence system for ROP that received breakthrough status from the US Food and Drug Administration. He received a BS in Electrical Engineering and Biology from Stanford University, an MD from the Harvard-MIT Division of Health Sciences and Technology, and an MA in Biomedical Informatics from Columbia University. He completed clinical training at the Johns Hopkins Wilmer Eye Institute. Between 2001 and 2010, he worked at Columbia University, where he was Anne S. Cohen Associate Professor of Ophthalmology & Biomedical Informatics, director of medical student education in ophthalmology, and director of the introductory graduate student course in biomedical informatics. From 2010 to 2020, he was Knowles Professor of Ophthalmology & Medical Informatics and Clinical Epidemiology, and Associate Director of the Casey Eye Institute, at the Oregon Health & Science University (OHSU) Casey Eye Institute. He has served as a member of the American Academy of Ophthalmology (AAO) Board of Trustees, Chair of the AAO IRIS Registry Data Analytics Committee, Chair of the AAO Task Force on Artificial Intelligence, Chair of the AAO Medical Information Technology Committee, and on numerous other national and local committees. He currently serves as an Associate Editor for *JAMIA*, and is on the Editorial Board for *Ophthalmology* and the *Asia-Pacific Journal of Ophthalmology*.

Recurrent Themes in Biomedical Informatics

Contents

Biomedical Informatics: The Science and the Pragmatics

Edward H. Shortliffe and Michael F. Chiang

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_1

1

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- Why is information and knowledge management a central issue in biomedical research, clinical practice, and public health?
- What are integrated information management environments, and how are they affecting the practice of medicine, the promotion of health, and biomedical research?
- What do we mean by the terms *biomedical informatics*, *medical computer science*, *medical computing*, *clinical informatics*, *nursing informatics*, *bioinformatics*, *public health informatics*, and *health informatics*?
- What is *translational research*, why is it being heavily promoted and supported, how does it depend on *translational bioinformatics* and *clinical research informatics*, and how do these all relate to *precision medicine*?
- Why should health professionals, life scientists, and students of the health professions learn about biomedical informatics concepts and informatics applications?
- How has the development of modern computing technologies and the Internet changed the nature of biomedical computing?
- How is biomedical informatics related to clinical practice, public health, biomedical engineering, molecular biology, decision science, information science, and computer science?
- How does information in clinical medicine and health differ from information in the basic sciences?
- How can changes in computer technology and the financing of health care influence the integration of biomedical computing into clinical practice?

1.1 The Information Revolution Comes to Medicine

After scientists had developed the first digital computers in the 1940s, society was told that these new machines would soon be serving routinely as memory devices, assisting with calculations and with information retrieval. Within the next decade, physicians and other health professionals had begun to hear about the dramatic effects that such technology would have on clinical practice.

More than seven decades of remarkable progress in computing have followed those early predictions, and many of the original prophesies have come to pass. Stories regarding the "information revolution", "artificial intelligence", and "big data" fill our newspapers and popular magazines, and today's children show an uncanny ability to make use of computers (including their handheld mobile versions) as routine tools for study, communication, and entertainment. Similarly, clinical workstations have been available on hospital wards and in outpatient offices for decades, and in some settings have been supplanted by mobile tablets with wireless connectivity.

Not long ago, the health care system was perceived as being slow to understand information technology and slow to exploit it for its unique practical and strategic functionalities. This is no longer the case. The enormous technological advances of the last four decades—personal computers and graphical interfaces, laptop machines, new methods for human-computer interaction, innovations in mass storage of data (both locally and in the "cloud"), mobile devices, personal health-monitoring devices, the Internet, wireless communications, social media, and more—have all combined to make use of computers by health workers and biomedical scientists part of today's routine. This new world is already upon us, but its greatest influence is yet to come as today's prominent innovations such as

electronic health records and decision-support software are further refined. This book will teach you about our present resources and accomplishments, and about gaps that need to be addressed in the years ahead.

When one considers today's penetration of computers and communication into our daily lives, it is remarkable that the first personal computers were introduced as recently as the late 1970s; local area networking has been available only since the 1980s; the World Wide Web dates only to the early 1990s; and smart phones, social networking, tablet computers, wearable devices, and wireless communication are even more recent. This dizzying rate of change, combined with equally pervasive and revolutionary changes in almost all international health care systems, makes it difficult for public-health planners and health-institutional managers to try to deal with both issues at once.

As new technologies have been introduced and adopted in health settings, unintended consequences have emerged, such as **ransomware** and other security challenges that can compromise the protection and privacy of patient data. Yet many observers now believe that rapid changes in both technology and health systems are inextricably related. We can see that planning for the new health care environments of the coming decades requires a deep understanding of the role that information technology is likely to play in those environments.

What might that future hold for the typical practicing clinician? As we discuss in detail in ▶ Chap. 14, no applied clinical computing topic is gaining more attention currently than is the issue of **electronic health records** (EHRs). Health care organizations have largely replaced their paper-based recording systems, recognizing that they need to have digital systems in place that create opportunities to facilitate patient care that is safe and effective, to answer questions that are crucially important for strategic planning, to support a better understanding of how they and their providers compare with other organizations in their local or regional competitive environment, and to support reporting to regulatory agencies.

In the past, administrative and financial data were the major elements required for planning, but in recent years comprehensive clinical data have also become important for institutional self-analysis and strategic planning. Furthermore, the inefficiencies and frustrations associated with the use of paper-based medical records are well accepted (Dick and Steen 1991 (Revised 1997)), especially when inadequate access to clinical information is one of the principal barriers that clinicians encounter when trying to increase their efficiency in order to meet productivity goals for their practices.

1.1.1 Integrated Access to Clinical Information

Encouraged by **health information technology** (**HIT**) vendors (and by the US government, as is discussed later), most health care institutions have or are developing integrated computer-based information-management environments. These underlie a clinical world in which computational tools assist not only with patient-care matters (e.g., reporting results of tests, allowing direct entry of orders or patient information by clinicians, facilitating access to transcribed reports, and in some cases supporting **telemedicine** applications or decision-support functions) but also with administrative and financial topics (e.g., tracking of patients within the hospital, managing materials and inventory, supporting personnel functions, and managing the payroll), with research (e.g., analyzing the outcomes associated with treatments and procedures, performing quality assurance, supporting clinical trials, and implementing various treatment protocols), with access to

scholarly information (e.g., accessing digital libraries, supporting bibliographic search, and providing access to drug information databases), and even with office automation (e.g., providing access to spreadsheets and document-management software). The key idea, however, is that at the heart of the evolving integrated environments lies an electronic health record that is intended to be accessible, confidential, secure, acceptable to clinicians and patients, and integrated with other types of useful information to assist in planning and problem solving.

1.1.2 Today's Electronic Health Record (EHR) Environment

The traditional paper-based medical record is now recognized as being woefully inadequate for meeting the needs of modern medicine. It arose in the nineteenth century as a highly personalized "lab notebook" that clinicians could use to record their observations and plans so that they could be reminded of pertinent details when they next saw the same patient. There were no regulatory requirements, no assumptions that the record would be used to support communication among varied providers of care, and few data or test results to fill up the record's pages. The record that met the needs of clinicians a century or so ago struggled mightily to adjust over the decades and to accommodate to new requirements as health care and medicine changed. Today the inability of paper charts to serve the best interests of the patient, the clinician, and the health system is no longer questioned (see ▶ Chaps. 14 and 16).

Most organizations have found it challenging (and expensive) to move to a paperless, electronic clinical record. This observation forces us to ask the following questions: "What is a health record in the modern world? Are the available products and systems well matched with the modern notions of a comprehensive health record? Do they meet the needs of individual users as well as the health systems themselves? Are they efficient, easy to use, and smoothly integrated into clinical workflow? How should our concept of the comprehensive health record evolve in the future, as technology creates unprecedented opportunities for innovation?"

The complexity associated with automating clinical-care records is best appreciated if one analyzes the processes associated with the creation and use of such records rather than thinking of the record as a physical object (such as the traditional paper chart) that can be moved around as needed within the institution. For example, on the input side (◘ Fig. 1.1), an electronic version of the paper chart requires the integration of processes for data capture and for merging information from diverse sources.

The contents of the paper record were traditionally organized chronologically—often a severe limitation when a clinician sought to find a specific piece of information that could occur almost anywhere within the chart. To be useful, the electronic record system has to make it easy to access and display needed data, to analyze them, and to share them among colleagues and with secondary users of the record who are not involved in direct patient care (◘ Fig. 1.2). Thus, the EHR, as an adaptation of the paper record, is best viewed not as an object, or a product, but rather as a set of processes that an organization puts into place, supported by technology (◘ Fig. 1.3).

Implementing electronic records is inherently a systems-integration task. It accordingly requires a custom-tailored implementation at each institution, given the differences in existing systems and practices that must be suitably integrated. Joint development and local adaptation are crucial, which implies that the institutions that purchase such systems must have local expertise that can oversee and facilitate an effective implementation process, including elements of process re-engineering and cultural change that are inevitably involved.

Experience has shown that clinicians are "horizontal" users of information technology (Greenes and Shortliffe 1990). Rather than becoming "power users" of a narrowly defined software package, they tend to seek broad functionality across a wide variety of systems and resources. Thus, routine use of comput-

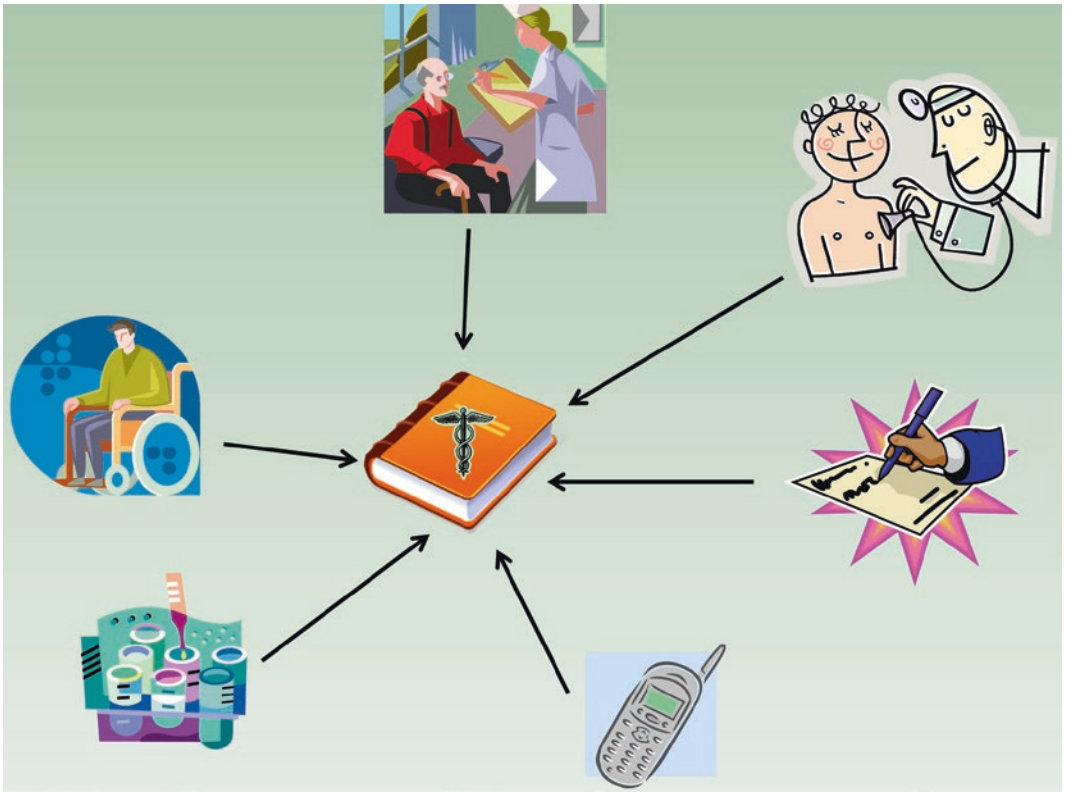


Fig. 1.1 Inputs to the clinical-care record. The traditional paper record was created by a variety of organizational processes that captured varying types of information (notes regarding direct encounters between health professionals and patients, laboratory or radiologic results, reports of telephone calls or prescriptions, and data obtained directly from patients). The paper record thus was a merged collection of such data, generally organized in chronological order

ers, and of EHRs, is most easily achieved when the computing environment offers a critical mass of functionality that makes the system both smoothly integrated with workflow and useful for essentially every patient encounter.

The arguments for automating clinical-care records are summarized in ▶ Chaps. 2 and 14 and in the now classic Institute of Medicine's report on **computer-based patient records** (**CPRs**) (Dick and Steen 1991 (Revised 1997).[1] One argument that warrants emphasis is the importance of the EHR in supporting **clinical trials**—experiments in which data from specific patient interactions are pooled and analyzed in order to learn about the safety and efficacy of new treatments or tests and to gain insight into disease processes that are not otherwise well understood. Medical researchers were constrained in the past by clumsy methods for identifying patients who met inclusion criteria for clinical trials as well as acquiring the data needed for the trials, generally relying on manual capture of information onto datasheets that were later transcribed into computer databases for statistical analysis (Fig. 1.4). The approach was labor-intensive, fraught with opportunities for error, and added to the high costs associated with randomized prospective research protocols.

The use of EHRs has offered many advantages to those carrying out clinical research (see ▶ Chap. 27). Most obviously, it helps to eliminate the manual task of extracting data from charts or filling out specialized data-

1 The Institute of Medicine, part of the National Academy of Sciences, is now known as the National Academy of Medicine.

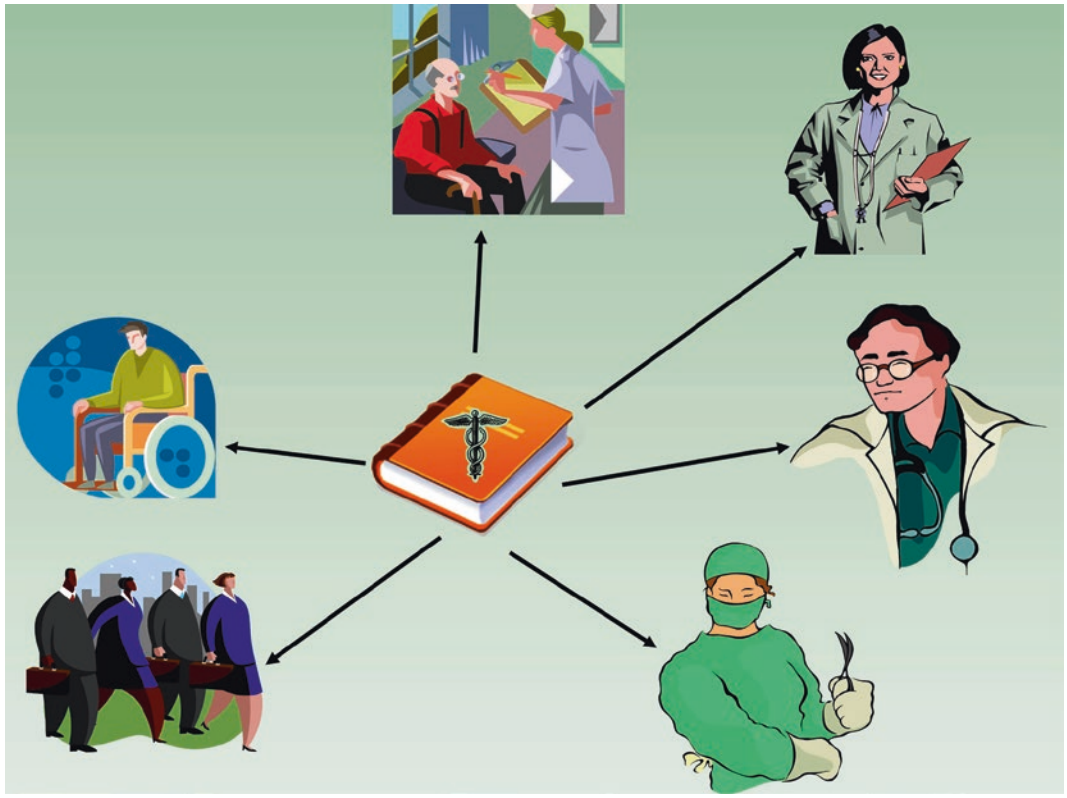


Fig. 1.2 Outputs from the clinical-care record. Once information was collected in the traditional paper chart, it needed to be provided to a wide variety of potential users of the information that it contained. These users included health professionals and the patients themselves, as well as "secondary users" (represented here by the individuals in business suits) who had valid reasons for accessing the record but who were not involved with direct patient care. Numerous providers are typically involved in a patient's care, so the chart also served as a means for communicating among them. The traditional mechanisms for displaying, analyzing, and sharing information from such records resulted from a set of processes that often varied substantially across several patient-care settings and institutions

sheets. The data needed for a study can often be derived directly from the EHR, thus making much of what is required for research data collection simply a by-product of routine clinical record keeping (Fig. 1.5). Other advantages accrue as well. For example, the record environment can help to ensure compliance with a research protocol, pointing out to a clinician when a patient is eligible for a study or when the protocol for a study calls for a specific management plan given the currently available data about that patient. We are also seeing the development of novel authoring environments for clinical trial protocols that can help to ensure that the data elements needed for the trial are compatible with the local EHR's conventions for representing patient descriptors.

Note that Fig. 1.5 represents a study at a single institution and often for a limited subset of the patients who receive care there. Yet much research is carried out with very large numbers of patients, such as within a regional health care system, statewide, or nationally. Accordingly, the size of research datasets can get very large, but analyzing across them introduces challenges related to data exchange and the standardization of the ways in which individual data elements are defined, identified, or stored (see ▶ Chap 8). Retrospective studies on data collected in the past typically cannot assume a prior standardization of the elements that will be needed, thereby requiring analyses that infer relationships among specific descriptors in different institutions represented in different ways. When the number of data elements

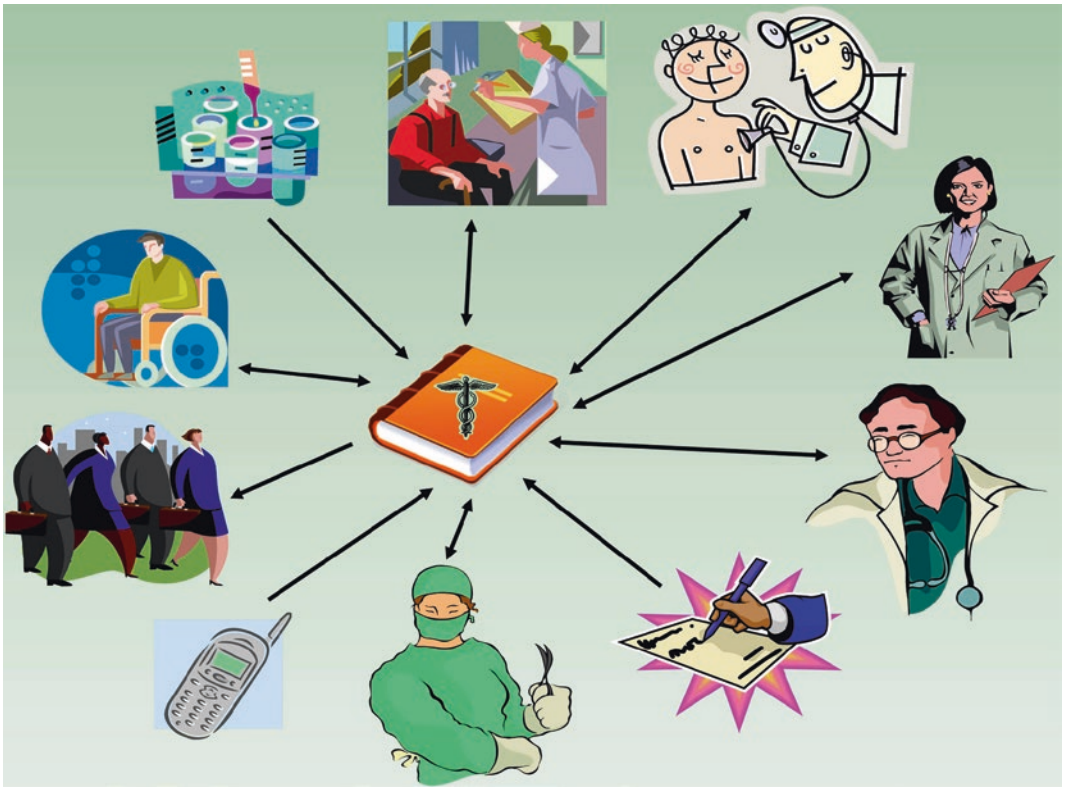


Fig. 1.3 Complex processes demanded of the record. As shown in Figs. 1.1 and 1.2, the paper chart evolved to become the incarnation of a complex set of organizational processes, which both gathered information to be shared and then distributed that information to those who had valid reasons for accessing it. Yet paper-based documents were severely limited in meeting the diverse requirements for data collection and information access that are implied by this diagram. These deficiencies accounted in large part for the effort to create today's electronic health records

is large, and the population being studied is also vast, the challenges are often described as "big data" analytics (James et al. 2013).

Another theme in the changing world of health care is the increasing investment in the creation of **standard order sets**, **clinical guidelines**, and **clinical pathways** (see ▶ Chap. 24), generally in an effort to reduce practice variability and to develop consensus approaches to recurring management problems. Several government and professional organizations, as well as individual provider groups, have invested heavily in guideline development, often putting an emphasis on using clear evidence from the literature, rather than expert opinion alone, as the basis for the advice. Despite the success in creating such **evidence-based guidelines**, there is a growing recognition that we need better methods for delivering the decision logic to the point of care. Guidelines that appear in monographs or journal articles tend to sit on shelves, unavailable when the knowledge they contain would be most valuable to practitioners. Computer-based tools for implementing such guidelines, and integrating them with the EHR, present a means for making high-quality advice available in the routine clinical setting. Many organizations are accordingly integrating decision-support tools with their EHR systems (see ▶ Chaps. 14 and 24), and there are highly visible commercial efforts underway to provide computer-based diagnostic decision support to practitioners.[2]

There are at least five major issues that have consistently constrained our efforts to build effective EHRs: (1) the need for stan-

2 ▶ https://ehrintelligence.com/news/top-clinical-decision-support-system-cdss-companies-by-ambulatory-inpatient; ▶ https://www.ibm.com/watson/health/. (Accessed 5/29/19/).

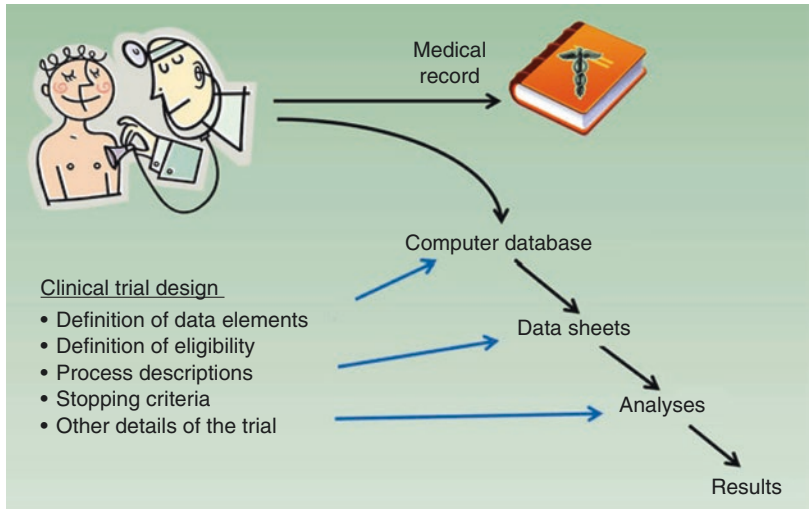


Fig. 1.4 Traditional data collection for clinical trials. Until the introduction of EHRs and similar systems, the gathering of research data for clinical studies was typically a manual task. Physicians who cared for patients enrolled in trials, or their research assistants, would be asked to fill out special datasheets for later transcription into computer databases. Alternatively, data managers were often hired to abstract the relevant data from the paper chart. The trials were generally designed to define data elements that were required and the methods for analysis, but it was common for the process of collecting those data in a structured format to be left to manual processes at the point of patient care

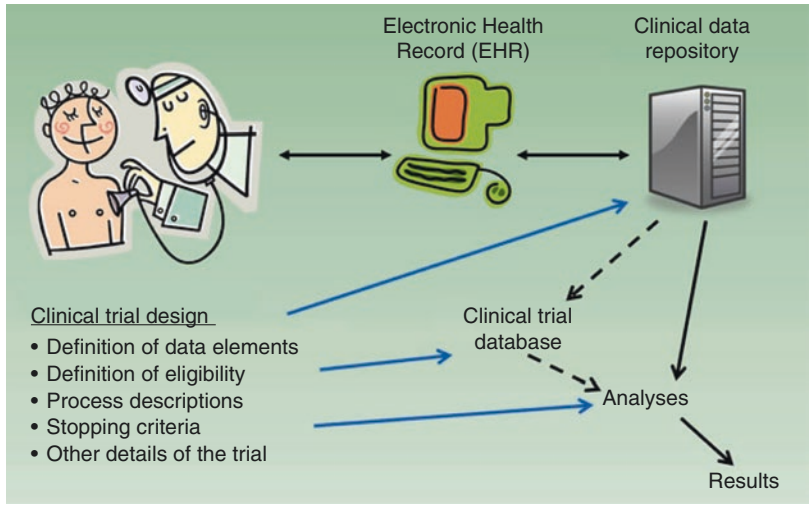


Fig. 1.5 Role of electronic health records (EHRs) in supporting clinical trials. With the introduction of EHR systems, the collection of much of the research data for clinical trials can become a by-product of the routine care of the patients. Research data may be analyzed directly from the clinical data repository, or a secondary research database may be created by downloading information from the online patient records. The manual processes in Fig. 1.4 are thereby largely eliminated. In addition, the interaction of the physician with the EHR permits two-way communication, which can greatly improve the quality and efficiency of the clinical trial. Physicians can be reminded when their patients are eligible for an experimental protocol, and the computer system can also remind the clinicians of the rules that are defined by the research protocol, thereby increasing compliance with the experimental plan

dards in the area of clinical terminology; (2) concerns regarding data privacy, confidentiality, and security; (3) challenges in data entry by physicians; (4) difficulties associated with the integration of record systems with other information resources in the health care setting, and (5) designing and delivering systems that are efficient, acceptable to clinicians, and intuitive to use. The first of these issues is discussed in detail in ▶ Chap. 7, and privacy is one of the central topics in ▶ Chap. 12. Issues of direct data entry by clinicians are discussed in ▶ Chaps. 2 and 14 and throughout many other chapters as well. ▶ Chapter 15 examines the fourth topic, focusing on recent trends in networked data integration, and offers solutions for the ways in which the EHR can be better joined with other relevant information resources and clinical processes, especially within communities where patients may have records with multiple providers and health care systems (Yasnoff et al. 2013). Finally, issues of the interface between computers and clinicians (or other users), with a cognitive emphasis, are the subject of ▶ Chap. 5.

1.1.3 Anticipating the Future of Electronic Health Records

One of the first instincts of software developers is to create an electronic version of an object or process from the physical world. Some familiar notion provides the inspiration for a new software product. Once the software version has been developed, however, human ingenuity and creativity often lead to an evolution that extends the software version far beyond what was initially contemplated. The computer can thus facilitate paradigm shifts in how we think about such familiar concepts.

Consider, for example, the remarkable difference between today's office automation software and the typewriter, which was the original inspiration for the development of "word processors". Although the early word processors were designed largely to allow users to avoid retyping papers each time a minor change was made to a document, the document-management software of today bears little resemblance to a typewriter. Consider all the powerful desktop-publishing facilities, integration of figures, spelling correction, grammar aids, "publishing" online, collaboration on individual documents by multiple users, etc. Similarly, today's spreadsheet programs bear little resemblance to the tables of numbers that we once created on graph paper. To take an example from the financial world, consider automatic teller machines (ATMs) and their facilitation of today's worldwide banking in ways that were never contemplated when the industry depended on human bank tellers.

It is accordingly logical to ask what the health record will become after it has been effectively implemented on computer systems and new opportunities for its enhancement become increasingly clear to us. It is clear that EHRs a decade from now will be remarkably different from the antiquated paper folders that used to dominate our health care environments. We might similarly predict that the state of today's EHR is roughly comparable to the status of commercial aviation in the 1930s. By that time air travel had progressed substantially from the days of the Wright Brothers, and air travel was becoming common. But 1930s air travel seems archaic by modern standards, and it is logical to assume that today's EHRs, albeit much better than both paper records and the early computer-based systems of the 1960s and 1970s, will be greatly improved and further modernized in the decades ahead.

If people had failed to use the early airplanes for travel, the quality and efficiency of airplanes and air travel would not have improved as they have. A similar point can be made about the importance of committing to the use of EHRs today, even though we know that they need to be much better in the future. We must also commit to assuring that those improvements are made, which suggests a dynamic interaction and interdependency among the researchers who address limitations in EHRs and their underlying methods and philosophy, the EHR compa-

nies that currently exist or will arise in the future, and the users who identify requirements and areas for improvement. These companies must look to creative researchers, both within their own companies and in academia, who will forge the changes that will encourage EHR users to embrace and appreciate the technology much more than they often do today.

1.2 Communications Technology and Health Data Integration

An obvious opportunity for changing the role and functionality of clinical-care records in the digital age is the power and ubiquity of the Internet. The Internet began in 1968 as a U.S. research activity funded by the Advanced Research Projects Agency (ARPA) of the Department of Defense. Initially known as the **ARPANET**, the network began as a novel mechanism for allowing a handful of defense-related mainframe computers, located mostly at academic institutions or in the research facilities of military contractors, to share data files with each other and to provide remote access to computing power at other locations. The notion of electronic mail arose soon thereafter, and machine-to-machine electronic mail exchanges quickly became a major component of the network's traffic. As the technology matured, its value for nonmilitary research activities was recognized, and by 1973 the first medically related research computer had been added to the network (Shortliffe 1998a, 2000).

During the 1980s, the technology began to be developed in other parts of the world, and the National Science Foundation took over the task of running the principal high-speed **backbone network** in the United States. Hospitals, mostly academic centers, began to be connected to what had by then become known as the Internet, and in a major policy move it was decided to allow commercial organizations to join the network as well. By April 1995, the Internet in the United States had become a fully commercialized operation, no longer depending on the U.S. government to support even the major backbone connections. Today, the Internet is ubiquitous, worldwide, accessible through mobile wireless devices, and has provided the invisible but mandatory infrastructure for social, political, financial, scientific, corporate, and entertainment ventures. Many people point to the Internet as a superb example of the facilitating role of federal investment in promoting innovative technologies. The Internet is a major societal force that arguably would never have been created if the research and development, plus the coordinating activities, had been left to the private sector.

The explosive growth of the Internet did not occur until the late 1990s, when the **World Wide Web** (which had been conceived initially by the physics community as a way of using the Internet to share preprints with photographs and diagrams among researchers) was introduced and popularized. Navigating the Web is highly intuitive, requires no special training, and provides a mechanism for access to multimedia information that accounts for its remarkable growth as a worldwide phenomenon. It is also accessible by essentially all digital devices—computers, tablets, smart phones, and a plethora of personal monitors and "smart home" tools—which is a tribute to its design and its compatibility with newer networking technologies, such as **Bluetooth** and **Wi-Fi**.

The societal impact of this communications phenomenon cannot be overstated, especially given the international connectivity that has grown phenomenally in the past two decades. Countries that once were isolated from information that was important to citizens, ranging from consumers to scientists to those interested in political issues, are now finding new options for bringing timely information to the desktop machines and mobile devices of individuals with an Internet connection.

There has in turn been a major upheaval in the telecommunications industry, with companies that used to be in different busi-

nesses (e.g., cable television, Internet services, and telephone) now finding that their activities and technologies have merged. In the United States, legislation was passed in 1996 to allow new competition to develop and new industries to emerge. We have subsequently seen the merging of technologies such as cable television, telephone, networking, and satellite communications. High-speed lines into homes and offices are widely available, wireless networking is ubiquitous, and inexpensive mechanisms for connecting to the Internet without using conventional computers (e.g., using cell phones or **set-top boxes**) have also emerged. The impact on everyone has been great and hence it is affecting the way that individuals seek health-related information while also enhancing how patients can gain access to their health care providers and to their clinical data.

The Internet also has exhibited unintended consequences, especially in the world of social media, which has created opportunities for promoting political unrest, social shaming, and dissemination of falsehoods. In the world of health care, the Internet has created opportunities for attacks on personal privacy, even while facilitating socially valuable exchanges of data among institutions and individuals. Many of these practical, legal, and ethical challenges are the subject of ▶ Chap. 12.

Just as individual hospitals and health care systems have come to appreciate the importance of integrating information from multiple clinical and administrative systems within their organizations (see ▶ Chap. 16), health planners and governments now appreciate the need to develop integrated information resources that combine clinical and health data from multiple institutions within regions, and ultimately nationally (see ▶ Chaps. 15 and 18). As you will see, the Internet and the role of digital communications has therefore become a major part of modern medicine and health. Although this topic recurs in essentially every chapter in this book, we introduce it in the following sections because of its importance to modern technical issues and policy directions.

1.2.1 A Model of Integrated Disease Surveillance[3]

To emphasize the role that the nation's networking infrastructure is playing in integrating clinical data and enhancing care delivery, consider one example of how disease surveillance, prevention, and care are increasingly being influenced by information and communications technology. The goal is to create an information-management infrastructure that will allow all clinicians, regardless of practice setting (hospitals, emergency rooms, small offices, community clinics, military bases, multispecialty groups, etc.) to use EHRs in their practices both to assist in patient care and to provide patients with counsel on illness prevention. The full impact of this use of electronic resources will occur when data from all such records are pooled in regional and national **registries** or surveillance databases (◘ Fig. 1.6), mediated through secure connectivity with the Internet. The challenge, of course, is to find a way to integrate data from such diverse practice settings, especially since there are multiple vendors and system developers active in the marketplace, competing to provide value-added capabilities that will excite and attract the practitioners for whom their EHR product is intended.

The practical need to pool and integrate clinical data from such diverse resources and systems emphasizes the practical issues that need to be addressed in achieving such functionality and resources. Interestingly, most of the barriers are logistical, political, and financial rather than technical in nature:

- *Encryption of data*: Concerns regarding privacy and data protection require that Internet transmission of clinical information occur only if those data are encrypted, with an established mechanism for identifying and authenticating individuals before they are allowed to decrypt the information for surveillance or research use.

3 This section is adapted from a discussion that originally appeared in (Shortliffe and Sondik 2004).

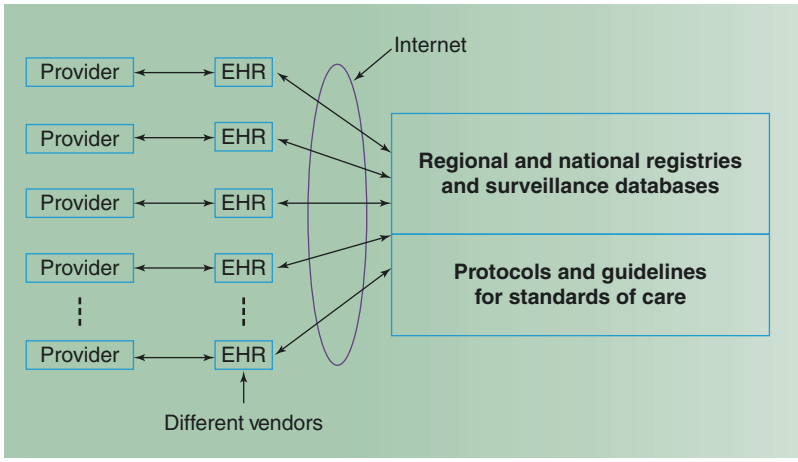


Fig. 1.6 A future vision of surveillance databases, in which clinical data are pooled in regional and national registries or repositories through a process of data submission that occurs over the Internet (with attention to privacy and security concerns as discussed in the text). When information is effectively gathered, pooled, and analyzed, there are significant opportunities for feeding back the results of derived insights to practitioners at the point of care. Thus the arrows indicate a bi-directional process. See also ▶ Chap. 15

- *Protection of stored clinical data*: Even when data are stored within an institution, there are opportunities for attack over the Internet, which can be an affront to patient privacy or, equally seriously, an opportunity for installing **malware** within an institution, resulting in rogue uses of data or even a lockout of valid users from crucially important functions or data. **Cybersecurity** has accordingly become a major topic of concern for health care institutions and other practice settings.[4]
- *HIPAA-compliant policies*: The privacy and security rules that resulted from the 1996 **Health Insurance Portability and Accountability Act (HIPAA)** do not prohibit the pooling and use of such data, but they do lay down policy rules and technical security practices that must be part of the solution in achieving the vision we are discussing here.
- *Standards for data transmission and sharing*: Sharing data over networks requires that all developers of EHRs and clinical databases adopt a single set of standards for communicating and exchanging information. The major enabling standard for such sharing, Health Level 7 (HL7), was introduced decades ago and, after years of work, has been uniformly adopted, implemented, and utilized. However, a uniform "envelope" for digital communication, such as HL7, does not assure that the contents of such messages will be understood or standardized. The pooling and integration of data requires the adoption of standards for clinical terminology and potentially for the schemas used to store clinical information in databases. Thus true interoperability of such systems requires additional standards to be adopted, many of which are discussed in ▶ Chap. 7.
- *Quality control and error checking*: Any system for accumulating, analyzing, and utilizing clinical data from diverse sources must be complemented by a rigorous approach to quality control and error checking. It is crucial that users have faith in the accuracy and comprehensiveness of the data that are collected in such repositories, because policies, guidelines, and a variety of metrics can be derived over time from such information.

4 ▶ https://www.theverge.com/2019/4/4/18293817/cybersecurity-hospitals-health-care-scan-simulation (Accessed 5/29/19).

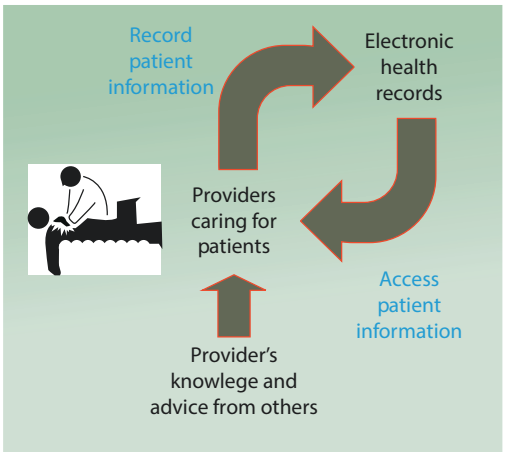
- *Regional and national registries and surveillance databases*: Any adoption of the model in ◘ Fig. 1.6 will require mechanisms for creating, funding, and maintaining the regional and national databases or registries that are involved (see ▶ Chap. 15). The growing amount of data that can be gathered in this way are naturally viewed as part of the "big data" problem that has characterized modern data science. The role of state and federal governments in gathering and curating such databases will need to be clarified, and the political issues addressed (including the concerns of some members of the populace that any government role in managing or analyzing their health data may have societal repercussions that threaten individual liberties, employability, and the like).

With the establishment of registries and surveillance databases, and a robust system of Internet integration with EHRs, summary information can flow back to providers to enhance their decision making at the point of care (◘ Fig. 1.6). This assumes standards that allow such information to be integrated into the vendor-supplied products that the clinicians use in their practice settings. These may be EHRs or their order-entry components that clinicians use to specify the actions that they want to have taken for the treatment or management of their patients (see ▶ Chaps. 14 and 16). Furthermore, as is shown in ◘ Fig. 1.6, the databases can help to support the creation of evidence-based guidelines, or clinical research protocols, which can be delivered to practitioners through the feedback process. Thus one should envision a day when clinicians, at the point of care, will receive integrated, non-dogmatic, supportive information regarding:

- Recommended steps for health promotion and disease prevention
- Detection of syndromes or problems, either in their community or more widely
- Trends and patterns of public health importance, a capability emphasized by the need for rapidly changing data on cases and deaths during the **COVID-19** pandemic in 2020.
- Clinical guidelines, adapted for execution and integration into patient-specific decision support rather than simply provided as text documents
- Opportunities for distributed (community-based) clinical research, whereby patients are enrolled in clinical trials and protocol guidelines are in turn integrated with the clinicians' EHR to support protocol-compliant management of enrolled patients

1.2.2 The Goal: A Learning Health System

We have been stressing the cyclical role of information—its capture, organization, interpretation, and ultimate use. You can easily understand the small cycle that is implied: patient-specific data and plans entered into an EHR and subsequently made available to the same practitioner or others who are involved in that patient's care (◘ Fig. 1.7). Although this view is a powerful contributor to improved data management in the care of patients, it fails to include a larger view of the societal value of the information that is contained in clinical-care records. In fact, such straightforward use of EHRs for direct patient care would not have met some of the



◘ **Fig. 1.7** There is a limited view of the role of EHRs that sees them as intended largely to support the ongoing care of the patient whose clinical data are stored in the record

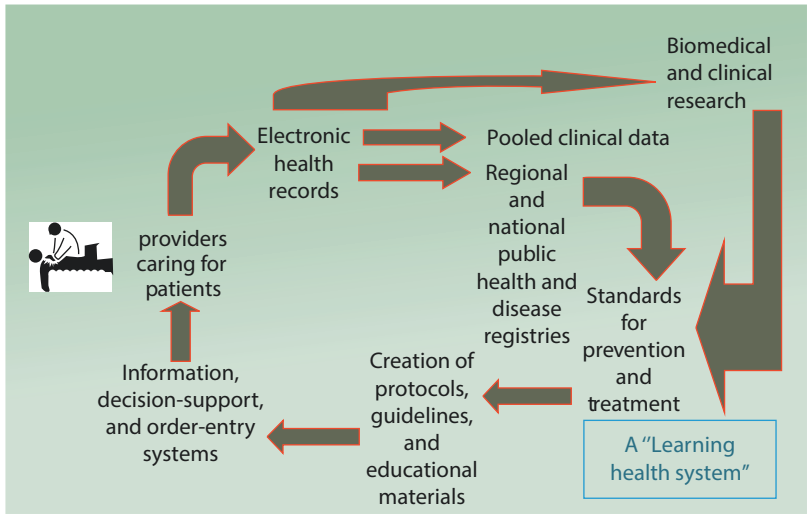


Fig. 1.8 The ultimate goal is to create a cycle of information flow, whereby data from local distributed electronic health records (EHRs) and their associated clinical datasets are routinely and effortlessly submitted to registries and research databases. The resulting new knowledge then can feed back to practitioners at the point of care, using a variety of computer-supported decision-support delivery mechanisms. This cycle of new knowledge, driven by experience, and fed back to clinicians, has been dubbed a "learning health system"

requirements that the US government specified after 2009 when determining eligibility for payment of incentives to clinicians or hospitals who implemented EHRs (see the discussion of the government HITECH program in ▶ Sect. 1.3).

Consider, instead, an expanded view of the health surveillance model introduced in ▶ Sect. 1.2.1 (Fig. 1.8). Beginning at the left of the diagram, clinicians caring for patients use electronic health records, both to record their observations and to gain access to information about the patient. Information from these records is then stored in local patient-care clinical databases and forwarded automatically to regional and national registries as well as to research databases that can support retrospective studies (see ▶ Chap. 15) or formal institutional or community-based clinical trials (see ▶ Chap. 27). The analyzed information from institutional datasets, registries and research studies can in turn be used to develop standards for prevention and treatment, with major guidance from biomedical research. Researchers can draw information either directly from the health records or from the pooled data in registries. The standards for treatment in turn can be translated into protocols, guidelines, and educational materials. This new knowledge and decision-support functionality can then be delivered over the network back to the clinicians so that the information informs patient care, where it is integrated seamlessly with EHRs and order-entry systems.

This notion of a system that allows us to learn from what we do, unlocking the experience that has traditionally been stored in unusable form in paper charts, is gaining wide attention now that we can envision an interconnected community of clinicians and institutions, building digital data resources using EHRs. The concept has been dubbed a **learning health system** and is an ongoing subject of study by the National Academy of Medicine (Daley 2013), which has published a series of reports on the topic.[5] It is also the organizing conceptual framework for a

5 ▶ https://nam.edu/programs/value-science-driven-health-care/learning-health-system-series/ (Accessed 05/29/19)

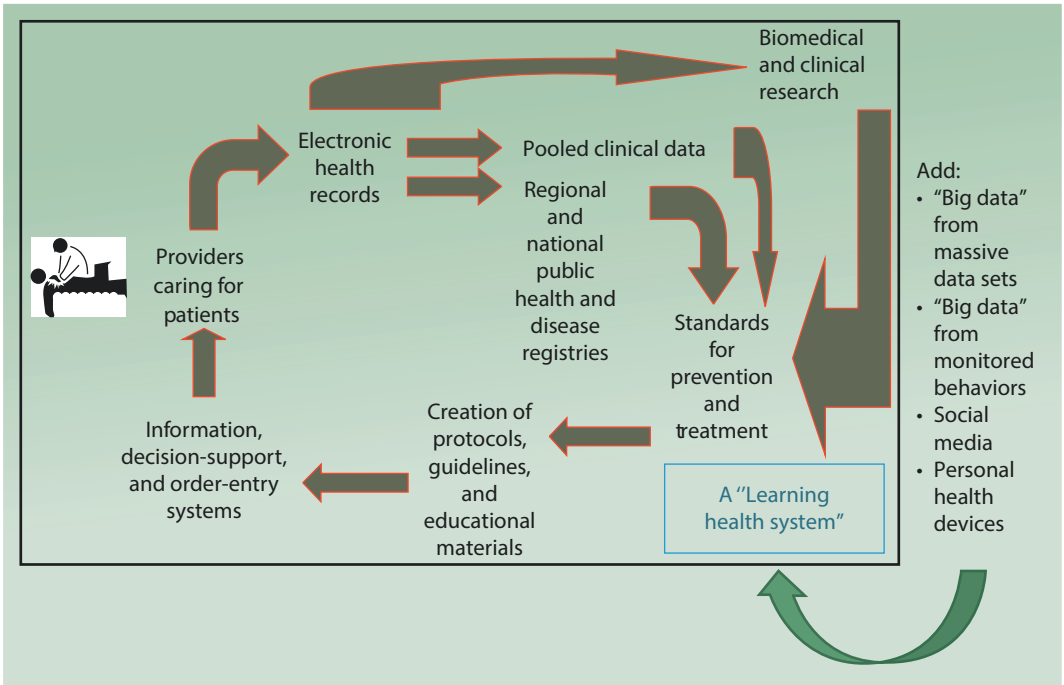


Fig. 1.9 Today the *learning health system* is increasingly embracing new forms of massive health-related data, often from outside the clinical care setting and derived from population activities that reflect individuals' health, activities, and attitudes

recently created department at the University of Michigan Medical School[6] and for a new scientific journal.[7]

Although the learning health system concept of Fig. 1.8 may at first seem expansive and all-inclusive, in recent years we have learned that there are other important inputs to the health care environment and these can have important implications for what we learn by analyzing what both patients and healthy individuals do. Some of these data sources are immense and are in line with the recent interest in "big data" analytics (Fig. 1.9). Consider, for example, the analysis of huge datasets associated with full human genome specifications for individuals and populations. Another approach for gathering massive amounts of relevant health-related data is to monitor the behavior of individuals as they use online information resources, searching for health-related information. Social media exchanges (e.g., Twitter, Facebook) have also been used to extract health-related information, such as complaints that suggest early stages of communicable diseases or expressed attitudes towards diseases and treatment. The explosive adoption of health monitoring devices (e.g., step counters, exercise analyzers, cardiac or sleep monitors) has also offered a useful source of large-scale information that is only beginning to be merged with other data in our learning health system.

1.2.3 Implications of the Internet for Patients

With the penetration of the Internet, patients, as well as healthy individuals, have turned to the Internet for health information. It is a rare North American physician who has not

6 ► https://medicine.umich.edu/dept/learning-health-sciences (Accessed 05/03/2020)

7 ► https://onlinelibrary.wiley.com/journal/23796146 (Accessed 05/03/2020)

encountered a patient who comes to an appointment armed with a question, or a stack of printouts, that arose due to medically related searches on the net. The companies that provide search engines for the Internet report that health-related sites are among the most popular ones being explored by consumers. As a result, physicians and other care providers have learned that they must be prepared to deal with information that patients discover on the net and bring with them when they seek care from clinicians. Some of the information is timely and excellent; in this sense, physicians can often learn about innovations from their patients and need to be open to the kinds of questions that this enhanced access to information will generate from patients in their practices.

On the other hand, much of the health information on the Web lacks peer review or is purely anecdotal. People who lack medical training can be misled by such information, just as they have been poorly served in the past by printed information in books and magazines dealing with fad treatments from anecdotal sources. This also creates challenges for health care providers, who often feel pressured to handle more issues in less time due to economic pressures. In addition, some sites provide personalized advice, sometimes for a fee, with all the attendant concerns about the quality of the suggestions and the ability to give valid advice based on an electronic-mail or Web-based interaction.

In a positive light, communications technologies offer clinicians creative ways to interact with their patients and to provide higher quality care. Years ago, medicine adopted the telephone as a standard vehicle for facilitating patient care, and we now take this kind of interaction with patients for granted. If we extend the audio channel to include our visual sense as well, typically relying on the Internet as our communication mechanism, the notion of **telemedicine** emerges (see ▶ Chap. 20). This notion of "medicine at a distance" arose early in the twentieth century (see ◘ Fig. 1.10), but the technology was too limited for much penetration of the idea beyond telephone conversations until the last 30–40 years. The use of telemedicine has subsequently grown rapidly, and there are settings in which it is already proving to be successful and cost-effective (e.g., rural care, international medicine, **teleradiology**, and video-based care of patients in prisons). Similarly, there are now a large number of **apps** (designed for smart phones, tablets, or desktop machines) that offer specialized medical care or advice or assist with health data management and communication with providers and support groups (see ▶ Chaps. 11 and 20).

1.2.4 Requirements for Achieving the Vision

Efforts that continue to push the state of the art in Internet technology all have significant implications for the future of health care delivery in general and of EHRs and their integration in particular (Shortliffe 1998b, 2000). But in addition to increasing speed, reliability, security, and availability of the Internet, there are many other areas that need attention if the vision of a learning health system is to be achieved.

1.2.4.1 Education and Training

There is a difference between computer literacy (familiarity with computers and their routine uses in our society) and knowledge of the role that computing and communications technology can and should play in our health system. We need to do a better job of training future clinicians in the latter area. Otherwise we will leave them poorly equipped for the challenges and opportunities they will face in the rapidly changing practice environments that surround them (Shortliffe 2010). Not only do they need to feel comfortable with the technology itself, but they need to understand the profound effect that it has had on the practice of medicine—with many more changes to come. Medicine, and other health professions, are being asked to adapt in ways that were not envisioned even a decade or two ago. Not all individuals embrace such change, but younger clinicians, who have grown up with technology in almost all aspects of their lives, have high

Fig. 1.10 "The Radio Doctor": long before television was invented, creative observers were suggesting how doctors and patients could communicate using advanced technologies. This 1924 example is from the cover of a popular magazine and envisions video enhancements to radio. (Source: "Radio News" 1924)

expectations for how digital systems and tools should enhance their professional experience. What is even more challenging, perhaps, is that assumptions that they have made about the field they have entered may no longer be valid in the coming years, as some skills are no longer required and new requirements are viewed as dramatically different from what health professionals have had to know in the past.

Furthermore, in addition to the implications for education of health professionals about computer-related topics, much of the future vision we have proposed here can be achieved only if educational institutions produce a cadre of talented individuals who are highly skilled in computing and communications technology but also have a deep understanding of the biomedical milieu and of the needs of practitioners and other health work-

ers. Computer science training alone is not adequate. Fortunately, there are increasing numbers of formal training programs in what has become known as **biomedical informatics** (see ► Sect. 1.4) that provide custom-tailored educational opportunities. Many of the trainees are life science researchers, physicians, nurses, pharmacists, and other health professionals who see the career opportunities and challenges at the intersections of biomedicine, information science, computer science, decision science, data science, cognitive science, and communications technologies. As has been clear for three decades (Greenes and Shortliffe 1990), however, the demand for such individuals far outstrips the supply, both for academic and industrial career pathways.[8, 9] We need more training programs,[10] expansion of those that already exist, plus support for junior faculty in health science schools who may wish to pursue additional training in this area.

1.2.4.2 Organizational and Management Change

Second, as implied above, there needs to be a greater understanding among health care leaders regarding the role of specialized multidisciplinary expertise in successful clinical systems implementation. The health care system provides some of the most complex organizational structures in society (Begun et al. 2003), and it is simplistic to assume that off-the-shelf products will be smoothly introduced into a new institution without major analysis, redesign, and cooperative joint-development efforts. Underinvestment and a failure to understand the requirements for process reengineering as part of software implementation, as well as problems with technical leadership and planning, account for many of the frustrating experiences that health care organizations report in their efforts to use computers more effectively in support of patient care and provider productivity.

The notion of a learning health system described previously is meant to motivate your enthusiasm for what lies ahead and to suggest the topics that need to be addressed in a book such as this one. Essentially all of the following chapters touch on some aspect of the vision of integrated systems that extend beyond single institutions. Before embarking on these topics, however, we must emphasize two points. First, the cyclical creation of new knowledge in a learning health care system will become reality only if individual hospitals, academic medical centers, and national coordinating bodies work together to provide the standards, infrastructure, and resources that are necessary. No individual system developer, vendor, or administrator can mandate the standards for connectivity, data pooling, and data sharing implied by a learning health care system. A national initiative of cooperative planning and implementation for computing and communications resources within and among institutions and clinics is required before practitioners will have routine access to the information that they need (see ► Chap. 15). A major federal incentive program for EHR implementation was a first step in this direction (see ► Sect. 1.3). The criteria that are required for successful EHR implementation are sensitive to the need for data integration, public-health support, and a learning health system.

Second, although our presentation of the learning health system notion has focused on the clinician's view of integrated information access, other workers in the field have similar needs that can be addressed in similar ways. The academic research community has already developed and made use of much of the technology that needs to be coalesced if the clinical user is to have similar access to data and information. There is also the

8 ► https://www.hcinnovationgroup.com/policy-value-based-care/staffing-professional-development/news/13024360/report-health-informatics-labor-market-lags-behind-demand-for-workers (Accessed 5/30/2019); ► https://www.bestvalueschools.com/faq/job-outlook-health-informatics-graduates/ (Accessed 5/30/2019).

9 ► https://www.burning-glass.com/wp-content/uploads/BG-Health_Informatics_2014.pdf (Accessed 5/30/2019).

10 A directory of some existing training programs is available at ► http://www.amia.org/education/programs-and-courses (Accessed 5/30/19).

patient's view, which must be considered in the notion of patient-centered health care that is now broadly accepted and encouraged (Ozkaynak et al. 2013).

1.3 The US Government Steps In

During the early decades of the evolution of clinical information systems for use in hospitals, patient care, and public health, the major role of government was in supporting the research enterprise as new methods were developed, tested, and formally evaluated. The topic was seldom mentioned by the nation's leaders, however, even during the 1990s when the White House was viewed as being especially tech savvy. It was accordingly remarkable when, in the President's State of the Union address in 2004 (and in each of the following years of his administration), President Bush called for universal implementation of electronic health records within 10 years. The Secretary of Health and Human Services, Tommy Thompson, was similarly supportive and, in May 2004, created an entity intended to support the expansion of the use of EHRs—the **Office of the National Coordinator for Health Information Technology** (initially referred to by the full acronym ONCHIT, but later shortened simply to ONC).

There was initially limited budget for ONC, although the organization served as a convening body for EHR-related planning efforts and the National Health Information Infrastructure (see ▶ Chaps. 14, 15 and 29). The topic of EHRs subsequently became a talking point for both major candidates during the Presidential election in 2008, with strong bipartisan support. Then, in early 2009, Congress enacted the American Recovery and Reinvestment Act (ARRA), also known as the economic "Stimulus Bill". One portion of that legislation was known as the **Health Information Technology for Economic and Clinical Health (HITECH) Act**. It was this portion of the bill that provided significant fiscal incentives for health systems, hospitals, and providers to implement EHRs in their practices and eventual financial penalties for lack of implementation. Such payments were made available, however, only when eligible organizations or individual practitioners implemented EHRs that were "certified" as meeting minimal standards and when they could document that they were making "meaningful use" of those systems. You will see references to such certification and **meaningful use** criteria in other chapters in this volume.

This volume also offers a discussion of HIT policy and the federal government in ▶ Chap. 29. Although the process of EHR implementation is approaching completion in the US, both in health systems and practices, the current status is largely due to this legislative program: because of the federal stimulus package, large numbers of hospitals, systems, and practitioners invested in EHRs and incorporated them into their practices. Furthermore, the demand for workers skilled in health information technology grew much more rapidly than did the general job market, even within health care (◘ Fig. 1.11). It is a remarkable example of how government policy and investment can stimulate major transitions in systems such as health care, where many observers had previously felt that progress had been unacceptably slow (Shortliffe 2005).

1.4 Defining Biomedical Informatics and Related Disciplines

With the previous sections of this chapter as background, let us now consider the scientific discipline that is the subject of this volume and has led to the development of many of the functionalities that need to be brought together in the integrated biomedical-computing environment of the future. The remainder of this chapter deals with biomedical informatics as a field and with biomedical and health information as a subject of study. It provides additional background needed to understand many of the subsequent chapters in this book.

Reference to the use of computers in biomedicine evokes different images depending on the nature of one's involvement in the field. To a hospital administrator, it might suggest

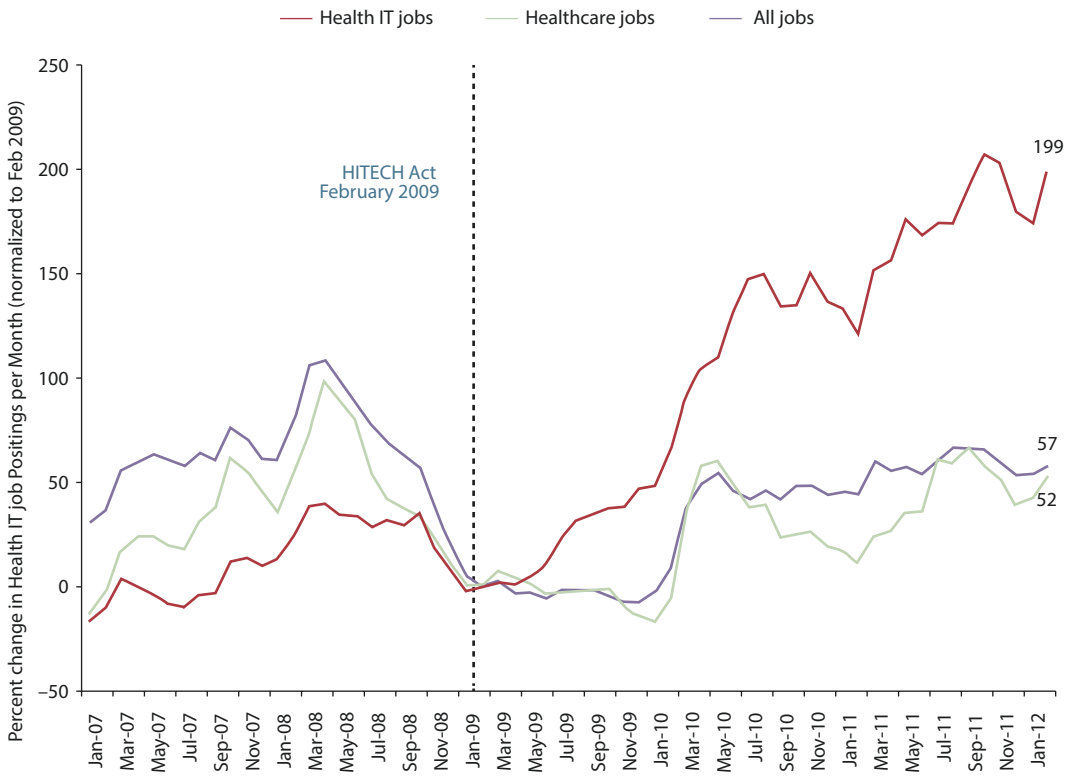


Fig. 1.11 Impact of the HITECH Act on health information technology (IT) employment. Percent change in online health IT job postings per month for first 3 years, relative to health care jobs and all jobs: normalized to February 2009 when ARRA passed. (Source: ONC analysis of data from O'Reilly Job Data Mart, ONC Data Brief, No. 2, May 2012 (▶ https://www.healthit.gov/sites/default/files/pdf/0512_ONCDataBrief2_JobPostings.pdf (Accessed 5/6/2019)))

the maintenance of clinical-care records using computers; to a decision scientist, it might mean the assistance by computers in disease diagnosis; to a basic scientist, it might mean the use of computers for maintaining, retrieving, and analyzing gene-sequencing information. Many physicians immediately think of office-practice tools for tasks such as patient billing or appointment scheduling, and of electronic health record systems for clinical documentation. Nurses often think of computer-based tools for charting the care that they deliver, or decision-support tools that assist in applying the most current patient-care guidelines. The field includes study of all these activities and a great many others too. More importantly, it includes the consideration of various external factors that affect the biomedical setting. Unless you keep in mind these surrounding factors, it may be difficult to understand how biomedical computing can help us to tie together the diverse aspects of health care and its delivery.

To achieve a unified perspective, we might consider four related topics: (1) the concept of biomedical information (why it is important in biological research and clinical practice and why we might want to use computers to process it); (2) the structural features of medicine, including all those subtopics to which computers might be applied; (3) the importance of evidence-based knowledge of biomedical and health topics, including its derivation and proper management and use; and (4) the applications of computers and communication methods in biomedicine and the scientific issues that underlie such efforts. We mention the first two topics briefly in this and the next chapter, and we provide references in the Suggested Readings section for readers who

wish to learn more. The third topic, knowledge to support effective decision making in support of human health, is intrinsic to this book and occurs in various forms in essentially every chapter. The fourth topic, however, is the principal subject of this book.

Computers have captured the imagination (and attention) of our society. Today's younger individuals have grown up in a world in which computers are ubiquitous and useful. Because the computer as a machine is exciting, people may pay a disproportionate amount of attention to it as such—at the expense of considering what the computer can do given the numbers, concepts, ideas, and cognitive underpinnings of fields such as medicine, health, and biomedical research. Computer scientists, philosophers, psychologists, and other scholars increasingly consider such matters as the nature of information and knowledge and how human beings process such concepts. These investigations have been given a sense of timeliness (if not urgency) by the simple existence of the computer. The cognitive activities of clinicians in practice probably have received more attention over the past three or four decades than in all previous history (see ► Chap. 4). Again, the existence of the computer and the possibilities of its extending a clinician's cognitive powers have motivated many of these studies. To develop computer-based tools to assist with decisions, we must understand more clearly such human processes as diagnosis, therapy planning, decision making, and problem solving in medicine. We must also understand how personal and cultural beliefs affect the way in which information is interpreted and decisions are ultimately made.

1.4.1 Terminology

Although, starting in the 1960s, a growing number of individuals conducting serious biomedical research or undertaking clinical practice had access to a computer system, there was initial uncertainty about what name should be used for the biomedical application of computer science concepts. The name computer science was itself new in 1960 and was only vaguely defined. Even today, the term computer science is used more as a matter of convention than as an explanation of the field's scientific content.

In the 1970s we began to use the phrase **medical computer science** to refer to the subdivision of computer science that applies the methods of the larger field to medical topics. As you will see, however, medicine has provided a rich area for computer science research, and several basic computing insights and methodologies have been derived from applied medical-computing research.

The term **information science**, which is occasionally used in conjunction with computer science, originated in the field of library science and is used to refer, somewhat generally, to the broad range of issues related to the management of both paper-based and electronically stored information. Much of what information science originally set out to be is now drawing evolving interest under the name **cognitive science**.

Information theory, in contrast, was first developed by scientists concerned about the physics of communication; it has evolved into what may be viewed as a branch of mathematics. The results scientists have obtained with information theory have illuminated many processes in communications technology, but they have had little effect on our understanding of human information processing.

The terms **biomedical computing** or **biocomputation** have been used for a number of years. They are non-descriptive and neutral, implying only that computers are employed for some purpose in biology or medicine. They are often associated with bioengineering applications of computers, however, in which the devices are viewed more as tools for a bioengineering application than as a primary focus of research.

In the 1970s, inspired by the French term for computer science (*informatique*), the English-speaking community began to use the term **medical informatics**. Those in the field were attracted by the word's emphasis on *information*, which they saw as more central to the field than the computer itself, and it gained momentum as a term for the discipline, espe-

cially in Europe, during the 1980s. The term is broader than **medical computing** (it includes such topics as medical statistics, record keeping, and the study of the nature of medical information itself) and deemphasizes the computer while focusing instead on the nature of the field to which computations are applied. Because the term *informatics* became widely accepted in the United States only in the late 1980s, **medical information science** was also used earlier in North America; this term, however, may be confused with library science, and it does not capture the broader implications of the European term. As a result, the name *medical informatics* appeared by the late 1980s to have become the preferred term, even in the United States. Indeed, this is the name of the field that we used in the first two editions of this textbook (published in 1990 and 2000), and it is still sometimes used in professional, industrial, and academic settings. However, many observers expressed concern that the adjective "medical" is too focused on physicians and disease, failing to appreciate the relevance of this discipline to other health and life-science professionals and to health promotion and disease prevention. Thus, the term **health informatics**, or health care informatics, gained some popularity, even though it has the disadvantage of tending to exclude applications to biomedical research (► Chaps. 9 and 26) and, as we shall argue shortly, it tends to focus the field's name on application domains (clinical care, public health, and prevention) rather than the basic discipline and its broad range of applicability.

Applications of informatics methods in biology and genetics exploded during the 1990s due to the human genome project[11] and the growing recognition that modern life-science research was no longer possible without computational support and analysis (see ► Chaps. 9 and 26). By the late 1990s, the use of informatics methods in such work had become widely known as **bioinformatics** and the director of the National Institutes of Health (NIH) appointed an advisory group called the Working Group on Biomedical Computing. In June 1999, the group provided a report[12] recommending that the NIH undertake an initiative called the **Biomedical Information Science and Technology Initiative (BISTI)**. With the subsequent creation of another NIH organization called the Bioinformatics Working Group, the visibility of informatics applications in biology was greatly enhanced. Today bioinformatics is a major area of activity at the NIH[13] and in many universities and biotechnology companies around the world. The explosive growth of this field, however, has added to the confusion regarding the naming conventions we have been discussing. In addition, the relationship between *medical informatics* and *bioinformatics* became unclear. As a result, in an effort to be more inclusive and to embrace the biological applications with which many medical informatics groups had already been involved, the name *medical informatics* gradually gave way to biomedical informatics (BMI). Several academic groups have changed their names, and a major medical informatics journal (*Computers and Biomedical Research,* first published in 1967) was reborn in 2001 as *The Journal of Biomedical Informatics*.[14]

Despite this convoluted naming history, we believe that the broad range of issues in biomedical information management does require an appropriate name and, beginning with the third edition of this book (2006), we used the term **biomedical informatics** for this purpose. It has become the most widely accepted term for the core discipline and should be viewed as encompassing broadly all areas of application in health, clinical practice, and biomedical research. When we speak specifically about computers and their use within biomedical informatics activities, we use the terms biomedical computer science (for the methodologic issues) or biomedical computing (to describe the activity itself).

11 ► https://www.ornl.gov/sci/techresources/Human_Genome/home.shtml (Accessed 5/31/2019).

12 Available at ► https://acd.od.nih.gov/documents/reports/060399_Biomed_Computing_WG_RPT.htm (Accessed 5/31/2019).

13 See ► http://www.bisti.nih.gov/. (Accessed 5/31/2019).

14 ► http://www.journals.elsevier.com/journal-of-biomedical-informatics (Accessed 5/30/19).

Note, however, that biomedical informatics has many other component sciences in addition to computer science. These include the decision sciences, statistics, cognitive science, data science, information science, and even management sciences. We return to this point shortly when we discuss the basic versus applied nature of the field when it is viewed as a basic research discipline.

Although labels such as these are arbitrary, they are by no means insignificant. In the case of new fields of endeavor or branches of science, they are important both in designating the field and in defining or restricting its contents. The most distinctive feature of the modern computer is the generality of its application. The nearly unlimited range of computer uses complicates the business of naming the field. As a result, the nature of computer science is perhaps better illustrated by examples than by attempts at formal definition. Much of this book presents examples that do just this for biomedical informatics as well.

The American Medical Informatics Association (AMIA), which was founded in the late 1980s under the former name for the discipline, has recognized the confusion regarding the field and its definition.[15] They accordingly appointed a working group to develop a formal definition of the field and to specify the core competencies that need to be acquired by students seeking graduate training in the discipline. The resulting definition, published in AMIA's journal and approved by the full board of the organization, identifies the focus of the field in a simple sentence and then adds four clarifying corollaries that refine the definition and the field's scope and content (► Box 1.1). We adopt this definition, which is very similar to the one we offered in previous editions of this text. It acknowledges that the emergence of biomedical informatics as a new discipline is due in large part to rapid advances in computing and communications technology, to an increasing awareness that the knowledge base of biomedicine is essentially unmanageable by traditional paper-based methods, and to a growing conviction that the process of informed decision making is as important to modern biomedicine as is the collection of facts on which clinical decisions or research plans are made.

Box 1.1: Definition of Biomedical Informatics

Biomedical informatics (*BMI*) *is the interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, driven by efforts to improve human health.*

Scope and breadth of discipline: BMI investigates and supports reasoning, modeling, simulation, experimentation, and translation across the spectrum from molecules to individuals and to populations, from biological to social systems, bridging basic and clinical research and practice and the health care enterprise.

Theory and methodology: BMI develops, studies, and applies theories, methods, and processes for the generation, storage, retrieval, use, management, and sharing of biomedical data, information, and knowledge.

Technological approach: BMI builds on and contributes to computer, telecommunication, and information sciences and technologies, emphasizing their application in biomedicine.

Human and social context: BMI, recognizing that people are the ultimate users of biomedical information, draws upon the social and behavioral sciences to inform the design and evaluation of technical solutions, policies, and the evolution of economic, ethical, social, educational, and organizational systems.

15 ► https://www.amia.org/about-amia/science-informatics (Accessed 5//27/19).

1

1.4.2 Historical Perspective

The modern digital computer grew out of developments in the United States and abroad during World War II, and general-purpose computers began to appear in the marketplace by the mid-1950s (Fig. 1.12). Speculation about what might be done with such machines (if they should ever become reliable) had, however, begun much earlier. Scholars, at least as far back as the Middle Ages, often had raised the question of whether human reasoning might be explained in terms of formal or **algorithmic processes**. Gottfried Wilhelm von Leibnitz, a seventeenth-century German philosopher and mathematician, tried to develop a calculus that could be used to simulate human reasoning. The notion of a "logic engine" was subsequently worked out by Charles Babbage in the mid nineteenth century.

The first practical application of automatic computing relevant to medicine was Herman Hollerith's development of a punched-card data-processing system for the 1890 U.S. census (Fig. 1.13). His methods were soon adapted to **epidemiologic** and public health surveys, initiating the era of electromechanical punched-card data-processing technology, which matured and was widely adopted during the 1920s and 1930s. These techniques were the precursors of the stored program and wholly electronic digital computers, which began to appear in the late 1940s (Collen 1995).

One early activity in biomedical computing was the attempt to construct systems that would assist a physician in decision making (see ▶ Chap. 24). Not all biomedical-computing programs pursued this course, however. Many of the early ones instead investigated the notion of a total **hospital information system** (HIS; see ▶ Chap. 16). These projects were perhaps less ambitious in that they were more concerned with practical applications in the short term; the difficulties they encountered, however, were still formidable. The earliest work on HISs in the United States was probably that associated with the MEDINET project at General Electric, followed by work at Bolt, Beranek, Newman in Cambridge, Massachusetts, and then at the Massachusetts General Hospital (MGH) in Boston. A number of hospital application programs were developed at MGH by Barnett and his associates over three decades beginning in the early 1960s. Work on similar systems was undertaken by Warner at Latter Day Saints (LDS) Hospital in Salt Lake City, Utah, by Collen at Kaiser Permanente in Oakland, California, by Wiederhold at

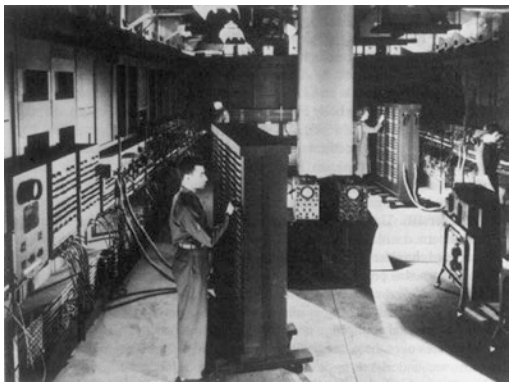


Fig. 1.12 The ENIAC. Early computers, such as the ENIAC, were the precursors of today's personal computers (PCs) and handheld calculators. (US Army photo. See also ▶ http://www.computersciencelab.com/ComputerHistory/HistoryPt4.htm (Accessed 5/31/2019))

Fig. 1.13 Tabulating machines. The Hollerith Tabulating Machine was an early data-processing system that performed automatic computation using punched cards. (Photograph courtesy of the Library of Congress)

Stanford University in Stanford, California, and by scientists at Lockheed in Sunnyvale, California.[16]

The course of HIS applications bifurcated in the 1970s. One approach was based on the concept of an integrated or monolithic design in which a single, large, *time-shared computer* would be used to support an entire collection of applications. An alternative was a distributed design that favored the separate implementation of specific applications on smaller individual computers—minicomputers—thereby permitting the independent evolution of systems in the respective application areas. A common assumption was the existence of a single shared database of patient information. The multimachine model was not practical, however, until network technologies permitted rapid and reliable communication among distributed and (sometimes) heterogeneous types of machines. Such distributed HISs began to appear in the 1980s (Simborg et al. 1983).

Biomedical-computing activity broadened in scope and accelerated with the appearance of the minicomputer in the early 1970s. These machines made it possible for individual departments or small organizational units to acquire their own dedicated computers and to develop their own application systems (Fig. 1.14). In tandem with the introduction of general-purpose software tools that provided standardized facilities to individuals with limited computer training (such as the UNIX operating system and programming environment), the minicomputer put more computing power in the hands of more biomedical investigators than did any other single development until the introduction of the microprocessor, a central processing unit (CPU) contained on one or a few chips (Fig. 1.15).

Everything changed radically in the late 1970s and early 1980s, when the microprocessor and the personal computer (PC) or microcomputer became available. Not only could hospital departments afford minicomputers but now individuals also could afford micro-

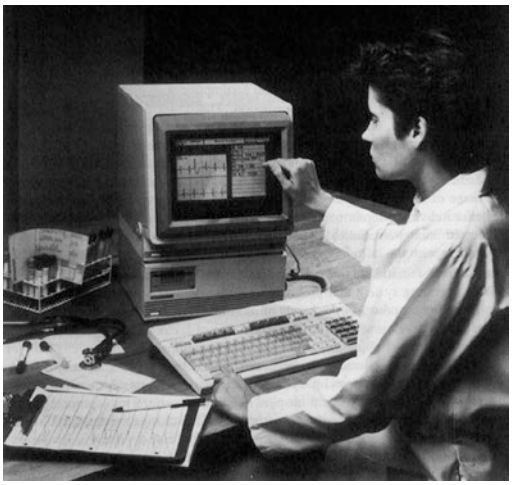


Fig. 1.14 Departmental system. Hospital departments, such as the clinical laboratory, were able to implement their own custom-tailored systems when affordable minicomputers became available. These departments subsequently used microcomputers to support administrative and clinical functions. (Copyright 2013 Hewlett-Packard Development Company, LP. Reproduced from ~1985 original with permission)

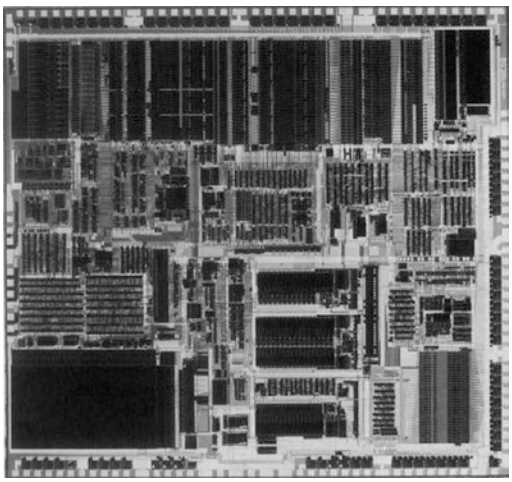


Fig. 1.15 Miniature computer. The microprocessor, or "computer on a chip," revolutionized the computer industry in the 1970s. By installing chips in small boxes and connecting them to a computer terminal, engineers produced the personal computer (PC)—an innovation that made it possible for individual users to purchase their own systems

16 The latter system was later taken over and further developed by the Technicon Corporation (subsequently TDS Healthcare Systems Corporation). Later the system was part of the suite of products available from Eclipsys, Inc. (which in turn was acquired by Allscripts, Inc in 2010).

computers. This change enormously broadened the base of computing in our society and gave rise to a new software industry. The first articles on computers in medicine had appeared in clinical journals in the late 1950s, but it was not until the late 1970s that the first use of computers in advertisements dealing with computers and aimed at physicians began to appear (Fig. 1.16). Within a few years, a wide range of computer-based information-management tools were available as commercial products; their descriptions began to appear in journals alongside the traditional advertisements for drugs and other medical products. Today individual physicians find it practical to employ PCs in a variety of settings, including for applications in patient care or clinical investigation.

Today we enjoy a wide range of hardware of various sizes, types, prices, and capabilities, all of which will continue to evolve in the decades ahead. The trend—reductions in size and cost of computers with simultaneous increases in power (Fig. 1.17)—shows no sign of slowing, although scientists foresee the

Fig. 1.16 Medical advertising. An early advertisement for a portable computer terminal that appeared in general medical journals in the late 1970s. The development of compact, inexpensive peripheral devices and personal computers (PCs) inspired future experiments in marketing directly to clinicians (Reprinted by permission of copyright holder Texas Instruments Incorporated © 1985)

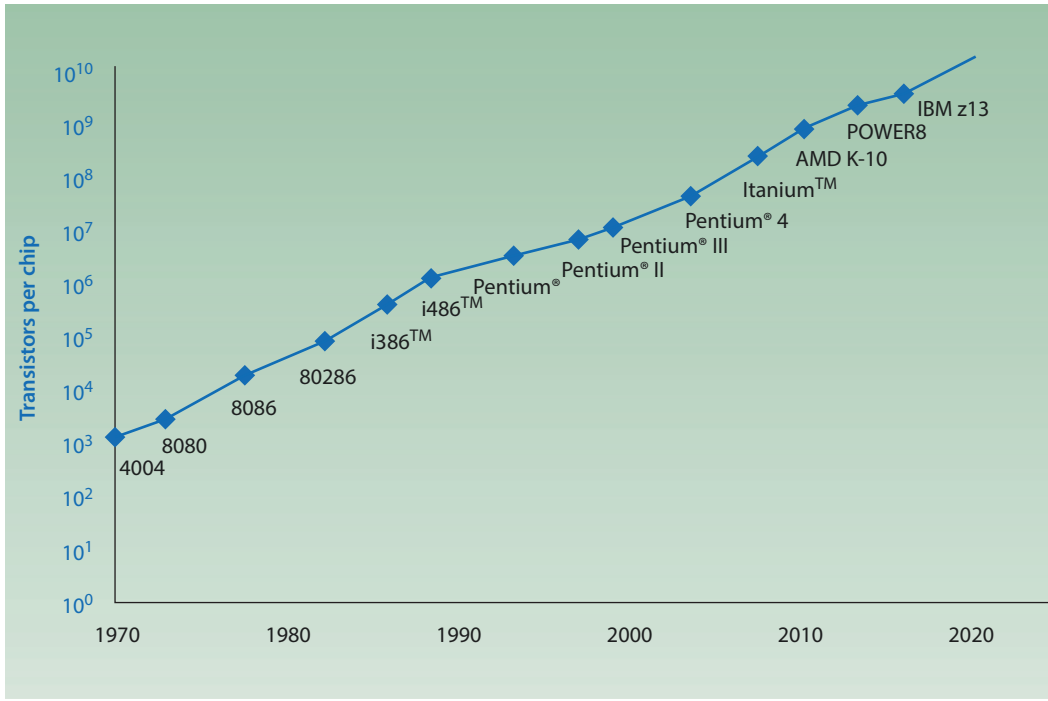


Fig. 1.17 Moore's Law. Former Intel chairman Gordon Moore is credited with popularizing the "law" that the size and cost of microprocessor chips will half every 18 months while they double in computing power. This graph shows the exponential growth in the number of transistors that can be integrated on a single microprocessor chip. The trend continues to this day. (Source: Wikipedia: ▶ https://en.wikipedia.org/wiki/Transistor_count)

Fig. 1.18 The National Library of Medicine (NLM). The NLM, on the campus of the National Institutes of Health (NIH) in Bethesda, Maryland, is the principal biomedical library for the nation (see ▶ Chap. 23). It is also a major source of support for research and training in biomedical informatics, both at NIH and in universities throughout the US. (Photograph courtesy of the National Library of Medicine)

ultimate physical limitations to the miniaturization of computer circuits.[17]

Progress in biomedical-computing research will continue to be tied to the availability of funding from either government or commercial sources. Because most biomedical-computing research is exploratory and is far from ready for commercial application, the federal government has played a key role in funding the work of the last four decades, mainly through the NIH and the Agency for Health Care Research and Quality (AHRQ). The National Library of Medicine (NLM) has assumed a primary role for biomedical informatics, especially with support for basic research in the field (Fig. 1.18). As increasing numbers of applications prove successful in the commercial marketplace, it is likely that more development work will shift to industrial settings and that university programs will focus increasingly on fundamental research problems viewed as too speculative for short-term commercialization – as has occurred in the field of computer science over the past several decades.

17 ▶ https://www.sciencedaily.com/releases/2008/01/080112083626.htm; ▶ https://arstechnica.com/science/2014/08/are-processors-pushing-up-against-the-limits-of-physics/ (Accessed 5/27/19).

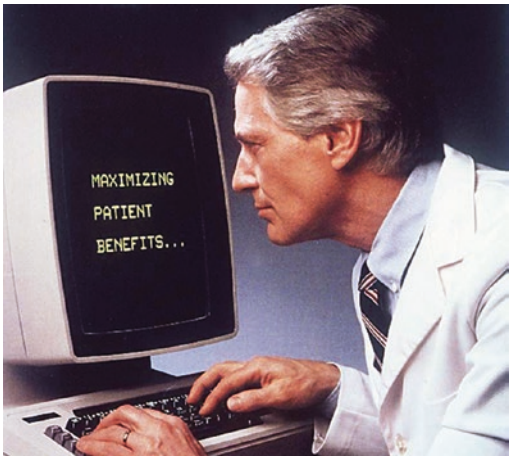


Fig. 1.19 Doctor of the future. By the early 1980s, advertisements in medical journals (such as this one for an antihypertensive agent) began to use computer equipment as props and even portrayed them in a positive light. The suggestion in this photograph seems to be that an up-to-date physician feels comfortable using computer-based tools in his practice. (Photograph courtesy of ICI Pharma, Division of ICI Americas, Inc)

1.4.3 Relationship to Biomedical Science and Clinical Practice

The exciting accomplishments of biomedical informatics, and the implied potential for future benefits to medicine, must be viewed in the context of our society and of the existing health care system. As early as 1970, an eminent clinician suggested that computers might in time have a revolutionary influence on medical care, on medical education, and even on the selection criteria for health-science trainees (Schwartz 1970). The subsequent enormous growth in computing activity has been met with some trepidation by health professionals. They ask where it will all end. Will health workers gradually be replaced by computers? Will nurses and physicians need to be highly trained in computer science or informatics before they can practice their professions effectively? Will both patients and health workers eventually revolt rather than accept a trend toward automation that they believe may threaten the traditional humanistic values in health care delivery (see ▶ Chap. 12) (Shortliffe 1993a)? Will clinicians be viewed as outmoded and backward if they do not turn to computational tools for assistance with information management and decision making (Fig. 1.19)?

1

Biomedical informatics is intrinsically entwined with the substance of biomedical science. It determines and analyzes the structure of biomedical information and knowledge, whereas biomedical science is constrained by that structure. Biomedical informatics melds the study data, information, knowledge, decision making, and supporting technologies with analyses of biomedical information and knowledge, thereby addressing specifically the interface between the science of information and knowledge management and biomedical science. To illustrate what we mean by the "structural" features of biomedical information and knowledge, we can contrast the properties of the information and knowledge typical of such fields as physics or engineering with the properties of those typical of biomedicine (see ► Box 1.2).

Box 1.2: The Nature of Medical Information

This material is adapted from a small portion of a classic book on this topic. It was written by Dr. Scott Blois, who coauthored the introductory chapter to this textbook in its 1st edition, which was published shortly after his death. Dr. Blois was a scholar who directed the informatics program at the University of California San Francisco and served as the first president of the American College of Medical Informatics (ACMI). [Blois, M. S. (1984). *Information and medicine: The nature of medical descriptions*. Berkeley: University of California Press].

From the material in this chapter, you might conclude that biomedical applications do not raise any unique problems or concerns. On the contrary, the biomedical environment raises several issues that, in interesting ways, are quite distinct from those encountered in most other domains of applied computing. Clinical information seems to be systematically different from the information used in physics, engineering, or even clinical chemistry (which more closely resembles chemical applications generally than it does medical ones). Aspects of biomedical information include an essence of uncertainty—we can never know all about a physiological process—and this results in inevitable variability among individuals. These differences raise special problems and some investigators suggest that biomedical computer science differs from conventional computer science in fundamental ways. We shall explore these differences only briefly here; for details, you can consult Blois' book on this subject (see Suggested Readings).

Let us examine an instance of what we will call a low-level (or readily formalized) science. Physics is a natural starting point; in any discussion of the hierarchical relationships among the sciences (from the fourth-century BC Greek philosopher Aristotle to the twentieth-century U.S. librarian Melvil Dewey), physics will be placed near the bottom. Physics characteristically has a certain kind of simplicity, or generality. The concepts and descriptions of the objects and processes of physics, however, are necessarily used in all applied fields, including medicine. The laws of physics and the descriptions of certain kinds of physical processes are essential in representing or explaining functions that we regard as medical in nature. We need to know something about molecular physics, for example, to understand why water is such a good solvent; to explain how nutrient molecules are metabolized, we talk about the role of electron-transfer reactions.

Applying a computer (or any formal computation) to a physical problem in a medical context is no different from doing so in a physics laboratory or for an engineering application. The use of computers in various **low-level processes** (such as those of physics or chemistry) is similar and is independent of the application. If we are talking about the solvent properties of water, it makes no difference whether we happen to be working in geology, engineering, or medicine. Such low-level processes of physics are particularly receptive to

mathematical treatment, so using computers for these applications requires only conventional numerical programming.

In biomedicine, however, there are other **higher-level processes** carried out in more complex objects such as organisms (one type of which is patients). Many of the important informational processes are of this kind. When we discuss, describe, or record the properties or behavior of human beings, we are using the descriptions of very high-level objects, the behavior of whom has no counterpart in physics or in engineering. The person using computers to analyze the descriptions of these high-level objects and processes encounters serious difficulties (Blois 1984).

One might object to this line of argument by remarking that, after all, computers are used routinely in commercial applications in which human beings and situations concerning them are involved and that relevant computations are carried out successfully. The explanation is that, in these commercial applications, the descriptions of human beings and their activities have been so highly abstracted that the events or processes have been reduced to low-level objects. In biomedicine, abstractions carried to this degree would be worthless from either a clinical or research perspective.

For example, one instance of a human being in the banking business is the customer, who may deposit, borrow, withdraw, or invest money. To describe commercial activities such as these, we need only a few properties; the customer can remain an abstract entity. In clinical medicine, however, we could not begin to deal with a patient represented with such skimpy abstractions. We must be prepared to analyze most of the complex behaviors that human beings display and to describe patients as completely as possible. We must deal with the rich descriptions occurring at high levels in the hierarchy, and we may be hard pressed to encode and process this information using the tools of mathematics and computer science that work so well at low levels. In light of these remarks, the general enterprise known as **artificial intelligence** (**AI**) can be aptly described as the application of computer science to high-level, real-world problems.

Biomedical informatics thus includes computer applications that range from processing of very low-level descriptions, which are little different from their counterparts in physics, chemistry, or engineering, to processing of extremely high-level ones, which are completely and systematically different. When we study human beings in their entirety (including such aspects as human cognition, self-consciousness, intentionality, and behavior), we must use these high-level descriptions. We will find that they raise complex issues to which conventional logic and mathematics are less readily applicable. In general, the attributes of low-level objects appear sharp, crisp, and unambiguous (e.g., "length," "mass"), whereas those of high-level ones tend to be soft, fuzzy, and inexact (e.g., "unpleasant scent," "good").

Just as we need to develop different methods to describe high-level objects, the inference methods we use with such objects may differ from those we use with low-level ones. In formal logic, we begin with the assumption that a given proposition must be either true or false. This feature is essential because logic is concerned with the preservation of truth value under various formal transformations. It is difficult or impossible, however, to assume that all propositions have truth values when we deal with the many high-level descriptions in medicine or, indeed, in everyday situations. Such questions as "Was Woodrow Wilson a good president?" cannot be answered with a "yes" or "no" (unless we limit the question to specific criteria for determining the goodness of presidents). Many common questions in biomedicine have the same property.

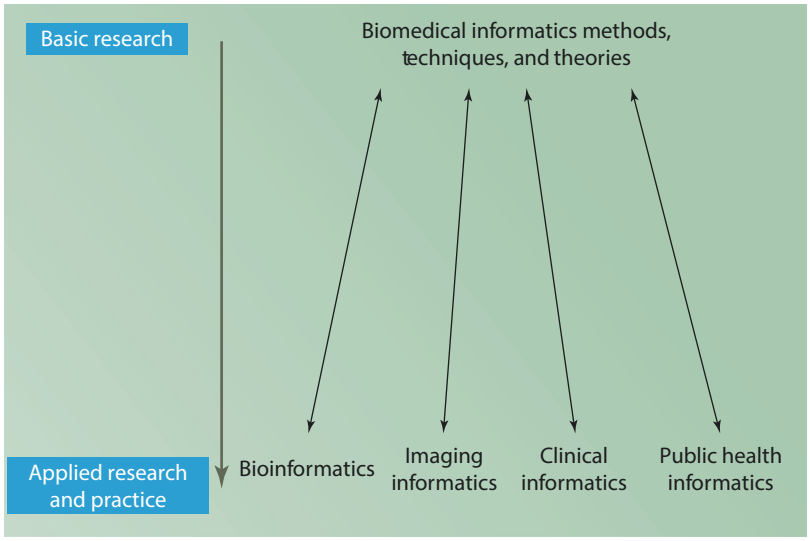


Fig. 1.20 Biomedical informatics as basic science. We view the term biomedical informatics as referring to the basic science discipline in which the development and evaluation of new methods and theories are a primary focus of activity. These core concepts and methods in turn have broad applicability in the health and biomedical sciences. The informatics subfields indicated by the terms across the bottom of this figure are accordingly best viewed as application domains for a common set of concepts and techniques from the field of biomedical informatics. Note that work in biomedical informatics is motivated totally by the application domains that the field is intended to serve (thus the two-headed arrows in the diagram). Therefore the basic research activities in the field generally result from the identification of a problem in the real world of health or biomedicine for which an informatics solution is sought (see text)

Biomedical informatics is perhaps best viewed as a basic biomedical science, with a wide variety of potential areas of application (Fig. 1.20). The analogy with other **basic sciences** is that biomedical informatics uses the results of past experience to understand, structure, and encode objective and subjective biomedical findings and thus to make them suitable for processing. This approach supports the integration of the findings and their analyses. In turn, the selective distribution of newly created knowledge can aid patient care, health planning, and basic biomedical research.

Biomedical informatics is, by its nature, an **experimental science**, characterized by posing questions, designing experiments, performing analyses, and using the information gained to design new experiments. One goal is simply to search for new knowledge, called **basic research**. A second goal is to use this knowledge for practical ends, called **applications (applied) research**. There is a continuity between these two endeavors (see Fig. 1.20). In biomedical informatics, there is an especially tight coupling between the application areas, broad categories of which are indicated at the bottom of Fig. 1.20, and the identification of basic research tasks that characterize the scientific underpinnings of the field. Research, however, has shown that there can be a very long period of time between the development of new concepts and methods in basic research and their eventual application in the biomedical world (Balas and Boren 2000). Furthermore (see Fig. 1.21), many discoveries are discarded along the way, leaving only a small percentage of basic research discoveries that have a practical influence on the health and care of patients.

Work in biomedical informatics (BMI) is inherently motivated by problems encountered in a set of applied domains in biomedicine. The first of these historically has been clinical care (including medicine, nursing, dentistry, and veterinary care), an area of activity that demands patient-oriented informatics applications. We refer to this area as **clinical informatics.**[18] It includes several sub-

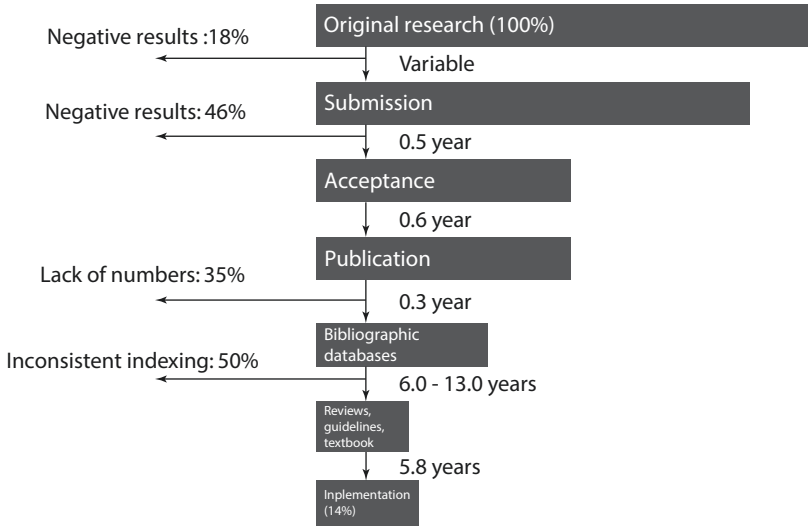


Fig. 1.21 Phases in the transfer of research into clinical practice. A synthesis of studies focusing on various phases of this transfer has indicated that it takes an average of 17 years to make innovation part of routine care (Balas and Boren 2000). Pioneering institutions often apply innovations much sooner, sometimes within a few weeks, but nationwide introduction is usually slow. National utilization rates of specific, well-substantiated procedures also suggest a delay of two decades in reaching the majority of eligible patients. For a well-documented study of such delays and their impact in an important area of clinical medicine, see (Krumholz et al. 1998). (Figure courtesy of Dr. Andrew Balas, used with permission)

topics and areas of specialized expertise, including patient-care foci such as **nursing informatics**, **dental informatics**, and even **veterinary informatics**. Furthermore, the former name of the discipline, **medical informatics**, is now reserved for those applied research and practice topics that focus on disease and the role of physicians. As was previously discussed, the term "medical informatics" is no longer used to refer to the discipline as a whole.

Closely tied to clinical informatics is **public health informatics** (Fig. 1.20), where similar methods are generalized for application to populations of patients rather than to single individuals (see ▶ Chap. 18). Thus clinical informatics and public health informatics share many of the same methods and techniques. The closeness of their relationship was amply demonstrated by the explosion in informatics research and applications that occurred in response to the COVID-19 pandemic.[19] By mid-2020, several articles had appeared to demonstrate the tight relationship between EHRs and public health informatics for management of the outbreak (Reeves et al. 2020).

Two other large areas of application overlap in some ways with clinical informatics and public health informatics. These include **imaging informatics** (and the set of issues developed around both radiology and other image management and image analysis domains such as pathology, dermatology, and molecular visualization—see ▶ Chaps. 10 and 22). Finally, there

18 Clinical informatics was approved in 2013 by the American Board of Medical Specialties as a formal subspecialty of medicine (Finnell and Dixon, 2015), with board certification examinations offered for eligible candidates by the American Board of Preventive Medicine (▶ https://www.theabpm.org/become-certified/subspecialties/clinical-informatics/ (Accessed 6/1/19)). AMIA is formulating a similar certification program, AMIA Health Informatics Certification (AHIC) for non-physicians who are working in the clinical informatics area (▶ https://www.amia.org/ahic, Accessed 1/5/2020).

19 ▶ https://www.amia.org/COVID19 (Accessed 05/03/2020)

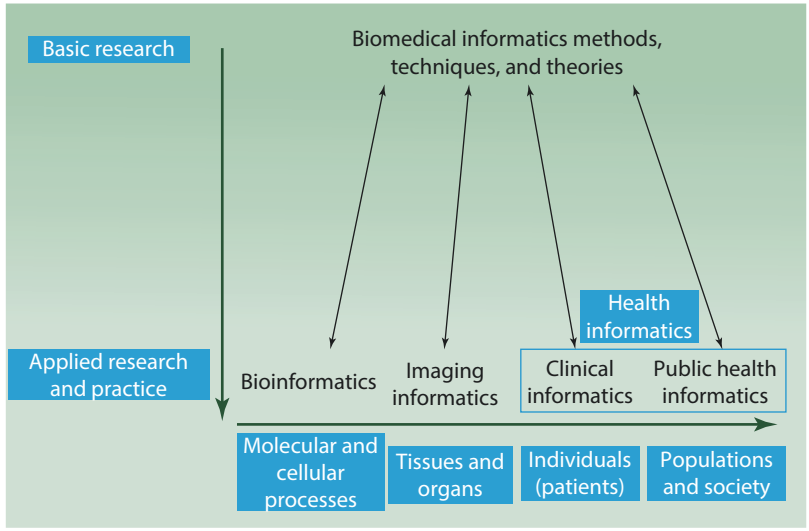


Fig. 1.22 Building on the concepts of Fig. 1.20, this diagram demonstrates the breadth of the biomedical informatics field. The relationship between biomedical informatics as a core scientific discipline and its diverse array of application domains that span biological science, imaging, clinical practice, public health, and others not illustrated (see text). Note that "health informatics" is the term used to refer to applied research and practice in clinical and public health informatics. It is not a synonym for the underlying discipline, which is "biomedical informatics"

is the burgeoning area of **bioinformatics**, which at the molecular and cellular levels is offering challenges that draw on many of the same informatics methods as well (see ▶ Chaps. 9 and 26).

As is shown in Fig. 1.22, there is a spectrum as one moves from left to right across these BMI application domains. In bioinformatics, workers deal with molecular and cellular processes in the application of informatics methods. At the next level, workers focus on tissues and organs, which tend to be the emphasis of imaging informatics work (also called **structural informatics** by some investigators). Progressing to clinical informatics, the focus is on individual patients, and finally to public health, where researchers address problems of populations and of society, including prevention. The core science of biomedical informatics has important contributions to make across that entire spectrum, and many informatics methods are broadly applicable across the full range of domains.

Note from Fig. 1.20 that biomedical informatics and bioinformatics are not synonyms and it is incorrect to refer to the scientific discipline as bioinformatics, which is, rather, an important area of application of BMI methods and concepts. Similarly, the term health informatics, which refers to applied research and practice in clinical and public-health informatics, is also not an appropriate name for the core discipline, since BMI is applicable to basic human biology as well as to health.

We acknowledge that the four major areas of application shown in Fig. 1.19 have "fuzzy" boundaries, and many areas of applied informatics research involve more than one of the categories. For example, **biomolecular imaging** involves both bioinformatics and imaging informatics concepts. Similarly, **personal** or **consumer health informatics** (see ▶ Chap. 11) includes elements of both clinical informatics and public-health informatics. Another important area of BMI research activities is **pharmacogenomics** (see ▶ Chap. 27), which is the effort to infer genetic determinants of human drug response. Such work requires the analysis of linked **genotypic** and **phenotypic** databases, and therefore lies at the intersection of bioinformatics and clinical informatics. Similarly, ▶ Chap. 28 presents the role of informatics in **precision medicine**, which relies heavily on both bioinformatics and clinical informatics concepts and systems.

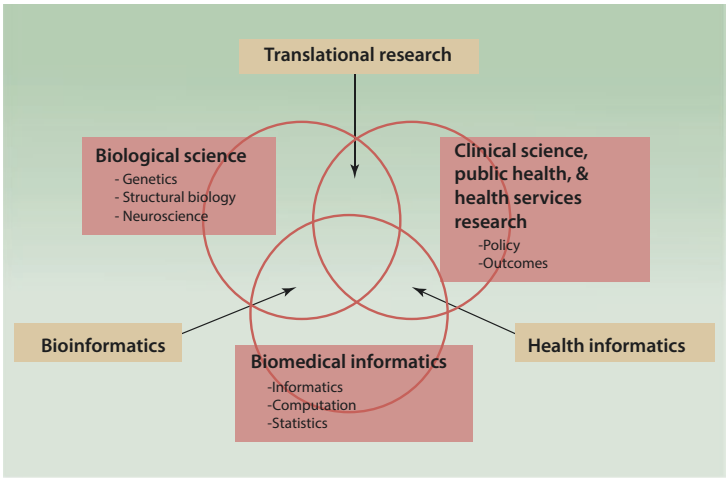


Fig. 1.23 A Venn diagram that depicts the relationships among the three major disciplines: biological research, clinical medicine / public health, and biomedical informatics. *Bioinformatics*, *Health Informatics*, and *Translational Research* lie at the intersections among pairs of these fields as shown. *Precision Medicine,* which relies on *Translational Bioinformatics* and *Clinical Research Informatics*, constitutes the area of common overlap among all three Venn circles. (Adapted with permission from a diagram developed by the Department of Biomedical Informatics at the Vanderbilt Medical Center, Nashville, TN)

Precision medicine is a product of the increasing emphasis on moving both data and concepts from basic science research into clinical science and ultimately into practice. Such efforts are typically characterized as **translational science**—a topic that has attracted major investments by the US National Institutes of Health (NIH) over the past two decades. Informatics scientists are engaged as collaborators in this translational work, which spans all four major categories of application shown in Fig. 1.20, pursuing work in **translational bioinformatics** (► Chap. 26) and **clinical research informatics** (► Chap. 27).[20] Accordingly, informatics was defined as a major component of the **Clinical and Translational Science Awards (CTSA) Program,**[21] support by the National Center for Advancing Translational Sciences (NCATS) at the NIH. AMIA sponsors an annual weeklong conference, known as the Informatics Summit, that presents new research results and applications in these areas.[22] The interactions among bioscience, clinical science, and informatics can be nicely captured by recognizing how informatics fields and translational science relate to one another (Fig. 1.23).

In general, BMI researchers derive their inspiration from one or two, rather than all, of the application areas, identifying fundamental methodologic issues that need to be addressed and testing them in system prototypes or, for more mature methods, in actual systems that are used in clinical or biomedical research settings. One important implication of this viewpoint is that the core discipline is identical, regardless of the area of application that a given individual is motivated to address, although some BMI methods have greater relevance to some domains than to others. This argues for unified BMI educational programs, ones that bring together students with a wide variety of application interests. Elective courses and internships in areas of specific

20 See also the diagram in (Kulikowski et al. 2012), which shows how these two disciplines span all areas of applied biomedical informatics.

21 ► https://ncats.nih.gov/ctsa (Accessed 6/2/2019).

22 ► https://www.amia.org/meetings-and-events (Accessed 6/2/2019)

interest are of course important complements to the core exposures that students should receive (Kulikowski et al. 2012), but, given the need for teamwork and understanding in the field, separating trainees based on the application areas that may interest them would be counterproductive and wasteful.[23]

The scientific contributions of BMI also can be appreciated through their potential for benefiting the education of health professionals (Shortliffe 2010). For example, in the education of medical students, the various cognitive activities of physicians traditionally have tended to be considered separately and in isolation—they have been largely treated as though they are independent and distinct modules of performance. One activity frequently emphasized is formal education regarding medical decision making (see ▶ Chap. 3). The specific content of this area continues to evolve, but the discipline's dependence on formal methods regarding the use of knowledge and information reveal that it is one aspect of biomedical informatics.

A particular topic in the study of medical decision making is **diagnosis**, which is often conceived and taught as though it were a free-standing and independent activity. Medical students may thus be led to view diagnosis as a process that physicians carry out in isolation before choosing therapy for a patient or proceeding to other modular tasks. A number of studies have shown that this model is oversimplified and that such a decomposition of cognitive tasks may be quite misleading (Elstein et al. 1978a; Patel and Groen 1986). Physicians seem to deal with several tasks at the same time. Although a diagnosis may be one of the first things physicians think about when they see a new patient, patient assessment (diagnosis, management, analysis of treatment results, monitoring of disease progression, etc.) is a process that never really terminates. A physician must be flexible and open-minded. It is generally appropriate to alter the original diagnosis if it turns out that treatment based on it is unsuccessful or if new information weakens the evidence supporting the diagnosis or suggests a second and concurrent disorder. ▶ Chapter 4 discusses these issues in greater detail.

When we speak of making a diagnosis, choosing a treatment, managing therapy, making decisions, monitoring a patient, or preventing disease, we are using labels for different aspects of medical care, an entity that has overall unity. The fabric of medical care is a continuum in which these elements are tightly interwoven. Regardless of whether we view computer and information science as a profession, a technology, or a science, there is no doubt about its importance to biomedicine. We can assume computers are here to stay as fundamental tools to be used in clinical practice, biomedical research, and health science education.

1.4.4 Relationship to Computer Science

During its evolution as an academic entity in universities, computer science followed an unsettled course as involved faculty attempted to identify key topics in the field and to find the discipline's organizational place. Many computer science programs were located in departments of electrical engineering, because major concerns of their researchers were computer architecture and design and the development of practical hardware components. At the same time, computer scientists were interested in programming languages and software, undertakings not particularly characteristic of engineering. Furthermore, their work with algorithm design, computability

23 Many current biomedical informatics training programs were designed with this perspective in mind. Students with interests in clinical, imaging, public health, and biologic applications are often trained together and are required to learn something about each of the other application areas, even while specializing in one subarea for their own research. Several such programs were described in a series of articles in the *Journal of Biomedical Informatics* in 2007 (Tarczy-Hornoch et al. 2007) and many more have been added since that time.

theory,[24] and other theoretical topics seemed more related to mathematics.

Biomedical informatics draws from all of these activities—development of hardware, software, and computer science theory. Biomedical computing generally has not had a large enough market to influence the course of major hardware developments; i.e., computers serve general purposes and have not been developed specifically for biomedical applications. Not since the early 1960s (when health-computing experts occasionally talked about and, in a few instances, developed special medical terminals) have people assumed that biomedical applications would use hardware other than that designed for general use.

The question of whether biomedical applications would require specialized programming languages might have been answered affirmatively in the 1970s by anyone examining the MGH Utility Multi-Programming System, known as the MUMPS language (Greenes et al. 1970; Bowie and Barnett 1976), which was specially developed for use in medical applications. For several years, MUMPS was the most widely used language for medical record processing. Under its subsequent name, M, it is still in widespread use and has been used to develop commercial electronic health record systems. New implementations have been developed for each generation of computers. M, however, like any programming language, is not equally useful for all computing tasks. In addition, the software requirements of medicine are better understood and no longer appear to be unique; rather, they are specific to the kind of task. A program for scientific computation looks pretty much the same whether it is designed for chemical engineering or for pharmacokinetic calculations.

How, then, does BMI differ from biomedical computer science? Is the new discipline simply the study of computer science with a "biomedical flavor"? If you return to the definition of biomedical informatics that we provided in ▶ Box 1.1, and then refer to ◘ Fig. 1.20, you will begin to see why biomedical informatics is more than simply the biomedical application of computer science.[25] The issues that it addresses not only have broad relevance to health, medicine, and biology, but the underlying sciences on which BMI professionals draw are inherently interdisciplinary as well (and are not limited to computer science topics). Thus, for example, successful BMI research will often draw on, and contribute to, computer science, but it may also be closely related to the decision sciences (probability theory, decision analysis, or the psychology of human problem solving), cognitive science, information sciences, or the management sciences (◘ Fig. 1.24). Furthermore, a biomedical informatics researcher will be tightly linked to some underlying problem from the real world of health or biomedicine. As ◘ Fig. 1.24 illustrates, for example, a biomedical informatics basic researcher or doctoral student will typically be motivated by one of the application areas, such as those shown at the bottom of ◘ Fig. 1.22, but a dissertation worthy of a PhD in the field will usually be identified by a generalizable scientific result that also contributes to one of the component disciplines (◘ Fig. 1.20) and on which other scientists can build in the future.

24 Many interesting problems cannot be computed in a reasonable time and require heuristics. Computability theory is the foundation for assessing the feasibility and cost of computation to provide the complete and correct results to a formally stated problem.

25 In fact, the multidisciplinary nature of biomedical informatics has led the informatics term to be borrowed in other disciplines, including computer science organizations, even though the English name for the field was first adopted in the biomedical context. Today we even have generic full departments of informatics in the US (e.g., see ▶ https://informatics.njit.edu, Accessed 11/28/2020) and in other parts of the world as well (e.g., ▶ http://www.sussex.ac.uk/informatics/. Accessed 1/5/2020). In the US, there are full schools with informatics in their title (e.g., ▶ https://luddy.indiana.edu/index.html. Accessed 1/5/2020) and even a School of Biomedical Informatics (▶ https://sbmi.uth.edu/. Accessed 1/2/2020).

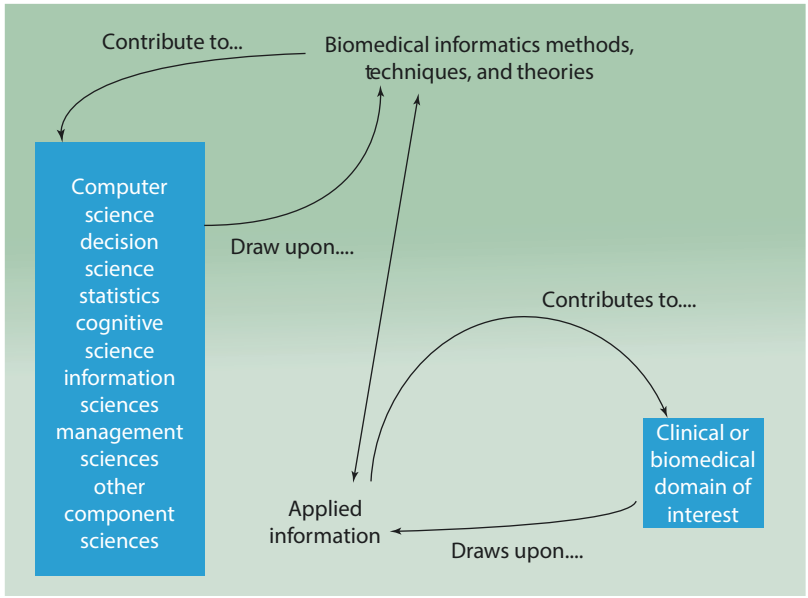


Fig. 1.24 Component sciences in biomedical informatics. An informatics application area is motivated by the needs of its associated biomedical domain, to which it attempts to contribute solutions to problems. Thus any applied informatics work draws upon a biomedical domain for its inspiration, and in turn often leads to the delineation of basic research challenges in biomedical informatics that must be tackled if the applied biomedical domain is ultimately to benefit. At the methodologic level, biomedical informatics draws on, and contributes to, a wide variety of component disciplines, of which computer science is only one. As Figs. 1.20 and 1.22 show explicitly, biomedical informatics is inherently multidisciplinary, both in its areas of application and in the component sciences on which it draws

1.4.5 Relationship to Biomedical Engineering

BMI is a relatively young discipline, whereas biomedical engineering (BME) is older and well-established. Many engineering and medical schools have formal academic programs in BME, often with departmental status and full-time faculty. Only in the last two or three decades has this begun to be true of biomedical informatics academic units. How does biomedical informatics relate to biomedical engineering, especially in an era when engineering and computer science are increasingly intertwined?

Biomedical engineering departments emerged in the late 1960s, when technology began to play an increasingly prominent role in medical practice.[26] The emphasis in such departments has tended to be research on, and development of, instrumentation (e.g., as discussed in ▶ Chaps. 21 and 22, advanced monitoring systems, specialized transducers for clinical or laboratory use, and imaging methods and enhancement techniques for use in radiology), with an orientation toward the

26 By the late 1960s the first BME departments were formed in the US at the University of Virginia, Case Western Reserve University, Johns Hopkins University, and Duke University (see ▶ https://navigate.aimbe.org/why-bioengineering/history/, Accessed 6/2/2019). Duke's undergraduate degree program in BMI was the first to be accredited by the Engineering Council for Professional Development (September 1972).

development of medical devices, **prostheses**, and specialized research tools. There is also a major emphasis on tissue engineering and related wet-bench research efforts. In recent years, computing techniques have been used both in the design and construction of medical devices and in the medical devices themselves. For example, the "smart" devices increasingly found in most medical specialties are all dependent on computational technology. Intensive care monitors that generate blood pressure records while calculating mean values and hourly summaries are examples of such "intelligent" devices.

The overlap between biomedical engineering and BMI suggests that it would be unwise for us to draw compulsively strict boundaries between the two fields. There are ample opportunities for interaction, and there are chapters in this book that clearly overlap with biomedical engineering topics—e.g., ▶ Chap. 21 on patient-monitoring systems and ▶ Chap. 22 on radiology systems. Even where they meet, however, the fields have differences in emphasis that can help you to understand their different evolutionary histories. In biomedical engineering, the emphasis is on medical devices and underlying methods; in BMI, the emphasis is on biomedical information and knowledge and on their management with the use of computers. In both fields, the computer is secondary, although both use computing technology. The emphasis in this book is on the informatics end of the spectrum of biomedical computer science, so we shall not spend much time examining biomedical engineering topics.

1.5 Integrating Biomedical Informatics and Clinical Practice

It should be clear from the material in this chapter that biomedical informatics is a remarkably broad and complex topic. We have argued that information management is intrinsic to both life-science research and clinical practice and that, in biomedical settings over a half century, the use of computers to aid in information management has grown from a futuristic notion to an everyday occurrence. In fact, the EHR and other information technology tools may now be the only kind of equipment that is used by every single health care professional, regardless of specialty or professional title. In this chapter and throughout the book, we emphasize the myriad ways in which computers are used in biomedicine to ease the burdens of information management and the means by which new technology is changing the delivery of health care. The degree to which such changes are positively realized, and their rate of occurrence, are being determined in part by external forces that influence the costs of developing and implementing biomedical applications and the ability of scientists, clinicians, patients, and the health care system to accrue the potential benefits.

We can summarize several global forces that are affecting biomedical computing and that will continue to influence the extent to which computers are assimilated into clinical practice: (1) new developments in communications plus computer hardware and software; (2) a further increase in the number of individuals who have been trained in both medicine, or another health profession, and in BMI; and (3) ongoing changes in health care financing designed to control the rate of growth of health-related expenditures.

We touched on the first of these factors in ▶ Sect. 1.4.2, when we described the historical development of biomedical computing and the trend from mainframe computers, to microcomputers and PCs, and to the mobile devices of today. The future view outlined in ▶ Sect. 1.1 similarly builds on the influence that the Internet has provided throughout society during the last decade. Hardware improvements have made powerful computers inexpensive and thus available to hospitals, to departments within hospitals, and even to individual physicians. The broad selection of computers of all sizes, prices, and capabilities makes computer applications both attractive and accessible. Technological advances in

information storage devices,[27] including the movement of files to the "cloud", are facilitating the inexpensive storage of large amounts of data, thus improving the feasibility of data-intensive applications, such as drawing inferences from human genome datasets (see ▶ Chaps. 9, 26, and 28) and the all-digital radiology department (▶ Chap. 22). Standardization of hardware and advances in network technology are making it easier to share data and to integrate related information-management functions within a hospital or other health care organization, although inadequacies in standards for encoding and sharing data continue to be challenging (▶ Chaps. 7, 14, 15, and 16).

The second factor is the frustratingly slow increase in the number of professionals who are being trained to understand the biomedical issues as well as the technical and engineering ones. Computer scientists who understand biomedicine are better able to design systems responsive to actual needs and sensitive to workflow and the clinical culture. Health professionals who receive formal training in BMI are likely to build systems using well-established techniques while avoiding the past mistakes of other developers. As more professionals are trained in the special aspects of both fields, and as the programs they develop are introduced, health care professionals are more likely to have useful and usable systems available when they turn to the computer for help with information management tasks.

The third factor affecting the integration of computing technologies into health care settings is our evolving health care system and the increasing pressure to control medical spending. The escalating tendency to apply technology to all patient-care tasks is a frequently cited phenomenon in modern medical practice. Mere physical findings no longer are considered adequate for making diagnoses and planning treatments. In fact, medical students who are taught by more experienced physicians to find subtle diagnostic signs by examining various parts of the body nonetheless often choose to bypass or deemphasize physical examinations in favor of ordering one test after another. Sometimes, they do so without paying sufficient attention to the ensuing cost. Some new technologies replace less expensive, but technologically inferior, tests. In such cases, the use of the more expensive approach is generally justified. Occasionally, computer-related technologies have allowed us to perform tasks that previously were not possible. For example, the scans produced with computed tomography or magnetic resonance imaging (see ▶ Chaps. 10 and 22) have allowed physicians to visualize cross-sectional slices of the body, and medical instruments in intensive care units perform continuous monitoring of patients' body functions that previously could be checked only episodically (see ▶ Chap. 21).

The development of expensive new technologies, and the belief that more technology is better, have helped to fuel rapidly escalating health care costs. In the 1970s and 1980s, such rising costs led to the introduction of managed care and **capitation**—changes in financing and delivery that were designed to curb spending. Today we are seeing a trend toward **value-based reimbursement**, which is predicated on the notion that payment for care of patients should be based on the demonstrated value received (as defined by high quality at low cost) rather than simply the existence of an encounter or procedure. Integrated computer systems can provide the means to capture data to help assess such value, while they also support detailed cost accounting, the analysis of the relationship of costs of care to the benefits of that care, evaluation of the quality of care provided, and identification of areas of inefficiency. Systems that improve the quality of care while reducing the cost of providing that care clearly will be favored. The effect of cost containment pressures on technologies that increase the cost of care while improving the quality are less clear. Medical technologies, including computers, will be

27 Technological progress in this area is occurring at a dizzying rate. Consider, for example, the announcement that scientists are advancing the notion of using DNA for data storage and can store as much as 704 terabytes of information in a gram of DNA. ▶ http://www.engadget.com/2012/08/19/harvard-stores-704tb-in-a-gram-of-dna; ▶ https://homes.cs.washington.edu/~bornholt/dnastorage-asplos16/ (Accessed 5/30/19).

embraced only if they improve the delivery of clinical care while either reducing costs or providing benefits that clearly exceed their costs.

Designers of medical systems must address satisfactorily many logistical and engineering questions before innovative solutions are integrated optimally into medical practice. For example, are the machines conveniently located? Should mobile devices further replace tethered workstations? Can users complete their tasks without excessive delays? Is the system reliable enough to avoid loss of data? Can users interact easily and intuitively with the computer? Does it facilitate rather than disrupt workflow? Are patient data secure and appropriately protected from prying eyes? In addition, cost-control pressures produce a growing reluctance to embrace expensive technologies that add to the high cost of health care. The net effect of these opposing trends is in large part determining the degree to which specific systems are embraced and effectively implemented in the health care environment.

In summary, rapid advances in communications, computer hardware, and software, coupled with an increasing computer literacy of health care professionals and researchers, favor the implementation of effective computer applications in clinical practice, public health, and life sciences research. Furthermore, in the increasingly competitive health care industry, providers have a greater need for the information management capabilities supplied by computer systems. The challenge is to demonstrate in persuasive and rigorous ways the financial and clinical advantages of these systems (see ► Chap. 13).

Suggested Readings

Blois, M. S. (1984b). Information and medicine: The nature of medical descriptions. Berkeley: University of California Press. In this classic volume, the author analyzes the structure of medical knowledge in terms of a hierarchical model of information. He explores the ideas of high- and low-level sciences and suggests that the nature of medical descriptions accounts for difficulties in applying computing technology to medicine. A brief summary of key elements in this book is included as Box 1.2 in this chapter.

Coiera, E. (2015). *Guide to health informatics* (3rd ed.). Boca Raton, FL: CRC Press. This introductory text is a readable summary of clinical and public health informatics, aimed at making the domain accessible and understandable to the non-specialist.

Collen, M. F., & Ball, M. J. (Eds.). (2015). *A history of medical informatics in the United States* (2nd ed.). London: Springer. This comprehensive book traces the history of the field of medical informatics, and identifies the origins of the discipline's name (which first appeared in the English-language literature in 1974). The original (1995) edition was being updated by Dr. Collen when he passed away shortly after his 100th birthday. Dr. Ball organized an effort to complete the 2nd edition, enlisting participation by many leaders in the field.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978b). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press. This classic collection of papers describes detailed studies that have illuminated several aspects of the ways in which expert and novice physicians solve medical problems. The seminal work described remains highly relevant to today's work on problem solving and clinical decision support systems.

Friedman, C. P., Altman, R. B., Kohane, I. S., McCormick, K. A., Miller, P. L., Ozbolt, J. G., Shortliffe, E. H., Stormo, G. D., Szczepaniak, M. C., Tuck, D., & Williamson, J. (2004). Training the next generation of informaticians: The impact of BISTI and bioinformatics. *Journal of American Medical Informatics Association, 11*, 167–172. This important analysis addresses the changing nature of biomedical informatics due to the revolution in bioinformatics and computational biology. Implications for training, as well as organization of academic groups and curriculum development, are discussed.

Hoyt, R. E., & Hersh. W. R. (2018). *Health informatics: Practical guide* (7th ed). Raleigh: Lulu.com. This introductory volume provides a broad view of informatics and is aimed especially at health professionals in management roles or IT professionals who are entering the clinical world.

Institute of Medicine[25]. (1991 [revised 1997]). *The computer-based patient record: An essential*

technology for health care. Washington, DC: National Academy Press. National Research Council (1997). For The Record: Protecting Electronic Health Information. Washington, DC: National Academy Press. National Research Council (2000). Networking Health: Prescriptions for the Internet. Washington, DC: National Academy Press. This set of three reports from branches of the US National Academies of Science has had a major influence on health information technology education and policy over the last 25 years.

Institute of Medicine25. (2000). To err is human: Building a safer health system. Washington, DC: National Academy Press. Institute of Medicine (2001). Crossing the Quality Chasm: A New Health Systems for the 21st Century. Washington, DC: National Academy Press. Institute of Medicine (2004). Patient Safety: Achieving a New Standard for Care. Washington, DC: National Academy Press. This series of three reports from the Institute of Medicine has outlined the crucial link between heightened use of information technology and the enhancement of quality and reduction in errors in clinical practice. Major programs in patient safety have resulted from these reports, and they have provided motivation for a heightened interest in health care information technology among policy makers, provider organizations, and even patients.

Kalet, I. J. (2013). *Principles of biomedical informatics* (2nd ed.). New York: Academic. This volume provides a technical introduction to the core methods in BMI, dealing with storage, retrieval, display, and use of biomedical data for biological problem solving and medical decision making. Application examples are drawn from bioinformatics, clinical informatics, and public health informatics.

National Academy of Medicine. (2018). *Procuring interoperability: Achieving high-quality, connected, and person-centered care*. Washington, DC: National Academy Press. National Academy of Medicine (2019). Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington, DC: National Academy Press. This series of two reports from the National Academy of Medicine outlines emerging issues in biomedical informatics: interoperability (which is discussed in greater detail in Chapter 8), and artificial intelligence (which is discussed in in many chapters throughout this volume).

National Academy of Medicine. (2019). *Taking action against clinician burnout: A systems approach to professional well-being*. Washington, DC: National Academy Press. This consensus study from the National Academy of Medicine discusses the problem of clinician burnout in the United States, including areas where health care information technology may contribute or reduce these problems.

Shortliffe, E. (1993b). Doctors, patients, and computers: Will information technology dehumanize health care delivery? *Proceedings of the American Philosophical Society, 137*(3), 390–398 In this paper, the author examines the frequently expressed concern that the introduction of computing technology into health care settings will disrupt the development of rapport between clinicians and patients and thereby dehumanize the therapeutic process. He argues, rather, that computers may eventually have precisely the opposite effect on the relationship between clinicians and their patients.

Questions for Discussion

1. How do you interpret the phrase "logical behavior"? Do computers behave logically? Do people behave logically? Explain your answers.
2. What do you think it means to say that a computer program is "effective"? Make a list of a dozen computer applications with which you are familiar. List the applications in decreasing order of effectiveness, as you have explained this concept. Then, for each application, indicate your estimate of how well human beings perform the same tasks (this will require that you determine what it means for a human being to be effective). Do you discern any pattern? If so, how do you interpret it?
3. Discuss three society-wide factors that will determine the extent to which computers are assimilated into clinical practice.
4. Reread the future vision presented in ▶ Sect. 1.1. Describe the characteristics of an integrated environment for

managing clinical information. Discuss two ways (either positive or negative) in which such a system could change clinical practice.

5. Do you believe that improving the technical quality of health care entails the risk of dehumanization? If so, is it worth the risk? Explain your reasoning.
6. Consider ◘ Fig. 1.20, which shows that bioinformatics, imaging informatics, clinical informatics, and public health informatics are all application domains of the biomedical informatics discipline because they share the same core methods and theories:
   (a) Briefly describe two examples of core biomedical informatics methods or theories that can be applied both to bioinformatics and clinical informatics.
   (b) Imagine that you describe ◘ Fig. 1.20 to a mathematics faculty member, who responds that "in that case, I'd also argue that statistics, computer science, and physics are all application domains of math because they share the same core mathematical methods and theories." In your opinion, is this a legitimate argument? In what ways is this situation similar to, and different from, the case of biomedical informatics?
   (c) Why is biomedical informatics *not* simply computer science applied to biomedicine, or to the practice of medicine, using computers?
   (d) How would you describe the relevance of psychology and cognitive science to the field of biomedical informatics? (Hint: See ◘ Fig. 1.24)
7. In 2000, a major report by the Institute of Medicine[28] entitled "To Err is Human: Building a Safer Health System" (see Suggested Readings) stated that up to 98,000 patient deaths were being caused by preventable medical errors in American hospitals each year.
   (a) It has been suggested that effective electronic health record (EHR) systems should mitigate this problem. What are three specific ways in which they could be reducing the number of adverse events in hospitals?
   (b) Are there ways in which computer-based systems could *increase* the incidence of medical errors? Explain.
   (c) Describe a practical experiment that could be used to examine the impact of an EHR system on patient safety. In other words, propose a study design that would address whether the computer-based system increases or decreases the incidence of preventable adverse events in hospitals – and by how much.
   (d) What are the limitations of the experimental design you proposed in (c)?
8. It has been argued that the ability to capture "nuance" in the description of what a clinician has seen when examining or interviewing a patient may not be as crucial as some people think. The desire to be able to express one's thoughts in an unfettered way (free text) is often used to argue against the use of structured data-entry methods using a controlled vocabulary and picking descriptors from lists when recording information in an EHR.
   (a) What is your own view of this argument? Do you believe that it is important to the quality and/or efficiency of care for clinicians to be able to record their observations, at least part of the time, using free text/natural language?
   (b) Many clinicians have been unwilling to use an EHR system requiring structured data entry

28 The Institute of Medicine (IOM), part of the former National Academy of Sciences (NAS) was reorganized in 2015 to become the National Academy of Medicine (NAM). The NAS is now known as the National Academies of Science, Engineering, and Medicine (NASEM).

1

because of the increased time required for documentation at the point of care and constraints on what can be expressed. What are two strategies that could be used to address this problem (other than "designing a better user interface for the system")?

References

Balas, E. A., & Boren, S. A. (2000). Managing clinical knowledge for health care improvement. In *Yearbook of medical informatics 2000: Patient-centered systems* (pp. 65–70). Stuttgart: Schattauer.

Begun, J. W., Zimmerman, B., & Dooley, K. (2003). Health care organizations as complex adaptive systems. In S. M. Mick & M. Wyttenbach (Eds.), *Advances in health care organization theory* (pp. 253–288). San Francisco: Jossey-Bass.

Blois, M. S. (1984). *Information and medicine: The nature of medical descriptions*. Berkeley: University of California Press.

Bowie, J., & Barnett, G. O. (1976). MUMPS: An economical and efficient time-sharing system for information management. *Computer Programs in Biomedicine, 6*, 11–22.

Collen, M. F. (1995). *A history of medical informatics in the United States, 1950 to 1990*. Bethesda: American Medical Informatics Association.

Daley, K. A. (2013). A continuously learning health care system in the United States. In *NAM perspectives*. Washington, DC: Commentary, National Academy of Medicine. https://doi.org/10.31478/201307b.

Dick, R., & Steen, E. (Eds.). (1991). *The computer – Based patient record : An essential technology for HealthCare*. (Rev. 1997). Washington, D.C.: Institute of Medicine, National Academy Press.

Elstein, K. A., Shulman, L. S., & Sprafka, S. A. (1978a). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Finnell, J. T., & Dixon, B. E. (2015). *Clinical informatics study guide: Text and review*. London: Springer.

Greenes, R. A., & Shortliffe, E. H. (1990). Medical informatics: An emerging academic discipline and institutional priority. *Journal of the American Medical Association, 263*(8), 1114–1120.

Greenes, R. A., Barnett, G. O., Klein, S. W., Robbins, A., & Prior, R. E. (1970). Recording, retrieval, and review of medical data by physician-computer interaction. *The New England Journal of Medicine, 282*(6), 307–315.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R, (Springer texts in statistics)*. New York: Springer.

Krumholz, H. M., Radford, M. J., Wang, Y., Chen, J., Heiat, A. & Marciniak, T. A. (1998). National use and effectiveness of beta-blockers for the treatment of elderly patients after acute myocardial infarction: National Cooperative Cardiovascular Project. *JAMA*, 19; *280*(7), 623–9.

Kulikowski, C. A., Shortliffe, E. H., et al. (2012). AMIA Board white paper: Definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *Journal of the American Medical Informatics Association, 19*(6), 931–938.

Ozkaynak, M., Brennan, P. F., Hanauer, D. A., Johnson, S., Aarts, J., Zheng, K., & Haque, S. N. (2013). Patient-centered care requires a patient oriented workflow model. *Journal of the American Medical Informatics Association: JAMIA, 20*(e1), e14–e16.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10*(1), 91–116.

Reeves, J. J., Hollandsworth, H. M., Torriani, F. J., et al. (2020). Rapid response to COVID-19: Health informatics support for outbreak management in an academic health system. *Journal of the American Medical Informatics Association: JAMIA*, ocaa037. https://doi.org/10.1093/jamia/ocaa037.

Schwartz, W. B. (1970). Medicine and the computer: The promise and problems of change. *The New England Journal of Medicine, 283*(23), 1257–1264.

Shortliffe, E. H. (1993a). Doctors, patients, and computers: Will information technology dehumanize health-care delivery? *Proceedings of the American Philosophical Society, 137*(3), 390–398.

Shortliffe, E. H. (1998a). The next generation internet and health care: A civics lesson for the informatics community. In *Proceedings of the AMIA annual fall symposium* (pp. 8–14). Orlando.

Shortliffe, E. H. (1998b). *The evolution of health-care records in the era of the internet, proceedings of Medinfo 98*. Seoul/Amsterdam: IOS Press.

Shortliffe, E. H. (2000). Networking health: Learning from others, taking the lead. *Health Affairs, 19*(6), 9–22.

Shortliffe, E. H. (2005). Strategic action in health information technology: Why the obvious has taken so long. *Health Affairs, 24*, 1222–1233.

Shortliffe, E. H. (2010). Biomedical informatics in the education of physicians. *Journal of the American Medical Association, 304*(11), 1227–1228.

Shortliffe, E. H., & Sondik, E. (2004). *The informatics infrastructure: Anticipating its role in cancer surveillance*. Proceedings of the C-change summit on cancer surveillance and information: The next decade, Phoenix, AZ.

Simborg, D. W., Chadwick, M., Whiting-O'Keefe, Q. E., Tolchin, S. G., Kahn, S. A., & Bergan, E. S. (1983). Local area networks and the hospital. *Computers and Biomedical Research, 16*(3), 247–259.

Tarczy-Hornoch, P., Markey, M. K., Smith, J. A., & Hiruki, T. (2007). Biomedical informatics and genomic medicine: Research and training. *Journal of Biomedical Informatics, 40*, 1–4.

Yasnoff, W. A., Sweeney, L., & Shortliffe, E. H. (2013). Putting health IT on the path to success. *Journal of the American Medical Association, 309*(10), 989–990.

Biomedical Data: Their Acquisition, Storage, and Use

Edward H. Shortliffe and Michael F. Chiang

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_2

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What are clinical data?
- How are clinical data used?
- What are the advantages and disadvantages of traditional paper medical records vs. electronic health records?
- What is the role of the computer in data storage, retrieval, and interpretation?
- What distinguishes a database from a knowledge base?
- How are data collection and hypothesis generation intimately linked in clinical diagnosis?
- What are the meanings of the terms *prevalence*, *predictive value*, *sensitivity*, and *specificity*?
- How are the terms related?
- What are the alternatives for entry of data into a clinical database?

2.1 What Are Clinical Data?

From earliest times, the ideas of ill health and its treatment have been wedded to those of the observation and interpretation of data. Whether we consider the disease descriptions and guidelines for management in early Greek literature or the modern physician's use of complex laboratory and X-ray studies, it is clear that gathering data and interpreting their meaning are central to the health care process. With the move toward the use of clinical and genomic information in assessing individual patients (their risks, prognosis, and likely responses to therapy), the sheer amounts of data that may be used in patient care have become huge. A textbook on biomedical informatics will accordingly refer time and again to issues in data collection, storage, and use. This chapter lays the foundation for this recurring set of issues that is pertinent to all aspects of the use of information, knowledge, and computers in biomedicine, both in the clinical world and in applications related to public health, biology and human genetics.

If data are central to all health care, it is because they are crucial to the process of decision making (as described in detail in ► Chaps. 3 and 4 and again in ► Chap. 26). In fact, simple reflection will reveal that all health care activities involve gathering, analyzing, or using data. Data provide the basis for categorizing the problems a patient may be having or for identifying subgroups within a population of patients. They also help a physician to decide what additional information is needed and what actions should be taken to gain a greater understanding of a patient's problem or most effectively to treat the problem that has been diagnosed.

It is overly simplistic to view data as the columns of numbers or the monitored waveforms that are a product of our technological health care environment. Although laboratory test results and other numeric data are often invaluable, a variety of more subtle types of data may be just as important to the delivery of optimal care: the awkward glance by a patient who seems to be avoiding a question during the medical interview, information about the details of a patient's symptoms or about his family or economic setting, or the subjective sense of disease severity that an experienced clinician will often have within a few moments of entering a patient's room. No clinician disputes the importance of such observations in decision making during patient assessment and management, yet the precise role of these data and the corresponding decision criteria are so poorly understood that it is difficult to record them in ways that convey their full meaning, even from one clinician to another. Despite these limitations, clinicians need to share descriptive information with others. When they cannot interact directly with one another, they often turn to the chart or electronic health record for communication purposes.

We consider a **clinical datum** to be any single observation of a patient—e.g., a temperature reading, a red blood cell count, a past history of rubella, or a blood pressure reading. As the blood pressure example shows, it is a matter of perspective whether a single observation is in fact more than one datum. A blood pressure of 120/80 might well be recorded as a single element in a setting where knowledge that a patient's blood pressure is normal is all that matters. If the difference between diastolic (while the heart cavities are beginning to fill)

and systolic (while they are contracting) blood pressures is important for decision making or for analysis, however, the blood pressure reading is best viewed as two pieces of information (systolic pressure = 120 mmHg, diastolic pressure = 80 mmHg). Human beings can glance at a written blood pressure value and easily make the transition between its unitary view as a single data point and the decomposed information about systolic and diastolic pressures. Such dual views can be much more difficult for computers, however, unless they are specifically allowed for in the design of the method for data storage and analysis. The idea of a *data model* for computer-stored medical data accordingly becomes an important issue in the design of medical data systems.

Clinical *data* may involve several different observations made concurrently, the observation of the same patient parameter made at several points in time, or both. Thus, a single datum generally can be viewed as defined by five elements:

1. The *patient* in question
2. The *parameter* being observed (e.g., liver size, urine sugar value, history of rheumatic fever, heart size on chest X-ray film)
3. The *value* of the parameter in question (e.g., weight is 70 kg, temperature is 98.6 °F, profession is steel worker)
4. The *time* of the observation (e.g., 2:30 A.M. on 14FEB2019[1])
5. The *method* by which the observation was made (e.g., patient report, thermometer, urine dipstick, laboratory instrument).

Time can particularly complicate the assessment and computer-based management of data. In some settings, the date of the observation is adequate—e.g., in outpatient clinics or private offices where a patient generally is seen infrequently and the data collected need to be identified in time with no greater accuracy than a calendar date. In others, minute-to-minute variations may be important—e.g., the frequent blood sugar readings obtained for a patient in diabetic ketoacidosis (acid production due to poorly controlled blood sugar levels) or the continuous measurements of mean arterial blood pressure for a patient in cardiogenic shock (dangerously low blood pressure due to failure of the heart muscle).

It may also be important to keep a record of the circumstances under which a data point was obtained. For example, was the blood pressure taken in the arm or leg? Was the patient lying or standing? Was the pressure obtained just after exercise? During sleep? What kind of recording device was used? Was the observer reliable? Such additional information, sometimes called contexts, methods, or modifiers, can be of crucial importance in the proper interpretation of data. Two patients with the same basic problem or symptom often have markedly different explanations for their problem, revealed by careful assessment of the modifiers of that problem.

A related issue is the uncertainty in the values of data. It is rare that an observation—even one by a skilled clinician—can be accepted with absolute certainty. Consider the following examples:

- An adult patient reports a childhood illness with fevers and a red rash in addition to joint swelling. Could he or she have had scarlet fever? The patient does not know what his or her pediatrician called the disease nor whether anyone thought that he or she had scarlet fever.
- A physician listens to the heart of an asthmatic child and thinks that she hears a heart murmur—but is not certain because of the patient's loud wheezing.
- A radiologist looking at a shadow on a chest X-ray film is not sure whether it represents overlapping blood vessels or a lung tumor.
- A confused patient is able to respond to simple questions about his or her illness, but under the circumstances the physician is uncertain how much of the history being reported is reliable.

As described in ▶ Chaps. 3 and 4, there are a variety of possible responses to deal with

1 Note that it was the tendency to record such dates in computers as "14FEB12" that led to the end-of-century complexities that were called the *Year 2K problem*. It was shortsighted to think that it was adequate to encode the year of an event with only two digits.

incomplete data, the uncertainty in them, and in their interpretation. One technique is to collect additional data that will either confirm or eliminate the concern raised by the initial observation. This solution is not always appropriate, however, because the costs of data collection must be considered. The additional observation might be expensive, risky for the patient, or wasteful of time during which treatment could have been instituted. The idea of trade-offs in data collection thus becomes extremely important in guiding health care decision making.

2.1.1 What Are the Types of Clinical Data?

The examples in the previous section suggest that there is a broad range of data types in the practice of medicine and the allied health sciences. They range from narrative, textual data to numerical measurements, genetic information, recorded signals, drawings, and photographs or other images.

Narrative data account for a large component of the information that is gathered in the care of patients. For example, the patient's description of his or her present illness, including responses to focused questions from the physician, generally is gathered verbally and is recorded as text in the medical record. The same is true of the patient's social and family history, the general review of systems that is part of most evaluations of new patients, and the clinician's report of physical examination findings. Such narrative data were traditionally handwritten by clinicians and then placed in the patient's medical record (Fig. 2.1a). Increasingly, however, the narrative summaries were dictated and then transcribed by typists who produced printed summaries or electronic copies for inclusion in paper or electronic medical records. Now, physicians and staff largely enter narrative text directly into electronic health records (EHRs), usually through keyboard, mouse-driven, or voice-driven interfaces (Fig. 2.1b). Electronic narrative data often include not only patient histories and physical examinations, but also other narrative descriptions such as reports of specialty consultations, surgical procedures, pathologic examinations of tissues, and hospitalization summaries when a patient is discharged.

Some narrative data are loosely coded with shorthand conventions known to health personnel, particularly data collected during the physical examination, in which recorded observations reflect the stereotypic examination process taught to all practitioners. It is common, for example, to find the notation "PERRLA" under the eye examination in a patient's medical record. This encoded form indicates that the patient's "Pupils are Equal (in size), Round, and Reactive to Light and Accommodation (the process of focusing on near objects)."

Note that there are significant problems associated with the use of such abbreviations. Many are not standard and can have different meanings depending on the context in which they are used. For example, "MI" can mean "mitral insufficiency" (leakage in one of the heart's valves) or "myocardial infarction" (the medical term for what is commonly called a heart attack). Many hospitals try to establish a set of "acceptable" abbreviations with meanings, but the enforcement of such standardization is often unsuccessful. Other hospitals approach this challenge by not permitting use of abbreviations in the medical record, and instead require use of full-length narrative descriptions.

Standard narrative expressions have often become loose standards of communication among medical personnel. Examples include "mild dyspnea (shortness of breath) on exertion," "pain relieved by antacids or milk," and "failure to thrive." Such standardized expressions are attempts to use conventional text notation as a form of summarization for otherwise heterogeneous conditions that together characterize a simple concept about a patient.

Many data used in medicine take on discrete numeric values. These include such parameters as laboratory tests, vital signs (such as temperature and pulse rate), and certain measurements taken during the physical examination. When such numerical data are interpreted, however, the issue of precision becomes important. Can physicians distinguish reliably between a 9-cm and a 10-cm liver span when they examine a

2

a

Present Illness: (date) June 3, 1989 Chief Complaint:

Admission Note

ID: 1st admission for this 42 y/o Mexican American ♀ who presents with

CC: headache for one week

HPI: On 5/25 pt noted the onset of myalgias, severe headache, nausea, neck pain, and shaking chills. She consulted her private MD for these problems, and he diagnosed migraines & prescribed a combination med (belladonna, alkaloids), phenobarbital, and ergotamine tartarate) plus meprobamate). However, her sx worsened over the next week until 6/3 when she presented to our ER. She denies photophobia, diplopia, & other neurologic symptoms. She has noted a nonproductive cough but is a nonsmoker and she denies hemoptysis. She denies exposure to diseased individuals, specifically including meningococcal disease or TB.

PMH: No hx of illnesses other than NCD's. Meds only as above. Allergies: ∅ Surgery ∅ One daughter, age 12, by NVD.

Social: Married 14 yrs. Works in home. Has never lived in San Joaquin Valley. Just travelled to Mexico by car in 1974.

ROS: Genl: well until 10 days PTA

Skin: ∅

Head: ∅ ↑ per HPI.

NARRATIVE PHYSICAL EXAMINATION

Fig. 2.1 Much of the information gathered during a physician–patient encounter is written in the medical record. This was traditionally done using **a** paper notes, and now increasingly using **b** electronic health records

patient's abdomen? Does it make sense to report a serum sodium level to two-decimal-place accuracy? Is a 1-kg fluctuation in weight from 1 week to the next significant? Was the patient weighed on the same scale both times (i.e., could the different values reflect variation between measurement instruments rather than changes in the patient)?

In some fields of medicine, analog data in the form of continuous signals are particularly important (see ▶ Chap. 23). Perhaps the best-known example is an electrocardiogram

b

Signed

Patient presents for followup following recent pneumonia. He is feeling much improved although still complains of mild fatigue and occasional cough productive of clear to cloudy white sputum. He lost about five pounds, but has gained about half of that back again.

Patient admits to being poorly compliant with his hypertension regimen over the past several months. He was concerned that the medication was making him tired and cut back his dosage to every other day. He has not been monitoring his blood pressure at home. He denies palpitations, anginal symptoms, or peripheral edema.

Patient indicates that he has been taking his thyroid medication regularly. Review of his prescriptions suggested he should have run out but he indicates that he had some extra tablets on hand. He is clinically euthyroid. Denies problems with constipation, dry skin, or unusual cold intolerance.

Current prescriptions:
CARDIZEM CD CPCR 240 MG/24HR OR, 1 CAPSULE DAILY, D: 60, R: 5
SYNTHROID TABS 0.1 MG OR, 1 TABLET DAILY, D: 60, R: 5
PEPCID TABS 20 MG OR, 1 TABLET AT BEDTIME PRN, D: 30, R: 3

Review of patient's allergies indicates:
Penicillins Hives

OBJECTIVE:
BP 150/98 | Pulse 88 | Temp 99.1 | Resp 16 | Wt 182 lbs (82.55 kg)
Alert, cooperative, and well hydrated. TMs clear. Nose clear rhinorrhea. Oropharynx displays moderate erythema. Neck supple without adenopathy. Once a few scattered expiratory rhonchi which partially clear with coughing. Heart regular rate and rhythm without extra signs or murmurs. Abdomen soft, nontender, no organomegaly.

Fig. 2.1 (continued)

(ECG), a tracing of the electrical activity from a patient's heart. When such data are stored in medical records, a graphical tracing frequently is included, with a written interpretation of its meaning. There are clear challenges in determining how such data are best managed in computer-based storage systems.

Visual images—acquired from machines or sketched by the physician—are another important category of data. Radiologic images or photographs of skin lesions are obvious examples. It has traditionally been common for physicians to draw simple pictures to represent abnormalities that they have observed; such drawings may serve as a basis for comparison when they or another physician next see the patient. For example, a sketch is a concise way of conveying the location and size of a nodule in the prostate gland (Fig. 2.2). In electronic health record systems, these hand drawings are increasingly being replaced in the medical record by text-based descriptions or photographs (Sanders et al. 2013).

As should be clear from these examples, the idea of data is inextricably bound to the idea of **data recording**. Physicians and other health care personnel are taught from the outset that it is crucial that they do not trust their memory when caring for patients. They must record their observations, as well as the actions they have taken and the rationales for those actions, for later communication to themselves and other people. A glance at a medical record will quickly reveal the wide variety of data-recording techniques that have evolved. The range goes from narrative text to commonly understood shorthand notation to cryptic symbols that only specialists can understand; for example, few physicians without specialized training know how to interpret the data-recording conventions of an ophthalmologist (Fig. 2.3). The notations may be highly structured records with brief text or numerical information, machine-generated tracings of analog signals, photographic images (of the patient or of radiologic or other studies), or drawings. This range of data-recording conventions presents significant challenges to the person implementing electronic health record systems.

2.1.2 Who Collects the Data?

Health data on patients and populations are gathered by a variety of health professionals. Although conventional ideas of the **healthcare team** evoke images of coworkers treating ill patients, the team has much broader responsibilities than treatment per se; data collection and recording are a central part of its task.

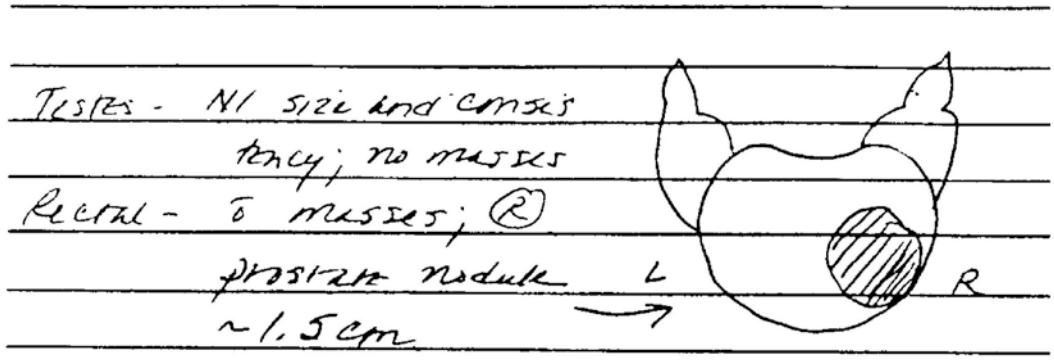


Fig. 2.2 A physician's hand-drawn sketch of a prostate nodule. Drawings may convey precise information more easily and compactly than a textual description, but are less common in electronic health records compared to paper charts

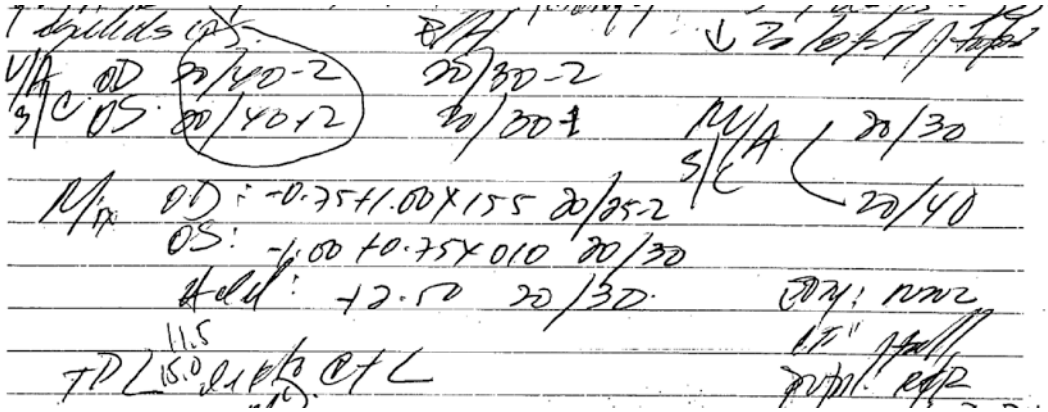


Fig. 2.3 An ophthalmologist's report of an eye examination. Most physicians trained in other specialties would have difficulty deciphering the symbols that the ophthalmologist has used. (Image courtesy of Nita Valikodath, MD, with permission)

Physicians are key players in the process of data collection and interpretation. They converse with a patient to gather narrative descriptive data on the chief complaint, past illnesses, family and social information, and the system review. They examine the patient, collecting pertinent data and recording them during or at the end of the visit. In addition, they generally decide what additional data to collect by ordering laboratory or radiologic studies and by observing the patient's response to therapeutic interventions (yet another form of data that contributes to patient assessment).

In both outpatient and hospital settings, nurses play a central role in making observations and recording them for future reference. The data that they gather contribute to nursing care plans as well as to the assessment of patients by physicians and by other health care staff. Thus, nurses' training includes instruction in careful and accurate observation, history taking, and examination of the patient. Because nurses typically spend more time with patients than physicians do, especially in the hospital setting, nurses often build relationships with patients that uncover information and insights that contribute to proper diagnosis, to understanding of pertinent psychosocial issues, or to proper planning of therapy or discharge management (Fig. 2.4). The role of

Fig. 2.4 Nurses often develop close relationships with patients. These relationships may allow the nurse to make observations that are missed by other staff. This ability is just one of the ways in which nurses play a key role in data collection and recording. (Photograph courtesy of Susan Ostmo, with permission)

information systems in contributing to patient care tasks such as care planning by nurses is the subject of ▶ Chap. 19.

Various other health care workers contribute to the data-collection process. Office staff and admissions personnel gather demographic and financial information. Physical or respiratory therapists record the results of their treatments and often make suggestions for further management. Laboratory personnel perform tests on biological samples, such as blood or urine, and record the results for later use by physicians and nurses. Radiology technicians perform X-ray examinations; radiologists interpret the resulting data and report their findings to the patients' physicians. Pharmacists may interview patients about their medications or about drug allergies and then monitor the patients' use of prescription drugs. Increasingly, health professionals such as physician assistants, nurse practitioners, nurse anesthetists, nurse midwives, psychologists, chiropractors, and optometrists are assuming patient care responsibilities. As these examples suggest, many different individuals employed in health care settings gather, record, and make use of patient data in their work.

Finally, there are the technological devices that generate data—laboratory instruments, imaging machines, monitoring equipment in intensive care units, and measurement devices that take a single reading (such as thermometers, ECG machines, sphygmomanometers for taking blood pressure, and spirometers for testing lung function). Sometimes such a device produces a paper report suitable for inclusion in a traditional medical record. Sometimes the device indicates a result on a gauge or traces a result that must be read by an operator and then recorded in the patient's chart. Sometimes a trained specialist must interpret the output. Increasingly, however, the devices feed their results directly into computer equipment so that the data can be analyzed or formatted for electronic storage in the electronic health record (see ▶ Chap. 16), thereby allowing access to information is through computer workstations, hand-held tablets, or even mobile devices.

2.2 Uses of Health Data

Health data are recorded for a variety of purposes. Clinical data may be needed to support the proper care of the patient from whom they were obtained, but they also may contribute to the good of society through the aggregation and analysis of data regarding populations of individuals (supporting clinical research or public health assessments; see ▶ Chaps. 20 and 28). Traditional data-recording techniques and a paper record may have worked reasonably well when care was given by a single physician over the life of a patient. However, given the increased complexity of modern health care, the broadly trained team of individuals who are involved in a patient's care, the need for multiple providers to access a patient's data and to communicate effectively with one another through the chart, and the need for aggregating clinical data from multiple individuals to support population health, the electronic health record has created new possibilities for improv-

ing the health care delivery process that were not feasible a generation ago. We will discuss these topics in more detail later in this chapter and in ▶ Chaps. 16 and 20.

2.2.1 Create the Basis for the Historical Record

Any student of science learns the importance of collecting and recording data meticulously when carrying out an experiment. Just as a scientific laboratory notebook provides a record of precisely what an investigator has done, the experimental data observed, and the rationale for intermediate decision points, medical records are intended to provide a detailed compilation of information about individual patients:

- What is the patient's history (development of a current illness; other diseases that coexist or have resolved; pertinent family, social, and demographic information)?
- What symptoms has the patient reported? When did they begin, what has seemed to aggravate them, and what has provided relief?
- What physical signs have been noted on examination?
- How have signs and symptoms changed over time?
- What laboratory results have been, or are now, available?
- What radiologic and other special studies have been performed?
- What medications are being taken and are there any allergies?
- What other interventions have been undertaken?
- What is the reasoning behind the management decisions?

Each new patient problem and its management can be viewed as a therapeutic experiment, inherently confounded by uncertainty, with the goal of answering three questions when the experiment is over:

1. What was the nature of the disease or symptom?
2. What was the treatment decision?
3. What was the outcome of that treatment?

As is true for all experiments, one purpose is to learn from experience through careful observation and recording of data. The lessons learned in a given encounter may be highly individualized (e.g., the physician may learn how a specific patient tends to respond to pain or how family interactions tend to affect the patient's response to disease). On the other hand, the value of some experiments may be derived only by pooling of data from many patients who have similar problems and through the analysis of the results of various treatment options to determine efficacy.

Although laboratory research has contributed dramatically to our knowledge of human disease and treatment, it is careful observation and recording by skilled health care personnel that has always been fundamental to the effective generation of new knowledge about patient care. We learn from the aggregation of information from large numbers of patients; thus, the historical record for individual patients is of inestimable importance to clinical research.

2.2.2 Support Communication Among Providers

A central function of structured data collection and recording in health care settings is to assist personnel in providing coordinated care to a patient over time. Most patients who have significant medical conditions are seen over months or years on several occasions for one or more problems that require ongoing evaluation and treatment. Given the increasing numbers of elderly patients in many cultures and health care settings, the care given to a patient is less oriented to diagnosis and treatment of a single disease episode and increasingly focused on management of one or more chronic disorders—possibly over many years.

It was once common for patients to receive essentially all their care from a single provider: the family doctor who tended both children and adults, often seeing the patient over many or all the years of that person's life. We tend to picture such physicians as having especially close relationships with their patients—knowing the family and sharing in many of the patient's life events, especially in smaller communities. Such

doctors nonetheless kept records of all encounters so that they could refer to data about past illnesses and treatments as a guide to evaluating future care issues.

In the world of modern medicine, the emergence of subspecialization and the increasing provision of care by teams of health professionals have placed new emphasis on the central role of the medical record. Over the past several decades, shared access to a paper chart (Fig. 2.5) has largely been replaced by clinicians accessing electronic records, sometimes conferring as they look at the same computer screen (Fig. 2.6). Now the record not only contains observations by a physician for reference on the next visit but also serves as a communication mechanism among physicians and other medical personnel, such as physical or respiratory therapists, nursing staff, radiology technicians, social workers, or discharge planners. In many outpatient settings, patients receive care over time from a variety of physicians—colleagues covering for the primary physician, or specialists to whom the patient has been referred, or a managed care organization's case manager. It is not uncommon to hear complaints from patients who remember the days when it was possible to receive essentially all their care from a single physician whom they had come to trust and who knew them well. Physicians are sensitive to this issue and therefore recognize the importance of the medical record in ensuring quality and **continuity of care** through adequate recording of the details and logic of past interventions and ongoing treatment plans. This idea is of particular importance in a health care system in which chronic diseases rather than care for trauma or acute infections increasingly dominate the basis for interactions between patients and their doctors.

Fig. 2.5 One role of the medical record: a communication mechanism among health professionals who work together to plan patient care. (Photograph courtesy of Janice Anne Rohn)

Fig. 2.6 Today similar communication sessions occur around a computer screen rather than a paper chart (see Fig. 2.5). (Photograph courtesy of Susan Ostmo with permission)

2.2.3 Anticipate Future Health Problems

Providing high-quality health care involves more than responding to patients' acute or chronic health problems. It also requires educating patients about the ways in which their environment and lifestyles can contribute to, or reduce the risk of, future development of disease. Similarly, data gathered routinely in the ongoing care of a patient may suggest that he or she is at high risk of developing a specific problem even though he or she may feel well and be without symptoms at present. Clinical

data therefore are important in screening for risk factors, following patients' risk profiles over time, and providing a basis for specific patient education or preventive interventions, such as diet, medication, or exercise. Perhaps the most common examples of such ongoing risk assessment in our society are routine monitoring for excess weight, high blood pressure, and elevated serum cholesterol levels. In these cases, abnormal data may be predictive of later symptomatic disease; optimal care requires early intervention before the complications have an opportunity to develop fully.

2.2.4 Record Standard Preventive Measures

The medical record also serves as a source of data on interventions that have been performed to prevent common or serious disorders. Sometimes the interventions involve counseling or educational programs (for example, regarding smoking cessation, measures for stopping drug abuse, safe sex practices, or dietary changes). Other important preventive interventions include immunizations: the vaccinations that begin in early childhood and continue throughout life, including special treatments administered when a person will be at particularly high risk (e.g., injections to protect people from certain highly communicable diseases, administered before travel to areas where such diseases are endemic). When a patient comes to his local hospital emergency room with a laceration, the physicians routinely check for an indication of when he most recently had a tetanus immunization. When easily accessible in the record (or from the patient), such data can prevent unnecessary treatments (in this case, a repeat injection) that may be associated with risk or significant cost.

2.2.5 Identify Deviations from Expected Trends

Data often are useful in medical care only when viewed as part of a continuum over time. An example is the routine monitoring of children for normal growth and development by pediatricians (Fig. 2.7). Single data points regarding height and weight may have limited use by themselves; it is the trend in such data points observed over months or years that may provide the first clue to a medical problem. It is accordingly common for such parameters to be recorded on special charts or forms that make the trends easy to discern at a glance. Women who want to have a child often keep similar records of body temperature. By measuring temperature daily and recording the values on special charts, women can identify the slight increase in temperature that accompanies ovulation and thus may discern the days of maximum fertility. Many physicians will ask a patient to keep such graphical records so that they can later discuss the data with the patient and include the scanned or photographed graph in the electronic record for ongoing reference. Such graphs are increasingly captured and displayed for viewing by clinicians as a feature of a patient's medical record.

2.2.6 Provide a Legal Record

Another use of health data, once they are charted and analyzed, is as the foundation for a legal record to which the courts can refer if necessary. The medical record is a legal document; the responsible individual must certify or sign most of the clinical information that is recorded. In addition, the chart generally should describe and justify both the presumed diagnosis for a patient and the choice of management.

We emphasized earlier the importance of recording data; in fact, data do not exist in a generally useful form unless they are recorded. The legal system stresses this point as well. Providers' unsubstantiated memories of what they observed or why they took some action are of little value in the courtroom. The medical record is the foundation for determining whether proper care was delivered. Thus, a well-maintained record is a source of protection for both patients and their physicians.

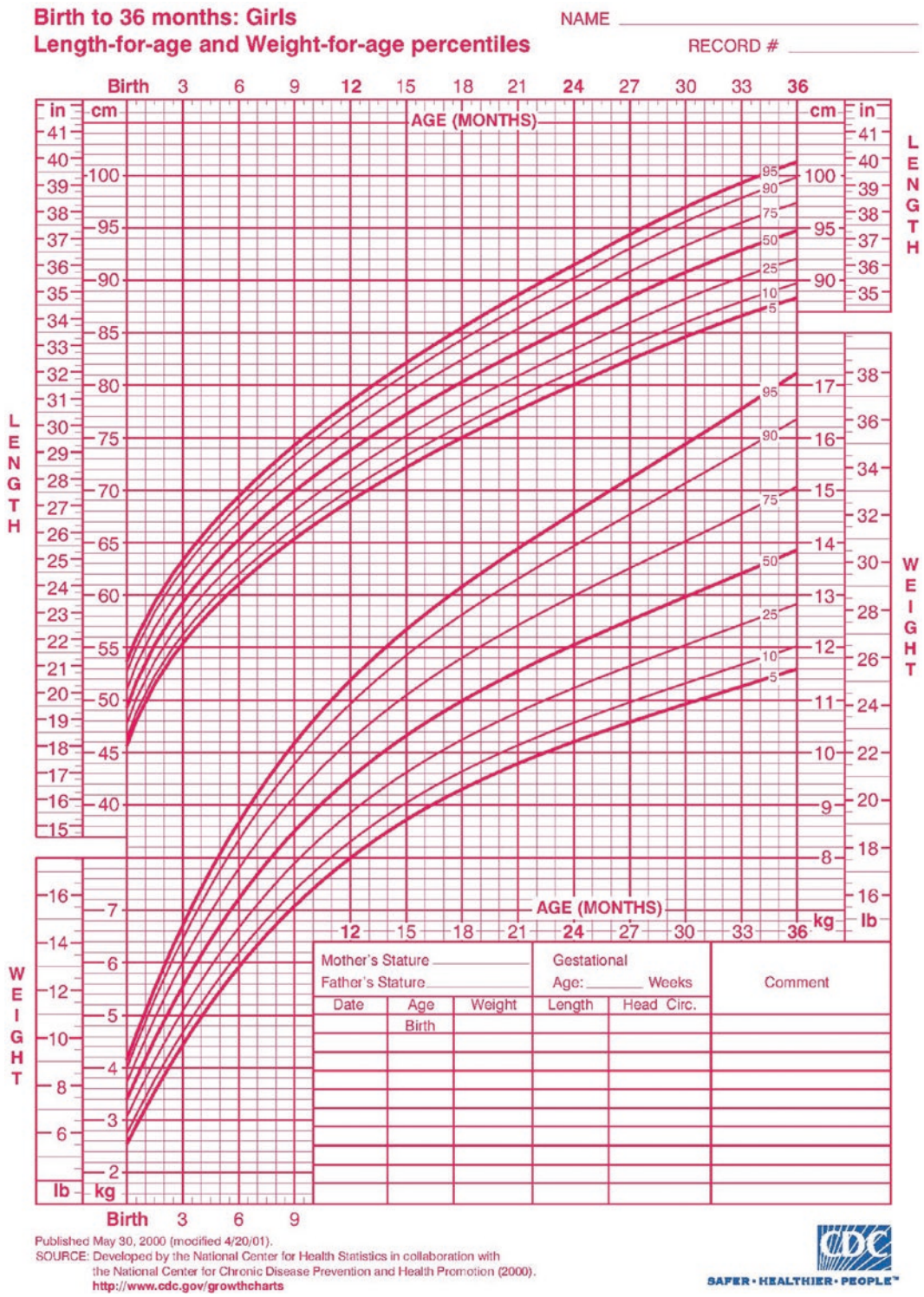


Fig. 2.7 A pediatric growth chart. Single data points would not be useful; it is the changes in values over time that indicate whether development is progressing normally. (Source: National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000). ▶ http://www.cdc.gov/growthcharts)

2.2.7 Support Clinical Research

Although experience caring for individual patients provides physicians with special skills and enhanced judgment over time, it is only by formally analyzing data collected from large numbers of patients that researchers can develop and validate new clinical knowledge of general applicability. Thus, another use of clinical data is to support research through the aggregation and statistical or other analysis of observations gathered from populations of patients (see ▶ Chap. 1).

A **randomized clinical trial** (**RCT**) (see also ▶ Chaps. 15 and 29) is a common method by which specific clinical questions are addressed experimentally. RCTs typically involve the random assignment of matched groups of patients to alternate treatments when there is uncertainty about how best to manage the patients' problem. The variables that might affect a patient's course (e.g., age, gender, weight, coexisting medical problems) are measured and recorded. As the study progresses, data are collected meticulously to provide a record of how each patient fared under treatment and precisely how the treatment was administered. By pooling such data, sometimes after years of experimentation (depending on the time course of the disease under consideration), researchers may be able to demonstrate a statistical difference among the study groups depending on precise characteristics present when patients entered the study or on the details of how patients were managed. Such results then help investigators to define the standard of care for future patients with the same or similar problems.

Medical knowledge also can be derived from the analysis of large patient data sets or registries, even when the patients were not specifically enrolled in an RCT, often referred to as **retrospective studies**. Much of the research in the field of epidemiology involves analysis of population-based data of this type. Our knowledge of the risks associated with cigarette smoking, for example, is based on irrefutable statistics derived from large populations of individuals with and without lung cancer, other pulmonary problems, and heart disease.

2.3 Rationale for the Transition from Paper to Electronic Documentation

The preceding description of medical data and their uses emphasizes the positive aspects of information storage and retrieval in the record. During the past several decades, the United States and many other countries have gradually transitioned from traditional paper records to electronic health records. The rationale for this transition has largely been to create the potential for enhancing the record's effectiveness for its intended uses, as summarized in the previous section.

2.3.1 Pragmatic and Logistical Issues

Recall, first, that data cannot effectively serve the delivery of health care unless they are recorded. Their optimal use depends on positive responses to the following questions:

- Can I find the data I need when I need them?
- Can I find the medical record in which they are recorded?
- Can I find the data within the record?
- Can I find what I need quickly?
- Can I read and interpret the data once I find them?
- Can I update the data reliably with new observations in a form consistent with the requirements for future access by me or other people?

The traditional paper record created situations in which people too often answered such questions in the negative. For example:

- The patient's paper chart was too often unavailable when the health care professional needed it. It could be in use by someone else at another location; it might have been misplaced despite the record-tracking system of the hospital, clinic, or office (◘ Fig. 2.8); or it might have been taken by someone unintentionally and is now buried on a desk.

Fig. 2.8 Storage room for paper-based medical records. These paper repositories have largely been replaced as EHRs have become more standard. (Photograph courtesy of Janice Anne Rohn)

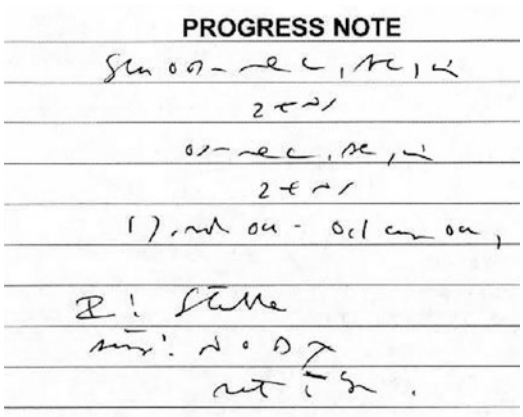


Fig. 2.9 Written entries were standard in paper records, yet handwritten notes could be illegible. Notes that cannot be interpreted by other people due to illegibility may cause delays in treatment or inappropriate care—an issue that is largely eliminated when EHRs are used. (Image courtesy of Emily Cole, MD, with permission)

- It could be difficult to find the information required in either the paper or electronic record. The data might have been known previously but never recorded due to an oversight by a physician or other health professional. Poor organization or sheer size of either the paper or electronic record may lead the user to spend an inordinate time searching for the data, especially for patients who have long and complicated histories.
- Paper records were notoriously difficult to read. It was not uncommon to hear one physician asking another as they peered together into a chart: "What is that word?" "Is that a two or a five?" "Whose signature is that?" Illegible and sloppy entries was too often a major obstruction to effective use of the paper chart (Fig. 2.9).
- When a paper chart was unavailable, the health care professional still had to provide patient care. Thus, providers would often make do without past data, basing their decisions instead on what the patient could tell them and on what their examination revealed. They then wrote a note for inclusion in the chart—when the chart was located! In a large institution with thousands of medical records, it is not surprising that such loose notes often failed to make it to the patient's chart or were filed out of sequence so that the actual chronology of management was disrupted in the record.
- When patients who have chronic or frequent diseases are seen over months or years, their paper records grew so large that the charts had to be broken up into multiple volumes. When a hospital clinic or emergency room ordered the patient's chart, only the most recent volume typically was provided. Old but pertinent data might have been in early volumes that were stored offsite or are otherwise unavailable. Alternatively, an early volume could be mistaken for the most recent volume, misleading its users and resulting in documents being inserted out of sequence.

▶ Chapter 16 describes approaches that electronic health record systems have taken toward addressing these practical problems in the use of the paper record. It is for this reason that almost all hospitals, health systems, and individual practitioners have implemented EHRs–further encouraged in the US by Federal incentive programs that helped to cover the costs of EHR acquisition and maintenance (see ▶ Chaps. 1 and 31). That said, one challenge is that electronic health records in the US have been criticized for being composed of bloated, lengthy documentation that

is often focused on billing and compliance over clinical care (▶ Chaps. 16 and 31).

2

2.3.2 Redundancy and Inefficiency

To be able to find data quickly in the medical record, health professionals developed a variety of techniques in paper documentation that provided redundant recording to match alternate modes of access. For example, the result of a radiologic study typically was entered on a standard radiology reporting form, which was filed in the portion of the chart labeled "X-ray." For complicated procedures, the same data often were summarized in brief notes by radiologists in the narrative part of the chart, which they entered at the time of studies because they knew that the formal report would not make it back to the chart for 1 or 2 days. In addition, the study results often were mentioned in notes written by the patient's admitting and consulting physicians and by the nursing staff. Although there may have been good reasons for recording such information multiple times in different ways and in different locations within the paper chart, the combined bulk of these notes accelerated the physical growth of the document and, accordingly, complicated the chart's logistical management. Furthermore, it became increasingly difficult to locate specific patient data as the chart succumbed to "obesity". The predictable result was that someone would write yet another redundant entry, summarizing information that it took hours to track down – and creating potential sources for transcription error.

A similar inefficiency occured because of a tension between opposing goals in the design of reporting forms used by many laboratories. Most health personnel preferred a consistent, familiar form, often with color-coding, because it helped them to find information more quickly (◘ Fig. 2.10). For example, a physician might know that a urinalysis report form is printed on yellow paper and records the bacteria count halfway down the middle column of the form. This knowledge allowed the physician to work backward quickly in the laboratory section of the chart to find the most recent urinalysis sheet and to check at a glance the bacterial count. The problem is that such forms typically stored only sparse information. It was clearly suboptimal if a rapidly growing physical chart was filled with sheets of paper that reported only a single data element.

	Ref Range & Units	03:55	1d ago	4d ago	2wk ago	2wk ago	2wk ago	2wk ago
WHITE CELL COUNT	3.50 - 10.80 K/cu mm	9.23	8.95	6.39	8.02	5.81	6.74	7.40
RED CELL COUNT	4.50 - 6.00 M/cu mm	2.19 ↓	2.19 ↓	2.13 ↓	2.53 ↓	2.46 ↓	2.48 ↓	2.48 ↓
HEMOGLOBIN	13.5 - 17.5 g/dL	6.6 ↓	6.7 ↓	6.5 ↓	7.7 ↓	7.5 ↓	7.7 ↓	8.0 ↓
HEMATOCRIT	41.0 - 53.0 %	22.0 ↓	22.5 ↓	22.5 ↓	25.4 ↓	24.5 ↓	24.5 ↓	24.6 ↓
MCV	80.0 - 100.0 fL	100.5 ↑	102.7 ↑	105.6 ↑	100.4 ↑	99.6	98.8	99.2
MCHC	32.0 - 36.0 g/dL	30.0 ↓	29.8 ↓	28.9 ↓	30.3 ↓	30.6 ↓	31.4 ↓	32.5
RDW SD	35.1 - 46.3 fL	63.5 ↑	60.4 ↑	62.2 ↑	59.5 ↑	60.0 ↑	59.4 ↑	57.2 ↑
PLATELET COUNT	150 - 400 K/cu mm	191	188	146 ↓	150	143 ↓	147 ↓	146 ↓
MPV	9.7 - 12.3 fL	11.0	11.1	11.8	11.3	11.4	11.4	11.4
NRBC%	0.0 - 0.3 %	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NRBC#	0.00 - 0.02 K/cu mm	0.00	0.00	0.00	0.00	0.00	0.00	0.00

◘ **Fig. 2.10** Laboratory reporting forms present medical data in a consistent, familiar format (in this case a complete blood count (CBC)). (Photograph courtesy of Jimy Chen, with permission)

2.3.3 Influence on Clinical Research

Anyone involved in a clinical research project based on retrospective data review from paper records can attest to the tediousness of flipping through myriad medical records. For all the reasons described in ▶ Chap. 1, it is arduous to sit with stacks of patient records, extracting data and formatting them for structured statistical analysis, and the process is vulnerable to transcription errors. Observers often wonder how much medical knowledge is sitting untapped in old paper medical records because there is no easy way to analyze experience across large populations of patients from the past without first extracting pertinent data from those charts.

Let's contrast such retrospective review with paper and electronic medical records. Suppose, for example, that physicians on a medical consultation service notice that patients receiving a certain common oral medication for diabetes (call it drug X) seem to be more likely to have significant postoperative hypotension (low blood pressure) than do surgical patients receiving other medications for diabetes. The doctors have based this hypothesis—that drug X influences postoperative blood pressure—on only a few recent observations, however, so they decide to look into existing hospital records to see whether this correlation has occurred with sufficient frequency to warrant a formal investigation. One efficient way to follow up on their theory from existing medical data would be to examine the hospital records of all patients who have diabetes and also have been admitted for surgery. The task would then be to examine those records (difficult and arduous with paper charts as will be discussed shortly, but subject to automated analysis in the case of EHRs) and to note for all patients (1) whether they were taking drug X when admitted and (2) whether they had postoperative hypotension. If the statistics showed that patients receiving drug X were more likely to have low blood pressure after surgery than were similar diabetic patients receiving alternate treatments, a controlled trial (prospective observation and data gathering) might well be appropriate.

Note the distinction between **retrospective chart review** to investigate a question that was not a subject of study at the time the data were collected and **prospective studies** in which the clinical hypothesis is known in advance and the **research protocol** is designed specifically to collect future data that are relevant to the question under consideration (see also ▶ Chaps. 15 and 29). Subjects are assigned **randomly** to different study groups to help prevent researchers—who are bound to be biased, having developed the hypothesis—from unintentionally skewing the results by assigning a specific class of patients all to one group. For the same reason, to the extent possible, the studies are **double blind**; i.e., neither the researchers nor the subjects know which treatment is being administered. Such blinding is of course impractical when it is obvious to patients or physicians what therapy is being given (such as surgical procedures versus drug therapy). Prospective, randomized, double-blind studies are considered the best method for determining optimal management of disease, but it is often impractical to carry out such studies, and then methods such as retrospective chart review may be used.

Returning to our example, consider the problems in paper chart review that the researchers used to encounter in addressing the postoperative hypotension question retrospectively. First, they would have to identify the charts of interest: the subset of medical records dealing with surgical patients who are also diabetic. In a hospital record room filled with thousands of charts, the task of chart selection was often overwhelming. Medical records departments generally did keep indexes of diagnostic and procedure codes cross-referenced to specific patients (see ▶ Sect. 2.5.1). Thus, it sometimes was possible to use such an index to find all charts in which the discharge diagnoses included diabetes and the procedure codes included major surgical procedures. The researcher might then have compiled a list of patient identification numbers and have the individual charts pulled from the file room for review.

The researchers' next task was to examine each paper chart serially to find out what treatment each patient was receiving for diabetes at the time of the surgery and to determine whether the patient had postoperative hypotension. Finding such information tended to be

extremely time-consuming. Where should the researcher look for it? The admission drug orders might have shown what the patient received for diabetes control, but it would also have been wise to check the medication sheets to see whether the therapy was also administered (as well as ordered) and the admission history to see whether a routine treatment for diabetes, taken right up until the patient entered the hospital, was not administered during the inpatient stay. Information about hypotensive episodes might be similarly difficult to locate. The researchers might start with nursing notes from the recovery room or with the anesthesiologist's datasheets from the operating room, but the patient might not have been hypotensive until after leaving the recovery room and returning to the ward. So the nursing notes from the ward would need to be checked too, as well as vital signs sheets, physicians' progress notes, and the discharge summary.

It should be clear from this example that retrospective paper chart review was a laborious and tedious process and that people performing it were prone to make transcription errors and to overlook key data. EHRs offer an enormous opportunity (▶ Chap. 16) to facilitate the chart review and clinical research process. They have obviated the need to retrieve hard copy charts; instead, researchers are increasingly using computer-based data retrieval and analysis techniques to do most of the work (finding relevant patients, locating pertinent data, and formatting the information for statistical analyses). Researchers can use similar techniques to harness computer assistance with data management in prospective clinical trials (▶ Chap. 29).

2.3.4 The Passive Nature of Paper Records

The traditional manual system has another limitation that would have been meaningless until the emergence of the computer age. A manual archival system is inherently passive; the charts sit waiting for something to be done with them. They are insensitive to the characteristics of the data recorded within their pages, such as legibility, accuracy, or implications for patient management. They cannot take an active role in responding appropriately to those implications.

EHR systems have changed our perspective on what health professionals can expect from the medical chart. Automated record systems introduce new opportunities for dynamic responses to the data that are recorded in them. As described in many of the chapters to follow, computational techniques for data storage, retrieval, and analysis make it feasible to develop record systems that (1) monitor their contents and generate warnings or advice for providers based on single observations or on logical combinations of data; (2) provide automated quality control, including the flagging of potentially erroneous data; or (3) provide feedback on patient-specific or population-based deviations from desirable standards.

2.4 New Kinds of Data and the Resulting Challenges

The revolution in human genetics that emerged with the **Human Genome Project** in the 1990s already has had a profound effect on the diagnosis, prognosis, and treatment of disease (Vamathevan and Birney 2017). The vast amounts of data that are generated in biomedical research (see ▶ Chaps. 11 and 28), and that can be pooled from patient datasets to support clinical research (▶ Chap. 29) and public health (▶ Chap. 20), have created new opportunities as well as challenges. Researchers are finding that the amount of data that they must manage and assess has become so large that they often find that they lack either the capabilities or expertise to handle the analytics that are required. This problem, sometimes dubbed the "big data" problem, has gathered the attention of government agencies as well.[2]

2 Big Data Senior Steering Group. The Federal Big Data Research and Development Strategic Plan. Available at: ▶ https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/bigdatardstrategicplan-nitrd_final-051916.pdf (Accessed 6/28/2019).

Some suggest that the genetic material itself will become our next-generation method for storing large amounts of data (Erlich and Zielinski 2017). Data analytics, and the management of large amounts of genomic/proteomic or clinical/public-health data, have accordingly become major research topics and key opportunities for new methodology development by biomedical informatics and **data scientists** (Adler-Milstein and Jha 2013; Brennan et al. 2018; Bycroft et al. 2018).

The issues that arise are practical as well as scientifically interesting. For example, developers of EHRs have begun to grapple with questions regarding how they might store an individual's personal genome within the electronic health record. New standards will be required, and tactical questions need answering regarding, for example, whether to store an entire genome or only those components (e.g., genetic markers) that are already reasonably well understood (Masys et al. 2012; Haendel et al. 2018). In cancer, for example, where mutations in cell lines can occur, an individual may actually have many genomes represented among his or her cells. These issues will undoubtedly influence the evolution of data systems and EHRs, as well as the growth of **precision medicine** (see ► Chap. 30), in the years ahead (Relling and Evans 2015).

2.5 The Structure of Clinical Data

Scientific disciplines generally develop a precise terminology or notation that is standardized and accepted by all workers in the field. Consider, for example, the universal language of chemistry embodied in chemical formulae, the precise definitions and mathematical equations used by physicists, the predicate calculus used by logicians, or the conventions for describing circuits used by electrical engineers. Medicine is remarkable for its failure to develop a widely accepted standardized vocabulary and **nomenclature**, and many observers believe that a true "scientific" basis for the field will be impossible until this problem is addressed (see ► Chap. 8). Other people argue that common references to the "art" of medicine reflect an important distinction between medicine and the "hard" sciences; these people question whether it is possible to introduce too much standardization into a field that prides itself in humanism.

The debate has been accentuated by the introduction of computers for data management, because such machines tend to demand conformity to data standards and definitions. Otherwise, issues of data retrieval and analysis are confounded by discrepancies between the meanings intended by the observers or recorders and those intended by the individuals retrieving information or doing data analysis. What is an "upper respiratory infection"? Does it include infections of the trachea or of the main stem bronchi? How large does the heart have to be before we can refer to "cardiomegaly"? How should we deal with the plethora of disease names based on eponyms (e.g., Alzheimer's disease, Hodgkin's disease) that are not descriptive of the illness and may not be familiar to all practitioners? What do we mean by an "acute abdomen"? Are the boundaries of the abdomen well agreed on? What are the time constraints that correspond to "acuteness" of abdominal pain? Is an "ache" a pain? What about "occasional" cramping?

Imprecision and the lack of a standardized vocabulary are particularly problematic when we wish to aggregate data recorded by multiple health professionals or to analyze trends over time. Without a controlled, predefined vocabulary, data interpretation is inherently complicated, and the automatic summarization of data may be impossible. For example, one physician might note that a patient has "shortness of breath." Later, another physician might note that she has "dyspnea." Unless these terms are designated as synonyms, an automated program will fail to indicate that the patient had the same problem on both occasions.

Regardless of arguments regarding the "artistic" elements in medicine, the need for health personnel to communicate effectively is clear both in acute care settings and when patients are seen over long periods. Both high-quality care and scientific progress depend on some standardization in terminology. Otherwise, differences in intended meaning or in defining criteria will lead to miscommunication, improper interpretation, and potentially negative consequences for the patients involved.

Given the lack of formal definitions for many medical terms, it is remarkable that medical workers communicate as well as they do. Only occasionally is the care for a patient clearly compromised by miscommunication. If EHRs are to become dynamic and responsive manipulators of patient data, however, their encoded logic must be able to presume a specific meaning for the terms and data elements entered by the observers. This point is discussed in greater detail in ▶ Chap. 8, which deals in part with the multiple efforts to develop healthcare computing standards, including a shared, controlled terminology for biomedicine.

2.5.1 Coding Systems

We are used to seeing figures regarding the growing incidences of certain types of tumors, deaths from influenza during the winter months, and similar health statistics that we tend to take for granted. How are such data accumulated? Their role in health planning and health care financing is clear, but electronic health records provide the infrastructure for aggregating individual patient data to learn more about the health status of the populations in various communities (see ▶ Chap. 20).

Because of the needs to know about health trends for populations and to recognize epidemics in their early stages, there are various health-reporting requirements for hospitals (as well as other public organizations) and practitioners. For example, cases of gonorrhea, syphilis, and tuberculosis generally must be reported to local public-health organizations, which code the data to allow trend analyses over time. The Centers for Disease Control and Prevention in Atlanta (CDC) then pool regional data and report national as well as local trends in disease incidence, bacterial-resistance patterns, etc.

Another kind of reporting involves the coding of all discharge diagnoses for hospitalized patients, plus coding of certain procedures (e.g., type of surgery) that were performed during the hospital stay. Such codes are reported to state and federal health-planning and analysis agencies and also are used internally at the institution for case-mix analysis (determining the relative frequencies of various disorders in the hospitalized population and the average length of stay for each disease category), for quality improvement, and for research. For such data to be useful, the codes must be well defined as well as uniformly applied and accepted.

The World Health Organization publishes a diagnostic coding scheme called the International Classification of Disease (ICD). The 10th revision of this standard, ICD-10-CM (clinical modification),[3] is currently in use in much of the world (see ▶ Chap. 8). ICD-10-CM is used by all nonmilitary hospitals in the United States for discharge coding, and must be reported on the bills submitted to most insurance companies (◘ Fig. 2.11). Pathologists have developed another widely used diagnostic coding scheme; originally known as Systematized Nomenclature of Pathology (SNOP), it was expanded to the Systematized Nomenclature of Medicine (SNOMED) and then merged with the Read Clinical Terms from Great Britain to become SNOMED-CT (Stearns et al. 2001; Lee et al. 2014). In recent years, support for SNOMED-CT was assumed by the International Health Terminology Standards Development Organization, based in Copenhagen, now renamed SNOMED International and relocated to London.[4] Another coding scheme, developed by the American Medical Association, is the Current Procedural Terminology (CPT) (Hirsch et al. 2015). It is similarly widely used in producing bills for services rendered to patients. More details on such schemes are provided in ▶ Chap. 8. What warrants emphasis here, however, is the motivation for the codes' development: health care personnel need standardized terms that can support pooling of data for analysis and can provide criteria for determining charges for individual patients.

The historical roots of a coding system reveal themselves as limitations or idiosyncrasies when the system is applied in more general clinical settings. For example, ICD-10-CM was derived from a classification scheme developed for epidemiologic reporting. Consequently, it has over 60 separate codes for describing tuber-

3 ▶ http://www.icd10data.com/ (Accessed 11/1/2019).

4 ▶ http://snomed.org/ (Accessed 5/6/2019).

J45 Asthma
Includes: allergic (predominantly) asthma, allergic bronchitis NOS, allergic rhinitis with asthma, atopic asthma, extrinsic allergic asthma, hay fever with asthma, idiosyncratic asthma, intrinsic nonallergic asthma, nonallergic asthma

Use additional code to identify: exposure to environmental tobacco smoke (Z77.22), exposure to tobacco smoke in the perinatal period (P96.81), history of tobacco use (Z87.891), occupational exposure to environmental tobacco smoke (Z57.31), tobacco dependence (F17.-), tobacco use (Z72.0)

Excludes: detergent asthma (J69.8), eosinophilic asthma (J82), lung diseases due to external agents (J60-J70), miner's asthma (J60), wheezing NOS (R06.2), wood asthma (J67.8), asthma with chronic obstructive pulmonary disease (J44.9), chronic asthmatic (obstructive) bronchitis (J44.9), chronic obstructive asthma (J44.9)

J45.2 Mild intermittent asthma
J45.20 Mild intermittent asthma, uncomplicated
Mild intermittent asthma NOS
J45.21 Mild intermittent asthma with (acute) exacerbation
J45.22 Mild intermittent asthma with status asthmaticus
J45.3 Mild persistent asthma
J45.30 Mild persistent asthma, uncomplicated
Mild persistent asthma NOS
J45.31 Mild persistent asthma with (acute) exacerbation
J45.32 Mild persistent asthma with status asthmaticus
J45.4 Moderate persistent asthma
J45.40 Moderate persistent asthma, uncomplicated
Moderate persistent asthma NOS
J45.41 Moderate persistent asthma with (acute) exacerbation
J45.42 Moderate persistent asthma with status asthmaticus
J45.5 Severe persistent asthma
J45.50 Severe persistent asthma, uncomplicated
Severe persistent asthma NOS
J45.51 Severe persistent asthma with (acute) exacerbation
J45.52 Severe persistent asthma with status asthmaticus
J45.9 Other and unspecified asthma
J45.90 Unspecified asthma
Asthmatic bronchitis NOS
Childhood asthma NOS
Late onset asthma
J45.901 Unspecified asthma with (acute) exacerbation
J45.902 Unspecified asthma with status asthmaticus
J45.909 Unspecified asthma, uncomplicated
Asthma NOS
J45.99 Other asthma
J45.990 Exercise induced bronchospasm
J45.991 Cough variant asthma
J45.998 Other asthma

Fig. 2.11 The subset of disease categories for asthma taken from ICD-10-CM. (Source: Centers for Medicare and Medicaid Services, US Department of Health and Human Services, ▶ https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs.html, accessed June 28, 2019)

culosis infections. SNOMED versions have long permitted coding of pathologic findings in exquisite detail but only in later years began to introduce codes for expressing the dimensions of a patient's functional status. In a particular clinical setting, none of the common coding schemes is likely to be completely satisfactory. In some cases, the granularity of the code will be too coarse; on the one hand, a hematologist (person who studies blood diseases) may want to distinguish among a variety of hemoglobinopathies (disorders of the structure and function of hemoglobin) lumped under a single code in ICD-10-CM. On the other hand, another practitioner may prefer to aggregate many individual codes—e.g., those for active tuberculosis—into a single category to simplify the coding and retrieval of data.

Such schemes cannot be effective unless health care providers accept them. There is an inherent tension between the need for a coding system that is general enough to cover many different patients and the need for precise and unique terms that accurately apply to a specific patient and do not unduly constrain physicians' attempts to describe what they observe. Yet if physicians view the EHR as a blank sheet of paper on which any unstructured information can be written, the data they record will be unsuitable for dynamic processing, clinical research, and health planning. The challenge is to learn how to meet all these needs. Researchers at many institutions worked for over two decades to develop a unified medical language system (UMLS), a common structure that ties together the various vocabularies that have been created. At the same time, the developers of specific terminologies are continually working to refine and expand their independent coding schemes (Humphreys et al. 1998) (see ► Chap. 8).

2.5.2 The Data-to-Knowledge Spectrum

A central focus in biomedical informatics is the information base that constitutes the "substance of medicine." Workers in the field have tried to clarify the distinctions among three terms frequently used to describe the content of computer-based systems: data, information, and knowledge (Blum 1986; Bernstam et al. 2010). These terms are often used interchangeably. In this volume, we shall refer to a **datum** as a single observational point that characterizes a relationship. It generally can be regarded as the value of a specific parameter for a particular object (e.g., a patient) at a given point in time. The term **information** refers to analyzed data that have been suitably curated and organized so that they have meaning. Data do not constitute information until they have been organized in some way, e.g., for analysis or display. **Knowledge**, then, is derived through the formal or informal analysis (or interpretation) of information that was in turn derived from data. Thus, knowledge includes the results of formal studies and also common sense facts, assumptions, heuristics (strategic rules of thumb), and models—any of which may reflect the experience or biases of people who interpret the primary data and the resulting information.

The observation that patient Brown has a blood pressure of 180/110 is a *datum*, as is the report that the patient has had a myocardial infarction (heart attack). When researchers pool such data, creating information, subsequent analysis may determine that patients with high blood pressure are more likely to have heart attacks than are patients with normal or low blood pressure. This analysis of organized data (information) has produced a piece of knowledge about the world. A physician's belief that prescribing dietary restriction of salt is unlikely to be effective in controlling high blood pressure in patients of low economic standing (because the latter are less likely to be able to afford special low-salt foods) is an additional personal piece of *knowledge*—a **heuristic** that guides physicians in their decision making. Note that the appropriate interpretation of these definitions depends on the context. Knowledge at one level of abstraction may be considered data at higher levels. A blood pressure of 180/110 mmHg is a raw piece of data; the statement that the patient has hypertension is an interpretation of several such data and thus represents a higher level of information. As input to a diagnostic decision aid, however, the presence or absence of hypertension may be requested, in which case the presence of hypertension is treated as a data item.

A **database** is a collection of individual observations without any summarizing analysis. An EHR system is thus primarily viewed as a database—the place where patient data are stored. When properly collated and pooled with other data, these elements in the EHR provide *information* about the patient. A **knowledge base**, on the other hand, is a collection of facts, heuristics, and models that can be used for problem solving and analysis of organized data (information). If the knowledge base provides sufficient structure, including semantic links among knowledge items, the computer itself may be able to apply that knowledge as an aid to case-based problem solving. Many decision-support systems have been called knowledge-based systems, reflect-

ing this distinction between knowledge bases and databases (see ▶ Chap. 26).

2.6 Strategies of Clinical Data Selection and Use

It is illusory to conceive of a "complete clinical data set." All medical databases, and medical records, are necessarily incomplete because they reflect the selective collection and recording of data by the health care personnel responsible for the patient. There can be marked interpersonal differences in both style and problem solving that account for variations in the way practitioners collect and record data for the same patient under the same circumstances. Such variations do not necessarily reflect good practices, however, and much of medical education is directed at helping physicians and other health professionals to learn what observations to make, how to make them (generally an issue of technique), how to interpret them, and how to decide whether they warrant formal recording.

An example of this phenomenon is the difference between the first medical history, physical examination, and summarizing report developed by a medical student and the similar process undertaken by a seasoned clinician examining the same patient. Medical students tend to work from comprehensive mental outlines of questions to ask, physical tests to perform, and additional data to collect. Because they have not developed skills of selectivity, the process of taking a medical history and performing a physical examination may take more than 1 h, after which students develop extensive reports of what they observed and how they have interpreted their observations. It clearly would be impractical, inefficient, and inappropriate for physicians in practice to spend this amount of time assessing every new patient. Thus, part of the challenge for the neophyte is to learn how to ask only the questions that are necessary, to perform only the examination components that are required, and to record only those data that will be pertinent in justifying the ongoing diagnostic approach and in guiding the future management of the patient.

What do we mean by **selectivity** in data collection and recording? It is precisely this process that often is viewed as a central part of the "art" of medicine, an element that accounts for individual styles and the sometimes marked distinctions among clinicians. As is discussed with numerous clinical examples in ▶ Chaps. 3 and 4, the idea of selectivity implies an ongoing decision-making process that guides data collection and interpretation. Attempts to understand how expert clinicians internalize this process, and to formalize the ideas so that they can better be taught and explained, are central in biomedical informatics research. Improved guidelines for such decision making, derived from research activities in biomedical informatics, not only are enhancing the teaching and practice of medicine (Shortliffe 2010) but also are providing insights that suggest methods for developing computer-based decision-support tools.

2.6.1 The Hypothetico-Deductive Approach

Studies of clinical decision makers have shown that strategies for data collection and interpretation may be imbedded in an iterative process known as the **hypothetico-deductive approach** (Elstein et al. 1978; Kassirer and Gorry 1978). As medical students learn this process, their data collection becomes more focused and efficient, and their medical records become more compact. The central idea is one of sequential, staged data collection, followed by data interpretation and the generation of hypotheses, leading to hypothesis-directed selection of the next most appropriate data to be collected. As data are collected at each stage, they are added to the growing database of observations and are used to reformulate or refine the active hypotheses. This process is iterated until one hypothesis reaches a threshold level of certainty (e.g., it is proved to be true, or at least the uncertainty is reduced to a satisfactory level). At that point, a management, disposition, or therapeutic decision can be made.

The diagram in ◘ Fig. 2.12 clarifies this process. As is shown, data collection begins

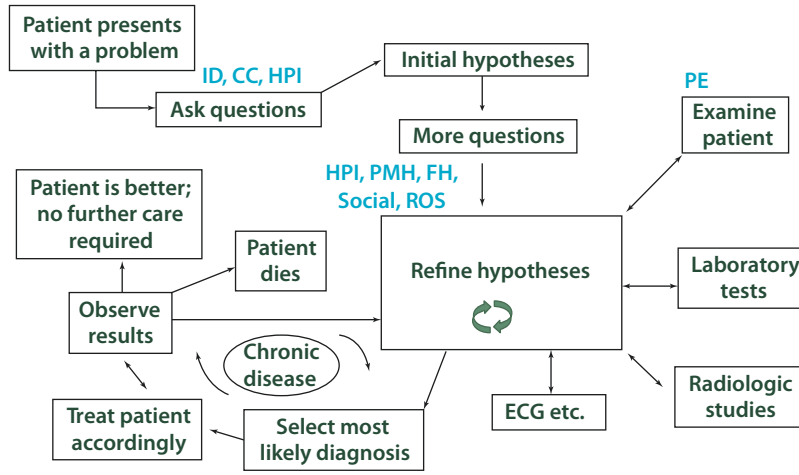


Fig. 2.12 A schematic view of the hypothetico-deductive approach. The process of medical data collection and treatment is intimately tied to an ongoing process of hypothesis generation and refinement. See text for full discussion. *ID* patient identification, *CC* chief complaint, *HPI* history of present illness, *PMH* past medical history, *FH* family history, *Social* social history, *ROS* review of systems, *PE* physical examination

when the patient presents to the physician with some issue (a symptom or disease, or perhaps the need for routine care). The physician generally responds with a few questions that allow one to focus rapidly on the nature of the problem. In the written report, the data collected with these initial questions typically are recorded as the patient identification, chief complaint, and initial portion of the history of the present illness. Studies have shown that an experienced physician will have an initial set of hypotheses (theories) in mind after hearing the patient's response to the first six or seven questions (Elstein et al. 1978). These hypotheses then serve as the basis for selecting additional questions. As shown in Fig. 2.12, answers to these additional questions allow the physician to refine hypotheses about the source of the patient's problem. Physicians refer to the set of active hypotheses as the **differential diagnosis** for a patient; the differential diagnosis comprises the set of possible diagnoses among which the physician must distinguish to determine how best to administer treatment.

Note that the question selection process is inherently heuristic; e.g., it is personalized and efficient, but it is not guaranteed to collect every piece of information that might be pertinent. Human beings use heuristics all the time in their decision making because it often is impractical or impossible to use an exhaustive problem-solving approach. A common example of heuristic problem solving is the playing of a complex game such as chess. Because it would require an enormous amount of time to define all the possible moves and countermoves that could ensue from a given board position, expert chess players develop personal heuristics for assessing the game at any point and then selecting a strategy for how best to proceed. Differences among such heuristics account in part for variations in observed expertise.

Physicians have developed safety measures, however, to help them to avoid missing important issues that they might not discover when collecting data in a hypothesis-directed fashion when taking the history of a patient's present illness (Pauker et al. 1976). These measures tend to be focused in four general categories of questions that follow the collection of information about the chief complaint: past medical history, family history, social history, and a brief **review of systems** in which the physician asks some general questions about the state of health of each of the major organ systems in the

body. Occasionally, the physician discovers entirely new problems or finds important information that modifies the hypothesis list or modulates the treatment options available (e.g., if the patient reports a serious past drug reaction or allergy).

When physicians have finished asking questions, the refined hypothesis list (which may already be narrowed to a single diagnosis) then serves as the basis for a focused physical examination. By this time, physicians may well have expectations of what they will find on examination or may have specific tests in mind that will help them to distinguish among still active hypotheses about diseases based on the questions that they have asked. Once again, as in the question-asking process, focused hypothesis-directed examination is augmented with general tests that occasionally turn up new abnormalities and generate hypotheses that the physician did not expect on the basis of the medical history alone. In addition, unexplained findings on examination may raise issues that require additional history taking. Thus, the asking of questions generally is partially integrated with the examination process.

When physicians have completed the physical examination, their refined hypothesis list may be narrowed sufficiently for them to undertake specific treatment. Additional data gathering may still be necessary, however. Such testing is once again guided by the current hypotheses. The options available include laboratory tests (of blood, urine, other body fluids, or biopsy specimens), radiologic studies (X-ray examinations, nuclear-imaging scans, computed tomography (CT) studies, magnetic resonance scans, sonograms, or any of a number of other imaging modalities), and other specialized tests (electrocardiograms (ECGs), electroencephalograms, nerve conduction studies, and many others), as well as returning to the patient to ask further questions or perform additional physical examination. As the results of such studies become available, physicians constantly revise and refine their hypothesis list.

Ultimately, physicians are sufficiently certain about the source of a patient's problem to be able to develop a specific management plan. Treatments are administered, and the patient is observed. Note data collected to measure response to treatment may themselves be used to synthesize information that affects the hypotheses about a patient's illness. If patients do not respond to treatment, it may mean that their disease is resistant to that therapy and that their physicians should try an alternate approach, or it may mean that the initial diagnosis was incorrect and that physicians should consider alternate explanations for the patient's problem.

The patient may remain in a cycle of treatment and observation for a long time, as shown in ◘ Fig. 2.12. This long cycle reflects the nature of chronic-disease management—an aspect of medical care that is accounting for an increasing proportion of the health care community's work (and an increasing proportion of health care cost). Alternatively, the patient may recover and no longer need therapy, or he or she may die. Although the process outlined in ◘ Fig. 2.12 is oversimplified in many regards, it is generally applicable to the process of data collection, diagnosis, and treatment in most areas of medicine.

Note that the hypothesis-directed process of data collection, diagnosis, and treatment is inherently knowledge-based. It is dependent not only on a significant fact base that permits proper interpretation of data and selection of appropriate follow-up questions and tests but also on the effective use of heuristic techniques that characterize individual expertise.

Another important issue, addressed in ▶ Chap. 3, is the need for physicians to balance financial costs and health risks of data collection against the perceived benefits to be gained when those data become available. It costs nothing but time to examine the patient at the bedside or to ask an additional question, but if the data being considered require, for example, X-ray exposure, coronary angiography, or a CT scan of the head (all of which have associated risks and costs), then it may be preferable to proceed with treatment in the absence of full information. Differences in the assessment of cost-benefit trade-offs in data collection, and variations among individuals in their willingness to make decisions under uncertainty, often account for differences of opinion among collaborating physicians.

2.6.2 The Relationship Between Data and Hypotheses

We wrote rather glibly in ▶ Sect. 2.6.1 about the "generation of hypotheses from data"; now we need to ask: What precisely is the nature of that process? As is discussed in ▶ Chap. 4, researchers with a psychological orientation have spent much time trying to understand how expert problem solvers evoke hypotheses (Elstein et al. 1978; Arocha et al. 2005) and the traditional probabilistic decision sciences have much to say about that process as well. We provide only a brief introduction to these ideas here; they are discussed in greater detail in ▶ Chaps. 3 and 4.

When an observation evokes a hypothesis (e.g., when a clinical finding makes a specific diagnosis come to mind), the observation presumably has some close association with the hypothesis. What might be the characteristics of that association? Perhaps the finding is almost always observed when the hypothesis turns out to be true. Is that enough to explain hypothesis generation? A simple example will show that such a simple relationship is not enough to explain the evocation process. Consider the hypothesis that a patient is pregnant and the observation that the patient is biologically female. Clearly, all pregnant patients are female. When a new patient is observed to be female, however, the possibility that the patient is pregnant is not immediately evoked. Thus, female gender is a highly sensitive indicator of pregnancy (there is a 100% certainty that a pregnant patient is female), but it is not a good predictor of pregnancy (most females are not pregnant). The idea of **sensitivity**—the likelihood that a given datum will be observed in a patient with a given disease or condition—is an important one, but it will not alone account for the process of hypothesis generation in medical diagnosis.

Perhaps the clinical manifestation seldom occurs unless the hypothesis turns out to be true; is that enough to explain hypothesis generation? This idea seems to be a little closer to the mark. Suppose a given datum is never seen unless a patient has a specific disease. For example, a Pap smear (a smear of cells swabbed from the cervix, at the opening to the uterus, treated with Papanicolaou's stain, and then examined under the microscope) with grossly abnormal cells (called class IV findings) is never seen unless the woman has cancer of the cervix or uterus. Such tests are called **pathognomonic**. Not only do they evoke a specific diagnosis but they also immediately prove it to be true. Unfortunately, there are few pathognomonic tests in medicine and they are often of relatively low sensitivity (that is, although having a particular test result makes the diagnosis, few patients with the condition may actually have that finding).

More commonly, a feature is seen in one disease or disease category more frequently than it is in others, but the association is not absolute. For example, there are few disease entities other than infections that elevate a patient's white blood cell count. Certainly it is true, for example, that leukemia can raise the white blood cell count, as can the use of certain medications, but most patients who do not have infections will have normal white blood cell counts. An elevated white count therefore does not prove that a patient has an infection, but it does tend to evoke or support the hypothesis that an infection is present. The word used to describe this relationship is **specificity**. An observation is highly specific for a disease if it is generally not seen in patients who do not have that disease. A pathognomonic observation is 100% specific for a given disease. When an observation is highly specific for a disease, it tends to evoke that disease during the diagnostic or data-gathering process.

By now, you may have realized that there is a substantial difference between a physician viewing test results that evoke a disease hypothesis and that physician being willing to act on the disease hypothesis. Yet even experienced physicians sometimes fail to recognize that, although they have made an observation that is highly specific for a given disease, it may still be more likely that the patient has other diseases (and does not have the suspected one) unless (1) the finding is pathognomonic or (2) the suspected disease is considerably more common than are the other diseases that can cause the observed abnormality. This mistake is one of the most common errors of intuition in the medical decision-making process. To explain the basis for this confusion in

more detail, we must introduce two additional terms: prevalence and predictive value.

The **prevalence** of a disease is simply the percentage of a population of interest that has the disease at any given time. A particular disease may have a prevalence of only 5% in the general population (1 person in 20 will have the disease) but have a higher prevalence in a specially selected subpopulation. For example, black-lung disease has a low prevalence in the general population but has a much higher prevalence among coal miners, who develop black lung from inhaling coal dust. The task of diagnosis therefore involves updating the probability that a patient has a disease from the **baseline rate** (the prevalence in the population from which the patient was selected) to a post-test probability that reflects the test results. For example, the probability that any given person in the United States has lung cancer is low (i.e., the prevalence of the disease is low), but the chance increases if his or her chest X-ray examination shows a possible tumor. If the patient were a member of the population composed of cigarette smokers in the United States, however, the prevalence of lung cancer would be higher. In this case, the identical chest X-ray report would result in an even higher updated probability of lung cancer than it would had the patient been selected from the population of all people in the United States.

The **predictive value (PV) of a test** is simply the post-test (updated) probability that a disease is present based on the results of a test. If an observation supports the presence of a disease, the PV will be greater than the prevalence (also called the pretest risk). If the observation tends to argue against the presence of a disease, the PV will be lower than the prevalence. For any test and disease, then, there is one PV if the test result is positive and another PV if the test result is negative. These values are typically abbreviated PV+ (the PV of a positive test) and PV− (the PV of a negative test).

The process of hypothesis generation in medical diagnosis thus involves both the evocation of hypotheses and the assignment of a likelihood (probability) to the presence of a specific disease or disease category. The PV of a positive test depends on the test's sensitivity and specificity, as well as the prevalence of the disease. The formula that describes the relationship precisely is:

$$\mathrm{PV+} = \frac{(\text{sensitivity})(\text{prevalence})}{(\text{sensitivity})(\text{prevalence}) + (1-\text{specificity})(1-\text{prevalence})}$$

There is a similar formula for defining PV− in terms of sensitivity, specificity, and prevalence. Both formulae can be derived from simple probability theory. Note that positive tests with high sensitivity and specificity may still lead to a low post-test probability of the disease (PV+) if the prevalence of that disease is low. You should substitute values in the PV+ formula to convince yourself that this assertion is true. It is this relationship that tends to be poorly understood by practitioners and that often is viewed as counterintuitive (which shows that your intuition can misguide you!). Note also (by substitution into the formula) that test sensitivity and disease prevalence can be ignored only when a test is pathognomonic (i.e., when its specificity is 100%, which mandates that PV+ be 100%). The PV+ formula is one of many forms of **Bayes' theorem**, a rule for combining probabilistic data that is generally attributed to the work of Reverend Thomas Bayes in the 1700s. Bayes' theorem is discussed in greater detail in ► Chap. 3.

2.6.3 Methods for Selecting Questions and Comparing Tests

We have described the process of hypothesis-directed sequential data collection and have asked how an observation might evoke or refine the physician's hypotheses about what abnormalities account for the patient's illness. The complementary question is: Given a set of current hypotheses, how does the physician decide what additional data should be collect-

ed? This question also has been analyzed at length (Elstein et al. 1978; Arocha et al. 2005) and is pertinent for computer programs that gather data efficiently to assist clinicians with diagnosis or with therapeutic decision making (see ▶ Chap. 26). Because understanding issues of test selection and data interpretation is crucial to understanding medical data and their uses, we devote ▶ Chap. 3 to these and related issues of medical decision making. In ▶ Sect. 3.6, for example, we discuss the use of decision-analytic techniques in deciding whether to treat a patient on the basis of available information or to perform additional diagnostic tests.

2.7 The Computer and Collection of Medical Data

Although this chapter has not directly discussed computer systems, the role of the computer in medical data storage, retrieval, and interpretation should be clear. Much of the rest of this book deals with specific applications in which the computer's primary role is data management. One question is pertinent to all such applications: What are the best approaches for getting data into the computer in the first place?

The need for data entry by physicians has posed a problem for medical-computing systems since the earliest days of the field. Awkward or nonintuitive interactions at computing devices—particularly ones requiring keyboard typing or confusing movement through multiple display screens by the physician—have perhaps done more to frustrate clinicians than have any other factor.

A variety of approaches have been used to try to finesse this problem. One is to design systems such that clerical staff can do essentially all the data entry and much of the data retrieval as well. Many clinical research systems (see ▶ Chap. 29) have taken this approach. Physicians may be asked to fill out structured paper datasheets, or such sheets may be filled out by data abstractors who review patient charts, but the actual entry of data into the database is done by paid transcriptionists. Other physicians have adopted "scribes" (staff whose role is to follow physicians in examination rooms and to enter data into the electronic health record) to reduce the data entry burden on physicians while they interact with patients.

In some applications, data are entered automatically into the computer by the device that measures or collects them. For example, monitors in intensive care or coronary care units, pulmonary function or ECG machines, and measurement equipment in the clinical chemistry laboratory can interface directly with a computer in which a database is stored. Certain data can be entered directly by patients; there are systems, for example, that take the patient's history by presenting on a computer screen or tablet multiple-choice questions that follow a branching logic. The patient's responses to the questions are used to generate electronic or hard copy reports for physicians and also may be stored directly in a computer database for subsequent use in other settings.

When physicians or other health personnel do use the machine themselves, specialized devices often allow rapid and intuitive operator–machine interaction. Most of these devices use a variant of the "point-and-select" approach—e.g., touch-sensitive computer screens, mouse-pointing devices, and increasingly the clinician's finger on a mobile tablet or smart phone (see ▶ Chaps. 5 and 6). When conventional computer workstations are used, specialized keypads can be helpful. Designers frequently permit logical selection of items from menus displayed on the screen so that the user does not need to learn a set of specialized commands to enter or review data. There were clear improvements when handheld tablets using pen-based or finger-based mechanisms for data entry were introduced. With ubiquitous wireless data services, such devices are allowing clinicians to maintain normal mobility (in and out of examining rooms or inpatient rooms) while accessing and entering data that are pertinent to a patient's care.

These issues arise in essentially all application areas, and, because they can be crucial to the successful implementation and use of a system, they warrant particular attention in system design. As more physicians are comfortable with computers in daily life, they will likely find the use of computers in their practice less of a hindrance. We encourage you to consider human–computer interaction, and

the cognitive issues that arise in dealing with computer systems (see ▶ Chap. 4), as you learn about the application areas and the specific systems described in later chapters.

Suggested Readings

Adler-Milstein, J., Zhao, W., Willard-Grace, R., Knox, M., & Grumbach, K. (2020). Electronic health records and burnout: Time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. Journal of the American Medical Informatics Association. https://doi.org/10.1093/jamia/ocz220. This paper examines the correlation between electronic health record use and clinician burnout, and concludes that two specific EHR usage measures (EHR time after hours and message volume) were associated with exhaustion.

Arocha, J. F., Wang, D., & Patel, V. L. (2005). Identifying reasoning strategies in medical decision making: A methodological guide. Journal of Biomedical Informatics, 38(2), 154–171. This paper illustrates the role of theory-driven psychological research and cognitive evaluation as they relate to medical decision making and the interpretation of clinical data. See also Chap. 4.

Bernstam, E. V., Smith, J. W., & Johnson, T. R. (2010). What is biomedical informatics? Journal of Biomedical Informatics, 43(1), 104–110. The authors discuss the transformation of data into information and knowledge, delineating the ways in which this focus lies at the heart of the field of biomedical informatics.

Brennan, P. F., Chiang, M. F., & Ohno-Machado, L. (2018). Biomedical informatics and data science: Evolving fields with significant overlap. Journal of the American Medical Informatics Association, 25(1), 2–3. This editorial introduces a special issue of the Journal of the American Medical Informatics Association, in which the rapidly evolving field of data science is the focus. There are 8 papers in this issue that involve applications such as secondary use of EHR data, repositories of data, and standardization of data representation.

Klasnja, P., & Pratt, W. (2012). Healthcare in the pocket: Mapping the space of mobile-phone health interventions. Journal of Biomedical Informatics, 45(1), 184–198. This review article describes the multiple ways in which both patients and providers are being empowered through the introduction of affordable mobile technologies that manage data and apply knowledge to generate advice.

Steinhubl, S. R., Muse, E. D., & Topol, E. J. (2015). The emerging field of mobile health. Science Translational Medicine, 7(283), 283rv3. The authors discuss the potential for mobile health (mHealth) to impact the delivery and quality of health care delivery and clinical research on a large scale. This paper includes a discuss of challenges to the field, as well as efforts to address those challenges.

Vamathevan, J., & Birney, E. (2017). A review of recent advances in translational bioinformatics: Bridges from biology to medicine. Yearbook of Medical Informatics, 26(1), 178–187. This articles reviews the latest trends and major developments in translational bioinformatics. This includes work applying findings from national genome sequencing initiatives to health care delivery. There is a discussion of current challenges and emerging technologies that bridge research with clinical care. See also Chap. 28.

Questions for Discussion

1. You check your pulse and discover that your heart rate is 100 beats per minute. Is this rate normal or abnormal? What additional information would you use in making this judgment? How does the context in which data are collected influence the interpretation of those data?
2. Given the imprecision of many medical terms, why do you think that serious instances of miscommunication among health care professionals are not more common? Why is greater standardization of terminology necessary if computers rather than humans are to manipulate patient data?
3. Based on the discussion of coding schemes for representing clinical information, discuss three challenges you foresee in

attempting to construct a standardized terminology to be used in hospitals, physicians' offices, and research institutions.

4. How would medical practice change if nonphysicians were to collect and enter all medical data into EHRs? What problems or unintended consequentes would you anticipate?
5. Consider what you know about the typical daily schedule of a busy clinician. What are the advantages of wireless devices, connected to the Internet, as tools for such clinicians? Can you think of disadvantages as well? Be sure to consider the safety and protection of information as well as workflow and clinical needs.
6. To decide whether a patient has a significant urinary tract infection, physicians commonly use a calculation of the number of bacterial organisms in a milliliter of the patient's urine. Physicians generally assume that a patient has a urinary tract infection if there are at least 10,000 bacteria per milliliter. Although laboratories can provide such quantification with reasonable accuracy, it is obviously unrealistic for the physician explicitly to count large numbers of bacteria by examining a milliliter of urine under the microscope. As a result, one article offers the following guideline to physicians: "When interpreting … microscopy of … stained centrifuged urine, a threshold of one organism per field yields a 95% sensitivity and five organisms per field a 95% specificity for bacteriuria [bacteria in the urine] at a level of at least 10,000 organisms per ml." (Senior Medical Review 1987, p. 4)

   (a) Describe an experiment that would have allowed the researchers to determine the sensitivity and specificity of the microscopy.
   (b) How would you expect specificity to change as the number of bacteria per microscopic field increases from one to five?
   (c) How would you expect sensitivity to change as the number of bacteria per microscopic field increases from one to five?
   (d) Why does it take more organisms per microscopic field to obtain a specificity of 95% than it does to achieve a sensitivity of 95%?

References

Adler-Milstein, J., & Jha, A. K. (2013). Healthcare's "big data" challenge. *American Journal of Managed Care, 19*(7), 537–538.

Arocha, J. F., Wang, D., & Patel, V. L. (2005). Identifying reasoning strategies in medical decision making: A methodological guide. *Journal of Biomedical Informatics, 38*(2), 154–171.

Bernstam, E. V., Smith, J. W., & Johnson, T. R. (2010). What is biomedical informatics? *Journal of Biomedical Informatics, 43*(1), 104–110.

Blum, B. I. (1986). Clinical information systems: A review. *Western Journal of Medicine, 145*(6), 791–797.

Brennan, P. F., Chiang, M. F., & Ohno-Machado, L. (2018). Biomedical informatics and data science: Evolving fields with significant overlap. *Journal of the American Medical Informatics Association, 25*(1), 2–3.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature, 562*(7726), 203–209.

Elstein, K. A., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Erlich, Y., & Zielinski, D. (2017). DNA Foundation enables a robust and efficient storage architecture. *Science, 355*, 950–954.

Haendel, M. A., Chute, C. G., & Robinson, P. N. (2018). Classification, ontology, and precision medicine. *New England Journal of Medicine, 379*, 1452–1462.

Hirsch, J. A., Leslie-Mazwi, T. M., Nocola, G. N., Barr, R. M., Bello, J. A., Donovan, W. D., et al. (2015). Current procedural terminology; a primer. *Journal of Neurointerventional Surgery, 7*(4), 309–312.

Humphreys, B., Lindberg, D., Schoolman, H., & Barnett, G. (1998). The unified medical language system: An informatics research collaboration. *Journal of the American Medical Informatics Association: JAMIA, 5*, 1–11.

Kassirer, J. P., & Gorry, G. A. (1978). Clinical problem solving: A behavioral analysis. *Annals of Internal Medicine, 89*(2), 245–255.

Lee, D., de Keizer, N., Lau, F., & Cornet, R. (2014). Literature review of SNOMED CT use. *Journal of the American Medical Informatics Association, 21*(e1), e11–e19.

Masys, D. R., Jarvik, G. P., Abernethy, N. F., Anderson, N. R., Papanicolaou, G. J., Paltoo, D. N., et al. (2012). Technical desiderata for the integration of genomic data into electronic health records. *Journal of Biomedical Informatics, 45*(3), 419–422.

Pauker, S. G., Gorry, G. A., Kassirer, J. P., & Scwartz, W. B. (1976). Towards the simulation of clinical cognition: Taking a present illness by computer. *American Journal of Medicine, 60*(7), 981–996.

Relling, M. V., & Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature, 526*(7573), 342–350.

Sanders, D. S., Lattin, D. J., Read-Brown, S., Tu, D. C., Wilson, D. J., Hwang, T. S., et al. (2013). Electronic health record systems in ophthalmology: Impact on clinical documentation. *Ophthalmology, 120*, 1745–1755.

Senior Medical Review (1987). Urinary tract infections. Senior Medical Review Newsletter.

Shortliffe, E. H. (2010). Biomedical informatics in the education of physicians. *Journal of the American Medical Association, 304*(11), 1227–1228.

Stearns, M. Q., Price, C., Spackman, K. A., & Wang, A. Y. (2001). SNOMED clinical terms: Overview of the development process and project status. *Proceedings of the AMIA Symposium*, 662–666.

Biomedical Decision Making: Probabilistic Clinical Reasoning

Douglas K. Owens, Jeremy D. Goldhaber-Fiebert, and Harold C. Sox

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_3

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- How is the concept of probability useful for understanding test results and for making medical decisions that involve uncertainty?
- How can we characterize the ability of a test to discriminate between disease and health?
- What information do we need to interpret test results accurately?
- What is expected-value decision making? How can this methodology help us to understand particular medical problems?
- What are utilities, and how can we use them to represent patients' preferences?
- What is a sensitivity analysis? How can we use it to examine the robustness of a decision and to identify the important variables in a decision?
- What are influence diagrams? How do they differ from decision trees?

3.1 The Nature of Clinical Decisions: Uncertainty and the Process of Diagnosis

Because clinical data are imperfect and outcomes of treatment are uncertain, health professionals often are faced with difficult choices. In this chapter, we introduce probabilistic medical reasoning, an approach that can help health care providers to deal with the uncertainty inherent in many medical decisions. Medical decisions are made by a variety of methods; our approach is neither necessary nor appropriate for all decisions. Throughout the chapter, we provide simple clinical examples that illustrate a broad range of problems for which probabilistic medical reasoning does provide valuable insight.

As discussed in ▶Chap. 2, medical practice is medical decision making. In this chapter, we look at the process of medical decision making. Together, ▶Chaps. 2 and 3 lay the groundwork for the rest of the book. In the remaining chapters, we discuss ways that computers can help clinicians with the decision-making process, and we emphasize the relationship between information needs and system design and implementation.

The material in this chapter is presented in the context of the decisions made by an individual clinician. The concepts, however, are more broadly applicable. Sensitivity and specificity are important parameters of laboratory systems that flag abnormal test results, of patient monitoring systems (▶Chap. 21), and of information-retrieval systems (▶Chap. 23). An understanding of what probability is and of how to adjust probabilities after the acquisition of new information is a foundation for our study of clinical decision-support systems (▶Chap. 24). The importance of probability in medical decision making was noted as long ago as 1922:

» [G]ood medicine does not consist in the indiscriminate application of laboratory examinations to a patient, but rather in having so clear a comprehension of the probabilities and possibilities of a case as to know what tests may be expected to give information of value (Peabody 1922).

▶ Example 3.1

You are the director of a blood bank. All potential blood donors are tested to ensure that they are not infected with the human immunodeficiency virus (HIV), the causative agent of acquired immunodeficiency syndrome (AIDS). You ask whether use of the polymerase chain reaction (PCR), a gene-amplification technique that can diagnose HIV, would be useful to identify people who have HIV. The PCR test is positive 98% of the time when antibody is present, and negative 99% of the time antibody is absent.[1] ◄

If the test is positive, what is the likelihood that a donor actually has HIV? If the test is negative, how sure can you be that the person does not have HIV? On an intuitive level, these questions

1 The test sensitivity and specificity used in ▶ Example 3.1 are consistent with the reported values of the sensitivity and specificity of the PCR test for diagnosis of HIV early in its development (Owens et al. 1996b); the test now has higher sensitivity and specificity.

do not seem particularly difficult to answer. The test appears accurate, and we would expect that, if the test is positive, the donated blood specimen is likely to contain the HIV. Thus, we are surprised to find that, if only one in 1000 donors actually is infected, the test is more often mistaken than it is correct. In fact, of 100 donors with a positive test, fewer than 10 would be infected. There would be ten wrong answers for each correct result. How are we to understand this result? Before we try to find an answer, let us consider a related example.

► Example 3.2

Ms. Kamala is a 66-year-old woman with coronary artery disease (narrowing or blockage of the blood vessels that supply the heart tissue). When the heart muscle does not receive enough oxygen (hypoxia) because blood cannot reach it, the patient often experiences chest pain (angina). Ms. Kamala has twice undergone coronary artery bypass graft (CABG) surgery, a procedure in which new vessels, often taken from the leg, are grafted onto the old ones such that blood is shunted past the blocked region. Unfortunately, she has again begun to have chest pain, which becomes progressively more severe, despite medication. If the heart muscle is deprived of oxygen, the result can be a heart attack (myocardial infarction), in which a section of the muscle dies. ◄

Should Ms. Kamala undergo a third operation? The medications are not working; without surgery, she runs a high risk of suffering a heart attack, which may be fatal. On the other hand, the surgery is hazardous. Not only is the surgical mortality rate for a third operation higher than that for a first or second one but also the chance that surgery will relieve the chest pain is lower than that for a first operation. All choices in ► Example 3.2 entail considerable uncertainty. Furthermore, the risks are grave; an incorrect decision may substantially increase the chance that Ms. Kamala will die. The decision will be difficult even for experienced clinicians.

These examples illustrate situations in which intuition is either misleading or inadequate. Although the test results in ►Example 3.1 are appropriate for the blood bank, a clinician who uncritically reports these results would erroneously inform many people that they had HIV—a mistake with profound emotional and social consequences. In ►Example 3.2, the decision-making skill of the clinician will affect a patient's quality and length of life. Similar situations are commonplace in medicine. Our goal in this chapter is to show how the use of probability and decision analysis can help to make clear the best course of action.

Decision making is one of the quintessential activities of the healthcare professional. Some decisions are made on the basis of deductive reasoning or of physiological principles. Many decisions, however, are made on the basis of knowledge that has been gained through collective experience: the clinician often must rely on empirical knowledge of associations between symptoms and disease to evaluate a problem. A decision that is based on these usually imperfect associations will be, to some degree, uncertain. In ►Sects. 3.1.1, 3.1.2 and 3.1.3, we examine decisions made under uncertainty and present an overview of the diagnostic process. As Smith (1985, p. 3) said: "Medical decisions based on probabilities are necessary but also perilous. Even the most astute physician will occasionally be wrong."

3.1.1 Decision Making Under Uncertainty

► Example 3.3

Ms. Kirk, a 33-year-old woman with a history of a previous blood clot (thrombus) in a vein in her left leg, presents with the complaint of pain and swelling in that leg for the past 5 days. On physical examination, the leg is tender and swollen to midcalf—signs that suggest the possibility of deep vein thrombosis.[2] A test (ultrasonography) is performed, and the flow of blood in the veins of Ms. Kirk's leg is evaluated. The blood flow is abnormal, but the radiologist cannot tell whether there is a new blood clot. ◄

2 In medicine, a sign is an objective physical finding (something observed by the clinician) such as a temperature of 101.2 °F. A symptom is a subjective experience of the patient, such as feeling hot or feverish. The distinction may be blurred if the patient's experience also can be observed by the clinician.

Should Ms. Kirk be treated for blood clots? The main diagnostic concern is the recurrence of a blood clot in her leg. A clot in the veins of the leg can dislodge, flow with the blood, and cause a blockage in the vessels of the lungs, a potentially fatal event called a pulmonary embolus. Of patients with a swollen leg, about one-half actually have a blood clot; there are numerous other causes of a swollen leg. Given a swollen leg, therefore, a clinician cannot be sure that a clot is the cause. Thus, the physical findings leave considerable uncertainty. Furthermore, in ▶Example 3.3, the results of the available diagnostic test are equivocal. The treatment for a blood clot is to administer anticoagulants (drugs that inhibit blood clot formation), which pose the risk of excessive bleeding to the patient. Therefore, clinicians do not want to treat the patient unless they are confident that a thrombus is present. But how much confidence should be required before starting treatment? We will learn that it is possible to answer this question by calculating the benefits and harms of treatment.

This example illustrates an important concept: Clinical data are imperfect. The degree of imperfection varies, but all clinical data—including the results of diagnostic tests, the history given by the patient, and the findings on physical examination—are uncertain.

3.1.2 Probability: An Alternative Method of Expressing Uncertainty

The language that clinicians use to describe a patient's condition often is ambiguous—a factor that further complicates the problem of uncertainty in medical decision making. Clinicians use words such as "probable" and "highly likely" to describe their beliefs about the likelihood of disease. These words have strikingly different meanings to different individuals. Because of the widespread disagreement about the meaning of common descriptive terms, there is ample opportunity for miscommunication.

The problem of how to express degrees of uncertainty is not unique to medicine. How is it handled in other contexts? Horse racing has its share of uncertainty. If experienced gamblers are deciding whether to place bets, they will find it unsatisfactory to be told that a given horse has a "high chance" of winning. They will demand to know the odds.

The odds are simply an alternate way to express a probability. The use of probability or odds as an expression of uncertainty avoids the ambiguities inherent in common descriptive terms.

3.1.3 Overview of the Diagnostic Process

In ▶Chap. 2, we described the hypothetico-deductive approach, a diagnostic strategy comprising successive iterations of hypothesis generation, data collection, and interpretation. We discussed how observations may evoke a hypothesis and how new information subsequently may increase or decrease our belief in that hypothesis. Here, we review this process briefly in light of a specific example. For the purpose of our discussion, we separate the diagnostic process into three stages.

The first stage involves making an initial judgment about whether a patient is likely to have a disease. After an interview and physical examination, a clinician intuitively develops a belief about the likelihood of disease. This judgment may be based on previous experience or on knowledge of the medical literature. A clinician's belief about the likelihood of disease usually is implicit; he or she can refine it by making an explicit estimation of the probability of disease. This estimated probability, made before further information is obtained, is the **prior probability** or **pretest probability** of disease.

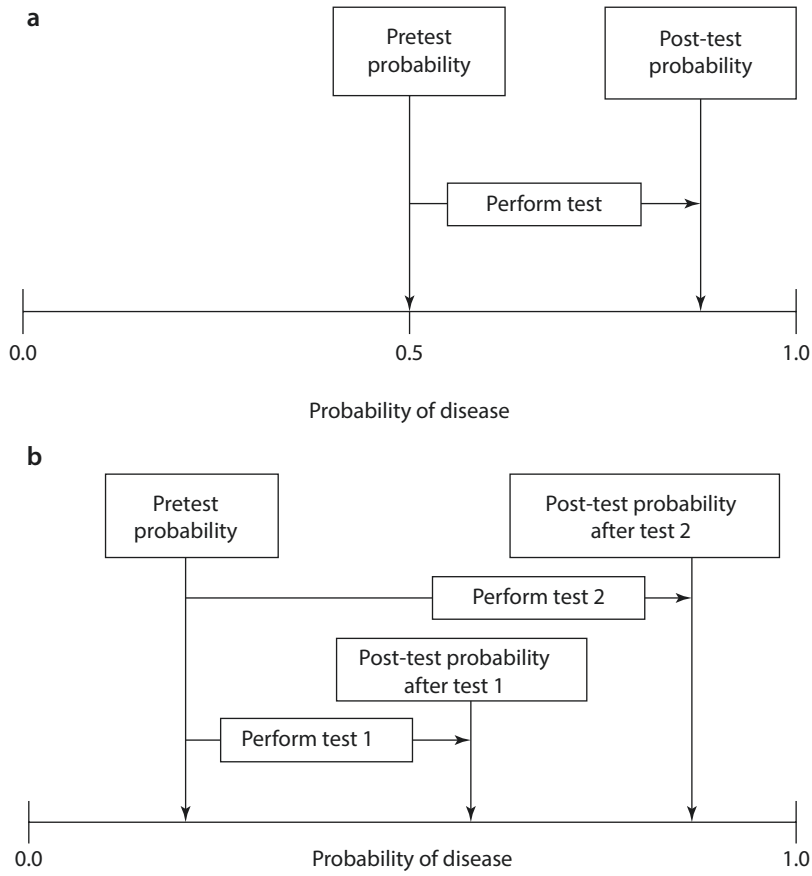
▶ Example 3.4

Mr. Smith, a 60-year-old man, complains to his clinician that he has pressure-like chest pain that occurs when he walks quickly. After taking his history and examining him, his clinician believes there is a high enough chance that he has heart disease to warrant ordering an exercise stress test. In the stress test, an electrocardiogram (ECG) is taken while Mr. Smith exercises. Because the heart must pump more blood per stroke and must beat faster (and thus

requires more oxygen) during exercise, many heart conditions are evident only when the patient is physically stressed. Mr. Smith's results show abnormal changes in the ECG during exercise—a sign of heart disease. ◄

How would the clinician evaluate this patient? The clinician would first talk to the patient about the quality, duration, and severity of his or her pain. Traditionally, the clinician would then decide what to do next based on his or her intuition about the etiology (cause) of the chest pain. Our approach is to ask the clinician to make his or her initial intuition explicit by estimating the pretest probability of disease. The clinician in this example, based on what he or she knows from talking with the patient, might assess the pretest or prior probability of heart disease as 0.5 (50% chance or 1:1 odds; see ▶Sect. 3.2). We explore methods used to estimate pretest probability accurately in ▶Sect. 3.2.

After the pretest probability of disease has been estimated, the second stage of the diagnostic process involves gathering more information, often by performing a diagnostic test. The clinician in ▶Example 3.4 ordered a test to reduce the uncertainty about the diagnosis of heart disease. The positive test result supports the diagnosis of heart disease, and this reduction in uncertainty is shown in ◘ Fig. 3.1a. Although the clinician in ▶Example 3.4 chose the exercise stress test, there are many tests available to diagnose heart disease, and the clinician would like to know which test he or she should order next. Some tests reduce uncertainty more than do others (see ◘ Fig. 3.1b), but may cost more. The more a test reduces uncertainty, the more useful it is. In ▶Sect. 3.3, we explore ways to measure



◘ **Fig. 3.1** The effect of test results on the probability of disease. **a** A positive test result increases the probability of disease. **b** Test 2 reduces uncertainty about presence of disease (increases the probability of disease) more than test 1 does

how well a test reduces uncertainty, expanding the concepts of test sensitivity and specificity first introduced in ▶Chap. 2.

Given new information provided by a test, the third step is to update the initial probability estimate. The clinician in ▶Example 3.4 must ask: "What is the probability of disease given the abnormal stress test?" The clinician wants to know the **posterior probability**, or **post-test probability**, of disease (see ◘ Fig. 3.1a). In ▶Sect. 3.4, we reexamine Bayes' theorem, introduced in ▶Chap. 2, and we discuss its use for calculating the post-test probability of disease. As we noted, to calculate post-test probability, we must know the pretest probability, as well as the sensitivity and specificity, of the test.[3]

3.2 Probability Assessment: Methods to Assess Pretest Probability

In this section, we explore the methods that clinicians can use to make judgments about the probability of disease before they order tests. **Probability** is our preferred means of expressing uncertainty. In this framework, probability (p) expresses a clinician's opinion about the likelihood of an event as a number between 0 and 1. An event that is certain to occur has a probability of 1; an event that is certain not to occur has a probability of 0.[4]

The probability of event A is written p[A]. The sum of the probabilities of all possible, collectively exhaustive outcomes of a chance event must be equal to 1. Thus, in a coin flip,

$$p[\text{heads}] + p[\text{tails}] = 1.0.$$

The probability of event A and event B occurring together is denoted by p[A&B] or by p[A,B].

Events A and B are considered **independent** if the occurrence of one does not influence the probability of the occurrence of the other. The probability of two independent events A and B both occurring is given by the product of the individual probabilities:

$$p[A,B] = p[A] \times p[B].$$

Thus, the probability of heads on two consecutive coin tosses is 0.5 × 0.5 = 0.25. (Regardless of the outcome of the first toss, the probability of heads on the second toss is 0.5).

The probability that event A will occur given that event B is known to occur is called the **conditional probability** of event A given event B, denoted by p[A|B] and read as "the probability of A given B." Thus a post-test probability is a conditional probability predicated on the test or finding. For example, if 30% of patients who have a swollen leg have a blood clot, we say the probability of a blood clot given a swollen leg is 0.3, denoted:

$$p[\text{blood clot}|\text{swollen leg}] = 0.3.$$

Before the swollen leg is noted, the pretest probability is simply the prevalence of blood clots in the leg in the population from which the patient was selected—a number likely to be much smaller than 0.3.

Now that we have decided to use probability to express uncertainty, how can we estimate probability? We can do so by either subjective or objective methods; each approach has advantages and limitations.

3.2.1 Subjective Probability Assessment

Most assessments that clinicians make about probability are based on personal experience. The clinician may compare the current problem to similar problems encountered previously and then ask: "What was the frequency of disease in similar patients whom I have seen?"

To make these subjective assessments of probability, people rely on several discrete, often unconscious mental processes that have been de-

3 Note that pretest and post-test probabilities correspond to the concepts of prevalence and predictive value. The latter terms were used in ▶Chap. 2 because the discussion was about the use of tests for screening populations of patients; in a population, the pretest probability of disease is simply that disease's prevalence in that population.

4 We assume a Bayesian interpretation of probability; there are other statistical interpretations of probability.

scribed and studied by cognitive psychologists (Tversky and Kahneman 1974). These processes are termed **cognitive heuristics**.

More specifically, a cognitive heuristic is a mental process by which we learn, recall, or process information; we can think of heuristics as rules of thumb. Knowledge of heuristics is important because it helps us to understand the underpinnings of our intuitive probability assessment. Both naive and sophisticated decision makers (including clinicians and statisticians) misuse heuristics and therefore make systematic—often serious—errors when estimating probability. So, just as we may underestimate distances on a particularly clear day (Tversky and Kahneman 1974), we may make mistakes in estimating probability in deceptive clinical situations. Three heuristics have been identified as important in estimation of probability:

1. Representativeness. One way that people estimate probability is to ask themselves: What is the probability that object A belongs to class B? For instance, what is the probability that this patient who has a swollen leg belongs to the class of patients who have blood clots? To answer, we often rely on the **representativeness** heuristic in which probabilities are judged by the degree to which A is representative of, or similar to, B. The clinician will judge the probability of the development of a blood clot (thrombosis) by the degree to which the patient with a swollen leg resembles the clinician's mental image of patients with a blood clot. If the patient has all the classic findings (signs and symptoms) associated with a blood clot, the clinician judges that the patient is highly likely to have a blood clot. Difficulties occur with the use of this heuristic when the disease is rare (very low prior probability, or prevalence); when the clinician's previous experience with the disease is atypical, thus giving an incorrect mental representation; when the patient's clinical profile is atypical; and when the probability of certain findings depends on whether other findings are present.
2. Availability. Our estimate of the probability of an event is influenced by the ease with which we remember similar events. Events more easily remembered are judged more probable; this rule is the **availability** heuristic, and it is often misleading. We remember dramatic, atypical, or emotion-laden events more easily and therefore are likely to overestimate their probability. A clinician who had cared for a patient who had a swollen leg and who then died from a blood clot would vividly remember thrombosis as a cause of a swollen leg. The clinician would remember other causes of swollen legs less easily, and he or she would tend to overestimate the probability of a blood clot in patients with a swollen leg.
3. Anchoring and adjustment. Another common heuristic used to judge probability is **anchoring and adjustment**. A clinician makes an initial probability estimate (the anchor) and then adjusts the estimate based on further information. For instance, the clinician in ▶ Example 3.4 makes an initial estimate of the probability of heart disease as 0.5. If he or she then learns that all the patient's brothers had died of heart disease, the clinician should raise the estimate because the patient's strong family history of heart disease increases the probability that he or she has heart disease, a fact the clinician could ascertain from the literature. The usual mistake is to adjust the initial estimate (the anchor) insufficiently in light of the new information. Instead of raising his or her estimate of prior probability to, say, 0.8, the clinician might adjust it to only 0.6.

Heuristics often introduce error into our judgments about prior probability. Errors in our initial estimates of probabilities will be reflected in the posterior probabilities even if we use quantitative methods to derive those posterior probabilities. An understanding of heuristics is thus important for medical decision making. The clinician can avoid some of these difficulties by using published research results to estimate probabilities.

3.2.2 Objective Probability Estimates

Published research results can serve as a guide for more objective estimates of probabilities. We can use the prevalence of disease in the popula-

tion or in a subgroup of the population, or clinical prediction rules, to estimate the probability of disease.

As we discussed in ▶ Chap. 2, the **prevalence** is the frequency of an event in a population; it is a useful starting point for estimating probability. For example, if you wanted to estimate the probability of prostate cancer in a 50-year-old man, the prevalence of prostate cancer in men of that age (5–14%) would be a useful anchor point from which you could increase or decrease the probability depending on your findings. Estimates of disease prevalence in a defined population often are available in the medical literature.

Symptoms, such as difficulty with urination, or signs, such as a palpable prostate nodule, can be used to place patients into a **clinical subgroup** in which the probability of disease is known. For patients referred to a urologist for evaluation of a prostate nodule, the prevalence of cancer is about 50%. This approach may be limited by difficulty in placing a patient in the correct clinically defined subgroup, especially if the criteria for classifying patients are ill-defined. A trend has been to develop guidelines, known as clinical prediction rules, to help clinicians assign patients to well-defined subgroups in which the probability of disease is known.

Clinical prediction rules are developed from systematic study of patients who have a particular diagnostic problem; they define how clinicians can use combinations of clinical findings to estimate probability. The symptoms or signs that make an independent contribution to the probability that a patient has a disease are identified and assigned numerical weights based on statistical analysis of the finding's contribution. The result is a list of symptoms and signs for an individual patient, each with a corresponding numerical contribution to a total score. The total score places a patient in a subgroup with a known probability of disease.

► Example 3.5

Ms. Troy, a 65-year-old woman who had a heart attack 4 months ago, has abnormal heart rhythm (arrhythmia), is in poor medical condition, and is about to undergo elective surgery. ◄

What is the probability that Ms. Troy will suffer a cardiac complication? Clinical prediction rules have been developed to help clinicians to assess this risk (Palda and Detsky 1997). ◘ Table 3.1 lists clinical findings and their corresponding diagnostic weights. We add the diagnostic weights for each of the patient's clinical findings to obtain the total score. The total score places the patient in a group with a defined probability of cardiac complications, as shown in ◘ Table 3.2. Ms. Troy receives a score of 20; thus, the clinician can estimate that the patient has a 27% chance of developing a severe cardiac complication.

Objective estimates of pretest probability are subject to error because of bias in the studies on which the estimates are based. For

◘ **Table 3.1** Diagnostic weights for assessing risk of cardiac complications from noncardiac surgery

Clinical finding	Diagnostic weight
Age greater than 70 years	5
Recent documented heart attack	
>6 months previously	5
<6 months previously	10
Severe angina	20
Pulmonary edema[a]	
Within 1 week	10
Ever	5
Arrhythmia on most recent ECG 5	
>5 PVCs	5
Critical aortic stenosis	20
Poor medical condition	5
Emergency surgery	10

ECG electrocardiogram, *PVCs* premature ventricular contractions on preoperative electrocardiogram
[a]Fluid in the lungs due to reduced heart function

Table 3.2 Clinical prediction rule for diagnostic weights in Table 3.1

Total score	Prevalence (%) of cardiac complications[a]
0–15	5
20–30	27
>30	60

[a]Cardiac complications defined as death, heart attack, or congestive heart failure

instance, published prevalence data may not apply directly to a particular patient. A clinical illustration is that early studies indicated that a patient found to have microscopic evidence of blood in the urine (microhematuria) should undergo extensive tests because a significant proportion of the patients would be found to have cancer or other serious diseases. The tests involve some risk, discomfort, and expense to the patient. Nonetheless, the approach of ordering tests for any patient with microhematuria was widely practiced for some years. A later study, however, suggested that the probability of serious disease in asymptomatic patients with only microscopic evidence of blood was only about 2%. In the past, many patients may have undergone unnecessary tests, at considerable financial and personal cost.

What explains the discrepancy in the estimates of disease prevalence? The initial studies that showed a high prevalence of disease in patients with microhematuria were performed on patients referred to urologists, who are specialists. The primary care clinician refers patients whom he or she suspects have a disease in the specialist's sphere of expertise. Because of this initial screening by primary care clinicians, the specialists seldom see patients with clinical findings that imply a low probability of disease. Thus, the prevalence of disease in the patient population in a specialist's practice often is much higher than that in a primary care practice; studies performed with the former patients therefore almost always overestimate disease probabilities. This example demonstrates **referral bias**. Referral bias is common because many published studies are performed on patients referred to specialists. Thus, one may need to adjust published estimates before one uses them to estimate pretest probability in other clinical settings.

We now can use the techniques discussed in this part of the chapter to illustrate how the clinician in ▶Example 3.4 might estimate the pretest probability of heart disease in his or her patient, Mr. Smith, who has pressure-like chest pain. We begin by using the objective data that are available. The prevalence of heart disease in 60-year-old men could be our starting point. In this case, however, we can obtain a more refined estimate by placing the patient in a clinical subgroup in which the prevalence of disease is known. The prevalence in a clinical subgroup, such as men with symptoms typical of coronary heart disease, will predict the pretest probability more accurately than would the prevalence of heart disease in a group that is heterogeneous with respect to symptoms, such as the population at large. We assume that large studies have shown the prevalence of coronary heart disease in men with typical symptoms of angina pectoris to be about 0.9; this prevalence is useful as an initial estimate that can be adjusted based on information specific to the patient. Although the prevalence of heart disease in men with typical symptoms is high, 10% of patients with this history do not have heart disease.

The clinician might use subjective methods to adjust his or her estimate further based on other specific information about the patient. For example, the clinician might adjust his or her initial estimate of 0.9 upward to 0.95 or higher based on information about family history of heart disease. The clinician should be careful, however, to avoid the mistakes that can occur when one uses heuristics to make subjective probability estimates. In particular, he or she should be aware of the tendency to stay too close to the initial estimate when adjusting for additional information. By combining subjective and objective methods for assessing pretest probability, the clinician can arrive at a reasonable estimate of the pretest probability of heart disease.

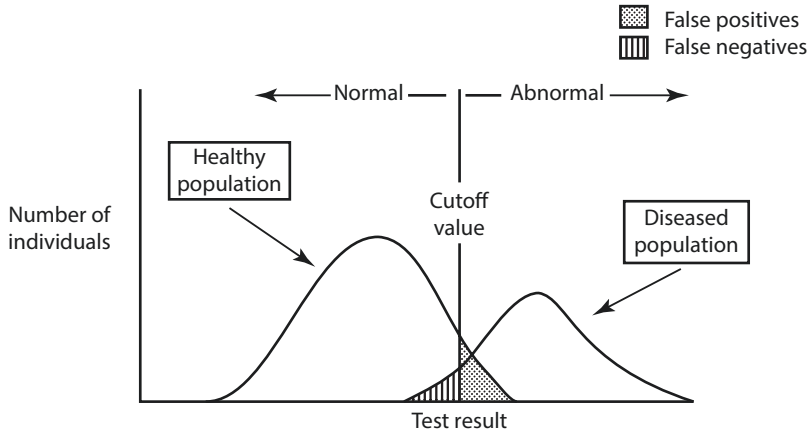


Fig. 3.2 Distribution of test results in healthy and diseased individuals. Varying the cutoff between "normal" and "abnormal" across the continuous range of possible values changes the relative proportions of false positives (FPs) and false negatives (FNs) for the two populations

In this section, we summarized subjective and objective methods to determine the pretest probability, and we learned how to adjust the pretest probability after assessing the specific subpopulation of which the patient is representative. The next step in the diagnostic process is to gather further information, usually in the form of formal diagnostic tests (laboratory tests, X-ray studies, etc.). To help you to understand this step more clearly, we discuss in the next two sections how to measure the accuracy of tests and how to use probability to interpret the results of the tests.

3.3 Measurement of the Operating Characteristics of Diagnostic Tests

The first challenge in assessing any test is to determine criteria for deciding whether a result is normal or abnormal. In this section, we present the issues that you need to consider when making such a determination.

3.3.1 Classification of Test Results as Abnormal

Most biological measurements in a population of healthy people are continuous variables that assume different values for different individuals. The distribution of values often is approximated by the normal (gaussian, or bell-shaped) distribution curve (Fig. 3.2). Thus, 95% of the population will fall within two standard deviations of the mean. About 2.5% of the population will be more than two standard deviations from the mean at each end of the distribution. The distribution of values for ill individuals may be normally distributed as well. The two distributions usually overlap (see Fig. 3.2).

How is a test result classified as abnormal? Most clinical laboratories report an "upper limit of normal," which usually is defined as two standard deviations above the mean. Thus, a test result greater than two standard deviations above the mean is reported as abnormal (or positive); a test result below that cutoff is reported as normal (or negative). As an example, if the mean cholesterol concentration in the blood is 180 mg/dl, a clinical laboratory might choose as the upper limit of normal 220 mg/dl because it is two standard deviations above the mean. Note that a cutoff that is based on an arbitrary statistical criterion may not have biological significance.

An ideal test would have no values at which the distribution of diseased and nondiseased people overlap. That is, if the cutoff value were set appropriately, the test would be normal in all healthy individuals and abnormal in all individuals with disease. Few tests meet this standard. If a test result is defined as

abnormal by the statistical criterion, 2.5% of healthy individuals will have an abnormal test. If there is an overlap in the distribution of test results in healthy and diseased individuals, some diseased patients will have a normal test (see Fig. 3.2). You should be familiar with the terms used to denote these groups:

- A **true positive** (TP) is a positive test result obtained for a patient in whom the disease is present (the test result correctly classifies the patient as having the disease).
- A **true negative** (TN) is a negative test result obtained for a patient in whom the disease is absent (the test result correctly classifies the patient as not having the disease).
- A **false positive** (FP) is a positive test result obtained for a patient in whom the disease is absent (the test result incorrectly classifies the patient as having the disease).
- A **false negative** (FN) is a negative test result obtained for a patient in whom the disease is present (the test result incorrectly classifies the patient as not having the disease).

Figure 3.2 shows that varying the cutoff point (moving the vertical line in the figure) for an abnormal test will change the relative proportions of these groups. As the cutoff is moved further up from the mean of the normal values, the number of FNs increases and the number of FPs decreases. Once we have chosen a cutoff point, we can conveniently summarize test performance—the ability to discriminate disease from nondisease—in a 2 × 2 **contingency table**, as shown in Table 3.3. The table summarizes the number of patients in each group: TP, FP, TN, and FN. Note that the sum of the first column is the total number of diseased patients, TP + FN. The sum of the second column is the total number of nondiseased patients, FP + TN. The sum of the first row, TP + FP, is the total number of patients with a positive test result. Likewise, FN + TN gives the total number of patients with a negative test result.

A perfect test would have no FN or FP results. Erroneous test results do occur, however, and you can use a 2 × 2 contingency table to define the measures of test performance that reflect these errors.

3.3.2 Measures of Test Performance

Measures of test performance are of two types: measures of agreement between tests or **measures of concordance**, and measures of disagreement or **measures of discordance**. Two types of **concordant test results** occur in the 2 × 2 table in Table 3.3: TPs and TNs. The relative frequencies of these results form the basis of the measures of concordance. These measures correspond to the ideas of the sensitivity and specificity of a test, which we introduced in ▶Chap. 2. We define each measure in terms of the 2 × 2 table and in terms of conditional probabilities.

The **true-positive rate** (TPR), or **sensitivity**, is the likelihood that a diseased patient has a positive test. In conditional-probability notation, sensitivity is expressed as the probability of a positive test given that disease is present:

$$p[\text{positive test}|\text{disease}].$$

Another way to think of the TPR is as a ratio. The likelihood that a diseased patient has a positive test is given by the ratio of diseased

Table 3.3 A 2 × 2 contingency table for test results

Results of test	Disease present	Disease absent	Total
Positive result	TP	FP	TP + FP
Negative result	FN	TN	FN + TN
	TP + FN	FP + TN	

TP true positive, *TN* true negative, *FP* false positive, *FN* false negative

patients with a positive test to all diseased patients:

$$\text{TPR} = \left(\frac{\text{number of diseased patients with positive test}}{\text{total number of diseased patients}} \right).$$

We can determine these numbers for our example from the 2 × 2 table (see Table 3.3). The number of diseased patients with a positive test is TP. The total number of diseased patients is the sum of the first column, TP + FN. So,

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

The **true-negative** rate (TNR), or **specificity**, is the likelihood that a nondiseased patient has a negative test result. In terms of conditional probability, specificity is the probability of a negative test given that disease is absent:

$$\text{p}[\text{negative test}|\text{no disease}].$$

Viewed as a ratio, the TNR is the number of nondiseased patients with a negative test divided by the total number of nondiseased patients:

$$\text{TNR} = \left(\frac{\text{Number of nondiseased patients with negative test}}{\text{Total number of nondiseased patients}} \right)$$

From the 2 × 2 table (see Table 3.3),

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

The measures of discordance—the **false-positive rate** (FPR) and the **false-negative rate** (FNR)—are defined similarly. The FNR is the likelihood that a diseased patient has a negative test result. As a ratio,

$$\text{FNR} = \left(\frac{\text{Number of diseased patients with negative test}}{\text{Total number of diseased patients}} \right) = \frac{\text{FN}}{\text{FN} + \text{TP}}.$$

Table 3.4 A 2 × 2 contingency table for HIV antibody EIA

EIA test result	Antibody present	Antibody absent	Total
Positive EIA	98	3	101
Negative EIA	2	297	299
	100	300	

EIA enzyme-linked immunoassay

The FPR is the likelihood that a nondiseased patient has a positive test result:

$$\text{FPR} = \left(\frac{\text{Number of nondiseased patients with positive test}}{\text{Total number of nondiseased patients}} \right) = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

► Example 3.6

Consider again the problem of screening blood donors for HIV. One test used to screen blood donors for HIV antibody is an enzyme-linked immunoassay (EIA). So that the performance of the EIA can be measured, the test is performed on 400 patients; the hypothetical results are shown in the 2 × 2 table in Table 3.4.[5] ◄

To determine test performance, we calculate the TPR (sensitivity) and TNR (specificity) of the EIA antibody test. The TPR, as defined previously, is:

$$\frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{98}{98 + 2} = 0.98$$

Thus, the likelihood that a patient with the HIV antibody will have a positive EIA test is

5 This example assumes that we have a perfect method (different from EIA) for determining the presence or absence of antibody. We discuss the idea of gold-standard tests in ► Sect. 3.3.4. We have chosen the numbers in the example to simplify the calculations. In practice, the sensitivity and specificity of the HIV EIAs are greater than 99%.

0.98. If the test were performed on 100 patients who truly had the antibody, we would expect the test to be positive in 98 of the patients. Conversely, we would expect two of the patients to receive incorrect, negative results, for an FNR of 2%. (You should convince yourself that the sum of TPR and FNR by definition must be 1: TPR + FNR = 1).

And the TNR is:

$$\frac{TN}{TN+FP} = \frac{297}{297+3} = 0.99$$

The likelihood that a patient who has no HIV antibody will have a negative test is 0.99. Therefore, if the EIA test were performed on 100 individuals who had not been infected with HIV, it would be negative in 99 and incorrectly positive in 1. (Convince yourself that the sum of TNR and FPR also must be 1: TNR + FPR = 1).

3.3.3 Implications of Sensitivity and Specificity: How to Choose Among Tests

It may be clear to you already that the calculated values of sensitivity and specificity for a continuous-valued test depend on the particular cutoff value chosen to distinguish normal and abnormal results. In Fig. 3.2, note that increasing the cutoff level (moving it to the right) would decrease significantly the number of FP tests but also would increase the number of FN tests. Thus, the test would have become more specific but less sensitive. Similarly, a lower cutoff value would increase the FPs and decrease the FNs, thereby increasing sensitivity while decreasing specificity. Whenever a decision is made about what cutoff to use in calling a test abnormal, an inherent philosophic decision is being made about whether it is better to tolerate FNs (missed cases) or FPs (nondiseased people inappropriately classified as diseased). The choice of cutoff depends on the disease in question and on the purpose of testing. If the disease is serious and if lifesaving therapy is available, we should try to minimize the number of FN results. On the other hand, if the disease in not serious and the therapy is dangerous, we should set the cutoff value to minimize FP results.

We stress the point that sensitivity and specificity are characteristics not of a test per se but rather of the test and a criterion for when to call that test abnormal. Varying the cutoff in Fig. 3.2 has no effect on the test itself (the way it is performed, or the specific values for any particular patient); instead, it trades off specificity for sensitivity. Thus, the best way to characterize a test is by the range of values of sensitivity and specificity that it can take on over a range of possible cutoffs. The typical way to show this relationship is to plot the test's sensitivity against 1 minus specificity (i.e., the TPR against the FPR), as the cutoff is varied and the two test characteristics are traded off against each other (Fig. 3.3). The resulting curve, known as a **receiver-operating characteristic (ROC) curve**, was originally described by researchers investigating methods of electromagnetic-signal detection and was later applied to the field of psychology (Peterson and Birdsall 1953; Swets 1973). Any given point along a ROC curve for a test corresponds to the test sensitivity and specificity for a given threshold of "abnormality." Similar curves can be drawn for any test used to associate ob-

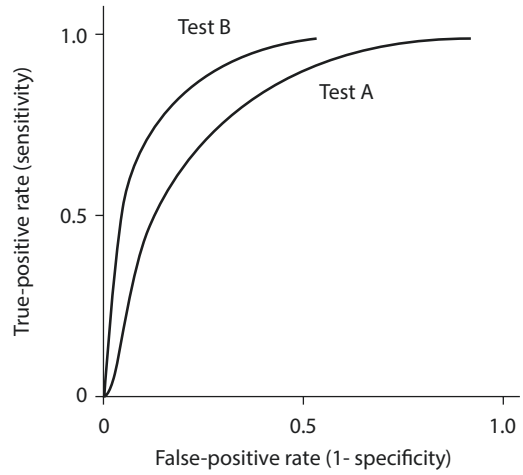


Fig. 3.3 Receiver operating characteristic (ROC) curves for two hypothetical tests. Test B is more discriminative than test A because its curve is higher (e.g., the false-positive rate (FPR) for test B is lower than the FPR for test A at any value of true-positive rate (TPR)). However, the more discriminative test may not always be preferred in clinical practice (see text)

served clinical data with specific diseases or disease categories.

Suppose a new test were introduced that competed with the current way of screening for the presence of a disease. For example, suppose a new radiologic procedure for assessing the presence or absence of pneumonia became available. This new test could be assessed for trade-offs in sensitivity and specificity, and an ROC curve could be drawn. As shown in Fig. 3.3, a test has better discriminating power than a competing test if its ROC curve lies above that of the other test. In other words, test B is more discriminating than test A when its specificity is greater than test A's specificity for any level of sensitivity (and when its sensitivity is greater than test A's sensitivity for any level of specificity).

Understanding ROC curves is important in understanding test selection and data interpretation. Clinicians should not necessarily, however, always choose the test with the most discriminating ROC curve. Matters of cost, risk, discomfort, and delay also are important in the choice about what data to collect and what tests to perform. When you must choose among several available tests, you should select the test that has the highest sensitivity and specificity, provided that other factors, such as cost and risk to the patient, are equal. The higher the sensitivity and specificity of a test, the more the results of that test will reduce uncertainty about probability of disease.

3.3.4 Design of Studies of Test Performance

In ▶ Sect. 3.3.2, we discussed measures of test performance: a test's ability to discriminate disease from no disease. When we classify a test result as TP, TN, FP, or FN, we assume that we know with certainty whether a patient is diseased or healthy. Thus, the validity of any test's results must be measured against a gold standard: a test that reveals the patient's true disease state, such as a biopsy of diseased tissue or a surgical operation. A **gold-standard test** is a procedure that is used to define unequivocally the presence or absence of disease. The test whose discrimination is being measured is called the **index test**. The gold-standard test usually is more expensive, riskier, or more difficult to perform than is the index test (otherwise, the less precise test would not be used at all).

The performance of the index test is measured in a small, select group of patients enrolled in a study. We are interested, however, in how the test performs in the broader group of patients in which it will be used in practice. The test may perform differently in the two groups, so we make the following distinction: the **study population** comprises those patients (usually a subset of the clinically relevant population) in whom test discrimination is measured and reported; the **clinically relevant population** comprises those patients in whom a test typically is used.

3.3.5 Bias in the Measurement of Test Characteristics

We mentioned earlier the problem of referral bias. Published estimates of disease prevalence (derived from a study population) may differ from the prevalence in the clinically relevant population because diseased patients are more likely to be included in studies than are nondiseased patients. Similarly, published values of sensitivity and specificity are derived from study populations that may differ from the clinically relevant populations in terms of average level of health and disease prevalence. These differences may affect test performance, so the reported values may not apply to many patients in whom a test is used in clinical practice.

▶ Example 3.7

In the early 1970s, a blood test called the carcinoembryonic antigen (CEA) was touted as a screening test for colon cancer. Reports of early investigations, performed in selected patients, indicated that the test had high sensitivity and specificity. Subsequent work, however, proved the CEA to be completely valueless as a screening blood test for colon cancer. Screening tests are used in unselected populations, and the differences between the study and clinically relevant populations were partly responsible for the original miscalculations of the CEA's TPR and TNR (Ransohoff and Feinstein 1978). ◀

The experience with CEA has been repeated with numerous tests. Early measures of test discrimination are overly optimistic, and subsequent test performance is disappointing. Problems arise when the TPR and TNR, as measured in the study population, do not apply to the clinically relevant population. These problems usually are the result of bias in the design of the initial studies—notably spectrum bias, test referral bias, or test interpretation bias.

Spectrum bias occurs when the study population includes only individuals who have advanced disease ("sickest of the sick") and healthy volunteers, as is often the case when a test is first being developed. Advanced disease may be easier to detect than early disease. For example, cancer is easier to detect when it has spread throughout the body (metastasized) than when it is localized to, say, a small portion of the colon. In contrast to the study population, the clinically relevant population will contain more cases of early disease that are more likely to be missed by the index test (FNs). Thus, the study population will have an artifactually low FNR, which produces an artifactually high TPR (TPR = 1 − FNR). In addition, healthy volunteers are less likely than are patients in the clinically relevant population to have other diseases that may cause FP results[6]; the study population will have an artificially low FPR, and therefore the specificity will be overestimated (TNR = 1 − FPR). Inaccuracies in early estimates of the TPR and TNR of the CEA were partly due to spectrum bias.

Test-referral bias (sometimes referred to as referral bias) occurs when a positive index test is a criterion for ordering the gold standard test. In clinical practice, patients with negative index tests are less likely to undergo the gold standard test than are patients with positive tests. In other words, the study population, comprising individuals with positive index–test results, has a higher percentage of patients with disease than does the clinically relevant population. Therefore, both TN and FN tests will be underrepresented in the study population. The result is overestimation of the TPR and underestimation of the TNR in the study population.

Test-interpretation bias develops when the interpretation of the index test affects that of the gold standard test or vice versa. This bias causes an artificial concordance between the tests (the results are more likely to be the same) and spuriously increases measures of concordance—the sensitivity and specificity—in the study population. (Remember, the relative frequencies of TPs and TNs are the basis for measures of concordance). To avoid these problems, the person interpreting the index test should be unaware of the results of the gold standard test.

To counter these three biases, you may need to adjust the TPR and TNR when they are applied to a new population. All the biases result in a TPR that is higher in the study population than it is in the clinically relevant population. Thus, if you suspect bias, you should adjust the TPR (sensitivity) downward when you apply it to a new population.

Adjustment of the TNR (specificity) depends on which type of bias is present. Spectrum bias and test interpretation bias result in a TNR that is higher in the study population than it will be in the clinically relevant population. Thus, if these biases are present, you should adjust the specificity downward when you apply it to a new population. Test-referral bias, on the other hand, produces a measured specificity in the study population that is lower than it will be in the clinically relevant population. If you suspect test referral bias, you should adjust the specificity upward when you apply it to a new population.

3.3.6 Meta-Analysis of Diagnostic Tests

Often, there are many studies that evaluate the sensitivity and specificity of the same diagnostic test. If the studies come to similar conclusions

6 Volunteers are often healthy, whereas patients in the clinically relevant population often have several diseases in addition to the disease for which a test is designed. These other diseases may cause FP test results. For example, patients with benign (rather than malignant) enlargement of their prostate glands are more likely than are healthy volunteers to have FP elevations of prostate-specific antigen (Meigs et al. 1996), a substance in the blood that is elevated in men who have prostate cancer. Measurement of prostate-specific antigen is often used to detect prostate cancer.

about the sensitivity and specificity of the test, you can have increased confidence in the results of the studies. But what if the studies disagree? For example, by 1995, over 100 studies had assessed the sensitivity and specificity of the PCR for diagnosis of HIV (Owens et al. 1996a, b); these studies estimated the sensitivity of PCR to be as low as 10% and to be as high as 100%, and they assessed the specificity of PCR to be between 40 and 100%. Which results should you believe? One approach that you can use is to assess the quality of the studies and to use the estimates from the highest-quality studies.

For evaluation of PCR, however, even the high-quality studies did not agree. Another approach is to perform a **meta-analysis**: a study that combines quantitatively the estimates from individual studies to develop a **summary ROC curve** (Moses et al. 1993; Owens et al. 1996a, b; Hellmich et al. 1999; Leeflang et al. 2008; Leeflang 2014). Investigators develop a summary ROC curve by using estimates from many studies, in contrast to the type of ROC curve discussed in ▶ Sect. 3.3.3, which is developed from the data in a single study. Summary ROC curves provide the best available approach to synthesizing data from many studies.

▶ Section 3.3 has dealt with the second step in the diagnostic process: acquisition of further information with diagnostic tests. We have learned how to characterize the performance of a test with sensitivity (TPR) and specificity (TNR). These measures reveal the probability of a test result given the true state of the patient. They do not, however, answer the clinically relevant question posed in the opening example: Given a positive test result, what is the probability that this patient has the disease? To answer this question, we must learn methods to calculate the post-test probability of disease.

3.4 Post-test Probability: Bayes' Theorem and Predictive Value

The third stage of the diagnostic process (see ◘ Fig. 3.1a) is to adjust our probability estimate to take into account the new information gained from diagnostic tests by calculating the post-test probability.

3.4.1 Bayes' Theorem

As we noted earlier in this chapter, a clinician can use the disease prevalence in the patient population as an initial estimate of the pretest risk of disease. Once clinicians begin to accumulate information about a patient, however, they revise their estimate of the probability of disease. The revised estimate (rather than the disease prevalence in the general population) becomes the pretest probability for the test that they perform. After they have gathered more information with a diagnostic test, they can calculate the post-test probability of disease with Bayes' theorem.

Bayes' theorem is a quantitative method for calculating post-test probability using the pretest probability and the sensitivity and specificity of the test. The theorem is derived from the definition of conditional probability and from the properties of probability (see the Appendix to this chapter for the derivation).

Recall that a conditional probability is the probability that event A will occur given that event B is known to occur (see ▶ Sect. 3.2). In general, we want to know the probability that disease is present (event A), given that the test is known to be positive (event B). We denote the presence of disease as D, its absence as − D, a test result as R, and the pretest probability of disease as p[D]. The probability of disease, given a test result, is written p[D|R]. Bayes' theorem is:

$$p[D|R] = \frac{p[D] \times p[R|D]}{p[D] \times p[R|D] + p[-D] \times p[R|-D]}$$

We can reformulate this general equation in terms of a positive test, (+), by substituting p[D|+] for p[D|R], p[+|D] for p[R|D], p[+| − D] for p[R| − D], and 1 − p[D] for p[− D]. From ▶ Sect. 3.3, recall that p[+|D] = TPR and p[+| − D] = FPR. Substitution provides Bayes' theorem for a positive test:

$$p[D|+] = \frac{p[D] \times TPR}{p[D] \times TPR + (1 - p[D]) \times FPR}$$

We can use a similar derivation to develop Bayes' theorem for a negative test:

$$p[D|-] = \frac{p[D] \times FNR}{p[D] \times FNR + (1 - p[D]) \times TNR}$$

► Example 3.8

We are now able to calculate the clinically important probability in ►Example 3.4: the post-test probability of heart disease after a positive exercise test. At the end of ►Sect. 3.2.2, we estimated the pretest probability of heart disease as 0.95, based on the prevalence of heart disease in men who have typical symptoms of heart disease and on the prevalence in people with a family history of heart disease. Assume that the TPR and FPR of the exercise stress test are 0.65 and 0.20, respectively. Substituting in Bayes' formula for a positive test, we obtain the probability of heart disease given a positive test result:

$$p[D|+] = \frac{0.95 \times 0.65}{0.95 \times 0.65 + 0.05 \times 0.20} = 0.98 \quad \blacktriangleleft$$

Thus, the positive test raised the post-test probability to 0.98 from the pretest probability of 0.95. The change in probability is modest because the pretest probability was high (0.95) and because the FPR also is high (0.20). If we repeat the calculation with a pretest probability of 0.75, the post-test probability is 0.91. If we assume the FPR of the test to be 0.05 instead of 0.20, a pretest probability of 0.95 changes to 0.996.

3.4.2 The Odds-Ratio Form of Bayes' Theorem and Likelihood Ratios

Although the formula for Bayes' theorem is straightforward, it is awkward for mental calculations. We can develop a more convenient form of Bayes' theorem by expressing probability as odds and by using a different measure of test discrimination. Probability and odds are related as follows:

$$\text{odds} = \frac{p}{1-p},$$

$$p = \frac{\text{odds}}{1+\text{odds}}.$$

Thus, if the probability of rain today is 0.75, the odds are 3:1. Thus, on similar days, we should expect rain to occur three times for each time it does not occur.

A simple relationship exists between pre-test odds and post-test odds:

$$\text{post-test odds} = \text{pretest odds} \times \text{likelihood ratio}$$

or

$$\frac{p[D|R]}{p[-D|R]} = \frac{p[D]}{p[-D]} \times \frac{p[R|D]}{p[R|-D]}.$$

This equation is the **odds-ratio form** of Bayes' theorem.[7] It can be derived in a straightforward fashion from the definitions of Bayes' theorem and of conditional probability that we provided earlier. Thus, to obtain the post-test odds, we simply multiply the pre-test odds by the **likelihood ratio** (LR) for the test in question.

The LR of a test combines the measures of test discrimination discussed earlier to give one number that characterizes the discriminatory power of a test, defined as:

$$LR = \frac{p[R|D]}{p[R|-D]}$$

or

$$LR = \frac{\text{probability of result in diseased people}}{\text{probability of result in nondiseased people}}$$

The LR indicates the amount that the odds of disease change based on the test result. We can use the LR to characterize clinical findings (such as a swollen leg) or a test result. We describe the performance of a test that has only two possible outcomes (e.g., positive or negative) by two LRs: one corresponding to a positive test result and the other corresponding to a negative test. These ratios are abbreviated LR+ and LR−, respectively.

$$LR+ = \left(\frac{\text{probability that test is positive in diseased people}}{\text{probability that test is positive in nondiseased people}} \right) = \frac{TPR}{FPR}$$

7 Some authors refer to this expression as the odds-likelihood form of Bayes' theorem.

In a test that discriminates well between disease and nondisease, the TPR will be high, the FPR will be low, and thus LR+ will be much greater than 1. A LR of 1 means that the probability of a test result is the same in diseased and nondiseased individuals; the test has no value. Similarly,

$$\text{LR}- = \left(\frac{\begin{array}{l}\text{probability that test}\\ \text{is negative in}\\ \text{diseased people}\end{array}}{\begin{array}{l}\text{probability that test}\\ \text{is negative in}\\ \text{nondiseased people}\end{array}} \right) = \frac{\text{FNR}}{\text{TNR}}$$

A desirable test will have a low FNR and a high TNR; therefore, the LR− will be much less than 1.

► Example 3.9

We can calculate the post-test probability for a positive exercise stress test in a 70 year-old woman whose pretest probability is 0.75. The pretest odds are:

$$\text{odds} = \frac{p}{1-p} = \frac{0.75}{1-0.75} = \frac{0.75}{0.25} = 3, \text{or } 3:1$$

The LR for the stress test is:

$$\text{LR}+ = \frac{\text{TPR}}{\text{FPR}} = \frac{0.65}{0.20} = 3.25$$

We can calculate the post-test odds of a positive test result using the odds-ratio form of Bayes' theorem:

$$\text{post-test odds} = 3 \times 3.25 = 9.75:1$$

We can then convert the odds to a probability:

$$\text{p} = \frac{\text{odds}}{1+\text{odds}} = \frac{9.75}{1+9.75} = 0.91$$ ◄

As expected, this result agrees with our earlier answer (see the discussion of ► Example 3.8).

The odds-ratio form of Bayes' theorem allows rapid calculation. The LR is a powerful method for characterizing the operating characteristics of a test: if you know the pretest odds, you can calculate the post-test odds in one step. The LR demonstrates that a useful test is one that changes the odds of disease.

3.4.3 Predictive Value of a Test

An alternative approach for estimation of the probability of disease in a person who has a positive or negative test is to calculate the predictive value of the test. The **positive predictive value** (PV+) of a test is the likelihood that a patient who has a positive test result also has disease. Thus, PV+ can be calculated directly from a 2 × 2 contingency table:

$$\text{PV}+ = \frac{\begin{array}{l}\text{number of diseased patients}\\ \text{with positive test}\end{array}}{\begin{array}{l}\text{total number of patients}\\ \text{with a positive test}\end{array}}$$

From the 2 × 2 contingency table in Table 3.3,

$$\text{PV}+ = \frac{\text{TP}}{\text{TP}+\text{FP}}$$

The **negative predictive value** (PV−) is the likelihood that a patient with a negative test does not have disease:

$$\text{PV}- = \frac{\begin{array}{l}\text{number of nondiseased patients}\\ \text{with negative test}\end{array}}{\begin{array}{l}\text{Total number of patients}\\ \text{with a negative test}\end{array}}$$

From the 2 × 2 contingency table in Table 3.3,

$$\text{PV}- = \frac{\text{TN}}{\text{TN}+\text{FN}}$$

► Example 3.10

We can calculate the PV of the EIA test from the 2 × 2 table that we constructed in ► Example 3.6 (see Table 3.4) as follows:

$$\text{PV}+ = \frac{98}{98+3} = 0.97$$

$$\text{PV}- = \frac{297}{297+2} = 0.99$$

The probability that antibody is present in a patient who has a positive index test (EIA) in this study is 0.97; about 97 of 100 patients with a positive test will have antibody. The likelihood that a patient with a negative index test does not have antibody is about 0.99. ◄

It is worth reemphasizing the difference between PV and sensitivity and specificity, given that both are calculated from the 2 × 2 table and they often are confused. The sensitivity and specificity give the probability of a particular test result in a patient who has a particular disease state. The PV gives the probability of true disease state once the patient's test result is known.

The PV+ calculated from Table 3.4 is 0.97, so we expect 97 of 100 patients with a positive index test actually to have antibody. Yet, in ▶ Example 3.1, we found that fewer than one of ten patients with a positive test were expected to have antibody. What explains the discrepancy in these examples? The sensitivity and specificity (and, therefore, the LRs) in the two examples are identical. The discrepancy is due to an extremely important and often overlooked characteristic of PV: the PV of a test depends on the prevalence of disease in the study population (the prevalence can be calculated as TP + FN divided by the total number of patients in the 2 × 2 table). The PV cannot be generalized to a new population because the prevalence of disease may differ between the two populations.

The difference in PV of the EIA in ▶ Example 3.1 and in ▶ Example 3.6 is due to a difference in the prevalence of disease in the examples. The prevalence of antibody was given as 0.001 in ▶ Example 3.1 and as 0.25 in ▶ Example 3.6. These examples should remind us that the PV+ is not an intrinsic property of a test. Rather, it represents the post-test probability of disease only when the prevalence is identical to that in the 2 × 2 contingency table from which the PV+ was calculated. Bayes' theorem provides a method for calculation of the post-test probability of disease for any prior probability. For that reason, we prefer the use of Bayes' theorem to calculate the post-test probability of disease.

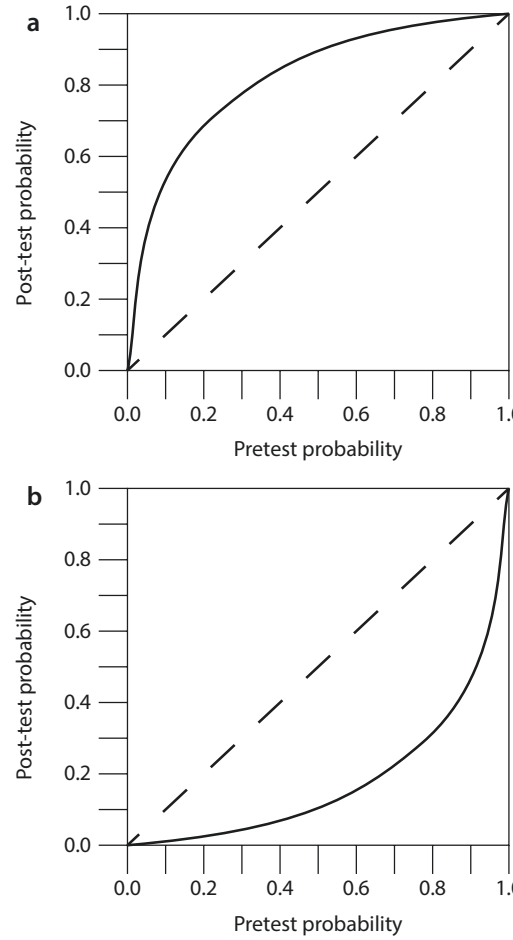


Fig. 3.4 Relationship between pretest probability and post-test probability of disease. The *dashed lines* correspond to a test that has no effect on the probability of disease. Sensitivity and specificity of the test were assumed to be 0.90 for the two examples. **a** The post-test probability of disease corresponding to a positive test result (*solid curve*) was calculated with Bayes' theorem for all values of pretest probability. **b** The post-test probability of disease corresponding to a negative test result (*solid curve*) was calculated with Bayes' theorem for all values of pretest probability. (Source: Adapted from Sox (1987), with permission)

3.4.4 Implications of Bayes' Theorem

In this section, we explore the implications of Bayes' theorem for test interpretation. These ideas are extremely important, yet they often are misunderstood.

Figure 3.4 illustrates one of the most essential concepts in this chapter: The post-test probability of disease increases as the pretest probability of disease increases. We produced Fig. 3.4a by calculating the post-test probability after a positive test result for all possible pretest probabilities of disease. We similarly derived Fig. 3.4b for a negative test result.

The 45-degree line in each figure denotes a test in which the pretest and post-test probability are equal (LR = 1), indicating a test that

is useless. The curve in Fig. 3.4a relates pretest and post-test probabilities in a test with a sensitivity and specificity of 0.9. Note that, at low pretest probabilities, the post-test probability after a positive test result is much higher than is the pretest probability. At high pretest probabilities, the post-test probability is only slightly higher than the pretest probability.

Figure 3.4b shows the relationship between the pretest and post-test probabilities after a negative test result. At high pretest probabilities, the post-test probability after a negative test result is much lower than is the pretest probability. A negative test, however, has little effect on the post-test probability if the pretest probability is low.

This discussion emphasizes a key idea of this chapter: the interpretation of a test result depends on the pretest probability of disease. If the pretest probability is low, a positive test result has a large effect, and a negative test result has a small effect. If the pretest probability is high, a positive test result has a small effect, and a negative test result has a large effect. In other words, when the clinician is almost certain of the diagnosis before testing (pretest probability nearly 0 or nearly 1), a confirmatory test has little effect on the posterior probability (see ▶ Example 3.8). If the pretest probability is intermediate or if the result contradicts a strongly held clinical impression, the test result will have a large effect on the post-test probability.

Note from Fig. 3.4a that, if the pretest probability is very low, a positive test result can raise the post-test probability into only the intermediate range. Assume that Fig. 3.4a represents the relationship between the pretest and post-test probabilities for the exercise stress test. If the clinician believes the pretest probability of coronary artery disease is 0.1, the post-test probability will be about 0.5. Although there has been a large change in the probability, the post-test probability is in an intermediate range, which leaves considerable uncertainty about the diagnosis. Thus, if the pretest probability is low, it is unlikely that a positive test result will raise the probability of disease sufficiently for the clinician to make that diagnosis with confidence. An exception to this statement occurs when a test has a very high specificity (or a large LR+); e.g., HIV antibody tests have a specificity greater than 0.99, and therefore a positive test is convincing. Similarly, if the pretest probability is very high, it is unlikely that a negative test result will lower the post-test probability sufficiently to exclude a diagnosis.

Figure 3.5 illustrates another important concept: test specificity affects primarily the in-

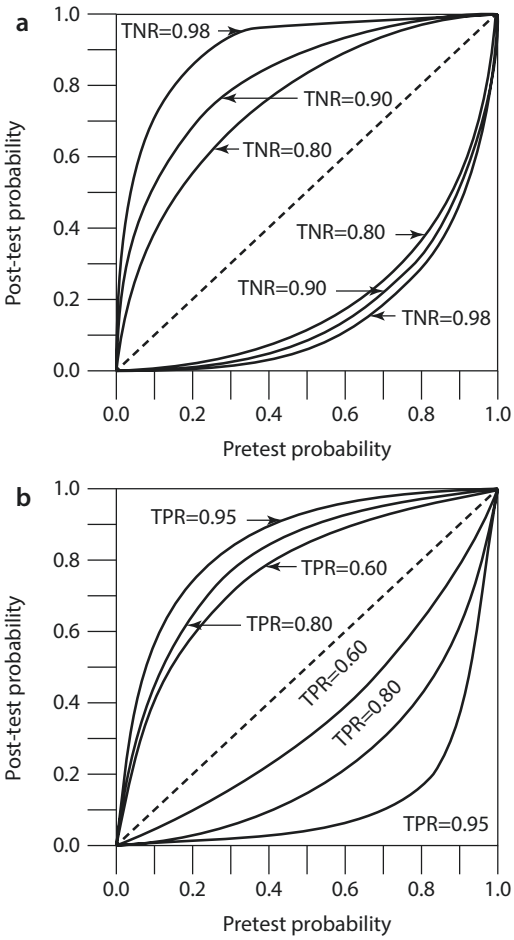


Fig. 3.5 Effects of test sensitivity and specificity on post-test probability. The *curves* are similar to those shown in Fig. 3.4 except that the calculations have been repeated for several values of the sensitivity (*TPR* true-positive rate) and specificity (*TNR* true-negative rate) of the test. **a** The sensitivity of the test was assumed to be 0.90, and the calculations were repeated for several values of test specificity. **b** The specificity of the test was assumed to be 0.90, and the calculations were repeated for several values of the sensitivity of the test. In both panels, the top family of curves corresponds to positive test results, and the bottom family of curves corresponds to negative test results. (Source: Adapted from Sox (1987), with permission)

terpretation of a positive test; test sensitivity affects primarily the interpretation of a negative test. In both parts (a) and (b) of Fig. 3.5, the top family of curves corresponds to positive test results and the bottom family to negative test results. Figure 3.5a shows the post-test probabilities for tests with varying specificities (TNR). Note that changes in the specificity produce large changes in the top family of curves (positive test results) but have little effect on the lower family of curves (negative test results). That is, an increase in the specificity of a test markedly changes the post-test probability if the test is positive but has relatively little effect on the post-test probability if the test is negative. Thus, if you are trying to rule in a diagnosis,[8] you should choose a test with high specificity or a high LR+. Figure 3.5b shows the post-test probabilities for tests with varying sensitivities. Note that changes in sensitivity produce large changes in the bottom family of curves (negative test results) but have little effect on the top family of curves. Thus, if you are trying to exclude a disease, choose a test with a high sensitivity or a high LR−.

3.4.5 Cautions in the Application of Bayes' Theorem

Bayes' theorem provides a powerful method for calculating post-test probability. You should be aware, however, of the possible errors you can make when you use it. Common problems are inaccurate estimation of pretest probability, faulty application of test-performance measures, and violation of the assumptions of conditional independence and of mutual exclusivity.

Bayes' theorem provides a means to adjust an estimate of pretest probability to take into account new information. The accuracy of the calculated post-test probability is limited, however, by the accuracy of the estimated pretest probability. Accuracy of estimated prior probability is increased by proper use of published prevalence rates, heuristics, and clinical prediction rules. In a decision analysis, as we shall see, a range of prior probability often is sufficient. Nonetheless, if the pretest probability assessment is unreliable, Bayes' theorem will be of little value.

A second potential mistake that you can make when using Bayes' theorem is to apply published values for the test sensitivity and specificity, or LRs, without paying attention to the possible effects of bias in the studies in which the test performance was measured (see ▶ Sect. 3.3.5). With certain tests, the LRs may differ depending on the pretest odds in part because differences in pretest odds may reflect differences in the spectrum of disease in the population.

A third potential problem arises when you use Bayes' theorem to interpret a sequence of tests. If a patient undergoes two tests in sequence, you can use the post-test probability after the first test result, calculated with Bayes' theorem, as the pretest probability for the second test. Then, you use Bayes' theorem a second time to calculate the post-test probability after the second test. This approach is valid, however, only if the two tests are conditionally independent. Tests for the same disease are **conditionally independent** when the probability of a particular result on the second test does not depend on the result of the first test, given (conditioned on) the disease state. Expressed in conditional probability notation for the case in which the disease is present,

8 In medicine, to *rule in* a disease is to confirm that the patient *does* have the disease; to *rule out* a disease is to confirm that the patient *does not* have the disease. A doctor who strongly suspects that his or her patient has a bacterial infection orders a culture to *rule in* his or her diagnosis. Another doctor is almost certain that his or her patient has a simple sore throat but orders a culture to rule out streptococcal infection (strep throat). This terminology oversimplifies a diagnostic process that is probabilistic. Diagnostic tests rarely, if ever, rule in or rule out a disease; rather, the tests raise or lower the probability of disease.

$$p[\text{second test positive}|\text{first test positive and disease present}]$$
$$= p[\text{second test positive}|\text{first test negative and disease present}]$$
$$= p[\text{second test positive}|\text{disease present}].$$

If the conditional independence assumption is satisfied, the post-test odds = pretest odds × LR1 × LR2. If you apply Bayes' theorem sequentially in situations in which conditional independence is violated, you will obtain inaccurate post-test probabilities.

The fourth common problem arises when you assume that all test abnormalities result from one (and only one) disease process. The Bayesian approach, as we have described it, generally presumes that the diseases under consideration are **mutually exclusive**. If they are not, Bayesian updating must be applied with great care.

We have shown how to calculate post-test probability. In ▶Sect. 3.5, we turn to the problem of decision making when the outcomes of a clinician's actions (e.g., of treatments) are uncertain.

3.5 Expected-Value Decision Making

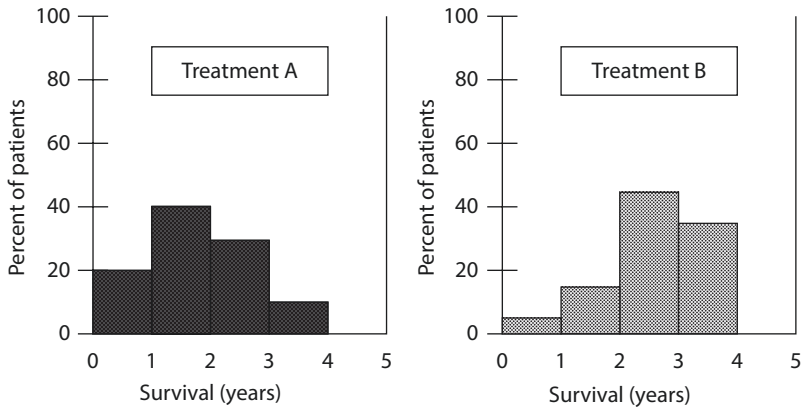
Medical decision-making problems often cannot be solved by reasoning based on pathophysiology. For example, clinicians need a method for choosing among treatments when the outcome of the treatments is uncertain, as are the results of a surgical operation. You can use the ideas developed in the preceding sections to solve such difficult decision problems. Here we discuss two methods: the decision tree, a method for representing and comparing the expected outcomes of each decision alternative; and the threshold probability, a method for deciding whether new information can change a management decision. These techniques help you to clarify the decision problem and thus to choose the alternative that is most likely to help the patient.

3.5.1 Comparison of Uncertain Prospects

Like those of most biological events, the outcome of an individual's illness is unpredictable. How can a clinician determine which course of action has the greatest chance of success?

▶ Example 3.11

There are two available therapies for a fatal illness. The length of a patient's life after either therapy is unpredictable, as illustrated by the frequency distribution shown in ◘ Fig. 3.6 and summarized in ◘ Table 3.5. Each therapy is associated with uncertainty: regardless of which therapy a patient receives, the patient will die by the end of the fourth year, but there is no way to know which



◘ **Fig. 3.6** Survival after therapy for a fatal disease. Two therapies are available; the results of either are unpredictable

Table 3.5 Distribution of probabilities for the two therapies in Fig. 3.7

	Probability of death	
Years after therapy	Therapy A	Therapy B
1	0.20	0.05
2	0.40	0.15
3	0.30	0.45
4	0.10	0.35

year will be the patient's last. Figure 3.6 shows that survival until the fourth year is more likely with therapy B, but the patient might die in the first year with therapy B or might survive to the fourth year with therapy A. ◄

Which of the two therapies is preferable? ► Example 3.11 demonstrates a significant fact: a choice among therapies is a choice among gambles (i.e., situations in which chance determines the outcomes). How do we usually choose among gambles? More often than not, we rely on hunches or on a sixth sense. How should we choose among gambles? We propose a method for choosing called expected-value decision making: we characterize each gamble by a number, and we use that number to compare the gambles.[9] In ► Example 3.11, therapy A and therapy B are both gambles with respect to duration of life after therapy. We want to assign a measure (or number) to each therapy that summarizes the outcomes such that we can decide which therapy is preferable.

The ideal criterion for choosing a gamble should be a number that reflects preferences (in medicine, often the patient's preferences) for the outcomes of the gamble. **Utility** is the name given to a measure of preference that has a desirable property for decision making: the gamble with the highest utility should be preferred. We shall discuss utility briefly (► Sect. 3.5.4), but you can pursue this topic and the details of decision analysis in other textbooks (see Suggested Readings at the end of this chapter).[10] We use the average duration of life after therapy (survival) as a criterion for choosing among therapies; remember that this model is oversimplified, used here for discussion only. Later, we consider other factors, such as the quality of life.

Because we cannot be sure of the duration of survival for any given patient, we characterize a therapy by the mean survival (average length of life) that would be observed in a large number of patients after they were given the therapy. The first step we take in calculating the mean survival for a therapy is to divide the population receiving the therapy into groups of patients who have similar survival rates. Then, we multiply the survival time in each group[11] by the fraction of the total population in that group. Finally, we sum these products over all possible survival values.

We can perform this calculation for the therapies in ► Example 3.11. Mean survival for therapy A = (0.2 × 1.0) + (0.4 × 2.0) + (0.3 × 3.0) + (0.1 × 4.0) = 2.3 years. Mean survival for therapy B = (0.05 × 1.0) + (0.15 × 2.0) + (0.45 × 3.0) + (0.35 × 4.0) = 3.1 years.

Survival after a therapy is under the control of chance. Therapy A is a gamble characterized by an average survival equal to 2.3 years. Therapy B is a gamble characterized by an average survival of 3.1 years. If length of life is our criterion for choosing, we should select therapy B.

3.5.2 Representation of Choices with Decision Trees

The choice between therapies A and B is represented diagrammatically in Fig. 3.7. Events that are under the control of chance can be represented by a **chance node**. By convention, a

9 Expected-value decision making had been used in many fields before it was first applied to medicine.

10 A more general term for expected-value decision making is expected utility decision making. Because a full treatment of utility is beyond the scope of this chapter, we have chosen to use the term expected value.

11 For this simple example, death during an interval is assumed to occur at the end of the year.

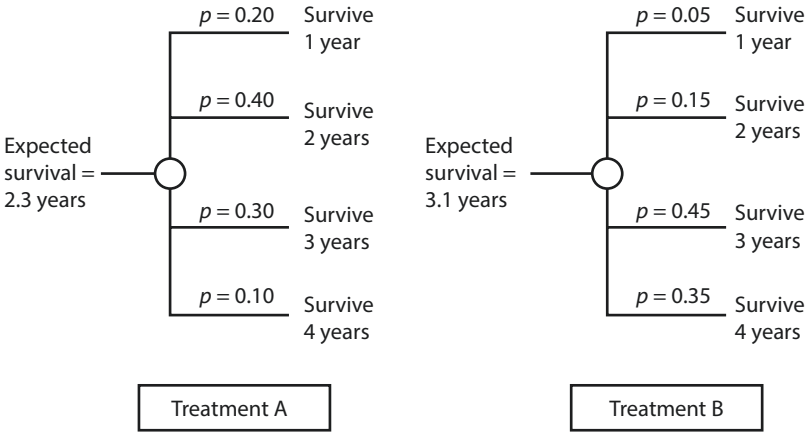


Fig. 3.7 A chance-node representation of survival after the two therapies in Fig. 3.6. The probabilities times the corresponding years of survival are summed to obtain the total expected survival

chance node is shown as a circle from which several lines emanate. Each line represents one of the possible outcomes. Associated with each line is the probability of the outcome occurring. For a single patient, only one outcome can occur. Some physicians object to using probability for just this reason: "You cannot rely on population data, because each patient is an individual." In fact, we often must use the frequency of the outcomes of many patients experiencing the same event to inform our opinion about what might happen to an individual. From these frequencies, we can make patient-specific adjustments and thus estimate the probability of each outcome at a chance node.

A chance node can represent more than just an event governed by chance. The outcome of a chance event, unknowable for the individual, can be represented by the **expected value** at the chance node. The concept of expected value is important and is easy to understand. We can calculate the mean survival that would be expected based on the probabilities depicted by the chance node in Fig. 3.7. This average length of life is called the expected survival or, more generally, the expected value of the chance node. We calculate the expected value at a chance node by the process just described: we multiply the survival value associated with each possible outcome by the probability that that outcome will occur. We then sum the product of probability times survival over all outcomes. Thus, if several hundred patients were assigned to receive either therapy A or therapy B, the expected survival would be 2.3 years for therapy A and 3.1 years for therapy B.

We have just described the basis of expected-value decision making. The term expected value is used to characterize a chance event, such as the outcome of a therapy. If the outcomes of a therapy are measured in units of duration of survival, units of sense of well-being, or dollars, the therapy is characterized by the expected duration of survival, expected sense of well-being, or expected monetary cost that it will confer on, or incur for, the patient, respectively.

To use expected-value decision making, we follow this strategy when there are therapy choices with uncertain outcomes: (1) calculate the expected value of each decision alternative and then (2) pick the alternative with the highest expected value.

3.5.3 Performance of a Decision Analysis

We clarify the concepts of expected-value decision making by discussing an example. There are four steps in decision analysis:

1. Create a decision tree; this step is the most difficult, because it requires formulating the decision problem, assigning probabilities, and measuring outcomes.

2. Calculate the expected value of each decision alternative.
3. Choose the decision alternative with the highest expected value.
4. Use sensitivity analysis to test the conclusions of the analysis.

Some health professionals hesitate when they first learn about the technique of decision analysis, because they recognize the opportunity for error in assigning values to both the probabilities and the utilities in a decision tree. They reason that the technique encourages decision making based on small differences in expected values that are estimates at best. The defense against this concern, which also has been recognized by decision analysts, is the technique known as sensitivity analysis. We discuss this important fourth step in decision analysis in ▶ Sect. 3.5.5. In addition, decision analysis helps make the assumptions underlying a decision explicit, so that the assumptions can be assessed carefully.

The first step in decision analysis is to create a **decision tree** that represents the decision problem. Consider the following clinical problem.

▶ Example 3.12

The patient is Mr. Danby, a 66-year-old man who has been crippled with arthritis of both knees so severely that, while he can get about the house with the aid of two canes, he must otherwise use a wheelchair. His other major health problem is emphysema, a disease in which the lungs lose their ability to exchange oxygen and carbon dioxide between blood and air, which in turn causes shortness of breath (dyspnea). He is able to breathe comfortably when he is in a wheelchair, but the effort of walking with canes makes him breathe heavily and feel uncomfortable. Several years ago, he seriously considered knee replacement surgery but decided against it, largely because his internist told him that there was a serious risk that he would not survive the operation because of his lung disease. Recently, however, Mr. Danby's wife had a stroke and was partially paralyzed; she now requires a degree of assistance that the patient cannot supply given his present state of mobility. He tells his doctor that he is reconsidering knee replacement surgery.

Mr. Danby's internist is familiar with decision analysis. She recognizes that this problem is filled with uncertainty: Mr. Danby's ability to survive the operation is in doubt, and the surgery sometimes does not restore mobility to the degree required by such a patient. Furthermore, there is a small chance that the prosthesis (the artificial knee) will become infected, and Mr. Danby then would have to undergo a second risky operation to remove it. After removal of the prosthesis, Mr. Danby would never again be able to walk, even with canes. The possible outcomes of knee replacement include death from the first procedure and death from a second mandatory procedure if the prosthesis becomes infected (which we will assume occurs in the immediate postoperative period, if it occurs at all). Possible functional outcomes include recovery of full mobility or continued, and unchanged, poor mobility. Should Mr. Danby choose to undergo knee replacement surgery, or should he accept the status quo? ◀

Using the conventions of decision analysis, the internist sketches the decision tree shown in ◘ Fig. 3.8. According to these conventions, a square box denotes a **decision node**, and each line emanating from a decision node represents an action that could be taken.

According to the methods of expected-value decision making, the internist first must assign a probability to each branch of each chance node. To accomplish this task, the internist asks several orthopedic surgeons for their estimates of the chance of recovering full function after surgery (p[full recovery] = 0.60) and the chance of developing infection in the prosthetic joint (p[infection] = 0.05). She uses her subjective estimate of the probability that the patient will die during or immediately after knee surgery (p[operative death] = 0.05).

Next, she must assign a value to each outcome. To accomplish this task, she first lists the outcomes. As you can see from ◘ Table 3.6, the outcomes differ in two dimensions: length of life (survival) and quality of life (functional status). To characterize each outcome accurately, the internist must develop a measure that takes into account these two dimensions. Simply using duration of survival is inadequate because Mr. Danby values 5 years of good health more than

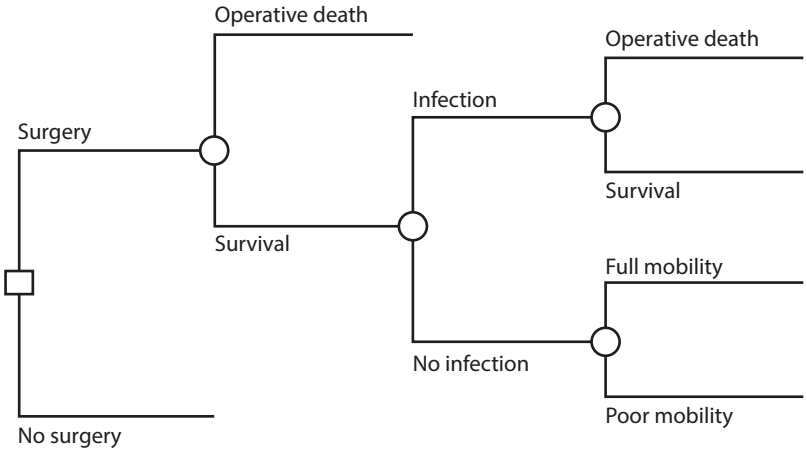


Fig. 3.8 Decision tree for knee replacement surgery. The *box* represents the decision node (whether to have surgery); the *circles* represent chance nodes

Table 3.6 Outcomes for ▶ Example 3.12

Survival (years)	Functional status	Years of full function equivalent to outcome
10	Full mobility (successful surgery)	10
10	Poor mobility (status quo or unsuccessful surgery)	6
10	Wheelchair-bound (the outcome if a second surgery is necessary)	3
0	Death	0

he values 10 years of poor health. The internist can account for this trade-off factor by converting outcomes with two dimensions into outcomes with a single dimension: duration of survival in good health. The resulting measure is called a **quality-adjusted life year** (QALY).[12]

She can convert years in poor health into years in good health by asking Mr. Danby to indicate the shortest period in good health (full mobility) that he would accept in return for his full expected lifetime (10 years) in a state of poor health (status quo). Thus, she asks Mr. Danby: "Many people say they would be willing to accept a shorter life in excellent health in preference to a longer life with significant disability. In your case, how many years with normal mobility do you feel is equivalent in value to 10 years in your current state of disability?" She asks him this question for each outcome. The patient's responses are shown in the third column of Table 3.6. The patient decides that 10 years of limited mobility are equivalent to 6 years of normal mobility, whereas 10 years of wheelchair confinement are equivalent to only 3 years of full function. Figure 3.9 shows the final decision tree—complete with probability estimates and utility values for each outcome.[13]

The second task that the internist must undertake is to calculate the expected value, in healthy years, of surgery and of no surgery. She calculates the expected value at each chance node, moving from right (the tips of

12 QALYs commonly are used as measures of utility (value) in medical decision analysis and in health policy analysis.

13 In a more sophisticated decision analysis, the clinician also would adjust the utility values of outcomes that require surgery to account for the pain and inconvenience associated with surgery and rehabilitation. Other approaches to assessing utility are available and may be preferable in some circumstances.

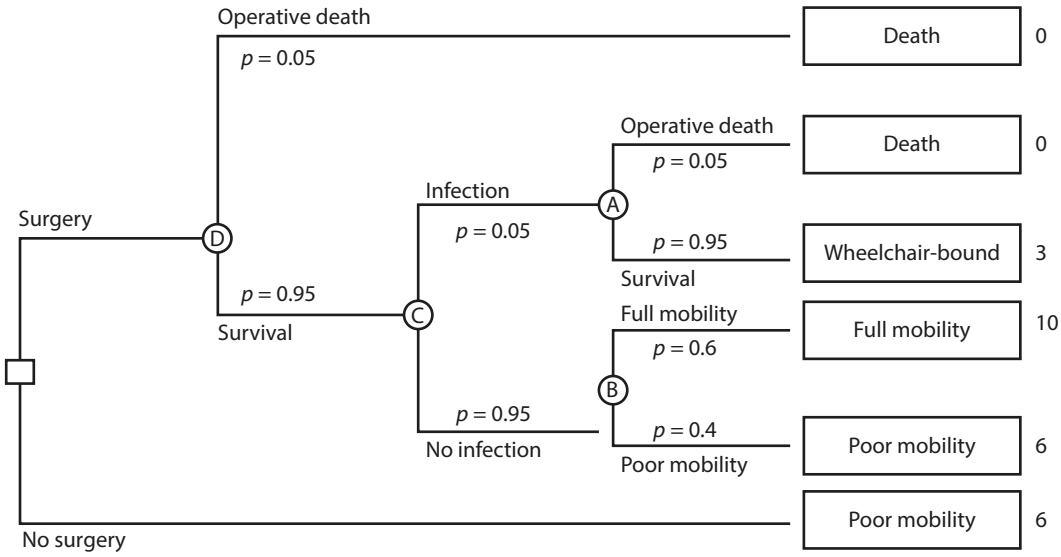


Fig. 3.9 Decision tree for knee-replacement surgery. Probabilities have been assigned to each branch of each chance node. The patient's valuations of outcomes (measured in years of perfect mobility) are assigned to the tips of each branch of the tree

the tree) to left (the root of the tree). Let us consider, for example, the expected value at the chance node representing the outcome of surgery to remove an infected prosthesis (Node A in Fig. 3.9). The calculation requires three steps:

1. Calculate the expected value of operative death after surgery to remove an infected prosthesis. Multiply the probability of operative death (0.05) by the QALY of the outcome—death (0 years): 0.05 × 0 = 0 QALY.
2. Calculate the expected value of surviving surgery to remove an infected knee prosthesis. Multiply the probability of surviving the operation (0.95) by the number of healthy years equivalent to 10 years of being wheelchair-bound (3 years): 0.95 × 3 = 2.85 QALYs.
3. Add the expected values calculated in step 1 (0 QALY) and step 2 (2.85 QALYs) to obtain the expected value of developing an infected prosthesis: 0 + 2.85 = 2.85 QALYs.

Similarly, the expected value at chance node B is calculated: (0.6 × 10) + (0.4 × 6) =8.4 QALYs. To obtain the expected value of surviving knee replacement surgery (Node C), she proceeds as follows:

1. Multiply the expected value of an infected prosthesis (already calculated as 2.85 QALYs) by the probability that the prosthesis will become infected (0.05): 2.85 × 0.05 = 0.143 QALYs.
2. Multiply the expected value of never developing an infected prosthesis (already calculated as 8.4 QALYs) by the probability that the prosthesis will not become infected (0.95): 8.4 × 0.95 = 7.98 QALYs.
3. Add the expected values calculated in step 1 (0.143 QALY) and step 2 (7.98 QALYs) to get the expected value of surviving knee replacement surgery: 0.143 + 7.98 = 8.123 QALYs.

The clinician performs this process, called **averaging out at chance nodes**, for node D as well, working back to the root of the tree, until the expected value of surgery has been calculated. The outcome of the analysis is as follows. For surgery, Mr. Danby's average life expectancy, measured in years of normal mobility, is 7.7. What does this value mean? It does not mean that, by accepting surgery,

Mr. Danby is guaranteed 7.7 years of mobile life. One look at the decision tree will show that some patients die in surgery, some develop infection, and some do not gain any improvement in mobility after surgery. Thus, an individual patient has no guarantees. If the clinician had 100 similar patients who underwent the surgery, however, the average number of mobile years would be 7.7. We can understand what this value means for Mr. Danby only by examining the alternative: no surgery.

In the analysis for no surgery, the average length of life, measured in years of normal mobility, is 6.0, which Mr. Danby considered equivalent to 10 years of continued poor mobility. Not all patients will experience this outcome; some who have poor mobility will live longer than, and some will live less than, 10 years. The average length of life, however, expressed in years of normal mobility, will be 6. Because 6.0 is less than 7.7, on average the surgery will provide an outcome with higher value to the patient. Thus, the internist recommends performing the surgery.

The key insight of expected-value decision making should be clear from this example: given the unpredictable outcome in an individual, the best choice for the individual is the alternative that gives the best result on the average in similar patients. Decision analysis can help the clinician to identify the therapy that will give the best results when averaged over many similar patients. The decision analysis is tailored to a specific patient in that both the utility functions and the probability estimates are adjusted to the individual. Nonetheless, the results of the analysis represent the outcomes that would occur on average in a population of patients who have similar utilities and for whom uncertain events have similar probabilities.

3.5.4 Representation of Patients' Preferences with Utilities

In ▶Sect. 3.5.3, we introduced the concept of QALYs, because length of life is not the only outcome about which patients care. Patients' preferences for a health outcome may depend on the length of life with the outcome, on the quality of life with the outcome, and on the risk involved in achieving the outcome (e.g., a cure for cancer might require a risky surgical operation). How can we incorporate these elements into a decision analysis? To do so, we can represent patients' preferences with utilities. The utility of a health state is a quantitative measure of the desirability of a health state from the patient's perspective. Utilities are typically expressed on a 0 to 1 scale, where 0 represents death and 1 represents ideal health. For example, a study of patients who had chest pain (angina) with exercise rated the utility of mild, moderate, and severe angina as 0.95, 0.92, and 0.82 (Nease et al. 1995), respectively. There are several methods for assessing utilities.

The **standard-gamble** technique has the strongest theoretical basis of the various approaches to utility assessment, as shown by Von Neumann and Morgenstern and described by Sox et al. (1988). To illustrate use of the standard gamble, suppose we seek to assess a person's utility for the health state of asymptomatic HIV infection. To use the standard gamble, we ask our subject to compare the desirability of asymptomatic HIV infection to those of two other health states whose utility we know or can assign. Often, we use ideal health (assigned a utility of 1) and immediate death (assigned a utility of 0) for the comparison of health states. We then ask our subject to choose between asymptomatic HIV infection and a gamble with a chance of ideal health or immediate death. We vary the probability of ideal health and immediate death systematically until the subject is indifferent between asymptomatic HIV infection and the gamble. For example, a subject might be indifferent when the probability of ideal health is 0.8 and the probability of death is 0.2. At this point of indifference, the utility of the gamble and that of asymptomatic HIV infection are equal. We calculate the utility of the gamble as the weighted average of the utilities of each outcome of the gamble $[(1 \times 0.8) + (0 \times 0.2)] = 0.8$. Thus in this example, the utility of asymptomatic HIV infection is 0.8. Use of the standard gamble enables an analyst to assess the utility of outcomes that differ in length or quality of life. Because the standard gamble involves chance events, it

also assesses a person's willingness to take risks—called the person's **risk attitude**.

A second common approach to utility assessment is the **time-trade-off** technique (Sox et al. 1988; Torrance and Feeny 1989). To assess the utility of asymptomatic HIV infection using the time-trade-off technique, we ask a person to determine the length of time in a better state of health (usually ideal health or best attainable health) that he or she would find equivalent to a longer period of time with asymptomatic HIV infection. For example, if our subject says that 8 months of life with ideal health was equivalent to 12 months of life with asymptomatic HIV infection, then we calculate the utility of asymptomatic HIV infection as $8 \div 12 = 0.67$. The time-trade-off technique provides a convenient method for valuing outcomes that accounts for gains (or losses) in both length and quality of life. Because the time trade-off does not include gambles, however, it does not assess a person's risk attitude. Perhaps the strongest assumption underlying the use of the time trade-off as a measure of utility is that people are risk neutral. A **risk-neutral** decision maker is indifferent between the expected value of a gamble and the gamble itself. For example, a risk-neutral decision maker would be indifferent between the choice of living 20 years (for certain) and that of taking a gamble with a 50% chance of living 40 years and a 50% chance of immediate death (which has an expected value of 20 years). In practice, of course, few people are risk-neutral. Nonetheless, the time-trade-off technique is used frequently to value health outcomes because it is relatively easy to understand.

Several other approaches are available to value health outcomes. To use the **visual analog scale**, a person simply rates the quality of life with a health outcome (e.g., asymptomatic HIV infection) on a scale from 0 to 100. Although the visual analog scale is easy to explain and use, it has no theoretical justification as a valid measure of utility. Ratings with the visual analog scale, however, correlate modestly well with utilities assessed by the standard gamble and time trade-off. For a demonstration of the use of standard gambles, time trade-offs, and the visual analog scale to assess utilities in patients with angina, see Nease et al. (1995); in patient living with HIV, see Joyce et al. (2009) and (2012). Other approaches to valuing health outcomes include the Quality of Well-Being Scale, the Health Utilities Index, and the EuroQoL (see Neumann et al. 2017, ch. 7). Each of these instruments assesses how people value health outcomes and therefore may be appropriate for use in decision analyses or cost-effectiveness analyses.

In summary, we can use utilities to represent how patients value complicated health outcomes that differ in length and quality of life and in riskiness. Computer-based tools with an interactive format have been developed for assessing utilities; they often include text and multimedia presentations that enhance patients' understanding of the assessment tasks and of the health outcomes (Sumner et al. 1991; Nease and Owens 1994; Lenert et al. 1995).

3.5.5 Performance of Sensitivity Analysis

Sensitivity analysis is a test of the robustness of the conclusions of an analysis over a wide range of assumptions about the probabilities and the values, or utilities. The probability of an outcome at a chance node may be the best estimate that is available, but there often is a wide range of reasonable probabilities that a clinician could use with nearly equal confidence. We use sensitivity analysis to answer this question: Do my conclusions regarding the preferred choice change when the probability and outcome estimates are assigned values that lie within a reasonable range?

The knee-replacement decision in ▶ Example 3.12 illustrates the power of sensitivity analysis. If the conclusions of the analysis (surgery is preferable to no surgery) remain the same despite a wide range of assumed values for the probabilities and outcome measures, the recommendation is trustworthy. ◘ Figures 3.10 and 3.11 show the expected survival in healthy years with surgery and without surgery under varying assumptions of the probability of operative death and the probability of attaining perfect mobility, respectively. Each point (value) on these lines represents one calculation of expected survival using the tree in

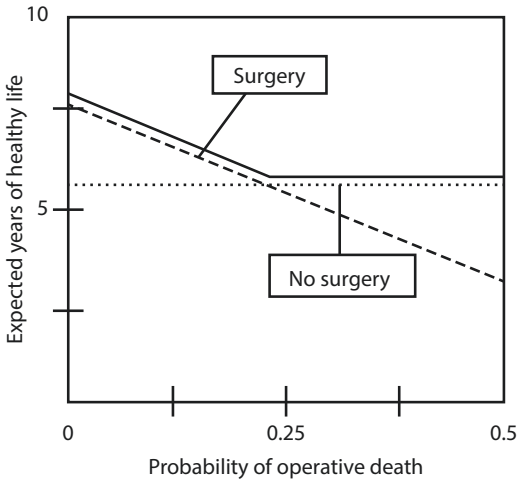


Fig. 3.10 Sensitivity analysis of the effect of operative mortality on length of healthy life (▶ Example 3.12). As the probability of operative death increases, the relative values of surgery versus no surgery change. The point at which the two lines cross represents the probability of operative death at which no surgery becomes preferable. The solid line represents the preferred option at a given probability

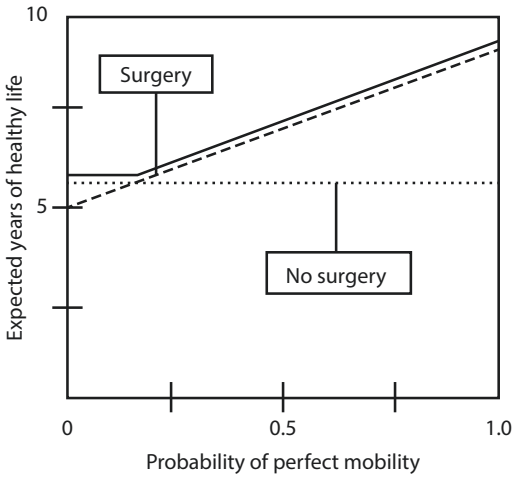


Fig. 3.11 Sensitivity analysis of the effect of a successful operative result on length of healthy life (▶ Example 3.12). As the probability of a successful surgical result increases, the relative values of surgery versus no surgery change. The point at which the two lines cross represents the probability of a successful result at which surgery becomes preferable. The *solid line* represents the preferred option at a given probability

Fig. 3.8. Figure 3.10 shows that expected survival is higher with surgery over a wide range of operative mortality rates. Expected survival is lower with surgery, however, when the operative mortality rate exceeds 25%. Figure 3.11 shows the effect of varying the probability that the operation will lead to perfect mobility. The expected survival, in healthy years, is higher for surgery as long as the probability of perfect mobility exceeds 20%, a much lower figure than is expected from previous experience with the operation. (In ▶ Example 3.12, the consulting orthopedic surgeons estimated the chance of full recovery at 60%). Thus, the internist can proceed with confidence to recommend surgery. Mr. Danby cannot be sure of a good outcome, but he has valid reasons for thinking that he is more likely to do well with surgery than he is without it.

Another way to state the conclusions of a sensitivity analysis is to indicate the range of probabilities over which the conclusions apply. The point at which the two lines in Fig. 3.10 cross is the probability of operative death at which the two therapy options have the same expected survival. If expected survival is to be the basis for choosing therapy, the internist and the patient should be indifferent between surgery and no surgery when the probability of operative death is 25%.[14] When the probability is lower, they should select surgery. When it is higher, they should select no surgery.

The approach to sensitivity analyses we have described enables the analyst to understand how uncertainty in one, two, or three parameters affects the conclusions of an analysis. But in a complex problem, a decision tree or decision model may have a 100 or more parameters. The analyst may have uncertainty about many parameters in a model. **Probabilistic sensitivity analysis** is an approach for understanding how the uncertainty in all (or a large number of) model parameters affects the conclusion of a decision analysis. To perform a probabilistic sensitivity analysis, the analyst must specify a probability distribution for each model parameter. The analytic software then chooses a value for each model parameter randomly from the

14 An operative mortality rate of 25% may seem high; however, this value is correct when we use QALYs as the basis for choosing treatment. A decision maker performing a more sophisticated analysis could use a utility function that reflects the patient's aversion to risking death.

parameter's probability distribution. The software then uses this set of parameter values and calculates the outcomes for each alternative. For each evaluation of the model, the software will determine which alternative is preferred. The process is usually repeated 10,000–100,000 times. From the probabilistic sensitivity analysis, the analyst can determine the proportion of times an alternative is preferred, accounting for all uncertainty in model parameters simultaneously. For more information on this advanced topic, see the article by Briggs and colleagues referenced at the end of the chapter.

3.5.6 Representation of Long-Term Outcomes with Markov Models

In ▶Example 3.12, we evaluated Mr. Danby's decision to have surgery to improve his mobility, which was compromised by arthritis. We assumed that each of the possible outcomes (full mobility, poor mobility, death, etc.) would occur shortly after Mr. Danby took action on his decision. But what if we want to model events that might occur in the distant future? For example, a patient with HIV infection might develop AIDS 10–15 years after infection; thus, a therapy to prevent or delay the development of AIDS could affect events that occur 10–15 years, or more, in the future. A similar problem arises in analyses of decisions regarding many chronic diseases: we must model events that occur over the lifetime of the patient. The decision tree representation is convenient for decisions for which all outcomes occur during a short time horizon, but it is not always sufficient for problems that include events that could occur in the future. How can we include such events in a decision analysis? The answer is to use Markov models (Beck and Pauker 1983; Sonnenberg and Beck 1993; Siebert et al. 2012).

To build a **Markov model**, we first specify the set of health states that a person could experience (e.g., Well, Cancer, and Death in Fig. 3.12). We then specify the **transition probabilities**, which are the probabilities that a person will transit from one of these health states to another during a specified time period. This period—often 1 month or 1 year—is the length of the **Markov cycle**. The Markov model then simulates the transitions among health states for a person (or for a hypothetical cohort of people) for a specified number of cycles; by using a Markov model, we can calculate the probability that a person will be in each of the health states at any time in the future. As an illustration, consider a simple Markov model that has three health states: Well, Cancer, and Death (see Fig. 3.12). We have specified each of the transition probabilities in Table 3.7 for the cycle length of 1 year. Thus, we note from Table 3.7 that a person who is in the well state will remain well with probability 0.9, will develop cancer with probability 0.06, and will die from non-cancer causes with probability 0.04 during 1 year. The calculations for a Markov model are performed by computer software. Based on the transition probabilities in Table 3.7, the probabilities that a person remains well, develops cancer, or dies from non-cancer causes over time is shown in Table 3.8. We can also determine from a Markov model the expected length of time that a person spends in each health state. Therefore, we can determine life expectancy, or quality-adjusted life expectancy, for any alternative represented by a Markov model.

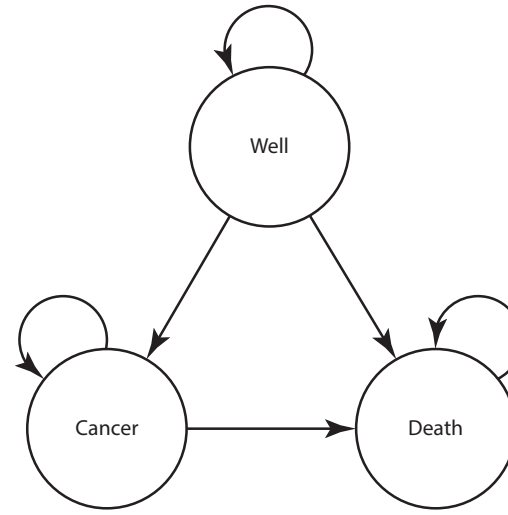


Fig. 3.12 A simple Markov model. The states of health that a person can experience are indicated by the *circles*; *arrows* represent allowed transitions between health states

In decision analyses that represent long-term outcomes, the analysts will often use a

Markov model in conjunction with a decision tree to model the decision (Owens et al. 1995; Salpeter et al. 1997; Sanders et al. 2005; Lin et al. 2018). The analyst models the effect of an intervention as a change in the probability of going from one state to another. For example, we could model a cancer-prevention intervention (such as screening for breast cancer with mammography) as a reduction in the transition probability from Well to Cancer in Fig. 3.12. (See the articles by Beck and Pauker (1983) and Sonnenberg and Beck (1993) for further explanation of the use of Markov models).

3.6 The Decision Whether to Treat, Test, or Do Nothing

The clinician who is evaluating a patient's symptoms and suspects a disease must choose among the following actions:

1. Do nothing further (neither perform additional tests nor treat the patient).
2. Obtain additional diagnostic information (test) before choosing whether to treat or do nothing.
3. Treat without obtaining more information.

When the clinician knows the patient's true state, testing is unnecessary, and the doctor needs only to assess the trade-offs among therapeutic options (as in ▶ Example 3.12). Learning the patient's true state, however, may require costly, time-consuming, and often risky diagnostic procedures that may give misleading FP or FN results. Therefore, clinicians often are willing to treat a patient even when they are not absolutely certain about a patient's true state. There are risks in this course: the clinician may withhold therapy from a person who has the disease of concern, or he may administer therapy to someone who does not have the disease yet may suffer undesirable side effects of therapy.

Deciding among treating, testing, and doing nothing sounds difficult, but you have already learned all the principles that you need to solve this kind of problem. There are three steps:

1. Determine the treatment threshold probability of disease.
2. Determine the pretest probability of disease.
3. Decide whether a test result could affect your decision to treat.

The **treatment threshold probability** of disease is the probability of disease at which you should be indifferent between treating and not treating (Pauker and Kassirer 1980). Below the treatment threshold, you should not treat. Above the treatment threshold, you should

Table 3.7 Transition probabilities for the Markov model in Fig. 3.13

Health state transition	Annual probability
Well to well	0.9
Well to cancer	0.06
Well to death	0.04
Cancer to well	0.0
Cancer to cancer	0.4
Cancer to death	0.6
Death to well	0.0
Death to cancer	0.0
Death to death	1.0

Table 3.8 Probability of future health states for the Markov model in Fig. 3.12

Health state	Probability of health state at end of year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
Well	0.9000	0.8100	0.7290	0.6561	0.5905	0.5314	0.4783
Cancer	0.0600	0.0780	0.0798	0.0757	0.0696	0.0633	0.0572
Death	0.0400	0.1120	0.1912	0.2682	0.3399	0.4053	0.4645

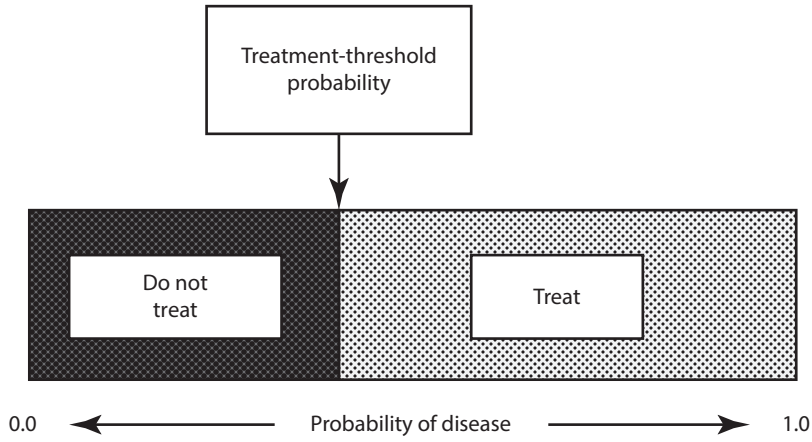


Fig. 3.13 Depiction of the treatment threshold probability. At probabilities of disease that are less than the treatment threshold probability, the preferred action is to withhold therapy. At probabilities of disease that are greater than the treatment threshold probability, the preferred action is to treat

treat (Fig. 3.13). Whether to treat when the diagnosis is not certain is a problem that you can solve with a decision tree, such as the one shown in Fig. 3.14.

You can use this tree to learn the treatment threshold probability of disease by leaving the probability of disease as an unknown, setting the expected value of surgery equal to the expected value for medical (i.e., nonsurgical, such as drugs or physical therapy) treatment, and solving for the probability of disease. (In this example, surgery corresponds to the "treat" branch of the tree in Fig. 3.14, and nonsurgical intervention corresponds to the "do not treat" branch). Because you are indifferent between medical treatment and surgery at this probability, it is the treatment threshold probability. Using the tree completes step 1. In practice, people often determine the treatment threshold intuitively rather than analytically.

An alternative approach to determination of the treatment threshold probability is to use the equation:

$$p^* = \frac{H}{H + B},$$

where p^* = the treatment threshold probability, H = the harm associated with treatment of a nondiseased patient, and B = the benefit associated with treatment of a diseased patient (Pauker and Kassirer 1980; Sox et al. 1988). We define B as the difference between the utility (U) of diseased patients who are treated and diseased patients who are not treated (U[D, treat] − U[D, do not treat], as shown in Fig. 3.14). The utility of diseased patients who are treated should be greater than that of diseased patients who are not treated; therefore, B is

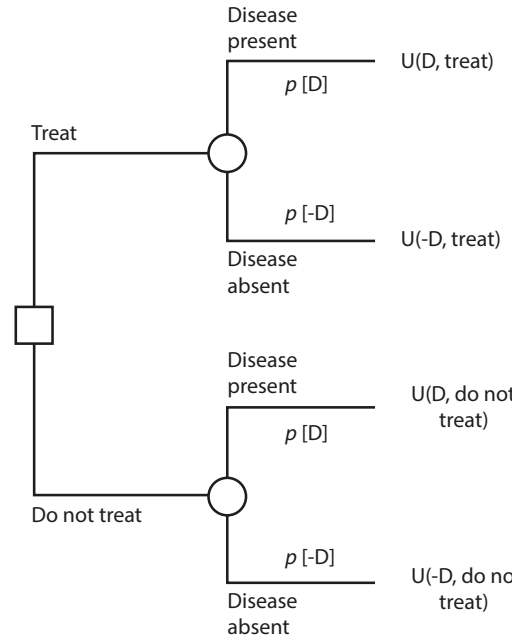


Fig. 3.14 Decision tree with which to calculate the treatment threshold probability of disease. By setting the utilities of the treat and do not treat choices to be equal, we can compute the probability at which the clinician and patient should be indifferent to the choice. Recall that p [−D] = 1 − p [D]

positive. We define H as the difference in utility of nondiseased patients who are not treated and nondiseased patients who are treated (U[−D, do not treat] − U[−D, treat], as shown in Fig. 3.14). The utility of nondiseased patients who are not treated should be greater than that of nondiseased patients who are treated; therefore, H is positive. The equation for the treatment threshold probability fits with our intuition: if the benefit of treatment is small and the harm of treatment is large, the treatment threshold probability will be high. In contrast, if the benefit of treatment is large and the harm of treatment is small, the treatment threshold probability will be low.

Once you know the pretest probability, you know what to do in the absence of further information about the patient. If the pretest probability is below the treatment threshold, you should not treat the patient. If the pretest probability is above the threshold, you should treat the patient. Thus, you have completed step 2.

One of the guiding principles of medical decision making is this: do not order a test unless it could change your management of the patient. In our framework for decision making, this principle means that you should order a test only if the test result could cause the probability of disease to cross the treatment threshold or lead to another test that would do so. Thus, if the pretest probability is above the treatment threshold, a negative test result must lead to a post-test probability that is below the threshold. Conversely, if the pretest probability is below the threshold probability, a positive result must lead to a post-test probability that is above the threshold. In either case, the test result would alter your decision of whether to treat the patient. This analysis completes step 3.

To decide whether a test could alter management, we simply use Bayes' theorem. We calculate the post-test probability after a test result that would move the probability of disease toward the treatment threshold. If the pretest probability is above the treatment threshold, we calculate the probability of disease if the test result is negative. If the pretest probability is below the treatment threshold, we calculate the probability of disease if the test result is positive.

► Example 3.13

You are a pulmonary medicine specialist. You suspect that a patient of yours has a pulmonary embolus (blood clot lodged in the vessels of the lungs). One approach is to do a computed tomography angiography (CTA) scan, a test in which a computed tomography (CT) of the lung is done after a radiopaque dye is injected into a vein. The dye flows into the vessels of the lung. The CT scan can then assess whether the blood vessels are blocked. If the scan is negative, you do no further tests and do not treat the patient. ◄

To decide whether this strategy is correct, you take the following steps:

1. Determine the treatment threshold probability of pulmonary embolus.
2. Estimate the pretest probability of pulmonary embolus.
3. Decide whether a test result could affect your decision to treat for an embolus.

First, assume you decide that the treatment threshold should be 0.10 in this patient. What does it mean to have a treatment threshold probability equal to 0.10? If you could obtain no further information, you would treat for pulmonary embolus if the pretest probability was above 0.10 (i.e., if you believed that there was greater than a 1 in 10 chance that the patient had an embolus), and would withhold therapy if the pretest probability was below 0.10. A decision to treat when the pretest probability is at the treatment threshold means that you are willing to treat nine patients without pulmonary embolus to be sure of treating one patient who has pulmonary embolus. A relatively low treatment threshold is justifiable because treatment of a pulmonary embolism with blood-thinning medication substantially reduces the high mortality of pulmonary embolism, whereas there is only a relatively small danger (mortality of less than 1%) in treating someone who does not have pulmonary embolus. Because the benefit of treatment is high and the harm of treatment is low, the treatment threshold probability will be low, as discussed earlier. You have completed step 1.

You estimate the pretest probability of pulmonary embolus to be 0.05, which is equal to a pretest odds of 0.053. Because the pretest probability is lower than the treatment thresh-

old, you should do nothing unless a positive CTA scan result could raise the probability of pulmonary embolus to above 0.10. You have completed step 2.

To decide whether a test result could affect your decision to treat, you must decide whether a positive CTA scan result would raise the probability of pulmonary embolism to more than 0.10, the treatment threshold. You review the literature and learn that the LR for a positive CTA scan is approximately 21 (Stein et al. 2006).

A negative CTA scan result will move the probability of disease away from the treatment threshold and will be of no help in deciding what to do. A positive result will move the probability of disease toward the treatment threshold and could alter your management decision if the post-test probability were above the treatment threshold. You therefore use the odds-ratio form of Bayes' theorem to calculate the post-test probability of disease if the lung scan result is reported as high probability.

$$\text{post-test odds} = \text{pretest odds} \times LR$$
$$= 0.053 \times 21 = 1.11.$$

A post-test odds of 1.1 is equivalent to a probability of disease of 0.53. Because the post-test probability of pulmonary embolus is higher than the treatment threshold, a positive CTA scan result would change your management of the patient, and you should order the lung scan. You have completed step 3.

This example is especially useful for two reasons: first, it demonstrates one method for making decisions and second, it shows how the concepts that were introduced in this chapter all fit together in a clinical example of medical decision making.

3.7 Alternative Graphical Representations for Decision Models: Influence Diagrams and Belief Networks

In ▶Sects. 3.5 and 3.6, we used decision trees to represent decision problems. Although decision trees are the most common graphical representation for decision problems, **influence diagrams** are an important alternative representation for such problems (Nease and Owens 1997; Owens et al. 1997).

As shown in Fig. 3.15, influence diagrams have certain features that are similar to decision trees, but they also have additional graphical elements. Influence diagrams represent decision nodes as squares and chance nodes as circles. In contrast to decision trees, however, the influence diagram also has arcs

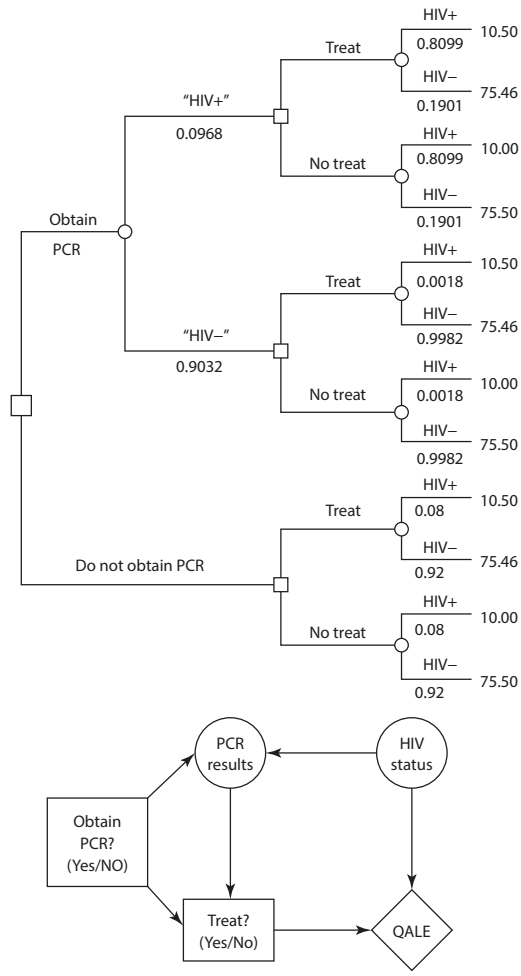
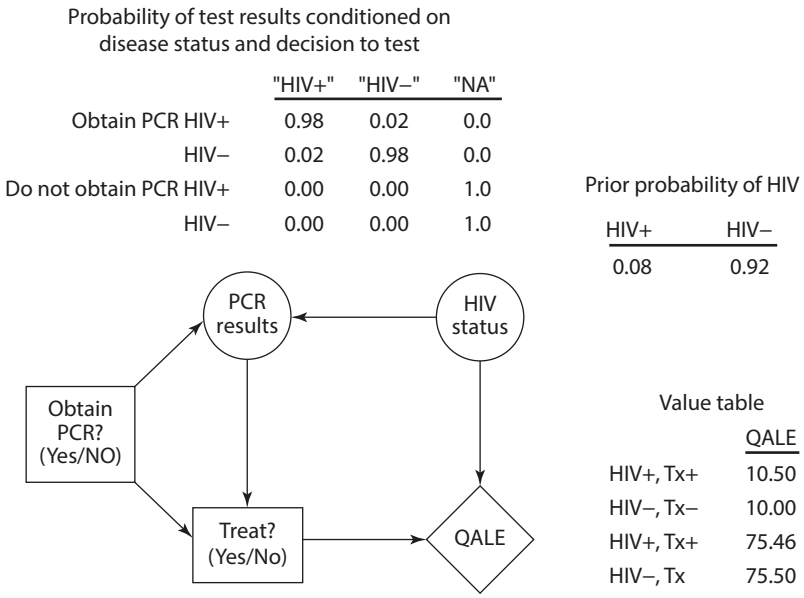


Fig. 3.15 A decision tree (*top*) and an influence diagram (*bottom*) that represent the decisions to test for, and to treat, HIV infection. The structural asymmetry of the alternatives is explicit in the decision tree. The influence diagram highlights probabilistic relationships. *HIV* human immunodeficiency virus, *HIV+* HIV infected, *HIV–* not infected with HIV, *QALE* quality-adjusted life expectancy, *PCR* polymerase chain reaction. Test results are shown in quotation marks ("HIV+"), whereas the true disease state is shown without quotation marks (HIV+). (Source: Owens et al. (1997). Reproduced with permission)

between nodes and a diamond-shaped value node. An arc between two chance nodes indicates that a probabilistic relationship may exist between the chance nodes (Owens et al. 1997). A **probabilistic relationship** exists when the occurrence of one chance event affects the probability of the occurrence of another chance event. For example, in ◘ Fig. 3.15, the probability of a positive or negative PCR test result (PCR result) depends on whether a person has HIV infection (HIV status); thus, these nodes have a probabilistic relationship, as indicated by the arc. The arc points from the **conditioning event** to the **conditioned event** (PCR test result is conditioned on HIV status in ◘ Fig. 3.15). The absence of an arc between two chance nodes, however, always indicates that the nodes are independent or conditionally independent. Two events are conditionally independent, given a third event, if the occurrence of one of the events does not affect the probability of the other event conditioned on the occurrence of the third event.

Unlike a decision tree, in which the events usually are represented from left to right in the order in which the events are observed, influence diagrams use arcs to indicate the timing of events. An arc from a chance node to a decision node indicates that the chance event has been observed at the time the decision is made. Thus, the arc from PCR result to Treat? in ◘ Fig. 3.15 indicates that the decision maker knows the PCR test result (positive, negative, or not obtained) when he or she decides whether to treat. Arcs between decision nodes indicate the timing of decisions: the arc points from an initial decision to subsequent decisions. Thus, in ◘ Fig. 3.15, the decision maker must decide whether to obtain a PCR test before deciding whether to treat, as indicated by the arc from Obtain PCR? to Treat?

The probabilities and utilities that we need to determine the alternative with the highest expected value are contained in tables associated with chance nodes and the value node (◘ Fig. 3.16). These tables contain the same information that we would use in a decision tree. With a decision tree, we can determine the expected value of each alternative by averaging out at chance nodes and folding back the tree (▶ Sect. 3.5.3).



◘ **Fig. 3.16** The influence diagram from ◘ Fig. 3.15, with the probability and value tables associated with the nodes. The information in these tables is the same as that associated with the branches and endpoints of the decision tree in ◘ Fig. 3.15. HIV human immunodeficiency virus, HIV+ HIV infected, HIV– not infected with HIV, QALE quality-adjusted life expectancy, PCR polymerase chain reaction, NA not applicable, TX+ treated, TX– not treated. Test results are shown in quotation marks ("HIV+"), and the true disease state is shown without quotation marks (HIV+). (Source: Owens et al. (1997). Reproduced with permission)

For influence diagrams, the calculation of expected value is more complex (Owens et al. 1997), and generally must be performed with computer software. With the appropriate software, we can use influence diagrams to perform the same analyses that we would perform with a decision tree. Diagrams that have only chance nodes are called **belief networks**; we use them to perform probabilistic inference.

Why use an influence diagram instead of a decision tree? Influence diagrams have both advantages and limitations relative to decision trees. Influence diagrams represent graphically the probabilistic relationships among variables (Owens et al. 1997). Such representation is advantageous for problems in which probabilistic conditioning is complex or in which communication of such conditioning is important (such as may occur in large models). In an influence diagram, probabilistic conditioning is indicated by the arcs, and thus the conditioning is apparent immediately by inspection. In a decision tree, probabilistic conditioning is revealed by the probabilities in the branches of the tree. To determine whether events are conditionally independent in a decision tree requires that the analyst compare probabilities of events between branches of the tree. Influence diagrams also are particularly useful for discussion with content experts who can help to structure a problem but who are not familiar with decision analysis. In contrast, problems that have decision alternatives that are structurally different may be easier for people to understand when represented with a decision tree, because the tree shows the structural differences explicitly, whereas the influence diagram does not. The choice of whether to use a decision tree or an influence diagram depends on the problem being analyzed, the experience of the analyst, the availability of software, and the purpose of the analysis. For selected problems, influence diagrams provide a powerful graphical alternative to decision trees.

3.8 Other Modeling Approaches

We have described decision trees, Markov models and influence diagrams. An analyst also can choose several other approaches to modeling. The choice of modeling approach depends on the problem and the objectives of the analysis. Although how to choose and design such models is beyond our scope, we note other type of models that analysts use commonly for medical decision making. **Microsimulation models** are individual-level health state transition models, similar to Markov models, that provide a means to model very complex events flexibly over time. They are useful when the clinical history of a problem is complex, such as might occur with cancer, heart disease, and other chronic diseases. They are also useful for modeling individual heterogeneity which may depend on combinations of individual characteristics (e.g. heterogeneity of response to treatment based on medical conditions or genetics). **Dynamic transmission models** are particularly well-suited for assessing the outcomes of infectious diseases. These models divide a population into compartments (for example, uninfected, infected, recovered, dead), and transitions between compartments are governed by differential or difference equations. The rate of transition between compartments depends in part on the number of individuals in the compartment, an important feature for infectious diseases in which the transmission may depend on the number of infected or susceptible individuals. **Discrete event simulation models** also are often used to model interactions between people. These models are composed of entities (a patient) that have attributes (clinical history), and that experience events (a heart attack). An entity can interact with other entities and use resources. Discrete event simulation models are also used when considering scarce resources such as queues for a diagnostic test or an operating room slot. For more information on these types of models, we suggest a recent series of papers on best modeling practices; the paper by Caro and colleagues noted in the suggested readings at the end of the chapter is an overview of this series of papers.

3.9 The Role of Probability and Decision Analysis in Medicine

You may be wondering how probability and decision analysis might be integrated smoothly into medical practice. An understanding of

probability and measures of test performance will prevent any number of misadventures. In ▶ Example 3.1, we discussed a hypothetical test that, on casual inspection, appeared to be an accurate way to screen blood donors for previous exposure to the AIDS virus. Our quantitative analysis, however, revealed that the hypothetical test results were misleading more often than they were helpful because of the low prevalence of HIV in the clinically relevant population. Fortunately, in actual practice, much more accurate tests are used to screen for HIV.

The need for knowledgeable interpretation of test results is widespread. The federal government screens civil employees in "sensitive" positions for drug use, as do many companies. If the drug test used by an employer had a sensitivity and specificity of 0.95, and if 10% of the employees used drugs, one-third of the positive tests would be FPs. An understanding of these issues should be of great interest to the public, and health professionals should be prepared to answer the questions of their patients.

Although we should try to interpret every kind of test result accurately, decision analysis has a more selective role in medicine. Not all clinical decisions require decision analysis. Some decisions depend on physiologic principles or on deductive reasoning. Other decisions involve little uncertainty. Nonetheless, many decisions must be based on imperfect data, and they will have outcomes that cannot be known with certainty at the time that the decision is made. Decision analysis provides a technique for managing these situations.

For many problems, simply drawing a tree that denotes the possible outcomes explicitly will clarify the question sufficiently to allow you to make a decision. When time is limited, even a "quick and dirty" analysis may be helpful. By using expert clinicians' subjective probability estimates and asking what the patient's utilities might be, you can perform an analysis quickly and learn which probabilities and utilities are the important determinants of the decision.

Health care professionals sometimes express reservations about decision analysis because the analysis may depend on probabilities that must be estimated, such as the pretest probability. A thoughtful decision maker will be concerned that the estimate may be in error, particularly because the information needed to make the estimate often is difficult to obtain from the medical literature. We argue, however, that uncertainty in the clinical data is a problem for any decision-making method and that the effect of this uncertainty is explicit with decision analysis. The method for evaluating uncertainty is sensitivity analysis: we can examine any variable to see whether its value is critical to the final recommended decision. Thus, we can determine, for example, whether a change in pretest probability from 0.6 to 0.8 makes a difference in the final decision. In so doing, we often discover that it is necessary to estimate only a range of probabilities for a particular variable rather than a precise value. Thus, with a sensitivity analysis, we can decide whether uncertainty about a particular variable should concern us.

The growing complexity of medical decisions, coupled with the need to control costs, has led to major programs to develop clinical practice guidelines. Decision models have many advantages as aids to guideline development (Eddy 1992; Habbema et al. 2014; Owens et al. 2016): they make explicit the alternative interventions, associated uncertainties, and utilities of potential outcomes. Decision models can help guideline developers to structure guideline-development problems (Owens and Nease 1993), to incorporate patients' preferences (Nease and Owens 1994; Owens 1998), and to tailor guidelines for specific clinical populations (Owens and Nease 1997). The U.S. Preventive Services Task Force, which develops national prevention guidelines, has used decision models in the development of guidelines on breast, lung, cervical, and colorectal cancer screening. In addition, Web-based interfaces for decision models can provide distributed decision support for guideline developers and users by making the decision model available for analysis to anyone who has access to the Web (Sanders et al. 1999).

We have not emphasized computers in this chapter, although they can simplify many aspects of decision analysis (see ▶ Chap. 24). MEDLINE and other bibliographic retrieval systems (see ▶ Chap. 23) make it easier to ob-

tain published estimates of disease prevalence and test performance. Computer programs for performing statistical analyses can be used on data collected by hospital information systems. Decision analysis software, available for personal computers, can help clinicians to structure decision trees, to calculate expected values, and to perform sensitivity analyses. Researchers continue to explore methods for computer-based automated development of practice guidelines from decision models and use of computer-based systems to implement guidelines (Musen et al. 1996). With the growing maturity of this field, there are now companies that offer formal analytical tools to assist with clinical outcome assessment and interpretation of population datasets.

Medical decision making often involves uncertainty for the clinician and risk for the patient. Most health care professionals would welcome tools that help them make decisions when they are confronted with complex clinical problems with uncertain outcomes. There are important medical problems for which decision analysis offers such aid.

3.10 Appendix A: Derivation of Bayes' Theorem

Bayes' theorem is derived as follows. We denote the conditional probability of disease, D, given a test result, R, p[D|R]. The prior (pretest) probability of D is p[D]. The definition of conditional probability is:

$$p[\mathrm{D}|\mathrm{R}] = \frac{p[\mathrm{R},\mathrm{D}]}{p[\mathrm{R}]} \tag{3.1}$$

The probability of a test result (p[R]) is the sum of its probability in diseased patients and its probability in nondiseased patients:

$$p[\mathrm{R}] = \mathrm{p}\,[\mathrm{R},\mathrm{D}] + p[\mathrm{R},-\mathrm{D}].$$

Substituting into Eq. 3.1, we obtain:

$$p[\mathrm{D}|\mathrm{R}] = \frac{p[\mathrm{R},\mathrm{D}]}{p[\mathrm{R},\mathrm{D}] + p[\mathrm{R},-\mathrm{D}]} \tag{3.2}$$

Again, from the definition of conditional probability,

$$p[R|\mathrm{D}] = \frac{p[\mathrm{R},\mathrm{D}]}{p[\mathrm{D}]} \text{ and } p[\mathrm{R}|-\mathrm{D}] = \frac{p[\mathrm{R},-\mathrm{D}]}{p[-\mathrm{D}]}$$

These expressions can be rearranged:

$$p[\mathrm{R},\mathrm{D}] = p[\mathrm{D}] \times p[\mathrm{R}|\ \mathrm{D}], \tag{3.3}$$

$$p[\mathrm{R},-\mathrm{D}] = p[-\mathrm{D}] \times p[\mathrm{R}\ |-\mathrm{D}]. \tag{3.4}$$

Substituting Eqs. 3.3 and 3.4 into Eq. 3.2, we obtain Bayes' theorem:

$$p[\mathrm{D}|\mathrm{R}] = \frac{p[\mathrm{D}] \times p[\mathrm{R}|\mathrm{D}]}{p[\mathrm{D}] \times p[\mathrm{R}|\mathrm{D}] + \mathrm{p}[-\mathrm{D}] \times p[\mathrm{R}|-\mathrm{D}]}$$

Suggested Readings

Briggs, A., Weinstein, M., Fenwick, E., Karnon, J., Sculpher, M., & Paltiel, A. (2012). Model parameter estimation and uncertainty analysis: A report of the ISPOR-SMDM modeling good research practices task force-6. *Medical Decision Making, 32*(5), 722–732. This article describes best practices for estimating model parameters and for performing sensitivity analyses, including probabilistic sensitivity analysis.

Caro, J., Briggs, A., Siebert, U., & Kuntz, K. (2012). Modeling good research practices – overview: A report of the ISPOR-SMDM modeling good research practices task force-1. *Value in Health, 15*, 796–803. This paper is an introduction to a series of papers that describe best modeling practices.

Hunink, M., Glasziou, P., Siegel, J., Weeks, J., Pliskin, J., Einstein, A., & Weinstein, M. (2001). *Decision making in health and medicine*. Cambridge: Cambridge University Press. This textbook addresses in detail most of the topics introduced in this chapter.

Nease, R. F., Jr., & Owens, D. K. (1997b). Use of influence diagrams to structure medical decisions. *Medical Decision Making, 17*(13), 263–275. This article provides a comprehensive introduction to the use of influence diagrams.

Neumann, P. J., Sanders, G. D., Russell, L. B., Siegel, J. E., & Ganiats, T. G. (Eds.). (2017b). *Cost-effectiveness in health and medicine* (2nd

ed.). New York: Oxford University Press. This book provides authoritative guidelines for the conduct of cost-effectiveness analyses. Chapter 7 discusses approaches for valuing health outcomes.

Owens, D. K., Schacter, R. D., & Nease, R. F., Jr. (1997b). Representation and analysis of medical decision problems with influence diagrams. *Medical Decision Making, 17*(3), 241–262. This article provides a comprehensive introduction to the use of influence diagrams.

Raiffa, H. (1970). *Decision analysis: Introductory lectures on choices under uncertainty*. Reading: Addison-Wesley. This now classic book provides an advanced, nonmedical introduction to decision analysis, utility theory, and decision trees.

Sox, H. C. (1986). Probability theory in the use of diagnostic tests. *Annals of Internal Medicine, 104*(1), 60–66. This article is written for clinicians; it contains a summary of the concepts of probability and test interpretation.

Sox, H. C., Higgins, M. C., & Owens, D. K. (2013). *Medical decision making*. Chichester: Wiley-Blackwell. This introductory textbook covers the subject matter of this chapter in greater detail, as well as discussing many other topics.

Tversky, A., & Kahneman, D. (1974b). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124. This now classic article provides a clear and interesting discussion of the experimental evidence for the use and misuse of heuristics in situations of uncertainty.

Questions for Discussion

1. Calculate the following probabilities for a patient about to undergo CABG surgery (see ▶ Example 3.2):
   (a) The only possible, mutually exclusive outcomes of surgery are death, relief of symptoms (angina and dyspnea), and continuation of symptoms. The probability of death is 0.02, and the probability of relief of symptoms is 0.80. What is the probability that the patient will continue to have symptoms?
   (b) Two known complications of heart surgery are stroke and heart attack, with probabilities of 0.02 and 0.05, respectively. The patient asks what chance he or she has of having both complications. Assume that the complications are conditionally independent, and calculate your answer.
   (c) The patient wants to know the probability that he or she will have a stroke given that he or she has a heart attack as a complication of the surgery. Assume that 1 in 500 patients has both complications, that the probability of heart attack is 0.05, and that the events are independent. Calculate your answer.
2. The results of a hypothetical study to measure test performance of a diagnostic test for HIV are shown in the 2 × 2 table in ◘ Table 3.9.
   (a) Calculate the sensitivity, specificity, disease prevalence, PV+, and PV–.
   (b) Use the TPR and TNR calculated in part (a) to fill in the 2 × 2 table in ◘ Table 3.10. Calculate the disease prevalence, PV+, and PV–.

◘ **Table 3.9** A 2 × 2 contingency table for the hypothetical study in problem 2

PCR test result	Gold standard test positive	Goldstandard test negative	Total
Positive PCR	48	8	56
Negative PCR	2	47	49
Total	50	55	105

PCR polymerase chain reaction

Table 3.10 A 2 × 2 contingency table to complete for problem 2b

PCR test result	Gold standard test positive	Gold standard test negative	Total
Positive PCR	x	x	x
Negative PCR	100	99,900	x
Total	x	x	x

PCR polymerase chain reaction
x quantities that the question ask students to calculate

3. You are asked to interpret the results from a diagnostic test for HIV in an asymptomatic patient whose test was positive when the patient volunteered to donate blood. After taking the patient's history, you learn that the patient has a history of intravenous-drug use. You know that the overall prevalence of HIV infection in your community is 1 in 500 and that the prevalence in people who have injected drugs is 20 times as high as in the community at large.
   (a) Estimate the pretest probability that this patient is infected with HIV.
   (b) The patient tells you that two people with whom the patient shared needles subsequently died of AIDS. Which heuristic will be useful in making a subjective adjustment to the pretest probability in part (a)?
   (c) Use the sensitivity and specificity that you worked out in 2(a) to calculate the post-test probability of the patient having HIV after a positive and negative test. Assume that the pretest probability is 0.10.
   (d) If you wanted to increase the post-test probability of disease given a positive test result, would you change the TPR or TNR of the test?
4. You have a patient with cancer who has a choice between surgery or chemotherapy. If the patient chooses surgery, he or she has a 2% chance of dying from the operation (life expectancy = 0), a 50% chance of being cured (life expectancy = 15 years), and a 48% chance of not being cured (life expectancy = 1 year). If the patient chooses chemotherapy, he or she has a 5% chance of death (life expectancy = 0), a 65% chance of cure (life expectancy = 15 years), and a 30% chance that the cancer will be slowed but not cured (life expectancy = 2 years). Create a decision tree. Calculate the expected value of each option in terms of life expectancy.
5. You are concerned that a patient with a sore throat has a bacterial infection that would require antibiotic therapy (as opposed to a viral infection, for which no treatment is available). Your treatment threshold is 0.4, and based on the examination you estimate the probability of bacterial infection as 0.8. A test is available (TPR = 0.75, TNR = 0.85) that indicates the presence or absence of bacterial infection. Should you perform the test? Explain your reasoning. How would your analysis change if the test were extremely costly or involved a significant risk to the patient?
6. What are the three kinds of bias that can influence measurement of test performance? Explain what each one is, and state how you would adjust the post-test probability to compensate for each.
7. How could a computer system ease the task of performing a complex decision analysis?
8. When you search the medical literature to find probabilities for patients similar to one you are treating, what is the most important question to consider? How should you adjust probabilities in light of the answer to this question?

9. Why do you think clinicians sometimes order tests even if the results will not affect their management of the patient? Do you think the reasons that you identify are valid? Are they valid in only certain situations? Explain your answers. See the January 1998 issue of Medical Decision Making for articles that discuss this question.
10. Explain the differences in three approaches to assessing patients' preferences for health states: the standard gamble, the time trade-off, and the visual analog scale.

Disclaimer The views presented are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors, or its Methodology Committee.

References

Beck, J. R., & Pauker, S. G. (1983). The Markov process in medical prognosis. *Medical Decision Making, 3*(4), 419–458.

Eddy, D. M. (1992). *A manual for assessing health practices and designing practice policies: The explicit approach*. Philadelphia: American College of Physicians.

Habbema, J. D. F., Wilt, T. J., Etzioni, R., Nelson, H. D., Schechter, C. B., Lawrence, W. F., Melnikow, J., Kuntz, K. M., Owens, D. K., & Feuer, E. J. (2014). Models in the development of clinical practice guidelines. *Annals of Internal Medicine, 161*, 812–818.

Hellmich, M., Abrams, K. R., & Sutton, A. J. (1999). Bayesian approaches to meta-analysis of ROC curves. *Medical Decision Making, 19*, 252–264.

Joyce, V. R., Barnett, P. G., Bayoumi, A. M., Griffin, S. C., Kyriakides, T. C., Yu, W., Sundaram, V., Holodniy, M., Brown, S. T., Cameron, W., Youle, M., Sculpher, M., Anis, A. H., & Owens, D. K. (2009). Health-related quality of life in a randomized trial of antiretroviral therapy for advanced HIV disease. *Journal of the Acquired Immunodeficiency Syndrome, 50*, 27–36.

Joyce, V. R., Barnett, P. G., Chow, A., Bayoumi, A. M., Griffin, S. C., Sun, H., Holodniy, M., Brown, S. T., Cameron, D. W., Youle, M., Sculpher, M., Anis, A. H., & Owens, D. K. (2012). Effect of treatment interruption and intensification of antiretroviral therapy on health-related quality of life in patients with advanced HIV: A randomized controlled trial. *Medical Decision Making, 32*, 70–82.

Leeflang, M. (2014). Systematic reviews and meta-analysis of diagnostic test accuracy. *Clinical Microbiology and Infection, 20*, 105–113.

Leeflang, M. M. G., Deeks, J. J., Gatsonis, C., Bossuyt, P. M., & on behalf of the Cochrane diagnostic test accuracy working group. (2008). Systematic reviews of diagnostic tests. *Annals of Internal Medicine, 149*, 889–897.

Lenert, L. A., Michelson, D., Flowers, C., & Bergen, M. R. (1995). *IMPACT: An object-oriented graphical environment for construction of multimedia patient interviewing software* (pp. 319–323). Washington, D.C.: Proceedings of the Annual Symposium of Computer Applications in Medical Care.

Lin, J. K., Lerman, B. J., Barnes, J. I., Bourisquot, B. C., Tan, Y. J., Robinson, A. Q. L., Davis, K. L., Owens, D. K., & Goldhaber-Fiebert, J. D. (2018). Cost effectiveness of chimeric antigen receptor T-cell therapy in relapsed or refractory pediatric B-cell acute lymphoblastic leukemia. *Journal of Clinical Oncology, 36*, 3192–3202.

Meigs, J., Barry, M., Oesterling, J., & Jacobsen, S. (1996). Interpreting results of prostate-specific antigen testing for early detection of prostate cancer. *Journal of General Internal Medicine, 11*(9), 505–512.

Moses, L. E., Littenberg, B., & Shapiro, D. (1993). Combining independent studies of a diagnostic test into a summary ROC curve: Data-analytic approaches and some additional considerations. *Statistics in Medicine, 12*(4), 1293–1316.

Musen, M. A., Tu, S. W., Das, A. K., & Shahar, Y. (1996). EON: A component-based approach to automation of protocol-directed therapy. *Journal of the American Medical Informatics Association: JAMIA, 3*(6), 367–388.

Nease, R. F., Jr., & Owens, D. K. (1994). A method for estimating the cost-effectiveness of incorporating patient preferences into practice guidelines. *Medical Decision Making, 14*(4), 382–392.

Nease, R. F., Jr., & Owens, D. K. (1997a). Use of influence diagrams to structure medical decisions. *Medical Decision Making, 17*(13), 263–275.

Nease, R. F., Jr., Kneeland, T., O'Connor, G. T., Sumner, W., Lumpkins, C., Shaw, L., Pryor, D., & Sox, H. C. (1995). Variation in patient utilities for the outcomes of the management of chronic stable angina. Implications for clinical practice guidelines. *Journal of the American Medical Association, 273*(15), 1185–1190.

Neumann, P. J., Sanders, G. D., Russell, L. B., Siegel, J. E., & Ganiats, T. G. (Eds.). (2017a). *Cost-effectiveness in health and medicine* (2nd ed.). New York: Oxford University Press.

Owens, D. K. (1998). Patient preferences and the development of practice guidelines. *Spine, 23*(9), 1073–1079.

Owens, D. K., & Nease, R. F., Jr. (1993). Development of outcome-based practice guidelines: A method for structuring problems and synthesizing evidence. *The Joint Commission Journal on Quality Improvement, 19*(7), 248–263.

Owens, D. K., & Nease, R. F., Jr. (1997). A normative analytic framework for development of practice guidelines for specific clinical populations. *Medical Decision Making, 17*(4), 409–426.

Owens, D., Harris, R., Scott, P., & Nease, R. F., Jr. (1995). Screening surgeons for HIV infection: A cost-effectiveness analysis. *Annals of Internal Medicine, 122*(9), 641–652.

Owens, D. K., Holodniy, M., Garber, A. M., Scott, J., Sonnad, S., Moses, L., Kinosian, B., & Schwartz, J. S. (1996a). The polymerase chain reaction for the diagnosis of HIV infection in adults: A meta-analysis with recommendations for clinical practice and study design. *Annals of Internal Medicine, 124*(9), 803–815.

Owens, D. K., Holodniy, M., McDonald, T. W., Scott, J., & Sonnad, S. (1996b). A meta-analytic evaluation of the polymerase chain reaction (PCR) for diagnosis of human immunodeficiency virus (HIV) infection in infants. *Journal of the American Medical Association, 275*(17), 1342–1348.

Owens, D. K., Shachter, R. D., & Nease, R. F., Jr. (1997a). Representation and analysis of medical decision problems with influence diagrams. *Medical Decision Making, 17*(3), 241–262.

Owens, D. K., Whitlock, E. P., Henderson, J., Pignone, M. P., Krist, A. H., Bibbins-Domingo, K., Curry, S. J., Davidson, K. W., Ebell, M., Gilman, M. W., Grossman, D. C., Kemper, A. R., Kurth, A. E., Maciosek, M., Siu, A. L., LeFevre, M. L., & for the U.S. Preventive Services Task Force. (2016). Use of decision models in the development of evidence-based clinical preventive services recommendations: Methods of the U.S. Preventive Services Task Force. *Annals of Internal Medicine, 165*, 501–508.

Palda, V. A., & Detsky, A. S. (1997). Perioperative assessment and management of risk from coronary artery disease. *Annals of Internal Medicine, 127*(4), 313–328.

Pauker, S. G., & Kassirer, J. P. (1980). The threshold approach to clinical decision making. *The New England Journal of Medicine, 34*(5 Pt 2), 1189–1208.

Peabody, G. (1922). The physician and the laboratory. *Boston Medical Surgery Journal, 187*, 324.

Peterson, W., & Birdsall, T. (1953). *The theory of signal detectability. (Technical Report No. 13.)*: Electronic Defense Group, University of Michigan, Ann Arbor.

Ransohoff, D. F., & Feinstein, A. R. (1978). Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *The New England Journal of Medicine, 299*(17), 926–930.

Salpeter, S. R., Sanders, G. D., Salpeter, E. E., & Owens, D. K. (1997). Monitored isoniazid prophylaxis for low-risk tuberculin reactors older than 35 years of age: A risk-benefit and cost-effectiveness analysis. *Annals of Internal Medicine, 127*(12), 1051–1061.

Sanders, G. D., Hagerty, C. G., Sonnenberg, F. A., Hlatky, M. A., & Owens, D. K. (1999). Distributed dynamic decision support using a web-based interface for prevention of sudden cardiac death. *Medical Decision Making, 19*(2), 157–166.

Sanders, G. D., Hlatky, M. A., & Owens, D. K. (2005). Cost effectiveness of the implantable cardioverter defibrillator (ICD) in primary prevention of sudden death. *The New England Journal of Medicine, 353*, 1471–1478.

Siebert, U., Alagoz, O., Bayoumi, A. M., Jahn, B., Owens, D. K., Cohen, D., et al. (2012). State-transition modeling: A report of the ISPOR-SMDM modeling good research practices task force-3. *Medical Decision Making, 32*, 690–700.

Smith, L. (1985). Medicine as an art. In J. Wyngaarden & L. Smith (Eds.), *Cecil textbook of medicine*. Philadelphia: W. B. Saunders.

Sonnenberg, F. A., & Beck, J. R. (1993). Markov models in medical decision making: A practical guide. *Medical Decision Making, 13*(4), 322–338.

Sox, H. C. (1987). Probability theory in the use of diagnostic tests: Application to critical study of the literature. In H. C. Sox (Ed.), *Common diagnostic tests: Use and interpretation* (pp. 1–17). Philadelphia: American College of Physicians.

Sox, H. C., Blatt, M. A., Higgins, M. C., & Marton, K. I. (1988). *Medical decision making*. Boston: Butterworth Publisher.

Stein, P. D., Fowler, S. E., Goodman, L. R., Gottschalk, A., Hales, C. A., et al. (2006). Multidetector computed tomography for acute pulmonary embolism. *The New England Journal of Medicine, 354*, 2317–2327.

Sumner, W., Nease, R. F., Jr., & Littenberg, B. (1991). U-titer: A utility assessment tool. In *Proceedings of the 15th annual symposium on computer applications in medical care* (pp. 701–705). Washington, DC.

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science, 182*, 990.

Torrance, G. W., & Feeny, D. (1989). Utilities and quality-adjusted life years. *International Journal of Technology Assessment in Health Care, 5*(4), 559–575.

Tversky, A., & Kahneman, D. (1974a). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Cognitive Informatics

Vimla L. Patel and David R. Kaufman

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_4

4

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- How can cognitive science theory meaningfully inform and shape the design, development, and assessment of healthcare information systems?
- How is cognitive science different from behavioral science?
- What are some of the ways in which we can characterize the structure of knowledge?
- What are some of the dimensions of difference between experts and novices?
- Why is it important to consider cognition and human factors in dealing with issues of patient safety?
- How does distributed cognition differ from other theories of human cognition?

4.1 Introduction

Enormous advances in health information technologies and more generally, in computing over the past several decades have begun to permeate diverse facets of clinical practice. The rapid pace of technological developments such as the Internet, wireless technologies, and mobile devices, in the last decade, affords significant opportunities for supporting, enhancing and extending user experiences, interactions and communications (Rogers 2004). These advances, coupled with a growing computer literacy among healthcare professionals, afford the potential for great improvements in healthcare. Yet many observers note that the healthcare system is slow to understand information technology and effectively incorporate it into the work environment (Shortliffe and Blois 2001; Karsh et al. 2010; Harrington 2015). Innovative technologies often produce profound cultural, social, and cognitive changes. These transformations necessitate adaptation at many different levels of aggregation from the individual to the larger institution, often causing disruptions of workflow and user dissatisfaction (Bloomrosen et al. 2011).

Similar to other complex domains, biomedical information systems embody ideals in design that often do not readily yield practical solutions in implementation. As computer-based systems infiltrate clinical practice and settings, the consequences often can be felt through all levels of the organization. This impact can have deleterious effects resulting in systemic inefficiencies and suboptimal practice, which can lead to frustrated healthcare practitioners, unnecessary delays in healthcare delivery, and even adverse events (Lin et al. 1998; Weinger and Slagle 2001). How can we manage change? How can we introduce systems that are designed to be more intuitive and also implemented efficiently to be confluent with everyday practice without compromising safety?

4.1.1 Introducing Cognitive Science

Cognitive science is a multidisciplinary domain of inquiry devoted to the study of cognition and its role in intelligent agency. The primary disciplines include cognitive psychology, artificial intelligence, neuroscience, linguistics, anthropology, and philosophy. From the perspective of informatics, cognitive science can provide a framework for the analysis and modeling of complex human performance in technology-mediated settings. Cognitive science incorporates basic science research focusing on fundamental aspects of cognition (e.g., attention, memory, reasoning, early language acquisition) as well as applied research. Applied cognitive research is focally concerned with the development and evaluation of useful and usable cognitive artifacts. **Cognitive artifacts** are human-made materials, devices, and systems that extend people's abilities in perceiving objects, encoding and retrieving information from memory, and problem-solving (Gillan and Schvaneveldt 1999). In this regard, applied cognitive research is closely aligned with the disciplines of **human-computer interaction** (HCI) and **human factors**. In everyday life, we interact with cognitive artifacts to receive and manipulate information to alter our thinking

processes and offload effort-intensive cognitive activity to the external world, thereby reducing mental workload.

The past three decades have produced a cumulative body of experiential and practical knowledge about system design and implementation that guide future initiatives. This practical knowledge embodies the need for sensible and intuitive user interfaces, an understanding of workflow, and the ways in which systems impact individual and team performance. However, experiential knowledge in the form of anecdotes and case studies is inadequate for producing robust generalizations or sound design and implementation principles. There is a need for a theoretical foundation. Biomedical informatics is more than the thin intersection of biomedicine and computing (Patel and Kaufman 1998). There is a growing role for the social sciences, including the cognitive and behavioral sciences, in biomedical informatics, particularly as they pertain to human-computer interaction and other areas such as **information retrieval** and **decision support** (Patel et al. 2017). In this chapter, we focus on the foundational role of cognitive science in biomedical informatics research and practice. Theories and methods from the cognitive sciences can illuminate different facets of design and implementation of information and knowledge-based systems. They can also play a larger role in characterizing and enhancing human performance on a wide range of tasks involving clinicians, patients, and healthy consumers of biomedical information. These tasks may include developing training programs and devising measures to reduce errors or increase efficiency. In this respect, cognitive science represents one of the basic component sciences of biomedical informatics (Shortliffe and Blois 2001; Patel and Kaufman 1998).

4.1.2 Cognitive Science and Biomedical Informatics

How can cognitive science theory meaningfully inform and shape design, development, and assessment of health-care information systems? Cognitive science provides insight into principles of system usability and *learnability*, the mediating role of technology in clinical performance, the process of medical judgment and decision-making, the training of healthcare professionals, patients, and health consumers, and the design of a safer workplace. The central argument is that it can inform our understanding of human performance in technology-rich healthcare environments (Carayon 2012; Patel et al. 2013b).

Precisely how will cognitive science theory and methods make a significant contribution towards these important objectives? The translation of research findings from one discipline into practical concerns that can be applied to another is rarely a straight-forward process (Rogers 2004). Furthermore, even when scientific knowledge is highly relevant in principle, making that knowledge actionable in a design context can be a significant challenge. In this chapter, we discuss (a) basic cognitive science research and theories that provide a foundation for understanding the underlying mechanisms guiding human performance (e.g., findings pertaining to the structure of human memory), and (b) research in the areas of **medical errors** and **patient safety** as they interact with health information technology),

As illustrated in Table 4.1, there are correspondences between basic cognitive science research, medical cognition and cognitive research in biomedical informatics along several dimensions. For example, theories of human memory and knowledge organization lend themselves to characterizations of expert clinical knowledge that can then be contrasted with the representations of such knowledge in clinical systems. Similarly, research in text comprehension has provided a theoretical framework for research in understanding biomedical texts. Additionally, theories of problem solving can be used to understand the processes and knowledge associated with diagnostic and therapeutic reasoning. This understanding provides a basis for developing medical artificial intelligence and decision support systems.

Cognitive research, theories, and methods can contribute to applications in informatics

Table 4.1 Correspondences between cognitive science, medical cognition and applied cognitive research in medical informatics

Cognitive Science	Medical Cognition	Biomedical Informatics
Knowledge organization and human memory	Organization of clinical and basic science knowledge	Development and use of medical knowledge bases
Problem solving, Heuristics/ reasoning strategies	Medical problem solving and decision making	Medical artificial intelligence/decision support systems/medical errors
Perception/attention	Radiologic and dermatologic diagnosis	Medical imaging systems
Text comprehension	Understanding medical texts Knowledge representation	Information retrieval/digital libraries/ health literacy
Conversational analysis	Medical discourse	Medical natural language processing
Distributed cognition	Collaborative practice and research in health care	Computer-based provider order entry systems
Coordination of theory and evidence	Diagnostic and therapeutic reasoning	Evidence-based clinical guidelines
Diagrammatic reasoning	Perceptual processing of patient data displays	Biomedical information visualization

in a number of ways including: (1) seed *basic research findings* that can illuminate dimensions of design (e.g., attention and memory, aspects of the visual system), (2) provide an *explanatory vocabulary* for characterizing how individuals process and communicate health information (e.g., various studies of medical cognition pertaining to doctor-patient interaction), (3) present an *analytic framework* for identifying problems and modeling certain kinds of user interactions, (4) characterize the relationship between health information technology, human factors and patient safety, (5) provide *rich descriptive accounts* of clinicians employing technologies in the context of work, and (6) furnish a generative approach for novel designs and productive applied research programs in informatics (e.g., intervention strategies for supporting low literacy populations in health information seeking).

Based on a review of articles published in the Journal of Biomedical Informatics between January 2001 and March 2014, Patel and Kannampallil (2015) identified 57 articles that focused on topics related to cognitive informatics. The topics ranged from characterizing the limits of clinician problem-solving and reasoning behavior, to describing coordination and communication patterns of distributed clinical teams, to developing sustainable and cognitively plausible interventions for supporting clinician activities.

The social sciences are constituted by multiple frameworks and approaches. **Behaviorism** constitutes a framework for analyzing and modifying behavior. It is an approach that has had an enormous influence on the social sciences. Cognitive science partially emerged as a response to the limitations of behaviorism. The next section of the chapter contains a brief history of the cognitive and behavioral sciences that emphasizes the points of difference between the two approaches. It also serves to introduce basic concepts in the study of cognition.

4.2 Cognitive Science: The Emergence of an Explanatory Framework

In this section, we sketch a brief history of the emergence of cognitive science in view to differentiate it with competing theoretical frameworks in the social sciences. The section also

serves to introduce core concepts that constitute an explanatory framework for cognitive science.

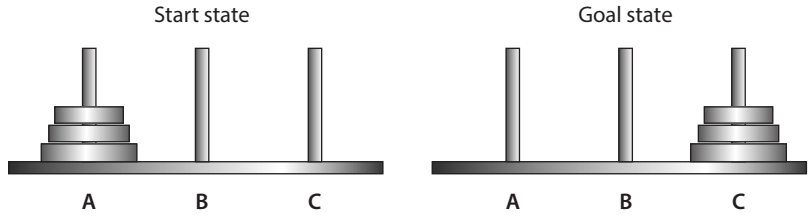
Behaviorism is the conceptual framework underlying a particular science of behavior (Zuriff 1985). It is not to be confused with the term behavioral science which names a large body of work across disciplines, but not a specific theoretical framework. Behaviorism dominated experimental and applied psychology as well as the social sciences for the better part of the twentieth century (Bechtel et al. 1998). Behaviorism represented an attempt to develop an objective, empirically-based science of behavior and more specifically, learning. Empiricism is the view that experience is the only source of knowledge (Hilgard and Bower 1975). Behaviorism endeavored to build a comprehensive framework of scientific inquiry around the experimental analysis of observable behavior. Behaviorists eschewed the study of thinking as an unacceptable psychological method because it was inherently subjective, error-prone, and could not be subjected to empirical validation. Similarly, hypothetical constructs (e.g., mental processes as mechanisms in a theory) were discouraged. All constructs had to be specified in terms of operational definitions, so they could be manipulated, measured, and quantified for empirical investigation (Weinger and Slagle 2001).

For reasons that go beyond the scope of this chapter, classical behavioral theories have been largely discredited as a comprehensive unifying theory of behavior. However, behaviorism continues to provide a theoretical and methodological foundation in a wide range of social science disciplines. For example, behaviorist tenets continue to play a central role in public health research. In particular, health behavior research emphasizes antecedent variables and environmental contingencies that serve to sustain unhealthy behaviors such as smoking (Sussman 2001). Around 1950, there was increasing dissatisfaction with the limitations and methodological constraints (e.g., the disavowal of the unobserved such as mental states) of behaviorism. In addition, developments in logic, information theory, cybernetics, and perhaps most importantly, the advent of the digital computer, aroused substantial interest in "information processing" (Gardner 1985).

Cognitive scientists placed "thought" and "mental processes" at the center of their explanatory framework. The "computer metaphor" provided a framework for the study of human cognition as the manipulation of "symbolic structures." It also provided the foundation for a model of memory, which was a prerequisite for an information processing theory (Atkinson and Shiffrin 1968). The implementation of models of human performance as computer programs provided a measure of objectivity and a *sufficiency test* of a theory and also served to increase the objectivity of the study of mental processes (Estes 1975).

Arguably, the landmark publication in the nascent field of cognitive science is Newell and Simon's "Human Problem Solving" (Newell and Simon 1972). This was the culmination of over 15 years of work on problem solving and research in artificial intelligence. It was a mature thesis that described a theoretical framework, extended a language for the study of cognition, and introduced protocol-analytic methods that have become ubiquitous in the study of high-level cognition. It laid the foundation for the formal investigation of symbolic-information processing (more specifically, problem solving). The development of models of human information processing also provided a foundation for the discipline of human-computer interaction and the first formal methods of analysis (Card et al. 1983).

The early investigations of problem solving focused primarily on investigations of experimentally contrived or toy-world tasks such as elementary deductive logic, the Tower of Hanoi, illustrated in Fig. 4.1, and mathematical word problems (Greeno and Simon 1988). These tasks required very little background knowledge and were well structured, in the sense that all the variables necessary for solving the problem were present in the problem statement. These tasks allowed for a complete description of the task environment, a step-by-step description of the sequential behavior of the subjects' performance, and the modeling of subjects' cognitive and overt



4

Fig. 4.1 Tower of Hanoi task illustrating a start state and a goal state

behavior in the form of a computer simulation. The Tower of Hanoi, in particular, served as an important test bed for the development of an explanatory vocabulary and framework for analyzing problem-solving behavior.

The Tower of Hanoi (TOH) is a relatively straight-forward task that consists of three pegs (A, B, and C) and three or more disks that vary in size. The goal is to move the three disks from peg A to peg C one at a time with the constraint that a larger disk can never rest on a smaller one. Problem solving can be construed as *search* in a **problem space**. A problem space has an *initial state*, a *goal state*, and a *set of operators*. Operators are any moves that transform a given state to a successor state. For example, the first move could be to move the small disk to peg B or peg C. In a three-disk TOH, there are a total of 27 possible states representing the complete problem space. TOH has 3^n states where n is the number of disks. The minimum number of moves necessary to solve a TOH is 2^{n-1}. Problem solvers will typically maintain only a small set of states at a time.

The search process involves finding a solution strategy that will minimize the number of steps. The metaphor of movement through a problem space provides a means for understanding how an individual can sequentially address the challenges they confront at each stage of a problem and the actions that ensue. We can characterize the problem-solving behavior of the subject at a local level in terms of state transitions or at a more global level in terms of *strategies*. For example, *means-ends analysis* is a commonly used strategy for reducing the difference between the start state and goal state. For instance, moving all but the largest disk from peg A to peg B is an interim goal associated with such a strategy. Although TOH bears little resemblance to the tasks performed by either clinicians or patients, the example illustrates the process of analyzing task demands and task performance in human subjects. The TOH helped lay the groundwork for cognitive task analyses that are performed today.

Protocol analysis[1] is among the most commonly used methods (Newell and Simon 1972). Protocol analysis refers to a class of techniques for representing verbal **think-aloud protocols** (Greeno and Simon 1988). Think-aloud protocols are the most common source of data used in studies of problem solving. In these studies, subjects are instructed to verbalize their thoughts as they perform an experimental task. Ericsson and Simon (1993) specify the conditions under which verbal reports are acceptable as legitimate data. For example, retrospective think-aloud protocols are viewed as somewhat suspect because the subject has had the opportunity to reconstruct the information in memory, and the verbal reports are inevitably distorted. Think-aloud protocols recorded in concert with observable behavioral data such as a subject's actions provide a rich source of evidence to characterize cognitive processes.

Cognitive psychologists and linguists have investigated the processes and properties of language and memory in adults and children for many decades. Early research focused on basic laboratory studies of list learning or processing of words and sentences (as in a sentence completion task) (Anderson 1985).

1 The term protocol refers to that which is produced by a subject during testing (e.g., a verbal record). It differs from the more common use of protocol as defining a code or set of procedures governing behavior or a situation.

van Dijk and Kintsch (1983) developed an influential method of analyzing the process of **text comprehension** based on the realization that text can be described at multiple levels from surface codes (e.g., words and syntax) to a deeper level of semantics. Comprehension refers to cognitive processes associated with understanding or deriving meaning from text, conversation, or other informational resources. It involves the processes that people use when trying to make sense of a piece of text, such as a sentence, a book, or a verbal utterance. It also involves the final product of such processes, which is, the mental representation of the text, essentially what people have understood.

Comprehension may often precede problem solving and decision making but is also dependent on perceptual processes that focus attention, the availability of relevant knowledge, and the ability to deploy knowledge in a given context. Some of the more important differences in medical problem solving and decision making arise from differences in knowledge and comprehension. Furthermore, many of the problems associated with decision making are the result of either a lack of knowledge or failure to understand the information appropriately.

The early investigations provided a well-constrained artificial environment for the development of the basic methods and principles of problem solving. They also provide a rich explanatory vocabulary (e.g., problem space), but were not fully adequate in accounting for cognition in knowledge-rich domains of greater complexity and involving uncertainty. In the mid to late 1970s, there was a shift in research to complex "real-life" knowledge-based domains of inquiry (Greeno and Simon 1988). Problem-solving research was studying performance in domains such as physics (Larkin et al. 1980), medical diagnoses (Elstein et al. 1978) and architecture (Akin 1982). Similarly, the study of text comprehension shifted from research on simple stories to technical and scientific texts in a range of domains, including medicine. This paralleled a similar change in artificial intelligence research from "toy programs" to addressing "real-world" problems and the development of expert systems (Clancey and Shortliffe 1984). The shift to real-world problems in cognitive science was spearheaded by research exploring the nature of expertise. Most of the early investigations on expertise involved laboratory experiments. However, the shift to knowledge-intensive domains provided a theoretical and methodological foundation to conduct both basic and applied research in real-world settings such as the workplace (Vicente 1999) and the classroom (Bruer 1993). These areas of application provided a fertile test bed for assessing and extending the cognitive science framework.

In recent years, the conventional information-processing approach has come under criticism for its narrow focus on the rational/cognitive processes of the solitary individual. One of the most compelling proposals has to do with a shift from viewing cognition as a property of the solitary individual to viewing cognition as distributed across groups, cultures, and artifacts. This claim has significant implications for the study of collaborative endeavors and human-computer interaction. We explore the concepts underlying *distributed cognition* in greater detail in a subsequent section.

4.3 Human Information Processing

It is well known that product design often fails to consider cognitive and physiological constraints adequately and imposes an unnecessary burden on task performance (Sharp et al. 2019). Fortunately, advances in theory and methods provide us with greater insight into designing systems for the human condition.

Cognitive science serves as a basic science and provides a framework for the analysis and modeling of complex human performance. A computational theory of mind provides the fundamental underpinning for most contemporary theories of cognitive science. The basic premise is that much of human cognition can be characterized as a series of operations or computations on mental representations. **Mental representations** are internal cognitive

states that have a certain correspondence with the external world. For example, they may reflect a clinician's hypothesis about a patient's condition after noticing an abnormal gait as he entered the clinic. These are likely to elicit further inferences about the patient's underlying condition and may direct the physician's information-gathering strategies and contribute to an evolving problem representation.

Two interdependent dimensions by which we can characterize cognitive systems are (1) architectural theories that endeavor to provide a unified theory for all aspects of cognition and (2) the different kinds of knowledge necessary to attain competency in a given domain. Individuals differ substantially in terms of their knowledge, experiences, and endowed capabilities. The architectural approach capitalizes on the fact that we can characterize certain regularities of the human information-processing system. These can be either structural regularities—such as the existence of and the relations between perceptual, attentional, and memory systems and memory capacity limitations—or processing regularities, such as processing speed, selective attention, or problem-solving strategies. Cognitive systems are characterized functionally in terms of the capabilities they enable (e.g., focused attention on selective visual features), the way they constrain human cognitive performance (e.g., limitations on memory), and their development during the lifespan. In regard to the lifespan issue, there is a growing body of literature on cognitive aging and how aspects of the cognitive system such as attention, memory, vision and motor skills change as a function of aging (Fisk et al. 2009). This basic science research is of growing importance to informatics as we seek to develop e-health applications for seniors, many of whom suffer from chronic health conditions such as arthritis and diabetes. A graphical user interface or more generally, a website designed for younger adults may not be suitable for older adults.

Differences in knowledge organization are a central focus of research into the nature of expertise. In medicine, the expert-novice paradigm has contributed to our understanding of the nature of medical expertise and skilled clinical performance.

4.3.1 Cognitive Architectures and Human Memory Systems

Fundamental research in perception, cognition, and psychomotor skills over the last 50 years has provided a foundation for design principles in human factors and human-computer interaction. Although cognitive guidelines have made significant inroads in the design community, there remains a significant gap in applying basic cognitive research (Gillan and Schvaneveldt 1999). Designers routinely violate basic assumptions about the human cognitive system. There are invariably challenges in applying basic research and theory to applications. More human-centered design and cognitive research can instrumentally contribute to such an endeavor (Zhang et al. 2004).

Over the last 50 years, there have been several attempts to develop a unified theory of cognition. The goal of such a theory is to provide a single set of mechanisms for all cognitive behaviors from motor skills, language, memory, to decision making, problem solving, and comprehension (Newell 1990). Such a theory provides a means to put together a voluminous and seemingly disparate body of human experimental data into a coherent form. Cognitive architecture represents unifying theories of cognition that are embodied in large-scale computer simulation programs. Although there is much plasticity evidenced in human behavior, cognitive processes are bound by biological and physical constraints. Cognitive architectures specify functional rather than biological constraints on human behavior (e.g., limitations on working memory). These constraints reflect the information-processing capacities and limitations of the human cognitive system. Architectural systems embody a relatively fixed permanent structure that is (more or less) characteristic of all humans and doesn't substantially vary over an individual's lifetime. It represents a scientific hypothesis about

those aspects of human cognition that are relatively constant over time and independent of the task (Carroll 2003). Cognitive architectures also play a role in providing blueprints for building future intelligent systems that embody a broad range of capabilities like those of humans (Duch et al. 2008). There are several large-scale cognitive architecture theories that embody computational models of cognition and have informed a substantial body of research in cognitive science and allied disciplines. ACT-R (short for "Adaptive Control of Thought-Rational") is perhaps, the most widely known cognitive architecture. It was developed by John R. Anderson and is sustained by a large global community of researchers centered at Carnegie Mellon University (Anderson 2013). It is a theory for simulating and understanding human cognition. It started more than 40 years ago as an architecture that could simulate basic tasks related to memory, language and problem solving. It has continued to evolve into a system that can perform an enormous range of human tasks (Ritter et al. 2019).

Cognitive architectures include short-term and long-term memories that store content about an individual's beliefs, goals, and knowledge, the representation of elements that are contained in these memories as well as their organization into larger-scale structures (Lieto et al. 2018). An extended discussion of architectural theories and systems is beyond the scope of this chapter. However, we employ the architectural frame of reference to introduce some basic distinctions in memory systems. Human memory is typically divided into at least two structures: **long-term memory** and **short-term/working memory**. Working memory is an emergent property of interaction with the environment. Long-term memory (LTM) can be thought of as a repository of all knowledge, whereas working memory (WM) refers to the resources needed to maintain information active during cognitive activity (e.g., text comprehension). The information maintained in working memory includes stimuli from the environment (e.g., words on a display) and knowledge activated from long-term memory. In theory, LTM is infinite, whereas WM is limited to five to ten "chunks" of information. A chunk is any stimulus or patterns of stimuli that have become familiar from repeated exposure and is subsequently stored in memory as a single unit (Larkin et al. 1980). Problems impose a variable **cognitive load** on working memory. This refers to an excess of information that competes for few cognitive resources, creating a burden on working memory (Chandler and Sweller 1991). For example, maintaining a seven-digit phone number in WM is not very difficult. However, to maintain a phone number while engaging in conversation is nearly impossible for most people. Multi-tasking is one factor that contributes to cognitive load. The structure of the task environment, for example, a crowded computer display is another contributor. High velocity/high workload clinical environments such as intensive care units also impose cognitive loads on clinicians carrying out the task.

4.3.2 The Organization of Knowledge

Architectural theories specify the structure and mechanisms of memory systems, whereas theories of knowledge organization focus on the content. There are several ways to characterize the kinds of knowledge that reside in LTM and that support decisions and actions. Cognitive psychology has furnished a range of domain-independent constructs that account for the variability of mental representations needed to engage the external world.

A central tenet of cognitive science is that humans actively construct and interpret information from their environment. Given that environmental stimuli can take a multitude of forms (e.g., written text, speech, music, images, etc.), the cognitive system needs to be attuned to different representational types to capture the essence of these inputs. For example, we process written text differently than we do mathematical equations. The power of cognition is reflected in the ability to form abstractions - to represent perceptions, experiences, and thoughts in some medium

1. 43-year-old white female who developed diarrhea after a brief period of 2 days of GI upset

1.1	female	ATT: Age (old); DEG: 43 year; ATT: white
1.2	develop	PAT: [she]; THM: diarrhea; TNS: past
1.3	period	ATT: brief; DUR: 2 days; THM: 1.4
1.4	upset	LOC: GI
1.5	TEM:ORD	[1.3], [1.2]

Fig. 4.2 Propositional analysis of a think-aloud protocol of a primary care physician

other than that in which they have occurred without extraneous or irrelevant information (Norman 1993). Representations enable us to remember, reconstruct, and transform events, objects, images, and conversations absent in space and time from our initial experience of the phenomena. Representations reflect states of knowledge.

Propositions are a form of natural language representation that captures the essence of an idea (i.e., semantics) or concept without explicit reference to linguistic content. For example, "hello", "hey", and "what's happening" can typically be interpreted as a greeting containing identical propositional content even though the literal semantics of the phrases may differ. These ideas are expressed as language and translated into speech or text when we talk or write. Similarly, we recover the propositional structure when we read or listen to verbal information. Numerous psychological experiments have demonstrated that people recover the gist of a text or spoken communication (i.e., propositional structure) not the specific words (Anderson 1985; van Dijk and Kintsch 1983). Studies have also shown the individuals at different levels of expertise will differentially represent a text (Patel and Kaufman 1998). For example, experts are more likely to selectively encode relevant propositional information that will inform a decision. On the other hand, non-experts will often remember more information, but much of the recalled information may not be relevant to the decision (Patel and Groen 1991a, b).

Propositional representations constitute an important construct in theories of comprehension. Propositional knowledge can be expressed using a predicate calculus formalism or as a semantic network. The predicate calculus representation is illustrated below. A subject's response, as given on Fig. 4.2, is divided into sentences or segments and sequentially analyzed. The formalism includes a head element of a segment and a series of arguments. For example, in proposition 1.1, the focus is on a female who has the attributes of being 43 years of age and white. The TEM:ORD or temporal order relation indicates that the events of 1.3 (GI upset) precede the event of 1.2 (diarrhea). The formalism is informed by an elaborate propositional language (Frederiksen 1975) and was first applied to the medical domain by Patel and her colleagues (Patel and Groen 1986). The method provides us with a detailed way to characterize the information subjects understood from reading a text, based on their summary or explanations.

Kintsch (1998) theorized that comprehension involves an interaction between what the text conveys and knowledge in long-term memory. Comprehension occurs when the reader uses prior knowledge to process the incoming information presented in the text. The text information is called the *textbase* (the propositional content of the text). For instance, in medicine, the textbase could consist of the representation of a patient problem as written in a patient chart. The situation model is constituted by the textbase representation plus the domain-specific and everyday knowledge that the reader uses to derive a broader meaning from the text. In medicine, the situation model would enable a physician to draw inferences from a patient's history leading to a diagnosis, therapeutic plan or prognosis (Patel and Groen 1991a, b). This situation model is typically derived from the general knowledge and specific knowledge acquired through medical teaching, readings

(e.g., theories and findings from biomedical research), clinical practice (e.g., knowledge of associations between clinical findings and specific diseases, knowledge of medications or treatment procedures that have worked in the past) and the textbase representation. Like other forms of knowledge representation, the situation model is used to "fit in" the incoming information (e.g., text, perception of the patient). Since the knowledge in LTM differs among physicians, the resulting situation model generated by any two physicians is likely to differ as well. Theories and methods of text comprehension have been widely used in the study of medical cognition and have been instrumental in characterizing the process of guideline development and interpretation (Peleg et al. 2006; Patel et al. 2014).

Schemata represent higher-level knowledge structures. They can be construed as data structures for representing categories of concepts stored in memory (e.g., fruits, chairs, geometric shapes, and thyroid conditions). There are schemata for concepts underlying situations, events, sequences of actions and so forth. To process information with the use of a schema is to determine which model best fits the incoming information. Schemata have constants (all birds have wings) and variables (chairs can have between one and four legs). The variables may have associated default values (e.g., birds fly) that represent the prototypical circumstance.

When a person interprets information, the schema serves as a "filter" for distinguishing relevant and irrelevant information. Schemata can be considered as generic knowledge structures that contain slots for particular kinds of propositions. For instance, a schema for myocardial infarction may contain the findings of "chest pain," "sweating," "shortness of breath," but not the finding of "goiter," which is part of the schema for thyroid disease.

The schematic and propositional representations reflect abstractions and don't necessarily preserve literal information about the external world. Imagine that you are having a conversation at the office about how to rearrange the furniture in your living room. To engage in such a conversation, one needs to be able to construct images of the objects and their spatial arrangement in the room. **Mental images** are a form of internal representation that captures perceptual information recovered from the environment. There is compelling psychological and neuropsychological evidence to suggest that mental images constitute a distinct form of mental representation (Bartolomeo 2008). Images play a particularly important role in domains of visual diagnosis such as dermatology and radiology.

Mental models are an analog-based construct for describing how individuals form internal models of systems. Mental models are designed to answer questions such as "how does it work?" or "what will happen if I take the following action?" "Analogy" suggests that the representation explicitly shares the structure of the world it represents (e.g., a set of connected visual images of a partial road map from your home to your work destination). This contrasts with an abstraction-based form such as propositions or schemas in which the mental structure consists of either the gist, an abstraction, or summary representation. However, like other forms of mental representation, mental models are always incomplete, imperfect, and subject to the processing limitations of the cognitive system. Mental models can be derived from perception, language, or one's imagination (Payne 2003). *Running* of a model corresponds to a process of mental simulation to generate possible future states of a system from observed or hypothetical state. For example, when one initiates a Google Search, one may reasonably anticipate that the system will return a list of relevant (and less than relevant) websites that correspond to the query. Mental models are a particularly useful construct in understanding human-computer interaction.

An individual's mental models provide predictive and explanatory capabilities of the function of a physical system. More often the construct has been used to characterize models that have a spatial and temporal context, as is the case in reasoning about the behavior of electrical circuits (White and Frederiksen 1990). The model can be used to simulate a process (e.g., predict the effects of network interruptions on getting cash from an ATM machine). Kaufman, Patel and Magder (1996)

characterized clinicians' mental models of the cardiovascular system (specifically, cardiac output). The study characterized the development of an understanding of the system as a function of expertise. The research also documented various conceptual flaws in subjects' models and how these flaws impacted subjects' predictions and explanations of physiological manifestations. Figure 4.3 illustrates the four chambers of the heart and blood flow in the pulmonary and cardiovascular systems. The claim is that clinicians and medical students have variably robust representations of the structure and function of the system. This model enables prediction and explanation of the effects of perturbations in the system on blood flow and on various clinical measures such as left ventricular ejection fraction.

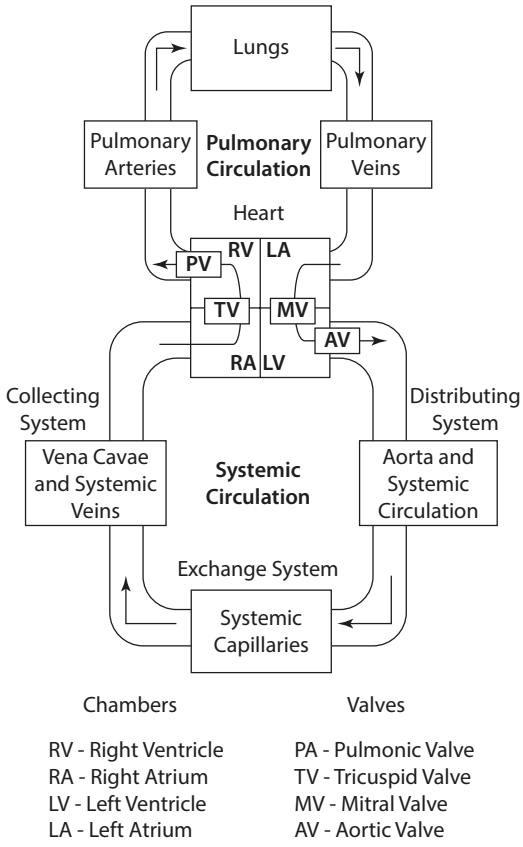


Fig. 4.3 Schematic model of circulatory and cardiovascular physiology. The diagram illustrates various structures of the pulmonary and systemic circulation system and the process of blood flow. The illustration is used to exemplify the concept of mental model and how it could be applied to explaining and predicting physiologic behavior

Thus far, we have only considered domain-general ways of characterizing the organization of knowledge. In view to understanding the nature of medical cognition, it is necessary to characterize the domain-specific nature of knowledge organization in medicine. Given the vastness and complexity of the domain of medicine, this can be a rather daunting task. There is no single way to represent all biomedical (or even clinical) knowledge, but it is an issue of considerable importance for research in biomedical informatics. Much research has been conducted in biomedical artificial intelligence to develop biomedical ontologies for use in knowledge-based systems (Ramoni et al. 1992). Patel et al. (1997) address this issue in the context of using empirical evidence from psychological experiments on medical expertise to test the validity of the AI systems. **Developers of biomedical taxonomies**, **nomenclatures,** and **vocabulary** systems such as UMLS or SNOMED are engaged in a similar pursuit (see ▶ Chap. 7).

We have employed an epistemological framework developed by Evans and Gadd (1989). They proposed a framework that serves to characterize the knowledge used for medical understanding and problem solving, and for differentiating the levels at which biomedical knowledge may be organized. This framework represents a formalization of biomedical knowledge as realized in textbooks and journals and can be used to provide us with insight into the organization of clinical practitioners' knowledge (see Fig. 4.4).

The framework consists of a hierarchical structure of concepts formed by *clinical observations* at the lowest level, followed by *findings*, *facets*, and *diagnoses*. Clinical observations are units of information that are recognized as potentially relevant in the problem-solving context. However, they do not constitute clinically useful facts. Findings are composed of observations that have potential clinical significance. Establishing a finding reflects a decision made by a physician that an array of data contains a significant cue or cues that need to be considered. Facets consist of clusters of findings that indicate an underlying

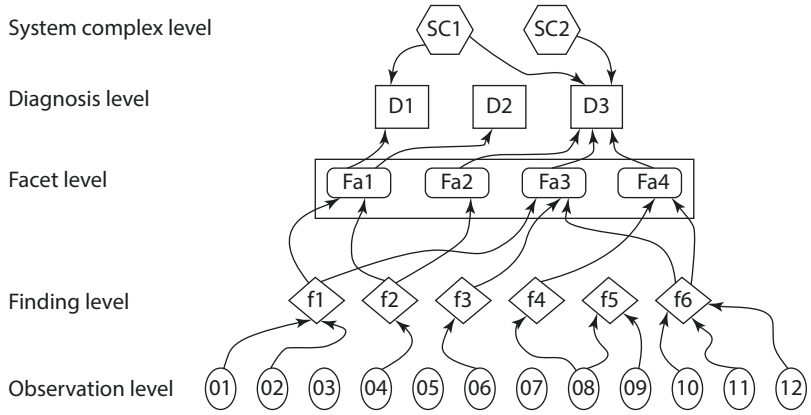


Fig. 4.4 Epistemological frameworks representing the structure of medical knowledge for problem solving

medical problem or class of problems. They reflect general pathological descriptions such as left-ventricular failure or thyroid condition. Facets resemble the kinds of constructs used by researchers in medical artificial intelligence to describe the partitioning of a problem space. They are interim hypotheses that serve to divide the information in the problem into sets of manageable sub-problems and to suggest possible solutions. Facets also vary in terms of their levels of abstraction. Diagnosis is the level of classification that subsumes and explains all levels beneath it. Finally, the systems level consists of information that serves to contextualize a problem, such as the ethnic background of a patient.

4.4 Medical Cognition

The study of expertise is one of the principal paradigms in problem-solving research, which has been documented in a number of volumes in literature (Sternberg and Ericsson 1996; Ericsson 2009; Ericsson et al. 2018). Comparing experts to novices provides us with the opportunity to explore the aspects of performance that undergo change and result in increased problem-solving skill (Glaser 2000). It also permits investigators to develop domain-specific models of competence that can be used for assessment and training purposes.

A goal of this approach has been to characterize expert performance in terms of the knowledge and cognitive processes used in comprehension, problem solving, and decision making, using carefully developed laboratory tasks (Chi and Glaser 1981), where deGroot's (1965) pioneering research in chess represents one of the earliest characterizations of expert-novice differences. In one of his experiments, subjects were allowed to view a chess board for 5–10 seconds and were then required to reproduce the position of the chess pieces from memory. The grandmaster chess players were able to reconstruct the mid-game positions with better than 90% accuracy, while novice chess players could only reproduce approximately 20% of the correct positions. When the chess pieces were placed on the board in a random configuration, not encountered in the course of a normal chess match, expert chess masters' recognition ability fell to that of novices. This result suggests that superior recognition ability is not a function of superior memory, but is a result of an enhanced ability to recognize typical situations (Chase and Simon 1973). This phenomenon is accounted for by a process known as "**chunking**." Patel and Groen (1991b) showed a similar phenomenon in medicine The expert physicians were able to reconstruct patient summaries in an accurate manner when patient information was collected out of order (e.g., history, physical exam, lab results), as long as the pattern of information, even out of sequence was famil-

iar. When the sentences were placed out of order in a way that the pattern was unfamiliar, the expert physicians' recognition ability was no better than the novices.

It is well known that knowledge-based differences impact the problem representation and determine the strategies a subject uses to solve a problem. Simon and Simon (1978) compared a novice subject with an expert subject in solving textbook physics problems. The results indicated that the expert solved the problems in one-quarter of the time required by the novice with fewer errors. The novice solved most of the problems by working *backward* from the unknown problem solution to the givens of the problem statement. The expert worked *forward* from the givens to solve the necessary equations and determine the quantities they are asked to solve for. Differences in the directionality of reasoning by levels of expertise has been demonstrated in diverse domains from computer programming (Perkins et al. 1990) to medical diagnosis (Patel and Groen 1986).

The expertise paradigm spans the range of content domains including physics (Larkin et al. 1980), sports (Allard and Starkes 1991), music (Sloboda 1991), and medicine (Patel et al. 1994). Edited volumes (Ericsson 2006; Chi et al. 1988 Ericsson et al. 2018; Ericsson and Smith 1991; Hoffman 1992) provide an informative general overview of the area. This research has focused on differences between subjects varying in levels of expertise in terms of memory, reasoning strategies, and in particular the role of domain-specific knowledge. Among the expert's characteristics uncovered by this research are the following: (1) experts are capable of perceiving large patterns of meaningful information in their domain, which novices cannot perceive; (2) they are fast at processing and at deployment of different skills required for problem solving; (3) they have superior short-term and long-term memories for materials (e.g., clinical findings in medicine) within their domain of expertise, but not outside of it; (4) they typically represent problems in their domain at deeper, more principled levels whereas novices show a superficial level of representation; (5) they spend more time assessing the problem prior to solving it, while novices tend to spend more time working on the solution itself and little time in problem assessment; (6) individual experts may differ substantially in terms of exhibiting these kinds of performance characteristics (e.g., superior memory for domain materials).

Usually, someone is designated as an expert based on a certain level of performance, as exemplified by Elo ratings in chess; by virtue of being certified by a professional licensing body, as in medicine, law, or engineering; on the basis of academic criteria, such as graduate degrees; or simply based on years of experience or peer evaluation (Hoffman et al. 1995). The concept of an expert, however, refers to an individual who surpasses competency in a domain (Sternberg and Horvath 1999). Although competent performers, for instance, may be able to encode relevant information and generate effective plans of action in a specific domain, they often lack the speed and the flexibility that we see in an expert. A domain expert (e.g., a medical practitioner) possesses an extensive, accessible knowledge base that is organized for use in practice and is tuned to the particular problems at hand. In the study of medical expertise, it has been useful to distinguish different types of expertise.

Patel and Groen (1991a, b) distinguished between general and specific expertise, a distinction supported by research indicating differences between subexperts (i.e., expert physicians who solve a case outside their field of specialization) and experts (i.e., domain specialist) with respect to reasoning strategies and organization of knowledge. General expertise corresponds to expertise that cuts across medical subdisciplines (e.g., general medicine). Specific expertise results from having extensive experience within a medical subdomain, such as cardiology or endocrinology. An individual may possess both or only generic expertise.

The development of expertise can follow a somewhat unusual trajectory. It is often assumed that the path from novice to expert goes through a steady process of gradual accumulation of knowledge and fine-tuning of skills. That is, as a person becomes more familiar with a domain, his or her level of per-

formance (e.g., accuracy, quality) gradually increases. However, research has shown that this assumption is often incorrect (Lesgold et al. 1988; Patel et al. 1994). Cross-sectional studies of experts, intermediates, and novices have shown that people at intermediate levels of expertise may perform more poorly than those at a lower level of expertise on some tasks.

Furthermore, there is a longstanding body of research on learning that has suggested that the learning process involves phases of error-filled performance followed by periods of stable, comparatively error-free performance. In other words, human learning does not consist of the gradually increasing accumulation of knowledge and fine-tuning of skills. Rather, it requires the arduous process of continually learning, re-learning, and exercising new knowledge, punctuated by periods of an apparent decrease in mastery and declines in performance, which may be necessary for learning to take place. Figure 4.5 presents an illustration of this learning and development phenomenon known as the **intermediate effect**.

The intermediate effect has been found in a variety of tasks and with a great number of performance indicators. The tasks used include comprehension and explanation of clinical problems, doctor-patient communication, recall and explanation of laboratory data, generation of diagnostic hypotheses, and problem solving (Patel and Groen 1991a, b). The performance indicators used have included recall and inference of medical-text information, recall, and inference of diagnostic hypotheses, generation of clinical findings from a patient in doctor-patient interaction, and requests for laboratory data, among others. The research has also identified developmental levels at which the intermediate phenomenon occurs, including senior medical students and residents. It is important to note, however, that in some tasks, the development is **monotonic**. For instance, in diagnostic accuracy, there is a gradual increase, with an intermediate exhibiting a greater degree of accuracy than the novice and the expert demonstrating a still greater degree than the intermediate. Furthermore, when the relevancy of the stimuli to a problem is considered, an appreciable monotonic phenomenon appears. For instance, in recall studies, novices, intermediates, and experts are assessed in terms of the total number of propositions recalled showing the typical non-monotonic effect. However, when propositions are divided in terms of their relevance to the problem (e.g., a clinical case), experts recall more relevant propositions than intermediates and novices, suggesting that intermediates have difficulty separating what is relevant from what is not.

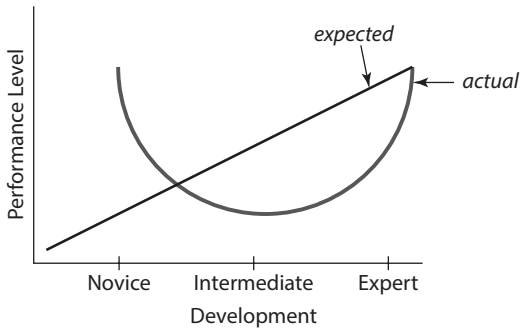


Fig. 4.5 Schematic representation of intermediate effect. The straight line gives a commonly assumed representation of performance development by level of expertise. The curved line represents the actual development from novice to expert. The Y-axis may represent any of a number of performance variables such as the number of errors made, number of concepts recalled, number of conceptual elaborations, or number of hypotheses generated in a variety of tasks

During the periods when the intermediate effect occurs, a reorganization of knowledge and skills takes place, characterized by shifts in perspectives or a realignment or creation of goals. The intermediate effect is also partly due to the unintended changes that take place as the person reorganizes for intended changes. People at intermediate levels typically generate a great deal of irrelevant information and seem incapable of discriminating what is relevant from what is not. As compared to a novice student (Fig. 4.6), the reasoning pattern of an intermediate student shows the generation of long chains of discussion evaluating multiple hypotheses and reasoning in haphazard direction (Fig. 4.7). A well-structured knowledge structure of a senior level student leads him more directly to a

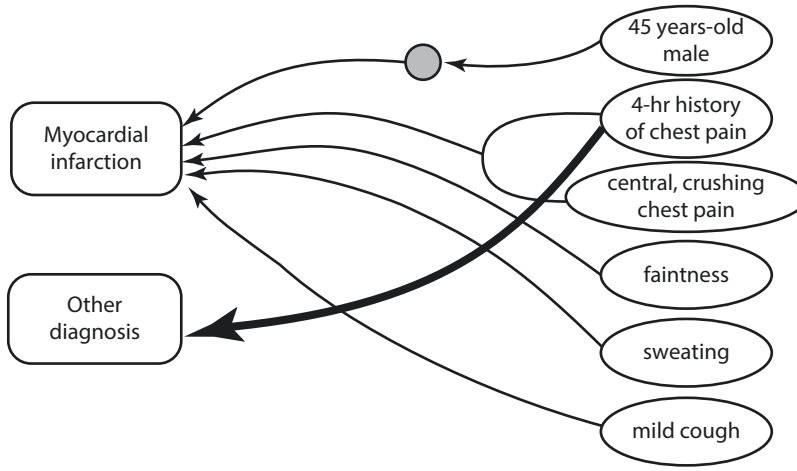


Fig. 4.6 Problem interpretations by a novice medical student. The given information from patient problem is represented on the right side of the figure and the new generated information is given on the left side, information in the box represents diagnostic hypothesis. Intermediate hypotheses are represented as solid dark circles (filled). Forward driven or data driven inference arrows are shown from left to right (solid dark line). Backward or hypothesis driven inference arrows are shown from right to left (solid light line). Thick solid dark line represents rule out strategy

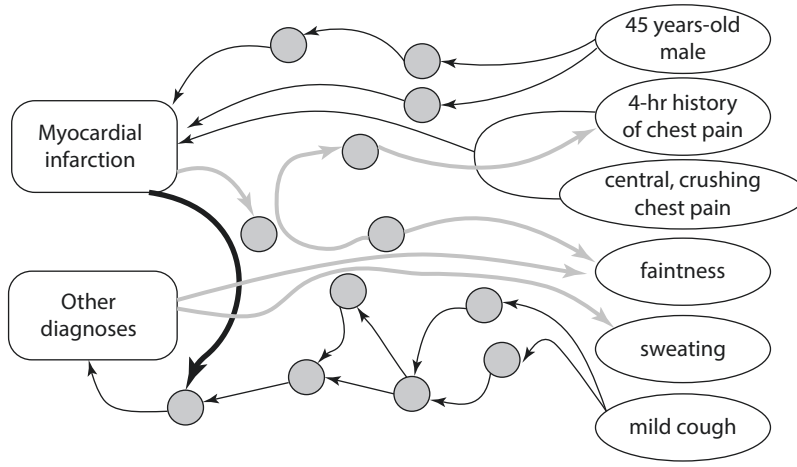


Fig. 4.7 Problem interpretations by an intermediate medical student

solution (Fig. 4.8). Thus, the intermediate effect can be explained as a function of the learning process, maybe as a necessary phase of learning. Identifying the factors involved in the intermediate effect may help in improving performance during learning (e.g., by designing decision-support systems or intelligent tutoring systems that help the user in focusing on relevant information).

The intermediate effect is not a one-time phenomenon. Rather, it repeatedly occurs at strategic points in a student or physician's training and follows periods in which large bodies of new knowledge or complex skills are acquired. These periods are followed by intervals in which there is a decrement in performance until a new level of mastery is achieved.

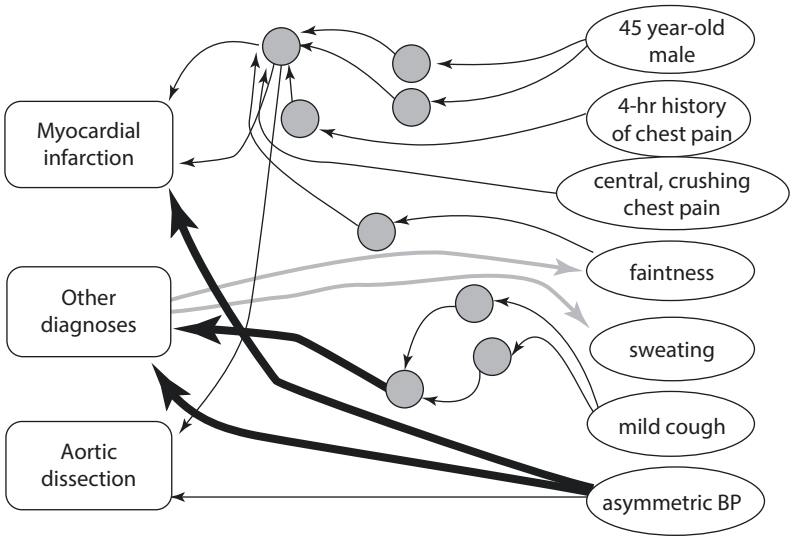


Fig. 4.8 Problem interpretations by a senior medical student

4.4.1 Expertise in Medicine

The systematic investigation of medical expertise began more than 60 years ago with research by Ledley and Lusted (1959) into the nature of clinical inquiry. They proposed a two-stage model of clinical reasoning involving a hypothesis-generation stage, followed by a hypothesis-evaluation stage. This latter stage is most amenable to formal decision analytic techniques. The earliest empirical studies of medical expertise can be traced to the works of Rimoldi (1961) and Kleinmuntz and McLean (1968) who conducted experimental studies of diagnostic reasoning by contrasting students with medical experts in simulated problem-solving tasks. The results emphasized the greater ability of expert physicians to attend to relevant information selectively and narrow the set of diagnostic possibilities (i.e., consider fewer hypotheses).

The origin of contemporary research on medical thinking is associated with the seminal work of Elstein, Shulman, and Sprafka (1978) who studied the problem-solving processes of physicians by drawing on then-contemporary methods and theories of cognition. This model of problem-solving has had a substantial influence both on studies of medical cognition and medical education. They were the first to use experimental methods and theories of cognitive science to investigate clinical competency.

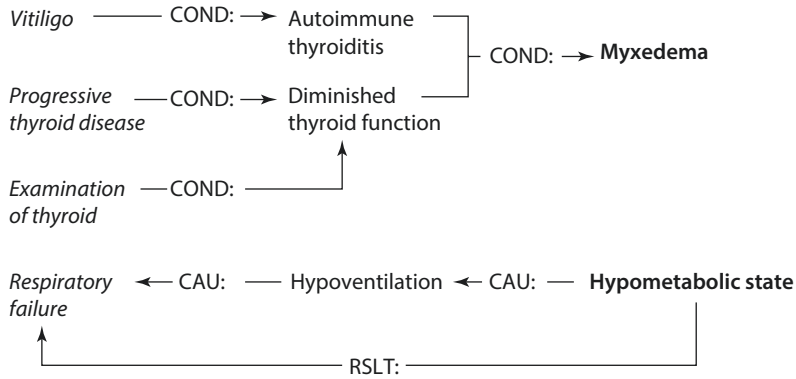
Their research findings led to the development of an elaborated model of **hypothetico-deductive reasoning**, which proposed that physicians reasoned by first generating and then testing a set of hypotheses to account for clinical data (i.e., reasoning from hypothesis to data). First, physicians generated a small set of hypotheses very early in the case, as soon as the first pieces of data became available. Second, physicians were selective in the data they collected, focusing only on the relevant data. Third, physicians made use of a hypothetico-deductive method of diagnostic reasoning (Elstein et al. 1978).

The previous research was largely modeled after early problem-solving studies in knowledge-lean tasks. Medicine is a knowledge-rich domain, and a different approach was needed. Feltovich, Johnson, Moller, and Swanson (1984), drawing on models of knowledge representation from medical artificial intelligence, characterized fine-grained differences in knowledge organization between subjects of different levels of expertise in the domain of pediatric cardiology. Patel and colleagues studied the knowledge-based solution strategies of expert cardiologists as evidenced by their pathophysiological explanations of a complex clinical problem (Patel and Groen 1986). The results indicated that subjects who accurately diagnosed the problem, employed a **forward-directed (data-driven) reasoning** strategy—

using patient data to lead toward a complete diagnosis (i.e., reasoning from data to hypothesis). This is in contrast to subjects who misdiagnosed or partially diagnosed the patient problem. They tended to use a **backward or hypothesis-driven reasoning** strategy.

Patel and Groen (1991a, b) investigated the nature and directionality of clinical reasoning in a range of contexts of varying complexity. The objectives of this research program were both to advance our understanding of medical expertise and to devise more effective ways of teaching clinical problem solving. It has been established that the patterns of data-driven and hypothesis-driven reasoning are used differentially by novices and experts. Experts tend to use data-driven reasoning, which depends on the physician possessing a highly organized knowledge base about the patient's disease (including sets of signs and symptoms). Because of their lack of substantive knowledge or their inability to distinguish relevant from irrelevant knowledge, novices and intermediates use more hypothesis-driven reasoning, often resulting in very complex reasoning patterns. The fact that experts and novices reason differently suggests that they might reach different conclusions (e.g., decisions or understandings) when solving medical problems. Similar patterns of reasoning have been found in other domains (Larkin et al. 1980). Due to their extensive knowledge base and the high-level inferences they make, experts typically skip steps in their reasoning.

Although experts typically use data-driven reasoning during clinical performance, this type of reasoning sometimes breaks down, and the expert must resort to hypothesis-driven reasoning. Although data-driven reasoning is highly efficient, it is often error-prone in the absence of adequate domain knowledge, since there are no built-in checks on the legitimacy of the inferences that a person makes. Pure data-driven reasoning is only successful in constrained situations, where one's knowledge of a problem can result in a complete chain of inferences from the initial problem statement to the problem solution, as illustrated in ◘ Fig. 4.9. In contrast, hypothesis-driven reasoning is slower and may make heavy demands on working memory, because one must keep track of goals and hypotheses. It is, therefore, most likely to be used when domain knowledge is inadequate, or the problem is complex. Hypothesis-driven reasoning is usually exemplary of a *weak method* of problem solving in the sense that is used in the absence of relevant prior knowledge and when there is uncertainty about problem solution. In problem-solving terms, strong methods engage knowledge, whereas weak methods refer to general strategies. Weak does not necessar-



◘ **Fig. 4.9** Diagrammatic representation of data-driven (top down) and hypothesis-driven (bottom-up) reasoning. From the presence of vitiligo, a prior history of progressive thyroid disease, and examination of the thyroid (clinical findings on the left side of figure), the physician reasons forward to conclude the diagnosis of Myxedema (right of figure). However, the anomalous finding of respiratory failure, which is inconsistent with the main diagnosis, is accounted for as a result of a hypometabolic state of the patient, in a backward-directed fashion. COND: refers to a conditional relation; CAU: indicates a causal relation; and RSLT: identifies a resultive relation

ily imply ineffectual in this context. However, hypothesis-driven reasoning may be more conducive to the novice learning experience in that it can guide the organization of knowledge (Patel et al. 1990).

In the more recent literature, described in a chapter by Patel and colleagues (2013a), two forms of human reasoning that are more widely accepted are deductive and inductive reasoning. **Deductive reasoning** is a process of reaching specific conclusions (e.g., a diagnosis) from a hypothesis or a set of hypotheses, whereas **inductive reasoning** is the process of generating possible conclusions based on available data, such as data from a patient. However, when reasoning in real-world clinical situations, it is too simplistic to think of reasoning with only these two strategies. A third form of reasoning, abductive, which combines deductive and inductive reasoning, was proposed (Peirce 1955). A physician developing and testing explanatory hypotheses based on a set of heuristics, may be considered abductive reasoning (Magnani 2001). Thus, an **abductive reasoning** process where a set of hypotheses are identified and then each of these hypotheses is evaluated on the basis of its potential consequences (Elstein et al. 1978; Ramoni et al. 1992). This makes abductive reasoning a data-driven process that relies heavily on the domain expertise of the person.

During the testing phase, hypotheses are evaluated by their ability to account for the current problem. Deduction helps in building up the consequences of each hypothesis, and this kind of reasoning is customarily regarded as a common way of evaluating diagnostic hypotheses (Joseph and Patel 1990; Kassirer 1989; Patel et al. 1994; Patel, Evans, and Kaufman 1989). All these types of inferences play different roles in the hypothesis generation and testing phases (Patel and Ramoni 1997; Peirce 1955). Our inherent ability to adapt to different kinds of knowledge domains, situations, and problems requires the use of a variety of reasoning modes, and this process describes the notion of abductive medical reasoning (Patel and Ramoni 1997). In contrast, novices and intermediate subjects (e.g., medical trainees) are more likely to employ more deliberative, effortful, and cognitively taxing forms of reasoning that can resemble hypothetico-deductive methods. As problems increase in complexity and uncertainness, expert clinicians' resort to hybrid forms of reasoning that may include substantial backward-directed reasoning.

The study of medical cognition has been summarized in a series of articles (Patel et al. 1994, 2018) and edited volumes (e.g., Evans and Patel 1989). In more recent times, medical cognition is discussed in the context of informatics and in the new field of investigation, **cognitive informatics** (Patel and Kannampallil 2015; Patel et al. 2014, 2015b, 2017). Furthermore, foundations of cognition also play a significant role in investigations of HCI, including human factors and patient safety. Details of HCI in biomedicine are covered in ▶ Chap. 5.

4.5 Human Factors Research and Patient Safety

» Human error in medicine and the adverse events which may follow are problems of psychology and engineering not of medicine "(Senders 1993)" (cited in (Woods et al. 2008).

Human factors research is a discipline devoted to the study of technology systems and how people work with them or are impacted by these technologies (Henriksen 2010). Human factors research discovers and applies information about human behavior, abilities, limitations, and other characteristics to the design of tools, machines, systems, tasks, and jobs, and environments for productive, safe, comfortable, and effective human use (Chapanis 1996). In the context of healthcare, human factors are concerned with the full complement of technologies and systems used by a diverse range of individuals including clinicians, hospital administrators, health consumers and patients (Flin and Patey 2009). Human factors work approaches the study of health practices from several perspectives or levels of analysis. A full exposition of human factors in medicine is beyond the scope of this chapter. For a detailed treatment of these

issues, the reader is referred to the Handbook of Human Factors and Ergonomics in Health Care and Patient Safety (Carayon et al. 2011). The focus in this chapter is on cognitive work in human factors and healthcare, particularly in relation to patient safety. We recognize that patient safety is a systemic challenge at multiple levels of aggregation beyond the individual. It is clear that understanding, predicting, and transforming human performance in any complex setting requires a detailed understanding of both the setting and the factors that influence performance (Woods et al. 2008).

Our objective in this section is to introduce a theoretical foundation, establish important concepts, and discuss illustrative research in **patient safety**. The field of human factors is guided by principles of engineering and applied cognitive psychology (Chapanis 1996). Human factors analysis applies knowledge about the strengths and limitations of humans to the design of interactive systems and their environment. The objective is to ensure their effectiveness, safety, and ease of use. Mental models and issues of decision making are central to human-factors analysis. Any system will be easier and less burdensome to use to the extent that it is co-extensive with users' mental models. The different dimensions of cognitive capacity, including memory, attention, and workload are central to human-factor analyses. Our perceptual system inundates us with more stimuli than our cognitive systems can process. Attentional mechanisms enable us to selectively prioritize and attend to certain stimuli and attenuate other ones. They also have the property of being sharable, which enables us to multitask by dividing our attention between two activities. For example, if we are driving on a highway, we can easily have a conversation with a passenger at the same time. However, as the skies get dark or the weather changes or suddenly you find yourself driving through winding mountainous roads, you will have to allocate more of your attentional resources to driving and less to the conversation.

Human factors research leverages theories and methods from cognitive engineering to characterize human performance in complex settings and challenging situations in aviation, industrial process control, military command control and space operations (Woods et al. 2008). The research has elucidated empirical regularities and provides explanatory concepts and models of human performance. This allows us to derive common underlying patterns in somewhat disparate settings.

4.5.1 Patient Safety

Patient safety refers to the prevention of healthcare errors, and the elimination or mitigation of patient injury caused by healthcare errors (Patel and Zhang 2007). It has been an issue of considerable concern for the past quarter-century, but the greater community was galvanized by the National Academy of Medicine report "To Err is Human," (Kohn et al. 2000) and by a follow-up report, "Improving Diagnosis in Health Care" (Balogh et al. 2015). The 2000 report communicated the surprising fact that up to 98,000 preventable deaths every single year in the United States are attributable to human error, which makes it the 8th leading cause of death in this country. Although one may argue over the specific numbers, there is no disputing that too many patients are harmed or die every year as a result of human actions or absence of action.

We can only analyze errors after they happened, and they often seem to be glaring blunders after the fact. This leads to the assignment of blame or searches for a single cause of the error. However, in hindsight, it is exceedingly difficult to recreate the situational context, stress, shifting attention demands, and competing goals that characterized a situation prior to the occurrence of an error. This sort of retrospective analysis is subject to hindsight bias. **Hindsight bias** masks the dilemmas, uncertainties, demands, and other latent conditions that were operative before the mishap. Too often the term 'human error' connotes blame and a search for the guilty culprits, suggesting some sort of human deficiency or irresponsible behavior. Human factors researchers recognized that this approach error is inherently incomplete and poten-

tially misleading. They argue for the need for a more comprehensive systems-centered approach that recognizes that error could be attributed to a multitude of factors as well as the interaction of these factors. Error is the failure of a planned sequence of mental or physical activities to achieve its intended outcome when these failures cannot be attributed to chance (Patel and Zhang 2007; Reason 1990). Reason (1990) introduced an important distinction between **latent** and **active failures**. Active failure represents the face of error. The effects of active failure are immediately felt. In healthcare, active errors are committed by providers such as nurses, physicians, or pharmacists who are actively responding to patient needs at the "sharp end". The latent conditions are less visible but equally important. Latent conditions are enduring systemic problems that may not be evident for some time, combine with other system problems to weaken the system's defenses and make errors possible. There is a lengthy list of potential latent conditions including poor interface design of important technologies, communication breakdown between key actors, gaps in supervision, inadequate training, and absence of a safety culture in the workplace—a culture that emphasizes safe practices and the reporting of any conditions that are potentially dangerous.

Zhang, Patel, Johnson, and Shortliffe (2004) developed a taxonomy of errors partially based on the distinctions proposed by Reason (1990). They further classified errors in terms of **slips** and **mistakes** (Reason 1990). A slip occurs when the actor selected the appropriate course of action, but it was executed inappropriately. A mistake involves an inappropriate course of action reflecting an erroneous judgment or inference (e.g., a wrong diagnosis or misreading of an x-ray). Mistakes may either be knowledge-based owing to factors such as incorrect knowledge, or they may be rule-based, in which case the correct knowledge was available, but there was a problem in applying the rules or guidelines. They further characterize medical errors as a progression of events. There is a period when everything is operating smoothly. Then an unsafe practice unfolds resulting in a kind of error, but not necessarily leading to an adverse event. For example, if there is a system of checks and balances that is part of routine practice or if there is a systematic supervisory process in place, the vast majority of errors will be trapped and defused in this middle zone. If these measures or practices are not in place, an error can propagate and cross the boundary to become an adverse event. At this point, the patient has been harmed. In addition, if an individual is subject to a heavy workload or intense time pressure, then that will increase the potential for an error, resulting in an adverse event.

The notion that human error should not be tolerated is prevalent in both the public and personal perception of the performance of most clinicians. However, researchers in other safety-critical domains have long since abandoned the quest for zero defect, citing it as an impractical goal, and choosing to focus instead on the development of strategies to enhance the ability to recover from error (Morel et al. 2008). Patel and her colleagues conducted empirical investigations into error detection and recovery by experts (attending physicians) and non-experts (resident trainees) in the critical care domain, using both laboratory-based and naturalistic approaches (Patel et al. 2011). These studies show that expertise is more closely tied to the ability to detect and recover from errors and not so much to the ability not to make errors. The study results show that both the experts and non-experts are prone to commit and recover from errors, but experts' ability to detect and recover from knowledge-based errors is better than that of trainees. Error detection and correction in complex real-time critical care situations appears to induce certain urgency for quick action in a high alert condition, resulting in rapid detection and correction. Studies on expertise and understanding of the limits and failures of human decision-making are important if we are to build robust decision-support systems to manage the boundaries of risk of error in decision making (Patel et al. 2015a; Patel and Cohen 2008). Research on situational complexity and medical errors is documented in a recent book by Patel, Kaufman, and Cohen (2014).

4.5.2 Unintended Consequences

It is widely believed that health information technologies have the potential to transform healthcare in a multitude of ways, including the reduction of errors. However, it is increasingly apparent that technology-induced errors are deeply consequential and have had deleterious consequences for patient safety.

There is evidence to suggest that a poorly designed user interface can present substantial challenges even for the well-trained and highly skilled user (Zhang et al. 2003). Lin et al. (1998) conducted a series of studies on a patient-controlled analgesic or PCA device, a method of pain relief that uses disposable or electronic infusion devices and allows patients to self-administer analgesic drugs as required. Lin and colleagues investigated the effects of two interfaces to a commonly used PCA device, including the original interface. Based on **cognitive task analysis**, they redesigned the original interface so that it was more in line with sound human factors principles. Based on the cognitive task analysis, they found the existing PCA interface to be problematic in several different ways. For example, the structure of many subtasks in the programming sequence was unnecessarily complex. There was a lack of information available on the screen to provide meaningful feedback and to structure the user experience (e.g., negotiating the next steps). For example, a nurse would not know that he or she was on the third of five screens or when they were half way through the task. Based on the CTA analysis, Lin et al. (1998) also redesigned the interface according to sound human factors principles and demonstrated significant improvements in efficiency, error rate, and reported workload.

Zhang and colleagues employed a modified heuristic evaluation method (see ▶ Sect. 4.5, above) to test the safety of two infusion pumps (Zhang et al. 2003). Based on an analysis by four evaluators, a total of 192 violations with the user interface design were documented. Consistency and visibility (the ease in which a user can discern the system state) were the most widely documented violations. Several of the violations were classified as problems of substantial severity. Their results suggested that one of the two pumps were likely to induce more medical errors than the other ones.

It is clear that usability problems are consequential and have the potential to impact patient safety. Kushniruk et al. (2005) examined the relationship between particular kinds of usability problems and errors in a handheld prescription writing application. They found that particular usability problems were associated with the occurrence of an error in entering the medication. For example, the problem of inappropriate default values automatically populating the screen was found to be correlated with errors in entering the wrong dosages of medications. In addition, certain types of errors were associated with mistakes (not detected by users) while others were associated with slips about unintentional errors. Horsky et al. (2005) analyzed a problematic medication order placed using a CPOE system that resulted in an overdose of potassium chloride being administered to an actual patient. The authors used a range of investigative methods including inspection of system logs, semi-structured interviews, the examination of the electronic health record, and cognitive evaluation of the order entry system involved. They found that the error was due to a confluence of factors including problems associated with the display, the labeling of functions, and ambiguous dating of the dates in which medication was administered. The poor interface design did not assist with the decision-making process, and in fact, its design served as a hindrance, where the interface was a poor fit for the *conceptual operators* utilized by clinicians when calculating medication dosage (i.e., based on volume, not duration).

Koppel et al. (2005) published an influential study examining how computer-provider order-entry systems (CPOE) facilitated medical errors. The study, which was published in JAMA (Journal of the American Medical Association), used a series of methods including interviews with clinicians, observations, and a survey to document the range of errors. According to the authors, the system facilitated 22 types of medication errors, and many

of them occurred with some frequency. The errors were classified into two broad categories: (1) information errors generated by fragmentation of data and failure to integrate the hospital's information systems and (2) human-machine interface flaws reflecting machine rules that do not correspond to work organization or usual behaviors.

The growing body of research on unintended consequences spurred the American Medical Informatics Association to devote a policy meeting to consider ways to understand and diminish their impact (Bloomrosen et al. 2011). The matter is especially pressing given the increased implementation of health information technologies nationwide, including ambulatory care practices that have little experience with health information technologies. The authors outline a series of recommendations, including a need for more cognitively-oriented research to guide the study of the causes and mitigation of unintended consequences resulting from health information technology implementations. These changes could facilitate improved management of those consequences, resulting in enhanced performance, patient safety, as well as greater user acceptance.

4.5.3 Distributed Cognition and Electronic Health Records

In this chapter, we have considered a classical model of information-processing cognition in which mental representations mediate all activity and constitute the central units of analysis. The analysis emphasizes how an individual formulates internal representations of the external world. To illustrate the point, imagine an expert user of a word processor who can effortlessly negotiate tasks through a combination of key commands and menu selections. The traditional cognitive analysis might account for this skill by suggesting that the user has formed an image or schema of the layout structure of each of eight menus, and retrieves this information from memory each time an action is to be performed. For example, if the goal is to "insert a clip art icon," the user would recall that this is subsumed under pictures that are the ninth item on the "Insert" menu and then execute the action, thereby achieving the goal. However, there are some problems with this model. Mayes, Draper, McGregor, and Koatley (1988) demonstrated that even highly skilled users could not recall the names of menu headers, yet they could routinely make fast and accurate menu selections. The results indicate that many or even most users relied on cues in the display to trigger the right menu selections. This suggests that the display can have a central role in controlling interaction in graphical user interfaces.

As discussed, the conventional information-processing approach has come under criticism for its narrow focus on the rational/cognitive processes of the solitary individual. In the previous section, we considered the relevance of external representations to cognitive activity. The emerging perspective of **distributed cognition** offers a more far-reaching alternative. The distributed view of cognition represents a shift in the study of cognition from being the sole property of the individual to being "stretched" across groups, material artifacts, and cultures (Hutchins 1995; Suchman 1987). This viewpoint is increasingly gaining acceptance in cognitive science and human-computer interaction research. In the distributed approach to HCI research, cognition is viewed as a process of coordinating distributed internal (i.e., knowledge) and external representations (e.g., visual displays, manuals). Distributed cognition has two central points of inquiry, one that emphasizes the inherently social and collaborative nature of cognition (e.g., doctors, nurses and technical support staff in neonatal care unit jointly contributing to a decision process), and one that characterizes the mediating effects of technology or other artifacts on cognition.

The mediating role of technology can be evaluated at several levels of analysis from the individual to the organization. Technologies, whether they be computer-based or an artifact in another medium, transform the ways individuals and groups think. They do not merely augment, enhance, or expedite perfor-

4

mance, although a given technology may do all of these things. The difference is not merely one of quantitative change, but one that is qualitative in nature.

In a distributed world, what becomes of the individual? We believe it is important to understand how technologies promote enduring changes in individuals. Salomon, Perkins and Globerson (1991) introduced an important distinction in considering the mediating role of technology on individual performance, the effects with technology and the effects of technology. The former is concerned with the changes in performance displayed by users while equipped with the technology. For example, when using an effective medical information system, physicians should be able to gather information more systematically and efficiently. In this capacity, medical information technologies may alleviate some of the cognitive load associated with a given task and permit physicians to focus on higher-order thinking skills, such as diagnostic hypothesis generation and evaluation. The effects of technology refer to enduring changes in general cognitive capacities (knowledge and skills) as a consequence of interaction with a technology. This effect is illustrated subsequently in the context of the enduring effects of an EHR (see ▶ Chap. 10).

We employed a pen-based EHR system, DCI (Dossier of Clinical Information), in several of our studies (see Kushniruk et al. 1996). Using the pen or computer keyboard, physicians can directly enter information into the EHR, such as the patient's chief complaint, past history, history of present illness, laboratory tests, and differential diagnoses. Physicians were encouraged to use the system while collecting data from patients (e.g., during the interview). The system allows the physician to record information about the patient's differential diagnosis, the ordering of tests, and the prescription of medication.

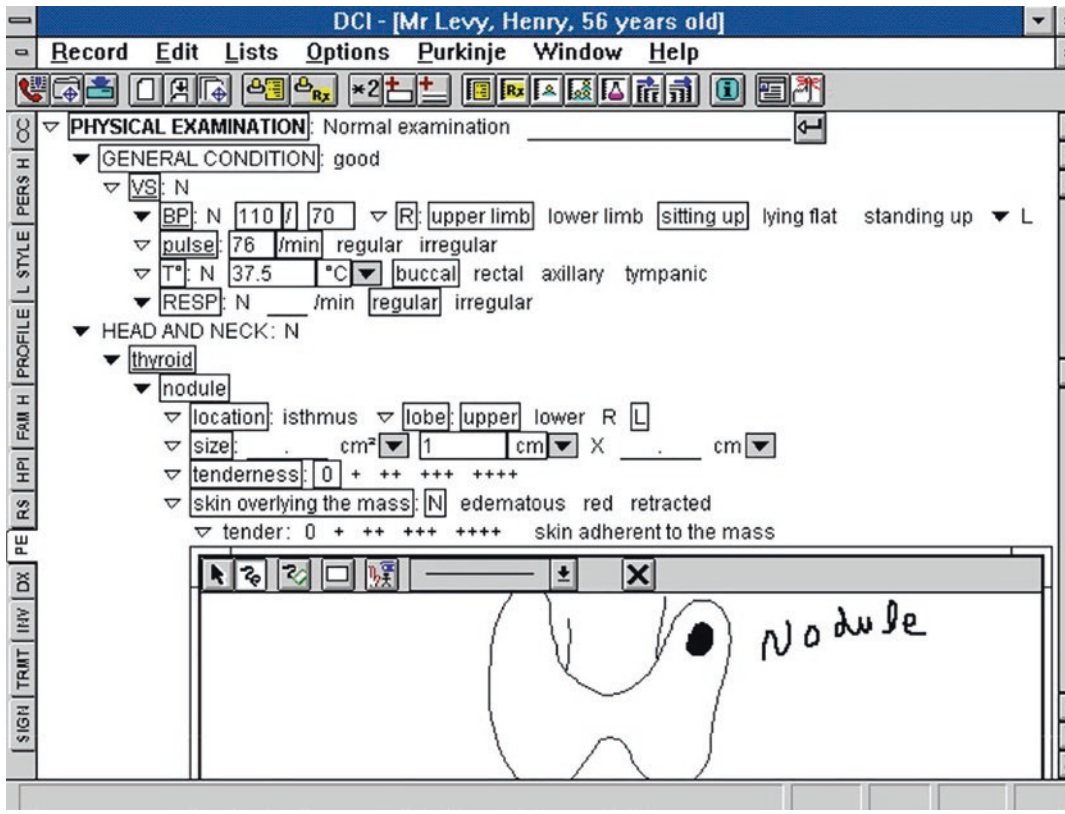


Fig. 4.10 Display of a structured electronic medical record with graphical capabilities

The graphical interface provides a highly structured set of resources for representing a clinical problem, as illustrated in Fig. 4.10.

We have studied the use of this EHR in both laboratory-based research (Kushniruk et al. 1996) and actual clinical settings using cognitive methods (Patel et al. 2000). The laboratory research included a simulated doctor-patient interview. We have observed two distinct patterns of EHR usage in the interactive condition, one in which the subject pursues information from the patient predicated on a hypothesis; the second strategy involves the use of the EHR display to guide asking the patient questions. In the screen-driven strategy, the clinician is using the structured list of findings in the order in which they appear on the display to elicit information from the patient. All experienced users of this system appear to have both strategies in their repertoire.

In general, a screen-driven strategy can enhance performance by reducing the cognitive load imposed by information-gathering goals and allow the physician to allocate more cognitive resources toward testing hypotheses and rendering decisions. On the other hand, this strategy can encourage a certain sense of complacency. We observed both effective as well as counter-productive uses of this screen-driven strategy. A more experienced user consciously used the strategy to structure the information-gathering process, whereas a novice user used it less discriminately. In employing this screen-driven strategy, the novice elicited almost all of the relevant findings in a simulated patient encounter. However, she also elicited numerous irrelevant findings and pursued incorrect hypotheses. In this particular case, the subject became too reliant on the technology and had difficulty imposing her own set of working hypotheses to guide the information-gathering and diagnostic-reasoning processes.

The use of a screen-driven strategy is evidence of how technology transforms clinical cognition, as manifested in clinicians' patterns of reasoning. Patel et al. (2000) extended this line of research to study the cognitive consequences of using the same EHR system in a diabetes clinic. The study considered the following questions (1) How do physicians manage information flow when using an EHR system? (2) What are the differences in the way physicians organize and represent this information using paper-based and EHR systems, and (3) Are there long-term, enduring effects of the use of EHR systems on knowledge representations and clinical reasoning? One study focused on an in-depth characterization of changes in knowledge organization in a single subject as a function of using the system. The study first compared the contents and structure of patient records produced by the physician using the EHR system and paper-based patient records, using ten pairs of records matched for variables such as patient age and problem type. After having used the system for six months, the physician was asked to conduct his/her next five patient interviews using only hand-written paper records.

The results indicated that the EHRs contained more information relevant to the diagnostic hypotheses. In addition, the structure and content of information were found to correspond to the structured representation of the particular medium. For example, EHRs were found to contain more information about the patient's past medical history, reflecting the query structure of the interface. The paper-based records appear to better preserve the integrity of the time course of the evolution of the patient problem, whereas, this is notably absent from the EHR. Perhaps, the most striking finding is that, after having used the system for six months, the structure and content of the physician's paper-based records bore a closer resemblance to the organization of information in the EHR than the paper-based records produced by the physician prior to exposure to the system. This finding is consistent with the enduring *effects of* technology even in the absence of the particular system (Salomon et al. 1991). The authors conclude that given these potentially enduring effects, the use of a particular EHR will almost certainly have a direct effect on medical decision making.

The previously discussed research demonstrates how information technologies can

mediate cognition and even produce enduring changes in how one performs a task. What dimensions of an interface contribute to such changes? What aspects of a display are more likely to facilitate efficient task performance, and what aspects are more likely to impede it? Norman (1986) argued that well-designed artifacts could reduce the need for users to remember large amounts of information, whereas poorly designed artifacts increased the knowledge demands on the user and the burden of working memory. In the distributed approach to HCI research, cognition is viewed as a process of coordinating distributed internal and external representations, and this, in effect, constitutes an indivisible information-processing system.

One of the appealing features of the distributed cognition paradigm is that it can be used to understand how properties of objects on the screen (e.g., links, buttons) can serve as external representations and reduce cognitive load. The distributed resource model proposed by Wright, Fields, and Harrison (2000) addresses the question of "what information is required to carry out some task and where should it be located: as an interface object or as something that is mentally represented to the user." The relative difference in the distribution of representations (internal and external) is central to determining the efficacy of a system designed to support a complex task. Wright, Fields, and Harrison (2000) were among the first to develop an explicit model for coding the kinds of resources available in the environment and how they are embodied on an interface.

Horsky, Kaufman, and Patel (2003a, b) applied the distributed resource model and analysis to a provider order entry system. The goal was to analyze specific order-entry tasks such as those involved in admitting a patient to a hospital and then to identify areas of complexity that may impede optimal recorded entries. The research consisted of two-component analyses: a cognitive walk-through evaluation that was modified based on the distributed resource model and a simulated clinical ordering task performed by seven physicians. The CW analysis revealed that the configuration of resources (e.g., very long menus, complexly configured displays) placed unnecessarily heavy cognitive demands on users, especially those who were new to the system. The resources model was also used to account for patterns of errors produced by clinicians. The authors concluded that the redistribution and reconfiguration of resources might yield guiding principles and design solutions in the development of complex interactive systems.

The distributed cognition framework has proved to be particularly useful in understanding the performance of teams or groups of individuals in a particular work setting (Hutchins 1995). Hazlehurst and colleagues (Hazlehurst et al. 2003, 2007) have drawn on this framework to illuminate how work in healthcare settings is constituted using shared resources and representations. The *activity system* is the primary explanatory construct. It is comprised of actors and tools, together with shared understandings among actors that structure interactions in a work setting. The "propagation of representational states through activity systems" is used to explain cognitive behavior and investigate the organization of the system and human performance. Following Hazlehurst et al. (2007, p. 540), "a **representational state** is a particular configuration of an information-bearing structure, such as a monitor display, a verbal utterance, or a printed label, that plays some functional role in a process within the system." The author has used the concept to explain the process of medication ordering in an intensive care unit and the coordinated communications of a surgical team in a heart room.

The framework for distributed cognition is still an emerging one in human-computer interaction. It offers a novel and potentially powerful approach for illuminating the kinds of difficulties users encounter and finding ways to better structure the interaction by redistributing the resources. Distributed cognition analyses may also provide a window into why technologies sometimes fail to reduce errors or even contribute to them.

4.6 Conclusion

Theories and methods from cognitive science can shed light on a range of issues about the design and implementation of health information technologies. They can also serve an instrumental role in understanding and enhancing the performance of clinicians and patients as they engage in a range of cognitive tasks related to health. We believe that fundamental studies in psychology and cognitive science in general, can provide general guiding principles to study these issues, and can be combined with field studies which serve to illuminate different facets and contextualize the phenomena observed in laboratory studies. The potential scope of applied cognitive research in biomedical informatics is very broad. Significant inroads have been made in areas such as EHRs and patient safety. However, there are promising areas of future cognitive research that remain largely uncharted. These include understanding how to capitalize on health information technology without compromising patient safety (particularly in providing adequate decision support), understanding how various visual representations/graphical forms mediate reasoning in biomedical informatics and how these representations can be used by patients and health consumers with varying degrees of literacy. These are only a few of the cognitive challenges related to harnessing the potential of cutting-edge technologies to improve patient safety.

Suggested Readings

Anderson, J. R. (2015). *Cognitive psychology and its implications*. New York: Worth Publishers.

Carayon, P., Alyousef, B., & Xie, A. (2012). Human factors and ergonomics in health care. In Handbook of human factors and ergonomics (pp. 1574–1595).

Patel, V. L., Kaufman, D. R., & Kannampallil, T. G. (2013c). Diagnostic reasoning and decision making in the context of health information technology. In D. Marrow (Ed.), Reviews of human factors and ergonomics (Vol. 8). Thousand Oaks, CA: SAGE Publications.

Patel, V. L., Kaufman, D. R., & Arocha, J. F. (2002). Emerging paradigms of cognition in medical decision-making. Journal of Biomedical Informatics, 35, 52–75. This relatively recent article summarizes new directions in decision-making research. The authors articulate a need for alternative paradigms for the study of medical decision making.

Patel, V. L., Yoskowitz, N. A., Arocha, J. F., & Shortliffe, E. H. (2009). Cognitive and learning sciences in biomedical and health instructional design: A review with lessons for biomedical informatics education. Journal of Biomedical Informatics, 42(1), 176–197. A review of learning and cognition with a particular focus on biomedical informatics.

Patel, V. L., Kannampallil, T. G., & Shortliffe, E. H. (2015c). Role of cognition in generating and mitigating clinical errors. BMJ Quality & Safety, 24, 468–474. https://doi.org/10.1136/bmjqs-2014-003482.

Middleton, B., Bloomrosen, M., Dente, M. A., Hashmat, B., Koppel, R., Overhage, J. M., et al. (2013). Enhancing patient safety and quality of care by improving the usability of electronic health record systems: Recommendations from AMIA. Journal of the American Medical Informatics Association, 20(e1), e2–e8.

Questions for Discussion

1. How can cognitive science theory meaningfully inform and shape design, development, and assessment of health-care information systems?
2. Describe two or three kinds of mental representations and briefly characterize their significance in understanding human performance.
3. What is the purpose and value of cognitive architectures?
4. Identify three ways in which novices differ from experts in medicine.
5. What are the limitations of interpreting retroactive data on medical errors?
6. Explain the difference between latent and active failures and their implications for patient safety?
7. How does the field of Cognitive Informatics capture the interaction of cognition and informatics in biomedicine and healthcare?

8. Explain the role inductive, deductive and abductive reasoning play in medical diagnostic reasoning?
9. Explain some ways in which technology-mediated errors can compromise patient safety.
10. What are some of the assumptions of the distributed cognition framework? What implications does this approach have for the evaluation of electronic health records?
11. Explain the difference between the *effects of technology* and the *effects with technology*? How can each of these effects contribute to improving patient safety and reducing medical error?
12. The use of electronic health records (EHR) has been shown to differentially affect clinical reasoning relative to paper charts. Briefly characterize the effects they have on reasoning, including those that persist after the clinician ceases to use the system.

References

Akin, O. (1982). *The psychology of architecture design*. London: Pion.

Allard, F., & Starkes, J. L. (1991). Motor-skill experts in sports, dance, and other domains. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 126–150). New York: Cambridge University Press.

Anderson, J. R. (1985). *Cognitive psychology and its implications* (2nd ed.). New York: Freeman.

Anderson, J. R. (2013). *The architecture of cognition*. New York: Psychology Press.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). New York: Academic.

Balogh, E. P., Miller, B. T., & Ball, J. R. (2015). *Improving diagnosis in health care*. Washington, DC: National Academy Press.

Bartolomeo, P. (2008). The neural correlates of visual mental imagery: An ongoing debate. *Cortex, 44*(2), 107–108. S0010-9452(07)00016-0 [pii].

Bechtel, W., Abrahamsen, A., & Graham, G. (1998). Part I: The life of cognitive science. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science Blackwell companions to philosophy* (Vol. 13, pp. 2–104). Malden: Blackwell.

Bloomrosen, M., Starren, J., Lorenzi, N. M., Ash, J. S., Patel, V. L., & Shortliffe, E. H. (2011). Anticipating and addressing the unintended consequences of health IT and policy: A report from the AMIA 2009 health policy meeting. *Journal of American Medical Informatics Association: JAMIA, 18*(1), 82–90. http://doi.org/18/1/82 [pii]10.1136/jamia.2010.007567.

Bruer, J. T. (1993). *Schools for thought: A science of learning in the classroom*. Cambridge: MIT Press.

Carayon, P., Karsh, B.-T., Gurses, A. P., Holden, R. J., Hoonakker, P., Hundt, A. S., Wetterneck, T. (2011). Macroergonomics in patient care and health care safety. In D. G. Morrow (Ed.), Reviews of human factors and ergonomics (Vol. 8). Santa Monica, CA: Human Factors and Ergonomics Society.

Carayon, P. (2012). *Handbook of human factors and ergonomics in health care and patient safety* (2nd ed.). Boca Raton: CRC Press.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale: L. Erlbaum Associates.

Carroll, J. M. (2003). *HCI models, theories, and frameworks: Toward a multidisciplinary science*. San Francisco: Morgan Kaufmann.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.

Chapanis, A. (1996). *Human factors in systems engineering*. New York: Wiley.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55–81.

Chi, M. T. H., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale: L. Erlbaum Associates.

Clancey, W. J., & Shortliffe, E. H. (1984). *Readings in medical artificial intelligence: The first decade*. Reading: Addison-Wesley.

DeGroot, A. T. (1965). *Thought and choice in chess*. The Hague: Mouton.

Duch, W., Oentaryo, R. J., & Pasquier, M. (2008). *Cognitive architectures: Where do we go from here?*. Paper presented at the proceeding of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference.

Elstein, K. A., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Ericsson, K. A. (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences sports and games*. Mahwah: Lawrence Erlbaum Associates.

Ericsson, K. A. (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge/New York: Cambridge University Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. New York: Cambridge University Press.

Ericsson, K. A. (2009). Enhancing the development of professional performance: Implications from the study of deliberate practice. In K. A. Ericsson (Ed.), The development of professional expertise: Toward measurement of expert performance and design of optimal learning environments (pp. 405–431). New York: Cambridge University Press.

Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (Eds.). (2018). *The Cambridge handbook of expertise and expert performance*. UK: Cambridge University Press.

Ericsson, K. A. (Ed.). (1996). The road to excellence: The acquisition of expert performance in the arts and sciences, sports. Hillsdale, NJ: Lawrence Erlbaum.

Estes, W. K. (1975). The state of the field: General problems and issues of theory and metatheory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 1). Hillsdale/New York: L. Erlbaum Associates.

Evans, D. A., & Gadd, C. S. (1989). Managing coherence and context in medical problem-solving discourse. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 211–255). Cambridge, MA: MIT Press.

Evans, D. A., & Patel, V. L. (1989). *Cognitive science in medicine*. Cambridge, MA: MIT Press.

Feltovich, P. J., Johnson, P. E., Moller, J. H., & Swanson, D. B. (1984). The role and development of medical knowledge in diagnostic expertise. In W. J. Clancey & E. H. Shortliffe (Eds.), *Readings in medical artificial intelligence: The first decade* (pp. 275–319). Reading: Addison Wesley.

Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2009). *Designing for older adults: Principles and creative human factors approaches*. Boca Raton: CRC Press.

Flin, R., & Patey, R. (2009). Improving patient safety through training in non-technical skills. *British Medical Journal, 339*, B3595. https://doi.org/10.1136/Bmj.B3595.

Frederiksen, C. H. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology, 7*(3), 371–458.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Gillan, D. J., & Schvaneveldt, R. W. (1999). Applying cognitive psychology: Bridging the gulf between basic research and cognitive artifacts. In F. T. Durso, R. Nickerson, R. Schvaneveldt, S. Dumais, M. Chi, & S. Lindsay (Eds.), *Handbook of applied cognition* (pp. 3–31). Chichester/New York: Wiley.

Glaser, R. (Ed.). (2000). *Advances in instructional psychology: Education design and cognitive science* (Vol. 5). Mahwah: Lawrence Erlbaum and Associates.

Greeno, J. G., & Simon, H. A. (1988). Problem solving and reasoning. In R. C. Atkinson & R. J. Herrnstein (Eds.), *Stevens' handbook of experimental psychology Vol 1: Perception and motivation; Vol 2: Learning and cognition* (Vol. 1, 2nd ed., pp. 589–672). New York: Wiley.

Harrington, L. (2015). Electronic health record workflow: Why more work than flow? *Advanced Critical Care AACN, 26*(1), 5–9.

Hazlehurst, B., McMullen, C., Gorman, P., & Sittig, D. (2003). How the ICU follows orders: Care delivery as a complex activity system. *AMIA Annual Symposium Proceedings, 2003*, 284–288. D030003599 [pii].

Hazlehurst, B., McMullen, C. K., & Gorman, P. N. (2007). Distributed cognition in the heart room: How situation awareness arises from coordinated communications during cardiac surgery. *Journal of Biomedical Informatics, 40*(5), 539–551. https://doi.org/10.1016/j.jbi.2007.02.001. S1532-0464(07)00008-1 [pii].

Henriksen, K. (2010). Partial truths in the pursuit of patient safety. *BMJ Quality & Safety Health Care, 19*(3), i3–i7.

Hilgard, E. R., & Bower, G. H. (1975). *Theories of learning* (4th ed.). Englewood Cliffs: Prentice-Hall.

Hoffman, R. R. (Ed.). (1992). *The psychology of expertise: Cognitive research and empirical AI*. Mahwah: Lawrence Erlbaum Associates.

Hoffman, R. R., Shadbolt, N. R., Burton, A. M., & Klein, G. (1995). Eliciting knowledge from experts – A methodological analysis. *Organizational Behavior and Human Decision Processes, 62*(2), 129–158.

Horsky, J., Kaufman, D. R., & Patel, V. L. (2003a). The cognitive complexity of a provider order entry interface. *AMIA Annual Symposium Proceedings*, 2013, 294–298. PMID: 14728181; PMCID: PMC1480200.

Horsky, J., Kaufman, D. R., Oppenheim, M. I., & Patel, V. L. (2003b). A framework for analyzing the cognitive complexity of computer-assisted clinical ordering. *Journal of Biomedical Informatics, 36*(1–2), 4–22. https://doi.org/10.1016/S1532-0464(03)00062-5.

Horsky, J., Kuperman, G. J., & Patel, V. L. (2005). Comprehensive analysis of a medication dosing error related to CPOE. *Journal of the American Medical Informatics Association: Journal of the American Medical Informatics Association JAMIA, 12*(4), 377–382. https://doi.org/10.1197/jamia.M1740. M1740 [pii].

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Joseph, G. M., & Patel, V. L. (1990). Domain knowledge and hypothesis generation in diagnostic reasoning. *Medical Decision Making, 10*, 31–46.

Karsh, B., Weinger, M. B., Abbott, P. A., & Wears, R. L. (2010). Health information technology: Fallacies and sober realities. *Journal of the American Medical Informatics Association, 17*(6), 617–623.

Kassirer, J. P. (1989). Diagnostic reasoning. *Annals of Internal Medicine, 110*, 893–900.

Kaufman, D. R., Patel, V. L., & Magder, S. (1996). The explanatory role of spontaneously generated analo-

gies in a reasoning about physiological concepts. *International Journal of Science Education, 18*, 369–386.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge/New York: Cambridge University Press.

Klienmuntz, B., & McLean, R. S. (1968). Diagnostic interviewing by digital computer. *Behavioral Science, 13*(1), 75–80.

Kohn, L. T., Corrigan, J., & Donaldson, M. S. (2000). *To err is human: Building a safer health system* (Vol. 6). Washington, DC: National Academy Press.

Koppel, R., Metlay, J. P., Cohen, A., Abaluck, B., Localio, A. R., Kimmel, S. E., & Strom, B. L. (2005). Role of computerized physician order entry systems in facilitating medication errors. *Journal of the American Medical Association JAMA, 293*(10), 1197–1203.

Kushniruk, A. W., Kaufman, D. R., Patel, V. L., Levesque, Y., & Lottin, P. (1996). Assessment of a computerized patient record system: A cognitive approach to evaluating medical technology. *MD Computing, 13*(5), 406–415.

Kushniruk, A. W., Triola, M. M., Borycki, E. M., Stein, B., & Kannry, J. L. (2005). Technology induced error and usability: The relationship between usability problems and prescription errors when using a handheld application. *International Journal of Medical Informatics, 74*(7–8), 519–526. https://doi.org/10.1016/j.ijmedinf.2005.01.003. S1386-5056(05)00011-0 [pii].

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208*(4450), 1335–1342.

Ledley, R. S., & Lusted, L. B. (1959). Probability, logic and medical diagnosis. *Science, 130*(3380), 892–930.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale: Lawrence Erlbaum Associates.

Lieto, A., Lebiere, C., & Oltramari, A. (2018). The knowledge level in cognitive architectures: Current limitations and possible developments. *Cognitive Systems Research, 48*, 39–55.

Lin, L., Isla, R., Doniz, K., Harkness, H., Vicente, K. J., & Doyle, D. J. (1998). Applying human factors to the design of medical equipment: Patient-controlled analgesia. *Journal of Clinical Monitoring and Computing, 14*(4), 253–263.

Magnani, L. (2001). *Abduction, reason, and science: Processes of discovery and explanation*. Dordrecht: Kluwer Academic.

Mayes, T. J., Draper, S. W., McGregor, A. M., & Koatley, K. (1988). *Information flow in a user interface*: *The effect of experience and context on the recall of MacWrite screens*. Paper presented at the conference on people and computers IV, Cambridge.

Morel, G., Amalberti, R., & Chauvin, C. (2008). Articulating the differences between safety and resilience: The decision-making process of professional sea-fishing skippers. *Human Factors, 50*(1), 1–16.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction* (pp. 31–61). Hillsdale: Lawrence Erlbaum Associates.

Norman, D. A. (1993). Things that make us smart: defending human attributes in the age of the machine. Reading, Mass.: Addison-Wesley Pub. Co.

Patel, V. L., & Cohen, T. (2008). New perspectives on error in critical care. *Current Opinions in Critical Care, 14*(4), 456–459.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10*(1), 91–116.

Patel, V. L., & Groen, G. J. (1991a). Developmental accounts of the transition from medical student to doctor: Some problems and suggestions. *Medical Education, 25*(6), 527–535.

Patel, V. L., & Groen, G. J. (1991b). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93–125). New York: Cambridge University Press.

Patel, V. L., & Kannampallil, T. G. (2015). Cognitive informatics in biomedicine and healthcare. *Journal of Biomedical Informatics, 53*, 3–14.

Patel, V. L., & Kaufman, D. R. (1998). Medical informatics and the science of cognition. *Journal of the American Medical Informatics Association JAMIA, 5*(6), 493–502.

Patel, V. L., & Ramoni, M. F. (1997). Cognitive models of directional inference in expert medical reasoning. In P. J. Feltovich, K. M. Ford, & R. R. Hoffman (Eds.), *Expertise in context: Human and machine* (pp. 67–99). Cambridge: The MIT Press.

Patel, V. L., & Zhang, J. (2007). Cognition and patient safety in healthcare. In F. T. Durso, R. S. Nickerson, S. Dumais, S. Lewandowsky, & T. Perfect (Eds.), *Handbook of applied cognition* (2nd ed., pp. 307–331). New York: Wiley.

Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task-difficulty. *Memory and Cognition, 18*(4), 394–406.

Patel, V. L., Arocha, J. F., & Kaufman, D. R. (1994). Diagnostic reasoning and medical expertise. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 31, pp. 187–252). San Diego: Academic Press.

Patel, V. L., Kushniruk, A. W., Yang, S., & Yale, J. F. (2000). Impact of a computer-based patient record system on data collection, knowledge organization, and reasoning. *Journal of the American Medical Informatics Association JAMIA, 7*(6), 569–585.

Patel, V. L., Cohen, T., Murarka, T., Olsen, J., Kagita, S., Myneni, S., et al. (2011). Recovery at the edge of error: Debunking the myth of the infallible expert. *Journal of Biomedical Informatics, 44*(3), 413–424.

Patel, V. L., Kaufman, D. R., & Kannampallil, T. G. (2013a). Diagnostic reasoning and decision making in the context of health information technology. In D. Marrow (Ed.), *Reviews of human factors and ergonomics* (Vol. 8). Thousand Oaks: SAGE Publications.

Patel, V. L., Kaufman, D. R., & Kannampallil, T. G. (2013b). Diagnostic reasoning and decision making in the context of health information technology. *Reviews of Human Factors and Ergonomics, 8*(1), 149–190.

Patel, V. L., Kaufman, D. R., & Cohen, T. (2014). *Cognitive informatics in health and biomedicine: Case studies on critical care, complexity and errors* (pp. 1–13). London: Springer.

Patel, V. L., Kannampallil, T. G., & Shortliffe, E. H. (2015a). Role of cognition in generating and mitigating clinical errors. *BMJ Quality and Safety, 24*, 468–474.

Patel, V. L., Kannampallil, T. G., & Kaufman, D. R. (Eds.). (2015b). *Cognitive informatics in health and biomedicine: Human computer interaction*. London: Springer.

Patel, V. L., Arocha, J. F., & Ancker, J. (2017). *Cognitive informatics in health and biomedicine: Understanding and modeling health behaviors*. London: Springer.

Patel, V. L., Kaufman, D. R., & Kannampallil, T. G. (2018). Diagnostic reasoning and expertise in healthcare. In P. Ward, J. M. Schraagen, J. Gore, & E. Roth (Eds.), *The Oxford handbook of expertise: Research & application*. UK: Oxford University Press.

Patel, V. L., Evans, D. A., & Kaufman, D. R. (1989). Cognitive framework for doctor-patient communication. In D. A. Evans & V. L. Patel (Eds.), Cognitive science in medicine: Biomedical modeling (pp. 257–312). Cambridge, MA: MIT Press..

Payne, S. J. (2003). Users' mental models: The very ideas. In J. M. Carroll (Ed.), *HCI models, theories, and frameworks: Toward a multidisciplinary science* (1st ed., pp. 135–156). San Francisco: Morgan Kaufmann.

Peirce, C. S. (1955). Philosophical writings of Peirce. Ed. by Justus Buchler. New York: Dover.

Peleg, M., Gutnik, L. A., Snow, V., & Patel, V. L. (2006). Interpreting procedures from descriptive guidelines. *Journal of Biomedical Informatics, 39*, 184–195.

Perkins, D. N., Schwartz, S., & Simmons, R. (1990). A view from programming. In M. Smith (Ed.), *Toward a unified theory of problem solving: Views from content domains* (pp. 45–67). Hillsdale: Lawrence Erlbaum Associates.

Ramoni, M., Stefanelli, M., Magnani, L., & Barosi, G. (1992). An epistemological framework for medical knowledge-based systems. *IEEE Transactions on Systems, Man, and Cybernetics, 22*(6), 1361–1375.

Reason, J. T. (1990). *Human error*. Cambridge/New York: Cambridge University Press.

Rimoldi, H. J. (1961). The test of diagnostic skills. Journal of Medical Education, 36, 73–79.

Ritter, F. E., Tehranchi, F., & Oury, J. D. (2019). ACT-R: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(3), e1488.

Rogers, Y. (2004). New theoretical approaches for HCI. *Annual Review of Information Science and Technology, 38*, 87–143.

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher, 20*(3), 2–9. https://doi.org/10.3102/0013189x020003002.

Senders JW. (1993) Theory and analysis of typical errors in a medical setting. Hospital Pharmacy. 1993 Jun;28(6):505–508.

Sharp, H., Preece, J., & Rogers, Y. (2019). *Interaction design: Beyond human-computer interaction*. Hoboken, NJ: Wiley.

Shortliffe, E. H., & Blois, M. S. (2001). The computer meets medicine and biology: Emergence of a discipline. In E. H. Shortliffe & L. E. Perreault (Eds.), *Medical informatics: Computer applications in health care and biomedicine* (2nd ed., pp. 3–40). New York: Springer.

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. S. Siegler (Ed.), Children's thinking: What develops? (Vol. xi, pp. 325–348). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sloboda, J. (1991). Musical expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 153–171). New York: Cambridge University Press.

Sternberg, R. J., & Horvath, J. A. (Eds.). (1999). *Tacit knowledge in professional practice: Researcher and practitioner*. Mahwah: Lawrence Erlbaum Associates.

Suchman, L. A. (1987). Understanding computers and cognition: A new foundation for design – Winograd, T., Flores, F. *Artificial Intelligence, 31*(2), 227–232.

Sussman, S. Y. (2001). *Handbook of program development for health behavior research & practice*. Thousand Oaks: Sage.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive & healthy computer-based work*. Mahwah: Lawrence Erlbaum Associates Publishers.

Weinger, M. B., & Slagle, J. (2001). *Human factors research in anesthesia patient safety*. Proceedings of the AMIA Symposium, 756–760. PMID: 11825287; PMCID: PMC2243459.

White, B. Y., & Frederiksen, J. R. (1990). Causal model progressions as a foundation for intelligent learning environments. In W. J. Clancey & E. Soloway (Eds.), *Artificial intelligence and learning environments spe-*

cial issues of "Artificial Intelligence: An International Journal" (pp. 99–157).

Woods, D. D., Patterson, E. S., & Cook, R. I. (2008). Behind human error: Taming complexity to improve patient safety. In P. Carayon (Ed.), *Handbook of human factors and ergonomics in health care and patient safety* (pp. 459–476). Mahwah: Lawrence Erlbaum Associates.

Wright, P. C., Fields, R. E., & Harrison, M. D. (2000). Analyzing human-computer interaction as distributed cognition: The resources model. *Human Computer Interaction, 15*(1), 1–41.

Zhang, J., Johnson, T. R., Patel, V. L., Paige, D. L., & Kubose, T. (2003). Using usability heuristics to evaluate patient safety of medical devices. *Journal of Biomedical Informatics, 36*(1–2), 23–30. S1532046403000601 [pii].

Zhang, J., Patel, V. L., Johnson, T. R., & Shortliffe, E. H. (2004). A cognitive taxonomy of medical errors. *Journal of Biomedical Informatics, 37*(3), 193–204.

Zuriff, G. E. (1985). *Behaviorism: A conceptual reconstruction*. New York: Columbia University Press.

Human-Computer Interaction, Usability, and Workflow

Vimla L. Patel, David R. Kaufman, and Thomas Kannampallil

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_5

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What are the major attributes of system usability?
- What are the methods that can be used to evaluate usability of a health information system?
- How does a poorly designed HIT implementation contribute to disruptions to clinical workflow?

5

5.1 Introduction to Human-Computer Interaction

Human-computer interaction (HCI) is a multifaceted discipline devoted to the *study and practice* of design and usability (Carroll 2003). The history of computing and more generally, that of artifact design, are rife with stories of dazzlingly powerful devices with remarkable capabilities that are thoroughly unusable by anyone except for the team of designers and their immediate families. In the often-cited book, Psychology of Everyday Things, Donald Norman (1988) describes a litany of poorly designed artifacts ranging from programmable VCRs to answering machines and water faucets that are inherently non-intuitive and difficult to use. Similarly, there have been numerous innovative and promising clinical information technologies that have yielded decidedly suboptimal results and resulted in deep user dissatisfaction. At a minimum, difficult interfaces result in steep learning curves and structural inefficiencies in task performance. At worst, problematic interfaces can have serious consequences for patient safety (Koppel et al. 2005; Lin et al. 1998; Zhang et al. 2004).

Myers and Rosson (1992) reported that nearly 50% of software code was devoted to the user interface, and a survey of developers indicated that, on average, 6% of their project budgets were spent on usability evaluation. Given the complexities of the modern graphical user interfaces (GUI), it is likely that more than 50% of the code is now devoted to the GUI. On the other hand, usability evaluations have greatly increased over the last 20 years (Jaspers 2009). There have been numerous books and articles devoted to promoting effective user interface design (Preece et al. 2015; Shneiderman et al. 2016), and the importance of enhancing the user experience has been widely acknowledged by both consumers and producers of information technology. Part of the impetus is that usability has been demonstrated to be highly cost effective. Karat (1994) reported that for every dollar a company invests in the usability of a product, it receives between $10 and $100 in benefits. Although much has changed in the world of computing since Karat's estimate (e.g., the flourishing of the World Wide Web and mobile apps), it is clear that investments in usability still yield substantial rates of return (Nielsen 2008). It remains far costlier to fix a problem after product release than in an early design phase. The concept of usability as well as the methods and tools to measure and promote it are now "touchstones in the culture of computing" (Carroll 2003).

HCI has spawned a professional orientation that focuses on practical matters concerning the integration and evaluation of applications of technology to support human activities. There are also active academic HCI communities that have contributed significant advances to the science of computing. HCI researchers have been devoted to the development of innovative design concepts such as **virtual reality**, **ubiquitous computing**, **multimodal interfaces**, **collaborative workspaces**, **mobile technologies**, and **immersive and virtual environments**. HCI research has been instrumental in transforming the software engineering process towards a more user-centered iterative system development (e.g., rapid prototyping). HCI research has also been focally concerned with the cognitive, social, and cultural dimensions of the computing experience. In this regard, it is concerned with developing analytic frameworks for characterizing how technology can be used more productively across a range of tasks, settings, and user populations.

In this chapter, we describe the foundations of the role of HCI in biomedical informatics with a specific focus on methods for usability evaluation and clinical workflow. We also discuss the implications of HCI and clinical workflow methods for future biomedical informatics research. This chapter is a companion to chapter 4 in this volume on cognitive informatics (Chap. 4)

5.2 Role of HCI in Biomedical Informatics

HCI research in healthcare emerged at a time when health information technology and electronic health records (EHRs) were becoming more central to the practice of medicine (Patel et al. 2015). Much HCI work has been devoted to creating or enhancing design in healthcare systems. However, the focus of most of our work has been on the cognitive mediation of technology in healthcare practice (Patel et al. 2015). Most of the early HCI research focused on the solitary user of technology. Although such research is still commonplace, the focus has extended to distributed health information systems (Hazlehurst et al. 2007; Horsky et al. 2003) and analysis of unintended sociotechnical consequences with a particular focus on computerized provider order entry systems (Koppel et al. 2005). HCI studies in biomedicine extend across clinical and consumer health informatics, addressing a range of user populations including providers, biomedical scientists, and patients. While the implications of HCI principles for the design of HIT are acknowledged, the adoption of the tools and techniques among clinicians, informatics researchers and developers of HIT are limited. There is a consensus that HIT has not realized its potential as a tool that facilitates clinical decision-making, coordination of care, and improvement of patient safety (Middleton et al. 2013).

The field of human computer interaction intersects behavioral, and computer and information science. Thus, this field involves the study of interaction between people and computers. Computing systems includes both software and hardware. In addition, devices from smartphones to glucose meters are devices that present usability challenges. In this chapter, the focus is on the software and the interface componants. Thus the major focus of HCI is with the evaluation of interactive computer systems for human use. In the healthcare environment, it is important to understand HCI to ensure the users and the computers interact successfully. Therefore, the goals of HCI are to deploy usable, useful and safe systems.

5.3 Theoretical Foundations

In recent years, there has been a significant growth in research and application regarding HCI and healthcare systems. They have produced a collective body of experiential and practical knowledge about user experience, adoption and implementation to guide future design work. Some of the work is not specifically guided by a theoretical foundation and these efforts have proven to be useful in elucidating problems and contributing to user-centered design efforts. Human-computer interaction work is at least partly an empirical science in which local knowledge derived from a small body of studies will suffice in solving a problem. However, it is also necessary that we extrapolate knowledge from one context to another. Concentrated efforts in HCI are time-consuming, tend to employ small numbers of subjects and are conducted in a limited number of settings. For example, it is simply not possible to conduct an HCI research project in many different hospitals or to thoroughly test every facet of an electronic health record system. Knowledge solely based on practical experience or empirical studies are not adequate to account for the immense variety of health information technologies and the rich array of contexts that constitute the practice of medicine (Kaufman et al. 2015).

There are many facets to technology use and a range of theories that address them. For example, the technology acceptance model that focuses on user's perceived usefulness and usage intentions has been widely used in healthcare research (Venkatesh 2000). **Sociotechnical systems theory** is very broad in scope. It views all organizations as having the

following elements that comprise its organizational design: technological (including the actual IT system, usability, and unintended consequences), social (doctors, staff, patients, etc.), and external environment (e.g., political, economic, cultural, and legal influences) (Hendrick and Kleiner 1999). These subsystems are intricately connected, such that changes to any one affects others, sometimes in unanticipated or dysfunctional ways (Aarts et al. 2007; Ash et al. 2004). One of the most influential theories in clinical informatics was offered by Sittig and Singh (2010). They proposed an 8-dimensional model of interrelated concepts that can be used to explain performance in complex adaptive systems in the healthcare arena. The model has been applied in a range of settings model to understand and improve HIT applications at various stages of development and implementation.

Cognitive engineering (CE) is an interdisciplinary approach to the development of principles, methods, and tools to assess and guide the design of systems to support human performance (Hettinger et al. 2017). The approach is rooted in both cognitive science and engineering and has been used to support design of displays, decision support and training in numerous high-risk domains (Kushniruk et al. 2004). A computational theory of mind provides the fundamental underpinning for most contemporary cognitive theories. The basic premise is that much of human cognition can be characterized as a series of operations, computations on mental representations. At a higher level of cognitive analysis, CE also focuses on the discrepancy between user's goals and the physical controls embodied in a system (Norman 1986). Interface design choices differentially mediates task performance and various methods of analysis including those described below endeavor to measure this impact.

Distributed cognition (DCog) represents a shift in the study of cognition from an exclusive focus on the mind of the individual to being "stretched" across groups, material artifacts and cultures (Hutchins 1995). This paradigm has gained substantial currency in HCI research. In the distributed approach, cognition is viewed as a process of coordinating distributed internal (i.e., what's in the mind) and external representations (e.g., visual displays, post-it notes). DCog has two lines of analysis, one that emphasizes the social and collaborative nature of cognition (e.g., surgeons, nurses and respiratory therapists in cardiothoracic surgical setting jointly contributing to a decision process), and one that characterizes the mediating effects of technology (e.g., EHRs, paper charts, mobile devices, apps) or other artifacts on cognition. DCog constitutes a family of interrelated theories rather than a single approach (Cohen et al. 2006). The approaches collectively offer a penetrating view of the complexities embodied in human-computer interaction. However, there is no "off-the-shelf" methodology for using it in research or as a practitioner (Furniss et al. 2015). The application of DCog theory and methods are complicated by the fact that there are no set of features to attend to and no checklist or prescribed method to follow (Rogers 2012). In addition, the analysis and abstraction requires a high level of skill and training. More in-depth reviews of DCog can be found in (Rogers 2004) and as applied to healthcare in (Hazlehurst et al. 2008; Kaufman et al. 2015). DCog approaches have been particularly useful in the analysis of teamwork and EHR-mediated workflow in complex environments (Blandford and Furniss 2006; Hazlehurst et al. 2007; Kaufman et al. 2009). It is not unusual for HCI researchers to engage multiple theories depending on the area of focus.

5.4 Usability of Health Information Technology[1]

Theories of cognitive science meaningfully inform and shape design, development and assessment of health-care information systems by providing insight into principles of

1 Parts of the section, have been adapted, *with permission*, from Kannampallil, T. G., & Abraham, J. (2015). Evaluation of health information technology: Methods, frameworks and challenges. In V. L. Patel, T. G. Kannampallil, & D. Kaufman (Eds.), *Cognitive informatics in health and biomedicine: Human computer interaction*. London: Springer.

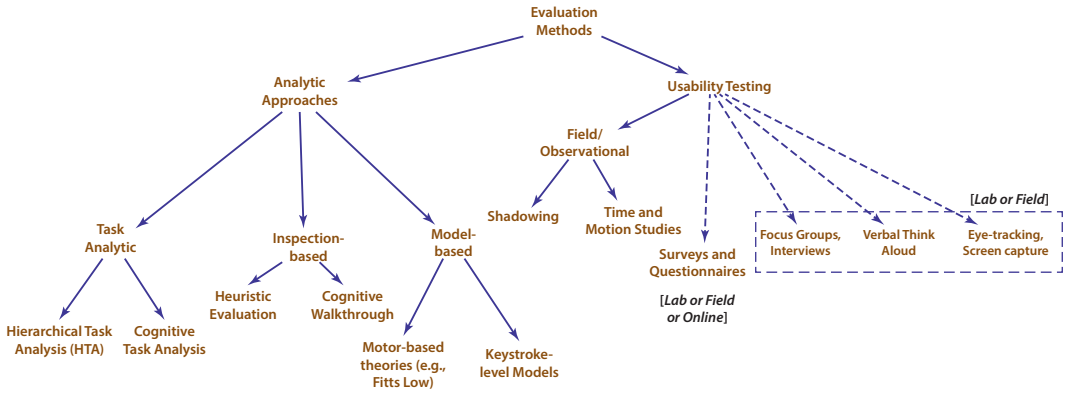


Fig. 5.1 Classification of evaluation methods

system ***usability*** and *learnability*, as well as the design of a safer workplace.

Usability methods, most often drawn from cognitive science, have been used to evaluate a wide range of medical information technologies including infusion pumps (Karat 1994), ventilator management systems, physician order entry (Ash et al. 2003; Horsky et al. 2003; Koppel et al. 2005), pulmonary graph displays (Wachter et al. 2003), information retrieval systems, and research web environments for clinicians (Elkin et al. 2002). In addition, usability techniques are increasingly used to assess patient-centered environments (Chan and Kaufman 2011; Cimino et al. 2000; Kaufman et al. 2003a, b). The methods include observations, focus groups, surveys and experiments. Collectively, these studies make a compelling case for the instrumental value of such research to improve efficiency, user acceptance and relatively seamless integration with current workflow and practices.

What do we mean by usability? Nielsen suggests that usability includes the following five attributes: (1) *learnability*: system should be relatively easy to learn, (2) *efficiency*: an experienced user can attain a high level of productivity, (3) *memorability*: features supported by the system should be easy to retain once learned, (4) *errors*: system should be designed to minimize errors and support error detection and recovery, and (5) *satisfaction*: the user experience should be subjectively satisfying.

The question then becomes how we evaluate and study the various attributes of usability. We classified usability evaluation methods into two categories: analytic evaluation approaches and usability testing. Analytic evaluation studies use experts as participants—usability experts, domain experts, software designers—or in some cases, are conducted without participants using task-analytic, inspection-based or model-based approaches and are conducted in laboratory-based settings.

We categorized usability testing into field-based studies that capture situated and contextual aspects of HIT use, and a general category of methods (e.g., interviews, focus groups, surveys) that solicit user opinions and can be administered in different modes (e.g., face-to-face or online). A brief categorization of the evaluation approaches can be found in Fig. 5.1. In the following sections, we provide a detailed description of each of the evaluation approaches along with research examples of its use.

5.4.1 Analytical Approaches

Analytical approaches rely on analysts' judgments and analytic techniques to perform evaluations on user interfaces, and often do not directly involve the participation of end users. These approaches employ experts—general usability, human factors, or soft-

ware—for conducting the studies. In general, analytical evaluation techniques involve task-analytic approaches, inspection-based methods, and predictive *model-based* methods (e.g., keystroke models, Fitts Law).

5.4.1.1 Task Analysis[2]

Task analysis is one of most commonly used techniques to evaluate "existing practices" in order to understand the rationale behind people's goals of performing a task, the motivations behind their goals, and how they perform these tasks (Preece et al. 1994). As described by Vicente (1999), task analysis is an evaluation of the "trajectories of behavior." There are several variants of task analysis—**hierarchical task analysis** (HTA) and **cognitive task analysis** (CTA) being the most commonly used in biomedical informatics research.

HTA is the simplest task analytic approach and involves the breaking down of a task into sub-tasks and smaller constituted parts (e.g., sub-sub-tasks). The tasks are organized according to specific goals. This method, originally designed to identify specific training needs, has been used extensively in the design and evaluation of interactive interfaces (Annett and Duncan 1967). The application of HTA can be explained with an example: consider the goal of printing a Microsoft Word document that is on your desktop. The sub-tasks for this goal would involve finding (or identifying) the document on your desktop, and then print it by selecting the appropriate printer. The HTA for this task can be organized as follows:

0. Print document on the desktop
1. Go to the desktop
2. Find the document
   - 2.1. Use "Search" function
   - 2.2. Enter the name of the document
   - 2.3. Identify the document
3. Open the document
4. Select the "File" menu and then "Print"
   - 4.1. Select relevant printer
   - 4.2. Click "Print" button

   Plan 0: do 1–3–4; if file cannot be located by a visual search, do 2–3–4
   Plan 2: do 2.1–2.2–2.3

In this task analysis, the task can be decomposed as follows: moving to your desktop, searching for the document (either visually or by using the search function and typing in the search criteria), selecting the document, opening and printing it using the appropriate printer. The order in which these tasks are performed may change based on specific situations. For example, if the document is not immediately visible on the desktop (or if the desktop has several documents making it impossible to identify the document visually), then a search function is necessary. Similarly, if there are multiple printer choices, then a relevant printer must be selected. The plans include a set of tasks that a user must undertake to achieve the goal (i.e., print the document). In this case, there are two plans: plan 0 and plan 2 (all plans are conditional on tasks having pertinent sub-tasks associated with it). For example, if the user cannot find a document on the desktop, plan 2 is instantiated, where a search function is used to identify the document (steps 2.1, 2.2 and 2.3). Figure 5.2 depicts the visual form of the HTA for this particular example.

HTA has been used in evaluating interfaces and medical devices. For example, Chung et al. (2003) used HTA to compare the differences between six infusion pumps. Using HTA, they identified potential sources for the generation of human errors during various tasks. While exploratory, their use of HTA provided insights into how the HTA can be used for evaluating human performance and for predicting potential sources of errors. Alternatively, HTA has been used to model information and clinical workflow in ambulatory clinics (Unertl et al. 2009). Unertl et al. (2009) used direct observations and semi-structured interviews to create a HTA of the workflows. The HTA was then used to identify the gaps in existing HIT functionality for supporting clinical workflows, and the needs of chronic disease care providers.

2 While GOMS (See ▶ Sect. 5.4.1.3) is considered a task-analytic approach, we have categorized it as a model-based approach for predictions of task completion times. It is based on a task analytic decomposition of tasks.

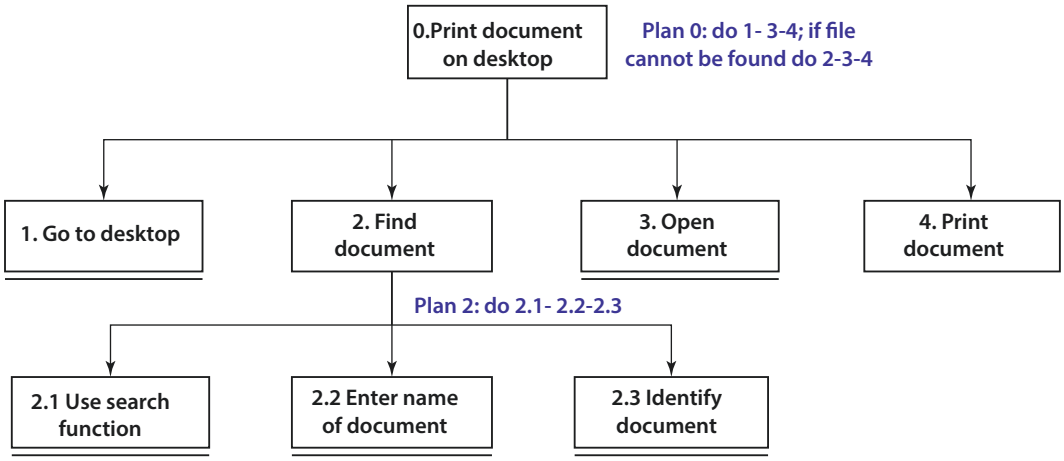


Fig. 5.2 Graphical representation of task analysis of printing a document: the tasks are represented in the boxes; the line underneath certain boxes represents the fact that there are no sub-tasks for these tasks

CTA is an extension of the general task analysis technique to develop a comprehensive understanding regarding the knowledge, cognitive/thought processes and goals that underlie observable task activities (Chipman et al. 2000). Although the focus is on knowledge and cognitive components of the task activities and performance, CTA relies on observable human activities to draw insights on the knowledge-based constraints and challenges that impair effective task performance.

CTA techniques are broadly classified into three groups based: (a) interviews and observations, (b) process tracing and (c) conceptual techniques (Cooke 1994). CTA using interviews and observations involve developing a comprehensive understanding of tasks through discussions with, and task observations of experts. For example, a researcher observes an expert physician performing the task of medication order entry into a CPOE (Computerized Physician Order Entry) system and asks to follow up questions regarding the specific aspects of the task. In a study on understanding providers' management of abnormal test results, Hysong et al. (2010) conducted CTA-based interviews with 28 primary care physicians on how and when they manage alerts, and how they use the various features on the EHR system to filter and sort their alerts.

CTA supported by process-tracing approaches relies on capturing task activities through direct (e.g., verbal think aloud) or indirect (e.g., unobtrusive screen recording) data capture methods. Whereas the process-tracing approach is generally used to capture expert behaviors, it has also been used to evaluate general users. In a study on experts' information seeking behavior in critical care, Kannampallil et al. (2013a) used the process-tracing approach to identify the nature of these activities including the information sources, cognitive strategies, and shortcuts used by critical care physicians in decision-making tasks. The CTA approach relied on the verbalizations of physicians, their access to various sources, and the time spent on accessing these sources to identify the strategies of information seeking.

Finally, CTA supported by conceptual techniques rely on the development of representations of a domain (and their related concepts) and the potential relationships between them. This approach is often used with experts and different methods are used for knowledge elicitation including concept elicitation, structured interviews, ranking approaches, card sorting, structural approaches such as multi-dimensional scaling, and graphical associations (Cooke 1994).

5.4.1.2 Inspection-Based Evaluation

Inspection methods involve experts appraising a system, playing the role of a user to identify potential usability and interaction issues with

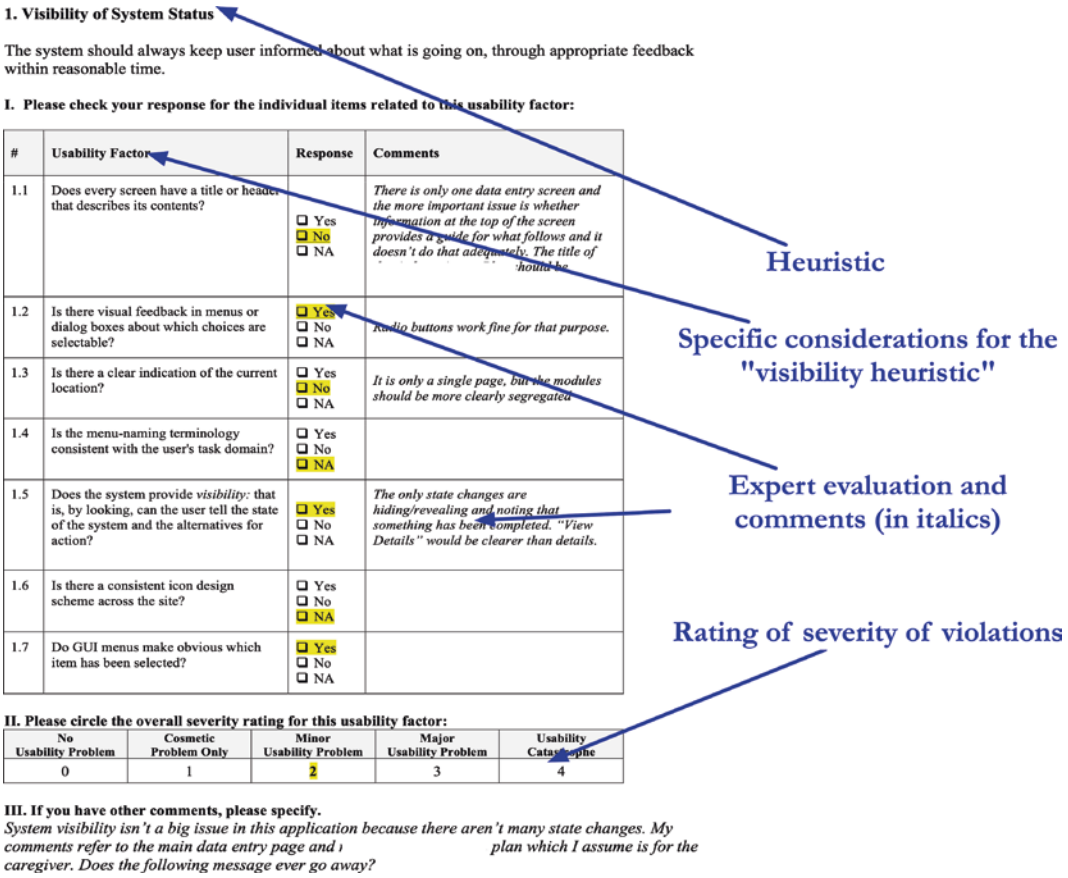
a system. Inspection methods are often conducted on fully developed systems or interfaces but may also be used for prototypes. Inspection methods rely on a usability expert, i.e., a person with significant training and experience in evaluating interfaces, to go through a system and identify whether the user interface elements conform to a predetermined set of usability guidelines and design requirements (or principles). The most commonly used inspection methods are **heuristic evaluations** (HE) and walkthroughs.

HE techniques utilize a small set of experts to evaluate a user interface (or a set of interfaces in a system) based on their understanding of a set of heuristic principles regarding interface design (Johnson et al. 2005). This technique was developed by Jakob Nielsen and colleagues (Nielsen and Molich 1990), and has been used extensively in the evaluation of user interfaces. The original set of heuristics was developed by Nielsen based on an abstraction of 249 usability problems. In general, the following ten heuristic principles (or a subset of these) are most often considered for HE studies: system status visibility; match between system and real world; user control and freedom; consistency and standards; error prevention; recognition rather than recall; flexibility and efficiency of use; aesthetic and minimalist design; help users recognize, diagnose and recover from errors; and help and documentation (retrieved from: ► http://www.nngroup.com/articles/ten-usability-heuristics/). Conducting an HE involves a usability expert going through an interface to identify potential violations to a set of usability principles (referred to as "heuristics"). These perceived violations could involve a variety of interface elements such as windows, menu items, links, navigation, and interaction.

Evaluators typically select a relevant subset of heuristics for evaluation (or add more based on the specific needs and context). The selection of heuristics is based on the type of system and interface being evaluated. For example, the relevant heuristics for evaluating an EHR interface would be different from that of an app on a mobile device. After selecting a set of applicable heuristics, one or more usability experts evaluate the user interface against the identified heuristics. After evaluating the heuristics, the potential violations are rated according to a severity score (1–5, where 1 indicates a cosmetic problem and 5 indicates a catastrophic problem). This process is iterative and continues until the expert feels that a majority (if not all) of the violations are identified. It is also generally recommended that a set of 4–5 usability experts are required to identify 95% of the perceived violations or problems with a user interface. However, it is not uncommon to employ fewer experts (e.g., 3). It should be acknowledged that the HE approach may not lead to the identification of all problems and the identified problems may be localized (i.e., specific to a particular interface in a system). An example of an HE evaluation form is shown in ◘ Fig. 5.3.

In the healthcare domain, HE has been used in the evaluation of medical devices and HIT interfaces. For example, Zhang et al. (2003) used a modified set of 14 heuristics to compare the patient safety characteristics of two 1-channel volumetric infusion pumps. Four independent usability experts evaluated both infusion pumps using the list of heuristics and identified 89 usability problems categorized as 192 heuristic violations for pump 1, and 52 usability problems categorized as 121 heuristic violations for pump 2. The heuristic violations were also classified based on their severity. In another study, Allen et al. (2006) developed a simplified list of heuristics to evaluate web-based healthcare interfaces (printouts of each interface). Multiple usability experts assigned severity ratings for each of the identified violations and the severity ratings were used to re-design the interface.

Walkthroughs are another inspection-based approach that relies on experts to evaluate the cognitive processes of users performing a task. It involves employing a set of potential stakeholders (designers, usability experts) to characterize a sequence of actions and goals for completing a task. Most commonly used walkthrough, referred to as **cognitive walkthrough** (CW), involves observing, recording and analyzing the actions and behaviors of users as they complete a scenario of use. CW is focused on identifying the usability and



1. Visibility of System Status

The system should always keep user informed about what is going on, through appropriate feedback within reasonable time.

I. Please check your response for the individual items related to this usability factor:

#	Usability Factor	Response	Comments
1.1	Does every screen have a title or header that describes its contents?	☐ Yes ☐ No ☐ NA	*There is only one data entry screen and the more important issue is whether information at the top of the screen provides a guide for what follows and it doesn't do that adequately. The title of should be*
1.2	Is there visual feedback in menus or dialog boxes about which choices are selectable?	☐ Yes ☐ No ☐ NA	*Radio buttons work fine for that purpose.*
1.3	Is there a clear indication of the current location?	☐ Yes ☐ No ☐ NA	*It is only a single page, but the modules should be more clearly segregated*
1.4	Is the menu-naming terminology consistent with the user's task domain?	☐ Yes ☐ No ☐ NA	
1.5	Does the system provide *visibility:* that is, by looking, can the user tell the state of the system and the alternatives for action?	☐ Yes ☐ No ☐ NA	*The only state changes are hiding/revealing and noting that something has been completed. "View Details" would be clearer than details.*
1.6	Is there a consistent icon design scheme across the site?	☐ Yes ☐ No ☐ NA	
1.7	Do GUI menus make obvious which item has been selected?	☐ Yes ☐ No ☐ NA	

II. Please circle the overall severity rating for this usability factor:

No Usability Problem	Cosmetic Problem Only	Minor Usability Problem	Major Usability Problem	Usability Catastrophe
0	1	2	3	4

III. If you have other comments, please specify.

System visibility isn't a big issue in this application because there aren't many state changes. My comments refer to the main data entry page and plan which I assume is for the caregiver. Does the following message ever go away?

Fig. 5.3 Example of an HE form (for visibility)

comprehensibility of a system (Polson et al. 1992). The aim of CW is to investigate and determine whether the user's knowledge and skills and the interface cues are sufficient to produce an appropriate goal-action sequence that is required to perform a given task (Kaufman et al. 2003a, b). CW is derived from the cognitive theory of how users work on computer-based tasks, using the exploratory learning approach, where system users continually appraise their goals and evaluate their progress against these goals (Kahn and Prail 1994).

While performing CW, the focus is on simulating the human-system interaction, and evaluating the fit between the system features and the user's goals. Conducting CW studies involves multiple steps. Potential participants (e.g., users, designers, usability experts) are provided a set of task sequences or scenarios for working with an interface or system. For example, for an interface for entering demographic and patient history details, participants (e.g., physicians) are asked to enter the age, gender, race and clinical history information. As the participants perform their assigned task, their task sequences, errors and other behavioral aspects are recorded. Often, follow up interviews or think aloud (described in a later section) are used to identify participants' interpretation of the tasks, how they make progress, and potential points of mismatches in the system. Detailed observations and recordings of these mismatches are documented for further analysis. While in most situations CWs are performed by individuals, sometimes groups of stakeholders perform the walkthrough together. For example, usability experts, designers and potential users could go through systems together to identify

the potential issues and drawbacks. Such group walkthroughs are often referred to as pluralistic walkthroughs.

In biomedical informatics domain, it must be noted that CW has been used extensively in evaluating situations other than human-computer interaction. For example, the CW method (and its variants) has been used to evaluate diagnostic reasoning, decision-making processes and clinical activities. For example, Kushniruk et al. (1996) used the CW method to perform an early evaluation of the mediating role of HIT in clinical practice. The CW was not only used to identify usability problems but was instrumental in the development of a coding scheme for subsequent usability testing. Hewing et al. (2013) used CW to evaluate an expert ophthalmologist's reasoning regarding retinal disease in infants. Using images, clinical experts were independently asked to rate the presence and severity of retinal disease and provide an explanation of how they arrived at their diagnostic decisions. Similar approaches were used by Kaufman et al. (2003a, b) to evaluate the usability of a home-based, telehealth system.

5.4.1.3 Model-Based Evaluation

Model-based evaluation approaches use predictive modeling approaches to characterize the efficiency of user interfaces. Model-based approaches are often used for evaluating routine, expert task performance. For example, how can the keys of a medical device interface be optimally organized such that the users can complete their tasks quickly and accurately? Similarly, predictive modeling can be used to compare the data entry efficiency between interfaces with different layouts and organization. We describe two commonly used predictive modeling techniques in the evaluation of interfaces.

Card et al. (1980) proposed the GOMS (Goals, Operators, Methods and Selection Rules) analytical framework for predicting human performance with interactive systems. Specifically, GOMS models predict the time taken to complete a task by a skilled/expert user based on "the composite of actions of retrieving plans from long-term memory, choosing among alternative available methods depending on features of the task at hand, keeping track of what has been done and what needs to be done, and executing the motor movements necessary for the keyboard and mouse" (Olson and Olson 2003). In other words, GOMS assumes that the execution of tasks can be represented as a serial sequence of cognitive operations and motor actions.

GOMS is used to describe an aggregate of the task and the user's knowledge regarding how to perform the task. This is expressed regarding the *G*oals, *O*perators, *M*ethods and *S*election rules. *Goals* are the expected outcomes that a user wants to achieve. For example, a goal for a physician could be documenting the details of a patient interaction on an EHR interface. *Operators* are the specific actions that can be performed on the user interface. For example, clicking on a text box or selecting a patient from a list in a dropdown menu. *Methods* are sequential combinations of operators and sub-goals that need to be achieved. For example, in the case of selecting a patient from a dropdown list, the user has to move the mouse over to the dropdown menu, click on the arrow using the appropriate mouse key to retrieve the list of patients. Finally, *selection* rules are used to ascertain which methods to choose when several choices are available. For example, using the arrow keys on the keyboard to scroll down a list versus using the mouse to select.

One of the simplest and most commonly used GOMS approaches is the Keystroke-Level Model (KLM), which was first described in Card et al. (1983). As opposed to the general GOMS model, the KLM makes several assumptions regarding the task. In KLM, methods are limited to keystroke level operations and task duration is predicted based on these estimates. For the KLM, there are six types of operators: *K* for pressing a key; *P* for pointing the mouse to a target; *H* for moving hands to the keyboard or pointing device; *D* for drawing a line segment; *M* for mental preparation for an action; and *R* for system response. Based on experimental data or other predictive models (e.g., Fitts Law), each of these operators is assigned a value or a param-

eterized estimate of execution time. We describe an example from Saitwal et al. (2010) on the use of the KLM approach.

In a study investigating the usability of EHR interfaces, Saitwal et al. (2010) used the KLM approach to evaluate the time is taken, and the number of steps required to complete a set of 14 EHR-based tasks. The purpose of the study was to characterize the issues with the user interface and also to identify potential areas for improvement. The evaluation was performed on the AHLTA (Armed Forces Health Longitudinal Technology Application) user interface. A set of 14 prototypical tasks was first identified. Sample tasks included entering the patient's current illness, history of present illness, social history and family history. KLM analysis was performed on each of the tasks: this involved breaking each of the tasks into its component goals, operators, methods and selection rules. The operators were also categorized as physical (e.g., move the mouse to a button) or mental (e.g., locate an item from a dropdown menu). For example, the selection of a patient name involved eight steps (M – mental operation; P – physical operation): (1) think of location on the menu [M, 1.2s], (2) move hand to the mouse [P, 0.4s], (3) move the mouse to "Go" in the menu [P, 0.4s], (4) extend the mouse to "Patient" [P, 0.4s], (5) retrieve the name of the patient [M, 1.2s], (6) locate patient name on the list [M, 1.2s], (7) move mouse to the identified patient [P, 0.4s] and (8) click on the identified patient [P, 0.4s]. In this case, there were a total of 8 steps that would take 5.2s to complete. Similarly, the number of steps and the time taken for each of the 14 considered AHLTA tasks were computed.

In addition, GOMS and its family of methods can be productively used to make comparisons regarding the efficiency of performing tasks interfaces. However, such approaches are approximations and have several disadvantages. Although GOMS provides a flexible and often reliable mechanism for predicting human performance in a variety of computer-based tasks, there are several potential limitations. A brief summary is provided here, and interested readers can find further details in Card et al. (1980). GOMS models can be applied only to the *error-free*, *routine* tasks of *skilled* users. Hence, it is not possible to make time predictions for non-skilled users, who are likely to take considerable time to learn to use a new system. For example, the use of the GOMS approach to predict the potential time spent by physicians in using a new EHR would be inaccurate owing to relative lack of knowledge of the physicians regarding the use of the various interfaces, and the learning curve required to be up-to-speed with the new system. The complexity of clinical work processes and tasks, and the variability of the user population create significant challenges for the effective use of GOMS in measuring the effectiveness of clinical tasks.

Fitts Law is used to predict human motor behavior; it is used to predict the time taken to acquire a target (Fitts 1954). On computer-based interfaces, it has been used to develop a predictive model of time it takes to acquire a target using a mouse (or another pointing device). The time taken to acquire a target depends on the distance between the pointer and target (referred to as amplitude, A) and the width of the target (W). The movement time (MT) is mathematically represented as follows:

$$\mathrm{MT} = k.\log_2\left(\frac{A}{W}+1\right)$$

where k is a constant, A – amplitude, W – width of the target.

In summary, based on Fitts law, one can say that the larger objects are easier to acquire while smaller, closely aligned objects are much more difficult to acquire with a pointing device. While the direct application of Fitts law is not often found in the evaluation studies of HIT or health interfaces in general, it has a profound influence in the design of interfaces. For example, the placement of menu items and buttons, such that a user can easily click on them for selection, are based on Fitts law parameters. Similarly, in the design of number keypads for medical devices, the size of the buttons and their location can be effectively predicted by Fitts law parameters.

In addition to the above-mentioned predictive models, there are several other less common models. While a detailed description of each of them or their use is beyond the scope of this chapter, we provide a brief introduction to another predictive approach: Hick-Hyman choice reaction time (Hick 1951; Hyman 1953). Choice reaction time, RT, can be predicted based on the number of available stimuli (or choices), n:

$$\mathrm{RT} = a + b.\log_2(n)$$

where a and b are constants.

Hick-Hyman law is particularly useful in predicting text entry rates for different keyboards (MacKenzie et al. 1999), and time required to select from different menus (e.g., a linear vs. a hierarchical menu). In particular, the method is useful to make decisions regarding the design and evaluation of menus. For example, consider two menu design choices: 9 items deep/3 items wide and 3 items deep/9 items wide. The RT for each of these can be calculated as follows: $(3 \times (a + b.\log_2(n)) < 9 \times (a + b.\log_2(n))$. This shows that the access to menus is more efficient when it is designed breadth-wise rather than depth-wise.

5.5 Usability Testing and User-Based Evaluation

In this section, we have grouped a range of approaches that are generally used for evaluating the usability of HIT systems. In general, we have classified them into field/observational studies and general approaches for usability evaluation that can be utilized in both field and laboratory settings. While formal usability testing is often conducted in laboratory settings where user performance (and other selected variables) are evaluated based on pre-selected tasks, we have loosely classified the evaluation techniques that utilize users in the evaluation process into general approaches (those that can be used in both field and laboratory-based studies) and field studies.

5.5.1 Interviews and Focus Groups

Interviews and focus groups are commonly used to elicit information about opinions and perspectives of participants and their work practices (Mason 2002). Interviews are viewed as an approach to elicit additional information and are often used in concert with other field study methods (e.g., observation or shadowing).

Individual interviews can be classified into three major categories based on the format and level of standardization of the interview questions – structured, semi-structured and narrative (or unstructured). During *structured interviews*, all interviewees are asked the same questions in the same order. This allows for comparisons between responses across interviewees, which can be analyzed using qualitative and quantitative methods. *Semi-structured interviews* are flexible and allow for probing of participants (i.e., with follow up questions) to discuss relevant issues (Denton et al. 2018).

Focus group is a type of interactive interviewing method that involves an in-depth discussion of a particular topic of interest with a small group of participants. Focus group method has been described as "a carefully planned discussion designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment" (Krueger 2009). The central elements of focus groups as highlighted by Vaughn et al. (1996) include: (a) the group is an informal assembly of target participants to discuss a topic; (b) the group is small, between 6 to 12 members and is relatively homogeneous; (c) the group conversation is facilitated by a trained moderator with prepared questions and probes; and (d) given that the primary goal of a focus group is to elicit the perceptions, feelings, attitudes, and ideas of participants about a selected topic, it can be used to generate hypotheses for further research (Krueger 2009).

Unlike individual interviews, focus group discussions allow the researcher to probe responses to a particular research topic while capturing the underlying group dynamics of the participants. According to Kitzinger

(1995), interaction is the crucial feature of focus groups because the interaction between participants views the group as a single unit and also captures their view of the world, the language they use about an issue and their values and beliefs about a situation (Gibbs 1997). For instance, a focus group involving usability experts, system designers and care providers can allow participants to share their varying perspectives on system design based on their work role. This will enable them to voice the key issues on the fit or (lack thereof) between the functionalities of the system and the clinical workflow.

Another important factor that plays a vital role in focus group sessions is the presence of a skilled moderator (or facilitator) (Burrows and Kendall 1997) who manages the conversations and interactions between participants. Moreover, scheduling a convenient time and location for administering focus group interviews can be very difficult, given the number of participants that are involved.

5.5.2 Verbal Think Aloud

Verbal think aloud (or simply "think aloud") is used to capture rich verbal data on the thought processes that underlie human actions. Analysis of these verbal reports can be used to characterize the underlying information and knowledge structures. Think aloud evaluations are generally characterized into two types: (1) concurrent and (2) retrospective (Ericsson and Simon 1980). A concurrent think aloud requires uninterrupted and direct verbalizations of participants as they perform a task, and is considered to be complete and consistent with their thought sequence. In contrast, a retrospective think aloud requires the researcher to ask and prompt subjects to recall their thought sequence while performing a task (or after completing a task). Ericsson and Simon (1984), the original proponents of the verbal think aloud method, suggested the value of think aloud data is based on the following assumptions: (1) the verbalizations capture only a subset of the cognitive processes underlying behavior; (2) human mind is an information processor; and (3) the verbalizations capture contents of working memory (i.e., information recently acquired is accessed).

Think aloud studies are typically conducted to identify and characterize cognitive processes such as reasoning, problem solving, and decision-making processes. For example, Patel and colleagues (Patel et al. 1994, 2001; Patel and Groen 1991a, b) have conducted several studies using verbal think aloud that investigated the nature of reasoning using electronic tools, its effects of expertise and decision-making. Most of these studies relied on verbalizations by a participant (e.g., a physician), and in-depth linguistic analysis of the verbalizations to identify inherent strategies in their reasoning and decision-making. Similarly, Fonteyn and Grobe (1994) utilized a think aloud study to understand the reasoning and decision-making behaviors of critical care nurses regarding unstable patients. Insights on the reasoning process of expert nurses informed the design of an expert system. Other examples of similar key studies can be found here (Fisher and Fonteyn 1995; Fowler 1997; Funkesson et al. 2007; Grobe et al. 1991; Simmons et al. 2003).

One of the concerns that have been raised in evaluation studies using verbal think aloud method is the issue of sample size. While many researchers have used a small sample size of five participants to focus on in-depth analysis of the cognitive processes, others have critiqued the sample size (e.g., Lewis 1994). Lundgrén-Laine and Salanterä (2010) have suggested that the characteristics of the study participants in terms of their verbalization skills and the appropriate application of the think aloud is more important than the sample size (Caulton 2001; Fonteyn et al. 1993; Hall et al. 2004). Measures of information and participant saturation are often used to determine study completion. A detailed description of the think aloud method and approaches for its analysis can be found here (Ericsson and Simon 1984).

5.5.3 Usability Surveys and Questionnaires

Surveys and questionnaires are widely used in usability evaluation studies. Their widespread use is related to ease of administration (through multiple modes: online, face-to-face) and limited time required to complete (especially those that use Likert scale measures). In terms of usability evaluation, there are several surveys that are commonly used. A list of the commonly used usability surveys are provided below:

(a) *QUIS* (Questionnaire for User Interface Satisfaction: ► http://lap.umd.edu/quis/): measure user interface interaction and subjective satisfaction;
(b) *SUMI* (Software Usability Measurement Inventory: ► http://sumi.ucc.ie/): assess usability of software;
(c) *PSSUQ* (Post-Study System Usability Questionnaire), and *ASQ* (After Scenario Questionnaire: ► http://hcibib.org/perlman/question.cgi?form=ASQ) (Lewis 1991): address global usability of a system along with specific scenarios of use;
(d) *SUS* (System Usability Scale – ► http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html) (Brooke 1996): a general survey of system usability;
(e) *Subjective workload assessment* (NASA-TLX Workload Instrument: ► http://humansystems.arc.nasa.gov/groups/tlx/paperpencil.html) (Hart and Staveland 1988): a multi-item scale to determine the physical, temporal, mental, effort, frustration and performance while working with interfaces.

Although most of the above-mentioned surveys are validated for their reliability, researchers often use a variety of self-created surveys and questionnaires. Questionnaires, as opposed to the surveys that use a specific scale (e.g., a scale of 1–7), often use open-ended questions to elicit responses from participants regarding system use (e.g., "Describe some of the challenges that you faced while using the system?").

Surveys are often used along with other data collection methods and are considered a complementary data collection method in HIT evaluation. For example, Karahoca and colleagues (Karahoca et al. 2010) used a generic survey along with system usage logs to characterize the usability of two mobile device prototypes. Similar open-ended questionnaires along with additional observational data was used by Holzinger and colleagues (Holzinger et al. 2011) to characterize patient interactions with a mobile interface. Dalai and colleagues (Dalai et al. 2014) used the SUS scale and the NASA-TLX scales for comparing the effectiveness of two interfaces for comprehending psychiatric clinical narratives. These survey scales were used in concert with an analysis of verbal reports to evaluate the effectiveness of presented interfaces.

5.5.4 Field/Observational Approaches

In contrast to the analytic evaluation techniques that often yield objective data, there are several qualitative approaches that focus on the subjective and contextual assessments of system design and user interactions within the *context* of a real work environment (Assila et al. 2014). These qualitative approaches are generally categorized as ethnographic-based methods and require an "immersion" in the field in order to understand the experiences and practices of the informants (Schatzberg 2008). Although field and observational methods are more central to human factors, they have also played an instrumental role in understanding how health information technologies mediate a range of decision-making, coordination and associated patient care activities. The next section provides examples of observational research in clinical settings within the context of clinical workflow and usability.

5.6 Clinical Workflow

Workflow is a "set of tasks grouped into chronologically ordered processes, plus the people and resources required to complete the

tasks and accomplish a desired goal" (Unertl et al. 2010). It is widely believed that workflow analysis is essential for ensuring successful design and implementation of health IT (Harrington 2015; Schumacher and Lowry 2010; Xie and Carayon 2015). **Electronic health record** workflow is a subset of workflow activities that are mediated by EHRs and related technologies (Harrington 2015). These activities are not viewed as discrete, but rather are embedded in a situational context and broader workflow. Clinical workflow, especially in high velocity clinical settings, is characterized by perpetual change, multiple providers with varying levels of communication, and a high volume of workload, multitasking, and interruptions (Harrington 2015). Workflow is further complicated by the need to negotiate complex and nonintuitive systems. When EHRs are well integrated into clinical workflow, it increased the likelihood of positive healthcare outcomes and diminished error rates (Carayon et al. 2010; Lau et al. 2012). In contrast, when health IT was not well integrated into clinical workflow in a way that supports clinicians' cognitive work, it can compromise patient safety (Carayon et al. 2010).

Clinical workflow has been extensively studied over the course of the last 20 years. A cursory search of the term "clinical workflow" in Google Scholar yields more than 16,000 articles and more than 8000 since 2014. The scope of workflow research is rather expansive incorporating analysis of individuals, work environments, human-system interfaces and organizational factors. There has also been a focus on workflow as a mediator of patient safety (Carayon et al. 2014; Middleton et al. 2013). The emphasis of this section is on EHR workflow, which names the subset of workflow mediated by EHRs and other health IT (Zheng et al. 2010). EHRs are known to lack flexibility and resist easy modification and are relatively independent of context. However, clinical workflow is variable and context dependent, and tends to resist "one-size-fits-all" solutions. Studies on EHR-mediated workflow tend to focus on task-performance of the individual clinician, for example, (1) the impact of usability on workflow (ref), (2) time and motion studies that endeavor to quantify how clinicians allocate their time, or (3) the role of the EHR in the coordination of care, for example, examining the ways the systems serve as either a facilitator or impediment to the delivery of care (Weir et al. 2011).

Although EHR has conferred significant advantages such as decision support (Ben-Assuli et al. 2015) and improved clinical note quality (Burke et al. 2014), it has also contributed to an onerous documentation burden which impacts workflow. According to AMIA's EHR 2020 Task Force's report, AMIA's report on the EHR 2020 Task Force's, clinician's time investment in patient care documentation has doubled in the last 20 years (Payne et al. 2015). A large scale survey of clinicians affiliated with the American College of Physicians found that clinicians reported a loss of time (relative to paper-based records) of 4 hours per week (McDonald et al. 2014). The authors concluded that this could decrease access and increase the cost of care. However, findings varied significantly across studies. Researchers have employed a range of methods to document burden including log files and time on task studies. Hripcsak et al. used audit logs to perform a detailed analysis of time spent reviewing and documenting clinical notes (Hripcsak et al. 2011). They found significant variation among clinicians with a range of 20–100 minutes documenting and 7 minutes to 56 reviewing notes. They also noted that a significant percent of notes (e.g., 38% of nursing notes) were never read by anyone. This impacted communication, for example, the transfer of information from nurses to physicians. In a recent study, Collins and colleague used log files to study flowsheet documentation at a highly granular level (Collins et al. 2018). They found that clinicians (mostly nurses) manually entered between 600–900 data points. The authors argue for the need for better automated device integration.

In summary, there is heterogeneity in findings in relation to documentation. Some can be attributable to differences specific to settings, for example, as reflected in the use of scribes and in relation to the execution of Meaningful Use mandates. Other differences

are reflected in the study methods. However, it is quite clear that EHRs typically add significantly to the documentation burden and that has resulted in frustration among users.

EHR usability problems and their impact on workflow are well documented. For example, the Healthcare Information and Management Systems Society (HIMSS) survey found that workflow was clinicians' number one EHR usability "pain point" (Ribitzky et al. 2010). Respondents also reported frustration due to numerous alerts and difficulties with navigation resulting from the need to negotiate too many displays in the EHR to access information (Carayon et al. 2010). As referred to previously, Saitwal and colleagues (Saitwal et al. 2010) employed a cognitive task analytic approach to evaluate an EHR interface and quantified interactive behavior (e.g., task duration, number of steps). Challenges of employing the user interface were: (a) large number of average total steps to complete routine tasks, (b) slow execution time, and (c) overall mental workload. Similarly, Carayon et al. (6) summarized a range of usability issues in their systematic review of the workflow literature, including the large number of mouse clicks to complete a task, difficulties in navigating between many screens to input and retrieve information and cluttered screens. These usability problems serve to increase cognitive load. Information overload, a form of cognitive load, is a problem that occurs when attentional, perceptual and cognitive capacity is exceeded by the quantity of data presented via an interface to the extent that errors occur in users' information processing (Zahabi et al. 2015).

Cognitive overload can be partially attributed to poor user design. Small interface differences can be consequential and have a significant impact on task efficiency. According to Gray and colleagues, interactive behavior is constrained by the design and configuration of displays "as well as by the ways in which elementary cognitive, perceptual, and motor operations can be combined" (Gray and Boehm-Davis 2000). Poorly configured interfaces can increase navigational complexity. Navigation in this context refers to route taken to complete a task including the action steps and the trajectory through space (e.g., sequence of tabs or display screens). In a systematic review, Roman et al. found that navigation actions (e.g., scrolling through a patient list) were often linked to specific usability heuristic violations, including flexibility and efficiency of use, and lack of an emphasis on recognition rather than recall (Roman et al. 2017).

Medication reconciliation (MedRec) tools are an essential part of a strategy to reduce medication errors and prevent adverse events (Agrawal 2009). MedRec tools enable clinicians to compare lists of medications in patients history and revise the lists so that they are up-to-date and accurate. There have been several recent studies that have applied cognitive methods of analysis to MedRec (Boockvar et al. 2011; Lesselroth et al. 2013). Horsky and colleagues et al. investigate the accuracy of two different medication reconciliation tools integrated into EHRs in a simulated study (Horsky et al. 2017). They found that the reconciled records were significantly more accurate when clinicians used the second tool. Specifically, the comparison showed clinicians made three times as many errors in EHRs with single column medication lists, as compared to using side-by-side lists. The authors concluded that the better outcome using the second tool was strongly facilitated by a design that was more effective in supporting a cognitively demanding task. The system made less demands on working memory than the first. Plaisant and colleagues similarly contrasted a conventional interface design (control) with a novel prototype (Twinlist) (Plaisant et al. 2015). The Twinlist interface divided information into five columns, while the control used two side-by-side lists. Evaluation showed that in Twinlist participants completed MedRec significantly faster and more accurately than the control. Both studies demonstrated the comparative advantage of having access to needed information on a single screen as opposed to having to toggle between two screens or tabs.

The studies described above employed simulated methods. Duncan compared

MedRec interfaces in three Mayo Clinic campuses in which different EHR systems were used (Duncan et al. 2018b). Although they support the same set of functions, the interfaces differed in important respects. Specifically, the steps to access the medication list and perform the addition of a last dose differ. System 2 necessitated a three-step process as opposed to a single step needed to execute the same reconciliation goal (system 1). The systems were compared using a predictive model (KLM) and with live observations captured to video. As described earlier in this chapter, the keystroke-level model (KLM) is a widely used analysis where execution time of routine tasks (performed without errors) are estimated. Duncan found that the KLM estimates for the MedRec task closely approximated the observations. Both methods found that the time required to reconcile a single medication was more than 2 seconds greater for system 2, which also required more mouse clicks and screen transitions. The two systems differ in terms of interactive complexity and demands on working memory. The difference highlights the importance of emphasizing recognition rather than recall to minimize the memory load on clinicians.

Duncan and colleagues employed a similar approach with the same EHR systems in relation to a vital signs documentation and medication administration record tasks (Duncan et al. 2018a; Duncan et al., 2020). Vital signs are used to gauge a patient's hemodynamic stability and, in this case, provide a point of reference prior to surgery. The objectives were to: (1) analyze aspects of vital signs charting interfaces and determine how these aspects differentially mediate task performance and (2) investigate variations in vital signs documentation across clinical sites. Analyses revealed that accessing displays and the organization of interface elements are often unintuitive and inefficient, creating unnecessary complexities when interacting with the system. The study documented the ways in which the systems differed in their modes of interaction, organization of patient information and cognitive support. The authors noted that identifying barriers to interface usability and bottlenecks in EHR-mediated workflow can lead to system redesigns that minimize cognitive load and may also improve patient safety and efficiency.

5.7 Future Directions

The role of HCI, and more specifically, usability, is likely to be more embedded within the biomedical informatics research paradigm over the next decade. This is primarily because of the role health information technologies play in transforming the practice of medicine. Such a transformation has led to the widespread use of technology in patient care (e.g., the use of EHRs) and the development of patient-facing applications for self-management of care (e.g., mobile devices). As a result, the role of usability is likely to have an increased focus in potentially two areas: development of new approaches for unobtrusive evaluation of user interactions, and in the evaluation of consumer-facing applications for health.

As previously described, usability and user interaction studies are expensive, in terms of time, and effort that are involved. Recent efforts on usability evaluation have focused on utilizing logs of user interactions—including audit trails, and other unobtrusively collected interaction data (e.g., key stroke or eye-tracker). One classic example of such data is the EHR-based audit logs. In a recent study on EHR use in an emergency department, Kannampallil and colleagues used user log files to track the physician interactions during patient care activities (Kannampallil et al. 2018b). Similar efforts on tracing and modeling interaction and navigation behaviors are ongoing. The availability of more powerful tools for data capture and analysis will create new opportunities for computer scientists, psychologists and clinicians to collaborate on HCI-related investigations of technology-mediated clinical practice. In a recent study of such a collaboration, Vankipuram and colleagues showed that the process of data-driven iterative workflow redesign using visualization of overlaid data from quantitative and qualitative sources could be used to identify

inefficiencies and bottlenecks in clinical workflow and potentially contribute to process improvement (Vankipuram et al., 2019).

The growth of wearable devices and mobile sensing technologies have led to the development and use of a large number of consumer-facing applications and tools. Although much of these are on mobile devices, other patient-facing applications such as portals and social networking tools have gained prominence in recent years. These applications are rapidly evolving and, in many cases, still require significant improvements for translation into a usable and sufficiently robust product.

5.8 Conclusion

The implementation and broad use of health information technologies have grown rapidly over the course of the last decade. Clinical applications and increasingly patient-facing systems are beginning to transform healthcare practices. They offer significant potential for transforming the quality of patient care as well as enabling patient to become agents of change in managing their own heath. Usability challenges continue to provide significant impediments to productive use of technology and efficient workflow. The focus of this chapter has been on methods of usability evaluation. There is a wealth of different methods available for researchers and professional usability analysts to deploy in view to optimize the user experience. The methods have contributed to a growing body of knowledge to inform user-centered design and best practices across the range of health technologies.

Suggested Readings

Carroll, J. M. (2003). HCI models, theories, and frameworks: Toward a multidisciplinary science. San Francisco: Morgan Kaufmann. An edited volume on the theoretical foundations of HCI.

Norman, D. A. (1993). Things that make us smart: Defending human attributes in the age of the machine. Reading: Addison-Wesley Pub. Co. This book addresses significant issues in human-computer interaction in a very readable and entertaining fashion.

Patel, V. L., Kannampallil, T., & Kaufman, D. (Eds.). (2015). Cognitive informatics in health and biomedicine: Human computer interaction. London: Springer. This edited book addresses the key gaps on the applicability of theories, models and evaluation frameworks of HCI and human factors for research in biomedical Informatics.

Preece, J., Rogers, Y., & Sharp, H. (2015). Interaction design: Beyond human-computer interaction (4th ed.). West Sussex: Wiley. A very readable and relatively comprehensive introduction to human-computer interaction. A new edition will be available in April, 2019.

Zheng, K., Westbrook, J., Kannampallil, T., & Patel, V. L. (Eds.). (2018). Cognitive informatics: Reengineering clinical workflow for more efficient and safer care. London: Springer. This edited book offers a comprehensive aspect of clinical workflow, supported by the theoretical, methodological, empirical, and pragmatic perspectives from experts in the field.

Questions for Discussion

1. What role do the theories of HCI and cognitive science play in providing insight into principles of system ***usability***, as well as the design of a safer workplace?
2. A large urban hospital is planning to implement a provider order entry system. You have been asked to advise them on system usability and to study the cognitive effects of the system on performance. Discuss the issues involved and suggests some of the steps you would take to study system usability.
3. What are the primary differences between analytic usability evaluation methods and usability testing?
4. What are some of the considerations for choosing analytic approaches for usability evaluation?
5. How does usability impact clinical workflow? How can we provide better cognitive support for workflow?

References

Aarts, J., Ash, J., & Berg, M. (2007). Extending the understanding of computerized physician order entry: Implications for professional collaboration, workflow and quality of care. *International Journal of Medical Informatics, 76*(Suppl 1), S4–S13.

Agrawal, A. (2009). Medication errors: Prevention using information technology systems. *British Journal of Clinical Pharmacology, 67*(6), 681–686.

Allen, M., Currie, L. M., Bakken, S., Patel, V. L., & Cimino, J. J. (2006). Heuristic evaluation of paper-based Web pages: A simplified inspection usability methodology. *Journal of Biomedical Informatics, 39*(4), 412–423. https://doi.org/10.1016/j.jbi.2005.10.004.

Annett, J., & Duncan, K. D. (1967). Task analysis and training design. *Occupational Psychology, 41*, 211–221.

Ash, J. S., Gorman, P. N., Lavelle, M., Payne, T. H., Massaro, T. A., Frantz, G. L., & Lyman, J. A. (2003). A cross-site qualitative study of physician order entry. *Journal of the American Medical Informatics Association, 10*(2), 188–200.

Ash, J. S., Berg, M., & Coiera, E. (2004). Some unintended consequences of information technology in health care: The nature of patient care information system-related errors. *Journal of the American Medical Informatics Association, 11*(2), 104–112. https://doi.org/10.1197/jamia.M1471.

Assila, A., de Oliveira, K. M., & Ezzedine, H. (2014). Towards qualitative and quantitative data integration approach for enhancing HCI quality evaluation. In M. Kurosu (Ed.), *Human-computer interaction. theories, methods, and tools* (pp. 469–480). Cham: Springer International Publishing.

Ben-Assuli, O., Shabtai, I., & Leshno, M. (2015). Using electronic health record systems to optimize admission decisions: The creatinine case study. *Health Informatics Journal, 21*(1), 73–88.

Blandford, D., & Furniss, D. (2006). DiCoT: A methodology for applying distributed cognition to the design of teamworking systems. In S. W. Gilroy & M. D. Harrison (Eds.), *Interactive systems: Design, specification, and verification* (pp. 26–38). Berlin: Springer.

Boockvar, K. S., Santos, S. L., Kushniruk, A., Johnson, C., & Nebeker, J. R. (2011). Medication reconciliation: Barriers and facilitators from the perspectives of resident physicians and pharmacists. *Journal of Hospital Medicine, 6*(6), 329–337.

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, & I. L. McClell (Eds.), *Usability evaluation in industry* (pp. 189–195). London: Taylor & Francis.

Burke, H. B., Sessums, L. L., Hoang, A., Becher, D. A., Fontelo, P., Liu, F., et al. (2014). Electronic health records improve clinical note quality. *Journal of the American Medical Informatics Association, 22*(1), 199–205.

Burrows, D., & Kendall, S. (1997). Focus groups: What are they and how can they be used in nursing and health care research? *Social Sciences in Health, 3*, 244–253.

Carayon, P., Karsh, B., Cartmill, R., Hoonakker, P., Hundt, A. S., Krueger, D., & Wetterneck, T. (2010). *Incorporating health information technology into workflow redesign-summary report*. Rockville: Agency for Healthcare Research and Quality.

Carayon, P., Xie, A., & Kianfar, S. (2014). Human factors and ergonomics as a patient safety practice. *BMJ Quality and Safety, 23*(3), 196–205.

Card, S. K., Moran, T. P., & Newell, A. l. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM, 23*(7), 396–410. https://doi.org/10.1145/358886.358895.

Card, S. K., Newell, A., & Moran, T. P. (1983). *The psychology of human-computer interaction*. Hillsdale: Lawrence Erlbaum Associates.

Carroll, J. M. (2003). *HCI models, theories, and frameworks: Toward a multidisciplinary science*. San Francisco: Morgan Kaufmann.

Caulton, D. A. (2001). Relaxing the homogeneity assumption in usability testing. *Behaviour & Information Technology, 20*, 1–7.

Chan, C. V., & Kaufman, D. R. (2011). A framework for characterizing eHealth literacy demands and barriers. *Journal of Medical Internet Research, 13*(4), e94.

Chipman, S. F., Schraagen, J. M., & Shalin, V. L. (2000). Introduction to cognitive task analysis. In J. M. Schraagen, S. F. Chipman, & V. J. Shute (Eds.), *Cognitive task analysis* (pp. 3–23). Mahwah: Lawrence Erlbaum Associates.

Chung, P., Zhang, J., Johnson, T., & Patel, V. (2003). An extended hierarchical task analysis for error prediction in medical devices. *AMIA Annu Symp Proc, 2003*, 165–169.

Cimino, J. J., Li, J., Mendonça, E. A., Sengupta, S., Patel, V. L., & Kushniruk, A. W. (2000). *An evaluation of patient access to their electronic medical records via the World Wide Web.* Paper presented at the Proceedings of the AMIA Symposium.

Collins, S., Couture, B., Dykes, P., Schnock, K., Knaplund, C., Frank Chang, F., & Cato, K. (2018). *Quantifying and visualizing nursing flowsheet documentation burden in acute and critical care.* Paper presented at the AMIA 2018 Annual Symposium, San Francisco.

Cohen, T., Blatter, B., Almeida, C., Shortliffe, E., & Patel, V. (2006). Distributed cognition in the Psychiatric Emergency Department: A cognitive blueprint of a collaboration in context. *Artificial Intelligence in Medicine; 37*:73–83.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies, 41*, 801–849.

Dalai, V. V., Khalid, S., Gottipati, D., Kannampallil, T. G., John, V., Blatter, B., et al. (2014). Evaluating the effects of cognitive support on psychiatric clini-

cal comprehension. *Aritificial Intelligence in Medicine, 62*(2), 91–104.
Denton, C. A., Soni, H. C., Kannampallil, T. G., Serrichio, A., Shapiro, J. S., Traub, S. J., & Patel, V. L. (2018). Emergency Physicians' Perceived Influence of EHR Use on Clinical Workflow and Performance Metrics. *Applied clinical informatics, 9*(03), 725–733.
Duncan, B., Zheng, L., Furniss, S., Solomon, A., Doebbeling, B., Grando, M., et al. (2018a). *In search of vital signs: A comparative study of EHR documentation*. Paper presented at the AMIA 2018 Annual Symposium, San Francisco.
Duncan, B. J., Zheng, L., Furniss, S. K., Doebbeling, B. N., Grando, A., Solomon, A. J., et al. (2018b). *Perioperative medication management: Reconciling differences across clinical sites.* Paper presented at the Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care.
Duncan, B. J., Kaufman, D. R., Zheng, L., Grando, A., Furniss, S. K., Poterack, K. A., ... & Doebbeling, B. N. (2020). A micro-analytic approach to understanding electronic health record navigation paths. Journal of Biomedical Informatics, 110, 103566.
Elkin, P. L., Sorensen, B., De Palo, D., Poland, G., Bailey, K. R., Wood, D. L., & LaRusso, N. F. (2002). Optimization of a research web environment for academic internal medicine faculty. *Journal of the American Medical Informatics Association, 9*(5), 472–478.
Ericsson, K. A., & Simon, H. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–250.
Ericsson, K. A., & Simon, H. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
Fisher, A., & Fonteyn, M. E. (1995). An exploration of an innovative methodological approach for examining nurses' heuristic use in clinical practice. *Scholarly Inquiry for Nursing Practice, 9*(3), 263–276.
Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381–391.
Fonteyn, M. E., & Grobe, S. J. (1994). Expert system development in nursing: Implications for critical care nursing practice. *Heart & Lung: The Journal of Critical Care, 23*(1), 80–87.
Fonteyn, M., Kuipers, B., & Grobe, S. (1993). A description of think aloud method and protocol analysis. *Qualitative Health Research, 3*, 430–441. https://doi.org/10.1177/104973239300300403.
Fowler, L. P. (1997). Clinical reasoning strategies used during care planning. *Clinical Nursing Research, 6*(4), 349–361.
Funkesson, K. H., Anbäcken, E.-M., & Ek, A.-C. (2007). Nurses' reasoning process during care planning taking pressure ulcer prevention as an example. A think-aloud study. *International Journal of Nursing Studies, 44*, 1109–1119.
Furniss, D., Masci, P., Curzon, P., Mayer, A., & Blandford, A. (2015). Exploring medical device design and use through layers of distributed cognition: How a glucometer is coupled with its context. *Journal of Biomedical Informatics, 53*, 330–341.
Gibbs, A. (1997). Focus groups. *Social Research Update, 19*, 1–7.
Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied, 6*(4), 322.
Grobe, S. J., Drew, J. A., & Fonteyn, M. E. (1991). A descriptive analysis of experienced nurses' clinical reasoning during a planning task. *Research in Nursing & Health, 14*(4), 305–314.
Hall, M., De Jong, M., & Steehouder, M. (2004). Cultural differences and usability evaluation: Individualistic and collectivistic participants compared. *Technical Communication, 51*(4), 489–503.
Harrington, L. (2015). Electronic health record workflow: Why more work than flow? *AACN Advanced Critical Care, 26*(1), 5–9.
Hart, S., & Staveland, L. (1988). Development of NASA TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload*. Amsterdam: North Holland Press.
Hazlehurst, B., McMullen, C. K., & Gorman, P. N. (2007). Distributed cognition in the heart room: How situation awareness arises from coordinated communications during cardiac surgery. *Journal of Biomedical Informatics, 40*(5), 539–551.
Hazlehurst, B., Gorman, P. N., & McMullen, C. K. (2008). Distributed cognition: An alternative model of cognition for medical informatics. *International Journal of Medical Informatics, 77*(4), 226–234.
Hendrick, H. W., & Kleiner, B. M. (1999). *Macroergonomics: An introduction to work system design*. Santa Monica: Human Factors and Ergonomics Society.
Hettinger, A. Z., Roth, E. M., & Bisantz, A. M. (2017). Cognitive engineering and health informatics: Applications and intersections. *Journal of Biomedical Informatics, 67*, 21–33.
Hewing, N. J., Kaufman, D. R., Chan, R. P., & Chiang, M. F. (2013). Plus disease in retinopathy of prematurity: Qualitative analysis of diagnostic process by experts. *JAMA Ophthalmology, 131*(8), 1026–1032.
Hick, W. E. (1951). A simple stimulus generator. *Quarterly Journal of Experimental Psychology, 3*, 94–95.
Holzinger, A., Kosec, P., Schwantzer, G., Debevc, M., Hofmann-Wellenhof, R., & Frühauf, J. (2011). Design and development of a mobile computer application to reengineer workflows in the hospital and the methodology to evaluate its effectiveness. *Journal of Biomedical Informatics, 44*(6), 968–977. https://doi.org/10.1016/j.jbi.2011.07.003.
Horsky, J., Kaufman, D. R., Oppenheim, M. I., & Patel, V. L. (2003). A framework for analyzing the cognitive complexity of computer-assisted clinical order-

ing. *Journal of Biomedical Informatics, 36*(1–2), 4–22.

Horsky, J., Drucker, E. A., & Ramelson, H. Z. (2017). Higher accuracy of complex medication reconciliation through improved design of electronic tools. *Journal of the American Medical Informatics Association, 25*(5), 465–475.

Hripcsak, G., Vawdrey, D. K., Fred, M. R., & Bostwick, S. B. (2011). Use of electronic clinical documentation: Time spent and team interactions. *Journal of the American Medical Informatics Association, 18*(2), 112–117. https://doi.org/10.1136/jamia.2010.008441.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45*, 188–196.

Hysong, S. J., Sawhney, M. K., Wilson, L., Sittig, D. F., Espadas, D., Davis, T., & Singh, H. (2010). Provider management strategies of abnormal test result alerts: A cognitive task analysis. *Journal of the American Medical Informatics Association, 17*(1), 71–77. https://doi.org/10.1197/jamia.M3200.

Jaspers, M. W. (2009). A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence. *International Journal of Medical Informatics, 78*(5), 340–353.

Johnson, C. M., Johnson, T. R., & Zhang, J. (2005). A user-centered framework for redesigning health care interfaces. *Journal of Biomedical Informatics, 38*, 75–87.

Kahn, M. J., & Prail, A. (1994). Formal usability inspections. In *Usability inspection methods* (pp. 141–171). New York: John Wiley & Sons, Inc.

Kannampallil, T. G., Franklin, A., Mishra, R., Cohen, T., Almoosa, K. F., & Patel, V. L. (2013a). Understanding the nature of information seeking behavior in critical care: Implications for the Design of Health Information Technology. *Artificial Intelligence in Medicine, 57*(1), 21–29.

Kannampallil, T. G., Denton, C. A., Shapiro, J. S., & Patel, V. L. (2018b). Efficiency of Emergency Physicians: Insights from an Observational Study using EHR Log Files. *Applied Clinical Informatics, 9*(01), 099–104.

Karahoca, A., Bayraktar, E., Tatoglu, E., & Karahoca, D. (2010). Information system design for a hospital emergency department: A usability analysis of software prototypes. *Journal of Biomedical Informatics, 43*(2), 224–232. https://doi.org/10.1016/j.jbi.2009.09.002.

Karat, C.-M. (1994). *A comparison of user interface evaluation methods.* Paper presented at the Usability Inspection Methods.

Kaufman, D. R., Patel, V. L., Hillman, C., Morin, P. C., Pevzner, J., Weinstock, R. S., et al. (2003a). Usability in the real world: Assessing medical information technologies in patients' homes. *Journal of Biomedical Informatics, 36*(1), 45–60.

Kaufman, D. R., Starren, J., Patel, V. L., Morin, P. C., Hilliman, C., Pevzner, J., et al. (2003b). *A cognitive framework for understanding barriers to the productive use of a diabetes home telemedicine system.* Paper presented at the AMIA Annual Symposium Proceedings.

Kaufman, D. R., Pevzner, J., Rodriguez, M., Cimino, J. J., Ebner, S., Fields, L., et al. (2009). Understanding workflow in telehealth video visits: Observations from the IDEATel project. *Journal of Biomedical Informatics, 42*(4), 581–592.

Kaufman, D. R., Kannampallil, T. G., & Patel, V. L. (2015). Cognition and human computer interaction in health and biomedicine. In *Cognitive informatics for biomedicine* (pp. 9–34). Cham: Springer.

Kitzinger, J. (1995). Introducing focus groups. *British Medical Journal, 311*, 299–302.

Koppel, R., Metlay, J. P., Cohen, A., Abaluck, B., Localio, A. R., Kimmel, S. E., & Strom, B. L. (2005). Role of computerized physician order entry systems in facilitating medication errors. *JAMA, 293*(10), 1197–1203.

Krueger, R. A. (2009). *Focus groups: A practical guide for applied research*. London: Sage.

Kushniruk, A. W., Kaufman, D. R., Patel, V. L., Levesque, Y., & Lottin, P. (1996). Assessment of a computerized patient record system: A cognitive approach to evaluating medical technology. *MD Computing, 13*, 406–415.

Kushniruk, A.W. & Patel, V.L. (2004). Cognitive and usability engineering methods for the evaluation of clinical information systems. Journal of Biomedical Informatics; 37(1): 56–76.

Lau, F., Price, M., Boyd, J., Partridge, C., Bell, H., & Raworth, R. (2012). Impact of electronic medical record on physician practice in office settings: A systematic review. *BMC Medical Informatics and Decision Making, 12*(1), 10.

Lesselroth, B. J., Adams, K., Tallett, S., Wood, S. D., Keeling, A., Cheng, K., et al. (2013). Design of admission medication reconciliation technology: A human factors approach to requirements and prototyping. *HERD: Health Environments Research & Design Journal, 6*(3), 30–48.

Lewis, J. R. (1991). Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. *ACM SIGCHI Bulletin, 23*(1), 78–81.

Lewis, J. R. (1994). Sample sizes for usability studies: Additional considerations. *Human Factors, 36*, 369–378.

Lin, L., Isla, R., Doniz, K., Harkness, H., Vicente, K. J., & Doyle, D. J. (1998). Applying human factors to the design of medical equipment: Patient-controlled analgesia. *Journal of Clinical Monitoring and Computing, 14*(4), 253–263.

Lundgrén-Laine, H., & Salanterä, S. (2010). Think-aloud technique and protocol analysis in clinical decision-making research. *Qualitative Health Research, 20*(4), 565–575. https://doi.org/10.1177/1049732309354278.

MacKenzie, I. S., Zhang, S. X., & Soukoreff, R. W. (1999). Text entry using soft keyboards. *Behaviour and Information Technology, 18*, 235–244.

Mason, J. (2002). *Qualitative researching*. London: Sage Publications.

McDonald, C. J., Callaghan, F. M., Weissman, A., Goodwin, R. M., Mundkur, M., & Kuhn, T. (2014). Use of internist's free time by ambulatory care Electronic Medical Record systems. *JAMA Internal Medicine, 174*(11), 1860–1863.

Middleton, B., Bloomrosen, M., Dente, M. A., Hashmat, B., Koppel, R., Overhage, J. M., et al. (2013). Enhancing patient safety and quality of care by improving the usability of electronic health record systems: Recommendations from AMIA. *Journal of the American Medical Informatics Association, 20*(e1), e2–e8.

Myers, B. A., & Rosson, M. B. (1992, June). Survey on user interface programming. In Proceedings of the SIGCHI conference on *Human factors in computing systems* (pp. 195–202).

Nielsen, J. (1994a). *Usability engineering*. San Francisco: Morgan Kaufmann.

Nielsen, J. (2008). *Usability ROI declining, but still strong*.

Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Seattle, WA.

Norman, D. A. (1986). Cognitive engineering. In D. A. Norman & S. W. Draper (Eds.), *User centered system design* (pp. 31–61). Hillsdale: Lawrence Erlbaum Associates.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Olson, G. M., & Olson, J. S. (2003). Human-computer interaction: Psychological aspects of the human use of computing. *Annual Review of Psychology, 54*(1), 491–516. https://doi.org/10.1146/annurev.psych.54.101601.145044.

Patel, V. L., & Groen, G. J. (1991a). Developmental accounts of the transition from medical student to doctor: Some problems and suggestions. *Medical Education, 25*(6), 527–535.

Patel, V. L., & Groen, G. J. (1991b). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93–125). New York: Cambridge University Press.

Patel, V. L., Arocha, J. F., & Kaufman, D. R. (1994). Diagnostic reasoning and medical expertise. In L. M. Douglas (Ed.), *Psychology of learning and motivation* (Vol. 31, pp. 187–252). San Diego: Academic Press.

Patel, V. L., Arocha, J. F., & Kaufman, D. R. (2001). A primer on aspects of cognition for medical informatics. *Journal of the American Medical Informatics Association, 8*(4), 324–343. https://doi.org/10.1136/jamia.2001.0080324.

Patel, V. L., Kannampallil, T. G., & Kaufman, D. R. (2015). *Cognitive informatics for biomedicine: Human computer interaction in healthcare*. Cham: Springer.

Payne, T. H., Corley, S., Cullen, T. A., Gandhi, T. K., Harrington, L., Kuperman, G. J., et al. (2015). Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. *Journal of the American Medical Informatics Association, 22*(5), 1102–1110.

Plaisant, C., Wu, J., Hettinger, A. Z., Powsner, S., & Shneiderman, B. (2015). Novel user interface design for medication reconciliation: An evaluation of Twinlist. *Journal of the American Medical Informatics Association, 22*(2), 340–349.

Polson, P. G., Lewis, C., Rieman, J., & Wharton, C. (1992). Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies, 36*(5), 741–773. https://doi.org/10.1016/0020-7373(92)90039-N.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-computer interaction*. Harlow: Addison-Wesley Longman Ltd.

Preece, J., Rogers, Y., & Sharp, H. (2015). *Interaction design: Beyond human-computer interaction*. Chichester: John Wiley & Sons.

Ribitzky, R., Sterling, M., & Bradley, V. (2010). *EHR usability pain points survey Q4 2009*. Paper presented at the Oral Presentation at the 2010 Annual HIMSS Conference & Exhibition.

Rogers, Y. (2004). New theoretical approaches for human-computer interaction. *Annual Review of Information Science and Technology, 38*(1), 87–143.

Rogers, Y. (2012). HCI theory: Classical, modern, and contemporary. *Synthesis Lectures on Human-Centered Informatics, 5*(2), 1–129.

Roman, L. C., Ancker, J. S., Johnson, S. B., & Senathirajah, Y. (2017). Navigation in the electronic health record: A review of the safety and usability literature. *Journal of Biomedical Informatics, 67*, 69–79.

Saitwal, H., Feng, X., Walji, M., Patel, V., & Zhang, J. (2010). Assessing performance of an electronic health record (EHR) using cognitive task analysis. *International Journal of Medical Informatics, 79*(7), 501–506.

Schatzberg, M. (2008). *Seeing the invisible, hearing silence, thinking the unthinkable: The advantages of ethnographic immersion*. Paper presented at the APSA 2008 Annual Meeting, Hynes Convention Center, Boston, MA.

Schumacher, R. M., & Lowry, S. Z. (2010). *NIST guide to the processes approach for improving the usability of electronic health records*. National Institute of Standards and Technology.

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). *Designing the user interface: Strategies for effective human-computer interaction*. Hoboken: Pearson.

Simmons, B., Lanuza, D., Fonteyn, M., Hicks, F., & Holm, K. (2003). Clinical reasoning in experienced nurses. *Western Journal of Nursing Research, 25*, 720–724.

Sittig, D. F., & Singh, H. (2010). A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Quality and Safety in Health Care, 19*(Suppl 3), i68–i74.

Unertl, K. M., Weinger, M. B., Johnson, K. B., & Lorenzi, N. (2009). Describing and modeling workflow and information flow in chronic disease care. *Journal of the American Medical Informatics Association, 16*(6), 826–836.

Unertl, K. M., Novak, L. L., Johnson, K. B., & Lorenzi, N. M. (2010). Traversing the many paths of workflow research: Developing a conceptual framework of workflow terminology through a systematic literature review. *Journal of the American Medical Informatics Association, 17*(3), 265–273.

Vankipuram A, Patel VL, Traub S, Shortliffe EH. (2019). Overlaying Multiple Sources of Data to Identify Bottlenecks in Clinical Workflow. Journal of Biomedical Informatics: X. Feb 28:100004.

Vaughn, S., Schumm, J. S., & Sinagub, J. (1996). *Focus group interviews in education and psychology*. Thousand Oaks: Sage.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research, 11*(4), 342–365.

Vicente, K. J. (1999). *Cognitive work analysis*. Mahwah: Lawrence Erlbaum Associates.

Wachter, S. B., Agutter, J., Syroid, N., Drews, F., Weinger, M. B., & Westenskow, D. (2003). The employment of an iterative design process to develop a pulmonary graphical display. *Journal of the American Medical Informatics Association, 10*(4), 363–372.

Weir, C. R., Hammond, K. W., Embi, P. J., Efthimiadis, E. N., Thielke, S. M., & Hedeen, A. N. (2011). An exploration of the impact of computerized patient documentation on clinical collaboration. *International Journal of Medical Informatics, 80*(8), e62–e71.

Xie, A., & Carayon, P. (2015). A systematic review of human factors and ergonomics (HFE)-based healthcare system redesign for quality of care and patient safety. *Ergonomics, 58*(1), 33–49.

Zahabi, M., Kaber, D. B., & Swangnetr, M. (2015). Usability and safety in electronic medical records interface design: A review of recent literature and guideline formulation. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 57*(5), 805–834.

Zhang, J., Johnson, T. R., Patel, V. L., Paige, D. L., & Kubose, T. (2003). Using usability heuristics to evaluate patient safety of medical devices. *Journal of Biomedical Informatics, 36*(1), 23–30. https://doi.org/10.1016/S1532-0464(03)00060-1.

Zhang, J., Patel, V. L., Johnson, T. R., & Shortliffe, E. H. (2004). A cognitive taxonomy of medical error. *Journal of Biomedical Informatics, 37*(3), 193–204.

Software Engineering for Health Care and Biomedicine

Adam B. Wilcox, David K. Vawdrey, and Kensaku Kawamoto

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_6

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What key functions do software applications perform in health care?
- How are the components of the software development lifecycle applied to health care?
- What are the trade-offs between purchasing commercial, off-the-shelf systems and developing custom applications?
- What are important considerations in comparing commercial software products?
- Why do systems in health care, both internally-developed and commercial purchased, require continued software development?

6.1 How Can a Computer System Help in Health Care?

In this chapter, we focus on the software applications and components of health care information systems, and describe how they are used and applied to support health care delivery. We give examples of some basic functions that may be performed by health information systems, and discuss important considerations in how the software may be acquired, implemented and used. This understanding of how a system gets put to use in health care settings will help as you read about the various specific applications in the chapters that follow.

Health care is an information-intensive field. Clinicians are constantly gathering, reviewing, analyzing and communicating information from many sources to make decisions. Humans are complex, and individuals have many different characteristics that are relevant to health care and that need to be considered in decision-making. Health care is also complex, with a huge body of existing knowledge that is expanding at an ever-increasing rate. Software for managing health information is intended to facilitate the use of this information at various points in the care delivery process. Software can determine the ways by which data are obtained, organized and processed to yield information. Software, in terms of design, development, acquisition, configuration and maintenance, is therefore a major component of the field of biomedical informatics. This chapter provides an introduction to some of the practical considerations regarding health information software, including both general software engineering principles, as well as the application of these principles to health care settings.

To this aim, we first describe the major software functions within a health care environment or health information system. While not all functions can be covered in detail, some specific examples are given to indicate the breadth of software applications as well as to provide an understanding of their relevance. We also describe the software development life cycle, with specific applications to health care. We then describe important considerations and strategies for acquiring and implementing software in health care settings. Finally, we discuss emerging trends influencing software engineering related to health information systems. Each system can be considered in regard to what it would take to make it functional in a health care system, and what advantages and disadvantages the software may have, based on how it was created and implemented. Understanding these principles will help you identify the risks and benefits of various applications, so that you can identify how to optimize the positive impact of health information systems.

6.2 Software Functions in Health Care

6.2.1 Case Study of Health Care Software

The following case study illustrates many important functions of health care software.

John Miller is a 42-year old man living in a medium-sized U.S. city. He is married and has

two children. He has type 2 diabetes, but it is currently well controlled and he has no other health concerns. There is some history of cardiovascular disease in his family. John has a primary care physician, Linda Stark, who practices at a clinic that is part of a larger health delivery network, Generation Healthcare System (GHS). GHS includes a physician group, primary and specialty care clinics, a tertiary care hospital and an affiliated health insurance plan.

John needs to make an appointment with Dr. Stark. He logs into the GHS ***patient portal*** *and uses an online scheduling application to request an appointment. While in the patient portal, John also reviews results from his most recent visit and prints a copy of his current medication list in order to discuss the addition of an over-the-counter supplement he recently started taking.*

Before John arrives for his visit, the clinic's scheduling system has already alerted the staff of John's appointment and the need to collect information related to his diabetes. Upon his arrival, Dr. Stark's nurse gathers the requested diabetes information and other vital signs data and enters these into the ***electronic health record*** *(EHR). In the exam room, Dr. Stark reviews John's history, the new information gathered during this visit, and recommendations and reminders provided by the EHR on a report tailored to her patient's medical history. They both go over John's medication list and Dr. Stark notes that, according to the EHR's drug-drug interaction tool, the supplement he is taking may have an interaction with one of his diabetes medications. One of the reminders suggests that John is due for a hemoglobin A1c (HbA1c) test, and Dr. Stark orders this in the EHR. Dr. Stark's nurse, who has been notified of the lab test order, draws a blood sample from John. Before the appointment ends, Dr. Stark completes and signs the clinic note and forwards a visit summary for John to review on the patient portal.*

A few days after his appointment, John receives an email from GHS that notifies him of an important piece of new information in his patient record. Logging into the patient portal application, John sees that his HbA1c test is back. The test indicates that the result is elevated. Dr. Stark has added a note to the result saying that she has reviewed the lab and would like to refer John to the GHS Diabetes Specialty Clinic for additional follow-up. John uses the messaging feature in the patient portal to respond to Dr. Stark and arrange for an appointment. John also clicks on an ***infobutton*** *next to the lab result to obtain more information about the abnormal value. He is linked to patient-focused material about HbA1c testing, common causes for elevated results, and ways this might be addressed. Lastly, John reviews the visit summary note from his appointment with Dr. Stark to remind him about suggestions she had for replacing his supplement.*

At his appointment with the Diabetes Specialty Clinic, John notes that they have access to all the information in his record. A diabetes care manager, Maria, reviews the important aspects of John's medical history. She suggests more frequent monitoring of his laboratory test results and evaluating whether he is able to control his diabetes without changes to his medications. Maria highlights diet and exercise suggestions in his patient portal record that have been shown to help similar patients. When the visit is complete, Maria sends an electronic summary of the visit to Dr. Stark.

A year later, John is experiencing greater difficulty controlling his diabetes. Dr. Stark and Maria have continued to actively monitor his HbA1c and other laboratory test results, and occasionally make changes to his treatment regimen. They use the EHR to visualize laboratory test results and correlate them with changes in medications. Due to a variety of personal and financial challenges, John struggles with adherence to his medication regimen, and he is not maintaining a healthy diet. As a result, his blood sugar has become seriously unstable, and the population health management module of the EHR flags John for urgent evaluation due to a dangerously high home blood glucose reading. Maria confirms the reading with John, collects additional information about his health status, and escalates the issue to Dr. Stark. Dr. Stark then recommends John go to the GHS hospital emergency department (ED) for urgent evaluation. Doctors in the ED access John's electronic record including his medication and lab history, as well as notes from Dr. Stark and Maria, which help them quickly assess his condition

and develop a treatment plan. John is admitted to the hospital, and physicians, nurses, and others caring for him access his longitudinal medical records and document new observations and treatments. They are also able to electronically reconcile his outpatient prescriptions with his inpatient medications to ensure continuity. After a brief hospital stay, John is stabilized and ready to be discharged, with an updated list of medications.

Because Dr. Stark is listed as John's primary care physician, she is notified electronically of both the hospital admission and discharge. She reviews his discharge summary in the EHR and instructs her staff to send a message through the patient portal to John, to let him know she reviewed his inpatient record and to schedule a follow-up appointment.

The GMS EHR is also part of a statewide ***health information exchange*** *(HIE), which allows medical records to be easily shared with health care providers outside the GMS system. This means that if John should need to visit a hospital, emergency department or specialty care clinic outside the GMS network, his record would be available for review and any information entered by these outside providers would be similarly available to Dr. Stark and others within the GMS network. The local and state health departments where John lives are also linked to the HIE. This allows clinics, hospitals and labs to electronically submit information to the health departments for disease surveillance and case reporting purposes.*

Back at home, John's wife, Gina, is able to view his medical records on the GHS patient portal because he has granted her proxy access to his account. This allows her to see notes from Dr. Stark and schedule appointments. Gina also views the hospital discharge instructions that were electronically sent to John's patient record. As she reviews the information about diabetes that GHS had automatically linked to John's record, Gina sees a notification about a clinical research study involving genetic links with diabetes. Concerned about their two children, Gina discusses the study with John, and they review more online materials about the study. Interested in the possible benefits of the research, John electronically volunteers to participate in the study, and he is later contacted by a study coordinator. Because GHS investigators are conducting the study, relevant parts of John's EHR are easily shared with the clinical trials management system.

This fictional case study highlights many of the current opportunities for improving health care delivery, including improved access to care, increased patient engagement, shared patient-provider decision-making, better care management, medication reconciliation, improved transitions of care, population health management, and research recruitment. In the case study, each of these goals required software to make health information accessible to the correct individuals at the proper time.

In today's health care system, few individuals enjoy the interaction with software depicted in the *John Miller* case study. Although the functions described in the scenario exist at varying levels of maturity, most health care delivery institutions have not connected all the functions together as described. The current role of software engineering in health care is therefore two-fold: to design and implement software applications that provide required functions, and to connect these functions in a seamless experience for both the clinicians and the patients.

The case study highlights the usefulness of several functions provided by health care software applications for clinicians, patients, and administrators. Some of these functions include:

1. Acquiring and storing data
2. Summarizing and displaying data
3. Facilitating communication and information exchange
4. Generating alerts, reminders, and other forms of decision support
5. Supporting educational, research, and public health initiatives

6.2.2 Acquiring and Storing Data

The amount of data needed to describe the health and health care of even a single person is huge. Health professionals require assistance with data acquisition to deal with the data that must be collected and processed. One of the first uses of computers in a medi-

cal setting was the automatic analysis of blood specimens and other body fluids by instruments that measure chemical concentrations or that count cells and organisms. These systems generated printed or electronic results to health care workers and identified values that were outside normal limits. Computer-based patient monitoring that collected physiological data directly from patients were another early application of computing technology (see ▶ Chap. 19). These systems provided frequent, consistent collection of vital signs, electrocardiograms (ECGs), and other indicators of patient status. More recently, researchers have developed medical imaging applications as described in ▶ Chaps. 9 and 20, including computed tomography (CT), magnetic resonance imaging (MRI), and digital subtraction angiography. The calculations for these computationally intensive applications cannot be performed manually; computers are required to collect and manipulate millions of individual observations.

Early computer-based medical instruments and measurement devices provided results only to human beings. Today, most instruments can transmit data directly into the EHR, although the interfaces can still be awkward and poorly standardized (see ▶ Chaps. 5 and 8). Computer-based systems that acquire information directly from patients are also data-acquisition systems; they free health professionals from the need to collect and enter demographic and health history information.

Various departments within a hospital use computer systems to store clinical data. For instance, clinical laboratories use information systems to keep track of orders and specimens and to report test results; most pharmacy and radiology departments use computers to perform analogous functions. Their systems may connect to outside services (e.g., pharmacy systems are typically connected to one or more drug distributors so that ordering and delivery are rapid and local inventories can be kept small). By automating processing in areas such as these, health care facilities are able to provide efficient service, reduce labor costs, and minimize errors.

6.2.3 Summarizing and Displaying Data

Computers are well suited to performing tedious and repetitive data-processing tasks, such as collecting and tabulating data, combining related data, and formatting and producing reports. They are particularly useful for processing large volumes of data.

Data acquired by computer systems can be detailed and voluminous. Data analysis systems must aid decision makers by reducing and presenting the intrinsic information in a clear and understandable form. Software can be used to create useful visualizations that facilitate trend analysis and compute secondary parameters (e.g., means, standard deviations, rates of change) to help identify abnormalities. Clinical research systems have modules for performing powerful statistical analyses over large sets of patient data. When employing such tools, research investigators should have insight into the methods being used. For clinicians, graphical displays are useful for interpreting data and identifying trends.

Fast retrieval of information is essential for any computer system. Data must be well organized and indexed so that information recorded in an EHR system can be easily retrieved. Here the variety of users must be considered. Obtaining recent information about a patient entering the office differs from the needs that a research investigator will have in accessing the same data. The query interfaces provided by EHRs and clinical research systems assist researchers in retrieving pertinent records from the huge volume of patient information. Recently, there has been increasing industry adoption of the **Health Level 7 International** (HL7) Fast Healthcare Interoperability Resources (FHIR) standard for sharing data on a patient-by-patient basis. The FHIR standard is being adapted for population-level data sharing through the FHIR Bulk Data initiative. As discussed in ▶ Chap. 21, bibliographic retrieval systems are also an essential component of health information services.

6.2.4 Facilitating Communication and Information Exchange

In hospitals and other large-scale health care institutions, myriad data are collected by multiple health professionals who work in a variety of settings; each patient receives care from a host of providers—nurses, physicians, technicians, pharmacists, and so on. Communication among the members of the team is essential for effective health care delivery. Data must be available to decision makers when and where they are needed, independent of when and where they were obtained. Computers help by storing, transmitting, sharing, and displaying those data. As described in ▶ Chaps. 2 and 12, the patient record is the primary vehicle for communication of clinical information. The limitation of the traditional paper-based patient record is the concentration of information in a single location, which prohibits simultaneous entry and access by multiple people. Hospital information systems (HISs; see ▶ Chap. 13) and EHR systems (▶ Chap. 12) allow distribution of many activities, such as admission, appointment, and resource scheduling; review of laboratory test results; and inspection of patient records to the appropriate sites.

Information necessary for specific decision-making tasks is rarely available within a single computer system. Clinical systems are installed and updated when needed, available, and affordable. Furthermore, in many institutions, inpatient, outpatient, and financial activities are supported by separate organizational units. Patient treatment decisions require inpatient and outpatient information. Hospital administrators must integrate clinical and financial information to analyze costs and to evaluate the efficiency of health care delivery. Similarly, clinicians may need to review data collected at other health care institutions, or they may wish to consult published biomedical information. Communication networks that permit sharing of information among independent computers and geographically distributed sites are now widely available. Actual integration of the information they contain requires additional software, adherence to standards, and operational staff to keep it all working as technology and systems evolve.

6.2.5 Generating Alerts, Reminders, and Other Forms of Decision Support

In the end, all the functions of storing, displaying and transmitting data support decision-making by health professionals, patients, and their caregivers. The distinction between decision-support systems and systems that monitor events and issue alerts is not clear-cut; the two differ primarily in the degree to which they interpret data and recommend patient-specific action. Perhaps the best-known examples of decision-support systems are the clinical consultation systems or event-monitoring systems that use population statistics or encode expert knowledge to assist physicians in diagnosis and treatment planning (see ▶ Chap. 22). Similarly, some nursing information systems help nurses to evaluate the needs of individual patients and thus assist their users in allocating nursing resources. ▶ Chapter 22 discusses systems that use algorithmic, statistical, or artificial-intelligence (AI) techniques to provide advice about patient care.

Timely reactions to data are crucial for quality in health care, especially when a patient has unexpected problems. Data overload, created by the ubiquity of information technology, is as detrimental to good decision making as is data insufficiency. Data indicating a need for action may be available but are easily overlooked by overloaded health professionals. Surveillance and monitoring systems can help people cope with all the data relevant to patient management by calling attention to significant events or situations, for example, by reminding doctors of the need to order screening tests and other preventive measures (see ▶ Chaps. 12 and 22) or by warning them when a dangerous event or constellation of events has occurred.

Laboratory systems routinely identify and flag abnormal test results. Similarly, when

patient-monitoring systems in intensive care units detect abnormalities in patient status, they sound alarms to alert nurses and physicians to potentially dangerous changes. A pharmacy system that maintains computer-based drug-profile records for patients can screen incoming drug orders and warn physicians who order a drug that interacts with another drug that the patient is receiving or a drug to which the patient has a known allergy or sensitivity. By correlating data from multiple sources, an integrated clinical information system can monitor for complex events, such as interactions among patient diagnosis, drug regimen, and physiological status (indicated by laboratory test results). For instance, a change in cholesterol level can be due to prednisone given to an arthritic patient and may not indicate a dietary problem.

6.2.6 Supporting Educational, Research, and Population and Public Health Initiatives

Rapid growth in biomedical knowledge and in the complexity of therapy management has produced an environment in which students cannot learn all they need to know during training—they must learn how to learn and must make a lifelong educational commitment. Today, physicians and nurses have available a broad selection of computer programs designed to help them to acquire and maintain the knowledge and skills they need to care for their patients. The simplest programs are of the drill-and-practice variety; more sophisticated programs can help students to learn complex problem-solving skills, such as diagnosis and therapy management (see ▶ Chap. 21). Computer-aided instruction provides a valuable means by which health professionals can gain experience and learn from mistakes without endangering actual patients. Clinical decision-support systems and other systems that can explain their recommendations also perform an educational function. In the context of real patient cases, they can suggest actions and explain the reasons for those actions.

As health care increasingly shifts to a mode of care based on population health rather than episodic health care transactions, there is increasing need for information systems to monitor and manage individuals' health outside the context of clinical visits. Surveillance also extends beyond the health care setting. Appearances of new infectious diseases, unexpected reactions to new medications, and environmental effects should be monitored. Thus the issue of data integration has a national or global scope (see the discussion of the National Health Information Infrastructure in ▶ Chaps. 1 and 16 that deals with public health informatics).

6.3 Software Development and Engineering

Clearly, software can be used in many different ways to manage and manipulate health information to facilitate health care delivery. However, just using a computer or a software program does not improve care. If critical information is unavailable, or if processes are not organized to operate smoothly, a computer program will only expose challenges and waste time of clinical staff that could be better applied in delivering care. To be useful, software must be developed with an understanding of its role in the care setting, geared to the specific functions that are required, and developed correctly. To be used, software must be integrated to support the users' workflow. We will discuss both aspects of software engineering – development and integration.

6.3.1 Software Development

Software development can be a complex, resource-intensive undertaking, particularly in environments like health care where safety and security provide added risk. The **software development life cycle** (SDLC) is a framework imposed over software development in order to better ensure a repeatable, predictable process that controls cost and improves quality of

the software product (usually an application). SDLC is a subset of the systems development life cycle, focusing on the software component of a larger system. In practice, and particularly in heath care, software development encompasses more than just the software, often stretching into areas such as process reengineering in order to maximize the benefits of the software product. Although SDLC most literally applies to an in-house development project, all or most of the life cycle framework is also relevant to shared development and even purchase of commercial off-the-shelf (COTS) software. The following is an overview of the phases of the SDLC.

6.3.1.1 Planning/Analysis

The software development life cycle begins with the formation of a project goal during the planning phase. This goal typically derives from an organization's or department's mission/vision, focusing on a particularly need or outcome. This is sometimes called project conceptualization. Planning includes some initial scoping of the project as well as resource identification (including funding). It is important to address what is not included in the project in order to create appropriate expectations for the final product. A detailed analysis of current processes and needs of the target users is often done. As part of the analysis, specific user requirements are gathered. Depending on the development process, this might include either detailed instructions on specific functions and operating parameters or more general user stories that explain in simple narrative the needs, expected workflow and outcomes for the software. It is important that users of the system are consulted, as well as those in the organization who will implement and maintain the software. The decision of whether to develop the software in-house, partner with a developer, or purchase a vendor system will likely determine the level of detail needed in the requirements. Vendors will want very specific requirements that allow them to properly scope and price their work. The requirements document will usually become part of a contract with a vendor and will be used to determine if the final product meets the agreed specification for the software. In-house development can have less detailed requirements, as the contract to build the software is with the organization itself, and can allow some evolution of the requirements as the project progresses. However, the more flexibility that is allowed and the longer changes or enhancements are permitted, the higher the likelihood of "scope creep" causing schedule and cost overruns.

Other tasks performed during analysis include an examination of existing products and potential alternative solutions, and, particularly for large projects, a cost/benefit analysis. A significant, and frequently overlooked, aspect of the planning and analysis phase is to determine outcome measures that can be used during the life cycle to demonstrate progress and evaluate success or failure of the project. These measures can be refined and details added as the project progresses. The planning and analysis phase typically ends when a decision to proceed is made, along with at least a rough plan of how to implement the next steps in the SDLC. If the organization decides to purchase a solution, a request for proposals (RFP) that contains the requirements document is released to the vendor community.

The planning and analysis stage of software development is perhaps both the most difficult and the most important stage in the development lifecycle as it is applied to health care. Requirements for software in health care are inherently difficult to define for many reasons. Health care practice is constantly changing, and as new therapies or approaches are discovered and validated, these new advancements can change the way care is delivered. In addition, the end-users of health care software are comparatively advanced relative to other industries. Unlike industries where front-line workers may be directed by supervisors with more advanced training and greater flexibility in decision making, in health care the front-line workers are often physicians, who are often the most highly-trained workers in the system (although not necessarily the most advanced with respect to computer literacy) and require the greatest flexibility for decisions. This flexibility makes it difficult to define workflows or even get indications of the workflows being followed,

since physicians will not always make explicit what actions or plans are being pursued. This flexibility is important for patient care, because it allows front-line clinicians to adapt appropriately to different settings, staffing levels, and specialties. The need for flexibility is such that defining requirements for software that could reduce flexibility is criticized as "cookbook" medicine, constituting a common reason for resistance to software adoption. However, this resistance is not just characteristic of software – clinical guidelines and other approaches to structured or formalized care processes can also be criticized, and the challenge of applying discovered-knowledge to clinical care processes remains difficult.

Over time, however, there have been some successful efforts that have defined standard requirements for health information software. Among the most notable efforts have been in EHRs, where organizations have created lists of requirements and certified systems that match those requirements. The Certification Commission for Health Information Technology (CCHIT) began in 2004 and defined criteria for electronic health records' functionality, interoperability and security (Leavitt and Gallagher 2006). Later, the certification approach was adopted by the Office of the National Coordinator of Health Information Technology (ONC) in 2010, when a list of EHR functions that were most related to "meaningful use" of EHRs was established (Blumenthal and Tavenner 2010). Such "meaningful use" of EHRs came with significant financial incentives administered by the Centers for Medicare and Medicaid Services (CMS), leading to a rapid increase in the adoption of EHRs meeting these requirements (Washington et al. 2017).

6.3.1.2 Design

During the Design phase, potential software solutions are explored. System architectures are examined for their abilities to meet the needs stated in the requirements. Data storage and interface technologies are assessed for appropriate fit. User front-end solutions are investigated to assess capabilities for required user input and data display functions. Other details, such as security, performance, and internationalization are also addressed during design. Analysts with domain knowledge in the target environment are often employed during this phase in order to translate user requirements into suitable proposals. Simple mock-ups of the proposed system may be developed, particularly for user-facing components, in order to validate the design and identify potential problems and missing information. Closely related to this, an integrated, automated testing architecture, with appropriate testing scripts/procedures, may be designed in this phase in order to ensure the software being developed meets quality standards and is responsive to the requirements. The depth and completeness of the design is contingent on the software development process, as well as other factors. In some cases, the entire design is completed before moving on to software coding. In other development strategies, a high-level system architecture is designed but the details of the software components are delayed until each component or component feature is being created. The pros and cons of these approaches are discussed later in this chapter. For vendor-developed systems, the purchasing organization will often hold design reviews and demonstrations of mock-ups or prototypes with the vendor to assess the solutions. In the case of COTS software, the purchasing organization relies on the vendor's system description and reviews from third parties, supplemented by system demonstrations, to determine the appropriateness of the design. As with the Analysis phase, it is important to include the target users and IT operations personnel in the design reviews.

Ideally, the software could be designed solely around the care requirements and the use of information. However, rarely are the clinical requirements of the use case the only consideration. In the design phase, other requirements are considered, such as the software cost and how it integrates with an existing health IT strategy of an organization. Resources applied to a development project are not available for other potential projects, so costs are always influential. The design phase must consider various alternatives to

meet the most important requirements, recognizing trade-offs and contingency approaches. Additional considerations are how the software will support long-term needs, not just the immediate requirements that have been identified. Clinicians and clinical workflow analysts are often the primary participants in the requirements analysis stage, whereas informaticians are more prominent in the design phase. This is because during this latter phase the clinical goals and strategies are considered together with what can be vastly different design approaches, and the ability to consider the various strengths and weaknesses of these different approaches is critical. Often, design considerations are between custom development, purchasing niche applications, or purchasing components of a monolithic EHR. The considerations of development versus COTS software is discussed in more detail in the Acquisition Strategy section (► 6.3.3.1) below.

6.3.1.3 Development

Coding of the software is done during the Development phase of the SDLC. The software engineers use the requirements and system designs as they program the code. Analysts help resolve questions about requirements and designs for the programmers when it is unclear how software might address a particular feature. The software process defines the pace and granularity of the development. In some cases, an entire software component or system is developed at once by the team. In other cases, the software is broken down into logical pieces and the programmers only work on the features that are relevant to the piece they are currently working on. As software components are completed, unit tests are run to confirm the component is free of known bugs and produces expected outputs or results.

In health care, development includes coding of custom software as well as configuration of COTS software. Health care practices across institutions (and even within larger organizations) are so variable that all software requires some – often substantial – configuration. Configuration can range from assigning local values to generic variables within the software, to complete development of documentation templates, order sets, clinical decision support rule, reports, and so on. In fact, configuration can be so considerable that institutions may use an internal brand name for the software and configuration project that is different from the name of the COTS software, which represents their local configuration. This configuration is often done using tools built specifically for the commercial software, which facilitate the integration of the configuration products into the software infrastructure. The tools can be complex, requiring significant training for developers. Typically, tools work well for basic configuration and may also have advanced functionality that can be used to configure more complicated functionality. The most intensive time investment for configuration is typically when the tools do not directly support certain configurations, and developers must find approaches to creatively adapt the development "around the tools."

6.3.1.4 Integration and Test

For complex software projects consisting of several components and/or interfaces with outside systems, an Integration phase in the SDLC is employed to tie together the various pieces. Some aspects of the software integration are likely done during the Development phase by simulating or mocking the outputs to, and inputs from, other systems. During Integration, these connections are finalized. Simulations are run to demonstrate functional integration of the various system components. Once the various components are integrated, a thorough testing regimen is conducted in order to prove the end-to-end operation of the entire software system. Specific test scenarios are run with known inputs and expected outputs. This is typically done in a safe, non-operational environment in order to avoid conflicts and unnecessary risk to production environments, although some inbound information from live systems may be used to verify scenarios that are difficult to simulate.

Testing and integration in health care are similar to other complex environments, in that it can be difficult to create a testing environment that matches the dynamics of the

real-world setting. Generally, testing is done around multiple use cases or case studies, using data to support the cases. In a production environment, however, there may be data and information that do not match the case studies, since both people and health care are complex. As a result, internally-developed applications are often provisionally used in a "pilot" phase as part of testing. For COTS software, companies may use simulation laboratories that try to mimic the clinical environment, or work with specific health care organizations as development and testing partners. Later, however, this can lead to challenges if data representing the dynamics of one organization are not easily transferable, and software must be further tested with new environments. Software transferability between institutions has been demonstrated in studies, even for specific applications (Hripcsak et al. 1998). Another challenge is that with current privacy laws, organizations are more reluctant to release data to vendors for testing.

6.3.1.5 Implementation

Once the software passes integration testing it moves to the implementation phase. In this phase, the software is installed in the live environment. In preparation for installation, server hardware, user devices, network infrastructure, changes specific to individual facilities, etc., may need to be implemented and tested as well. In addition, user training will be performed in the weeks before the software goes live. Any changes to policies and procedures required by the software will also be implemented in the build-up to installation.

Health care presents interesting considerations in each phase of the software development cycle, but the challenges have been more visible in implementation than any other phase. This may be because health IT, while intended to facilitate more efficient workflows with information, is still disruptive. Disruption happens most during implementation, when clinicians actually begin using the software, and studies have shown that during this time clinical productivity does decline (Shekelle et al. 2006). If users do not perceive that the benefits are sufficient to justify this disruption, or if the efficiency does not improve quickly enough after the initial implementation, they may choose to disregard the software or even revolt against its implementation. There have been prominent examples in biomedical informatics of software implementations failing during implementation (Bates 2006; Smelcer et al. 2009; Sullivan 2017), and even studies demonstrating harm (Han et al. 2005). Because of these risks, health IT professionals need to be flexible in implementation, and adapt the implementation strategies to how the system is adopted. Users have been shown to use health IT software in different ways for different benefits, and may need incentives or prodding to advance to different levels of use.

6.3.1.6 Verification and Validation

To ensure that the software satisfies the original requirements for the system and meets the need of the organization, a formal verification and validation of the software is performed. The implementing organization *verifies* that the software has the features and performs all the functions specified in the requirements document. The software is also *validated* to show that it performs according to specified operational requirements, that it produces valid outputs, and that it can be operated in a safe manner. For purchased software, the verification and validation phase is used by the purchasing organization in order to officially accept the software.

Since clinicians often use software at different levels or in different ways, tracking patterns of use can be an important approach for verification and validation of software in health care. Additionally, because they have experience working in complicated environments, users can be good at identifying inconsistencies in data or software functions. Two approaches that have been used and can be successful for validation are monitoring use, and facilitating user feedback.

6.3.1.7 Operations and Maintenance

Software eventually enters an operations and maintenance (O&M) phase where it is being regularly used to support the operational needs of the organization. During this phase,

an O&M team will ensure that the software is operating as desired and will be fielding the support needs of the users. Updates may need to be installed as new versions of the software are released. This may require new integration and testing, implementation, and verification and validation steps. Ongoing training will be required for new users and system updates. The O&M team may conduct regular security reviews of the system and its use. Data repositories and software interfaces will be monitored for proper operation and continued information validity. Software bugs and feature enhancement requests will be collected. These may drive an entire new development life cycle as new requirements persuade an organization to explore significant upgrades to its current software or even an entirely new system.

Maintenance is a demanding task in health information software. It involves correcting errors; adapting configurations and software to growth, new standards, and new regulations; and linking to other information sources. Maintenance tasks can exceed by more than double the initial acquisition costs, making it a substantial consideration that should affect software design. COTS suppliers often provide maintenance services for 15–30% of the purchase price annually, but custom development or configuration maintenance must be supported by the purchasing organization. If the software is not maintained, it can quickly become unusable in a health care setting. Indeed, optimization of COTS EHRs is a central and ongoing focus of applied clinical informatics, and this is likely to continue for the foreseeable future.

6.3.1.8 Evaluation

An important enhancement to the SDLC suggested by Thompson et al. (1999) is the inclusion of an evaluation process during each of the phases of the life cycle. The evaluation is influenced by risk factors that may affect a particular SDLC segment. An organization might perform formative evaluations during each phase, depending on specific needs, in order to assess the inputs, processes and resources employed during development. During Verification and Validation or O&M, a summative evaluation may be performed to assess the outcome effects, organizational impact, and cost-benefit of the software solution.

Health IT is considered an intervention into the health care delivery system, so evaluations have been done and published as comparative studies in the clinical literature (Bates et al. 1998; Campanella et al. 2016; Evans et al. 1998; Hunt et al. 1998; Jones et al. 2014). These evaluations, and syntheses of multiple studies, have identified areas of impact and areas where the effect of health IT software is inconsistent. Researchers have also noted that most of these studies have occurred in institutions where software was developed internally, with disproportionate under-representation of COTS software systems in evaluations, especially considering that most health care institutions use COTS rather than internal development (Chaudhry et al. 2006). It is hoped that the existing evaluations can be a model for software evaluations of COTS, to clarify their impact on care.

6.3.2 Software Development Models

Different software development processes or methods can be used in an SDLC. The **software development process** describes the day-to-day methodology followed by the development team, while the life cycle describes a higher-level view that encompasses aspects that take place well before code is ever written and after an application is in use. The following are two of the most common examples of different development processes in clinical information systems development.

6.3.2.1 Waterfall Model

The Waterfall model of software development suggests that each step in the process happens sequentially, as shown in ◘ Fig. 6.1. The term "Waterfall" refers to the analogy of water cascading downward in stages. A central concept of the Waterfall methodology is to solidify all of the requirements, establish complete functional specifications, and create the final soft-

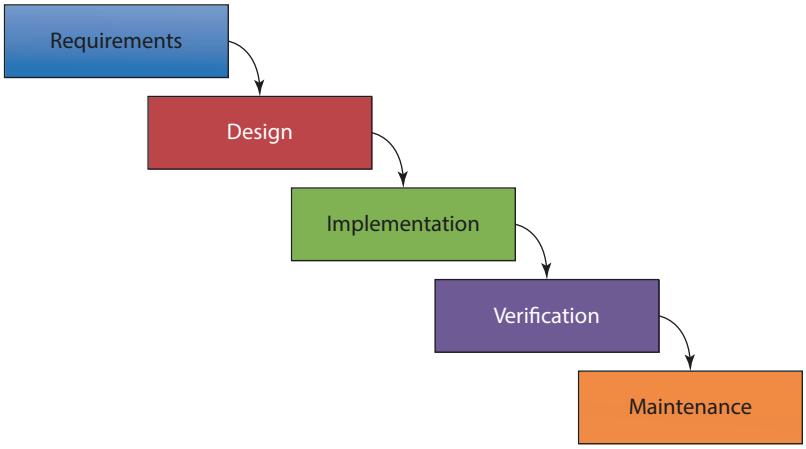


Fig. 6.1 The Waterfall model of software development

ware design prior to performing programming tasks. This concept is referred to as "Big Design Up Front," and reflects the thinking that time spent early-on making sure requirements and design are correct saves considerable time and effort later. Steve McConnell, an expert in software development, estimated that "...a requirements defect that is left undetected until construction or maintenance will cost 50–200 times as much to fix as it would have cost to fix at requirements time" (McConnell 1996).

The waterfall model provides a structured, linear approach that is easy to understand. Application of the model is best suited to software projects with stable requirements that can be completely designed in advance. In practice, it may not be possible to create a complete design for software a priori. Requirements and design specifications can change even late in the development process. Clients may not know exactly what requirements they need before reviewing a working prototype. In other cases, software developers may identify problems during the implementation that necessitate reworking the design or modifying the requirements.

6.3.2.2 Agile Models

In contrast to the Waterfall model, modern software development approaches have attempted to provide more flexibility, particularly in terms of involving the customer throughout the process. In 2001, a group of software developers published the Manifesto for Agile Software Development, which emphasizes iterative, incremental development and welcomes changes to software requirements even late in the development process (Beck et al. 2001).

Agile development eschews long-term planning in favor of short iterations that usually last from 1 to 4 weeks. During each iteration, a small collaborative team (typically 5–10 people) conducts planning, requirements analysis, design, coding, unit testing, and acceptance testing activities with direct involvement of a customer representative. Multiple iterations are required to release a product, and larger development efforts involve several small teams working toward a common goal. The agile method is value-driven, meaning that customers set priorities at the beginning of each iteration based on perceived business value.

Agile methods emphasize face-to-face communication over written documents. Frequent communication exposes problems as they arise during the development process. Typically, a formal meeting is held each morning during which team members report to each other what they did the previous day, what they intend to do today, and what their roadblocks are. The brief meeting, sometimes called a "stand-up," "scrum," or "huddle," usually lasts 5–15 minutes, and includes the

development team, customer representatives and other stakeholders. A common implementation of agile development is Extreme Programming.

6.3.3 Software Engineering

The software development life cycle can be used to actually create the software, and understanding it is critical for those developing software in biomedical informatics. However, as the field has expanded, software has matured to the point that it is developed by and available from commercial companies, so that software development has become less of a concern for most of the field. A more important consideration in biomedical informatics has been the strategy of whether to develop and how to develop. Software vendors can spread development costs over multiple organizations, rather than one organization having to fund the full development, which can make purchasing software economically advantageous. On the other hand, the core requirements for software continue to change, and sometimes organizations need specific capabilities that are not met by existing vendor software options. In addition to software development, informaticians often participate in software acquisition, as well as in subsequent enhancements to acquired software.

6.3.3.1 Software Acquisition

In health care information technology applications, a significant question is whether to develop the software internally or purchase an existing system from a vendor. Whether to "build vs. buy" is a core decision in planning and implementation.

Considerations for purchasing software begin with how the software will be selected. Software can be a component of a monolithic vendor system, be a secondary application sold by the same vendor as the EHR, or be "best-of-breed," meaning the software that meets the requirements best, independent of its architecture or source. Another consideration is whether the software needs to integrate with other applications. Some specialty applications require minimal data sharing with other software, while other applications must be tightly integrated with existing systems to achieve a benefit. Two examples are a picture archiving and communications system (PACS) and a medication reconciliation tool. Perhaps the most important requirement for a PACS is to provide access to images for a radiologist, who can then "read" the image and document a report which can be transferred into the EHR as a static document. On the other hand, a medication reconciliation tool may need substantial integration with medication ordering and administration modules in an EHR to support workflows of the care team. Another consideration, related to integration, is the storage mechanism. A standalone system will likely have a separate database, while an integrated system may be able to store and retrieve data using a common data repository. User interface deployment is also important, and possibilities include Web-based clients, **thin clients** (e.g., Citrix), and locally-installed **thick-client** applications. Greater functionality may exist with a thick-client application, but Web-based and thin-client tools are easier to update and distribute to users. Finally, security and privacy considerations are critical in health care, and can influence both the requirements and design of software. Security considerations can include whether **user authentication** is shared with other applications, or what data access events are audited for identifying potential security threats.

Most healthcare delivery organizations today use commercial – as opposed to locally developed – EHRs. But in reality, there is still a mix between building and buying health information technology. As mentioned, organizations using commercial systems require substantial local configuration that ranges from application-specific parameter configuration to coordinating multiple software applications to link together. Even when there is a commitment to limit local configuration, there may still be separate systems, local configuration or even development with data warehousing and analytics solutions for the EHR data. There is no single solution, commercial or internally-developed, that meets all

the health information needs of most health care organizations, and many implementations involve a mixture of software from multiple vendors. While there can be advantages to allowing best-of-breed, a current trend among organizations is to consolidate as much functionality as possible with one vendor. Another observed trend is for organizations that build systems to consider purchasing COTS, due to the substantial maintenance costs associated with in-house development and the increased functionality often available with vendor solutions. At present, virtually all health care organizations that utilize an EHR in the United States use a COTS solution or are in the process of migrating to such a solution.

Usually, if vendor software exists, it is more cost-efficient to purchase the software than build comparable capabilities internally for use at a single organization. This is because the vendor can spread development costs over multiple organizations, rather than one organization having to fund the full development. In fact, few organizations have the existing infrastructure and personnel to consider internal development for anything other than small applications. However, those few institutions with developed health information systems are notable for the success of their software. So while the costs may be higher for internal development, the benefits may also be higher. Furthermore, such solutions may be potentially licensed to other organizations, thereby spreading the cost of development across multiple organizations. Still, these institutions have invested many years to build an infrastructure that makes these benefits possible, and it is unlikely that many organizations can afford the time and resource investment to follow the same model. Even within historically internally-developing organizations, buying systems that can integrate with the existing system is oftentimes more efficient than development. An appropriate general guide is therefore, "Buy where you can, build where you can't."

Once an organization decides to acquire a health information system, there are many other decisions beyond whether to build or buy. In fact, since the costs in time and money are oftentimes prohibitive for internal development, the decision to build is typically the easiest decision to make for a large health information system such as the EHR. Coupled with the Meaningful Use EHR Incentive Program, adoption of at least basic EHRs in the United States is very high, exceeding 90% (see ◘ Figs. 6.2, 6.3 and 6.4).[1]

Once a decision to purchase a commercial system is made, the next decision is what system to purchase. There is a wide variation in the functionalities between different EHR systems, even though certification efforts have defined basic functions that each system should have (see ◘ Fig. 6.3). Even systems with the same certified functions may approach the functions so differently that some implementations will be incongruent to an organization.[2] Key factors an organization should consider when choosing a system include (a) the core functionality of the software, including integration with other systems, (b) total system cost, (c) the service experience of other customers, and (d) the system's certification status. Some organizations have performed systematic reviews of different commercial software offerings that can be a helpful start to identify possible vendors and understand variations between systems. For example, KLAS Research publishes periodic assessments of both software functions and vendor performance that can be used to identify potential software products. However, since systems are complex, it is important to meet with and discuss experiences with actual organizations that have used the software. This is typically done through site visits to existing customer organizations. It is also common for organizations to make a broad request of vendors for proposals to address a specific software need, especially

1 ► https://v.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php and ► https://dashboard.healthit.gov/quickstats/pages/FIG-Hospital-EHR-Adoption.php, from ► https://dashboard.healthit.gov/quickstats/quickstats.php

2 ► https://dashboard.healthit.gov/quickstats/pages/FIG-Vendors-of-EHRs-to-Participating-Hospitals.php (last accessed June 3, 2020).

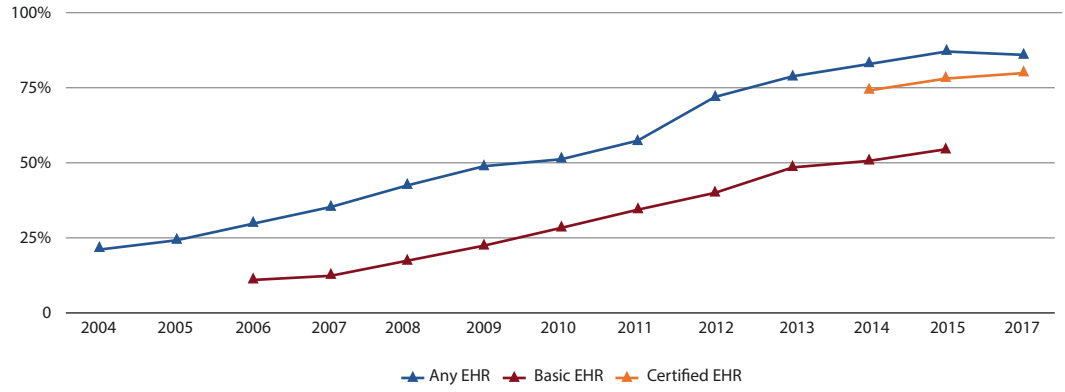


Fig. 6.2 Office-based physician EHR adoption, from ONC (2019a)

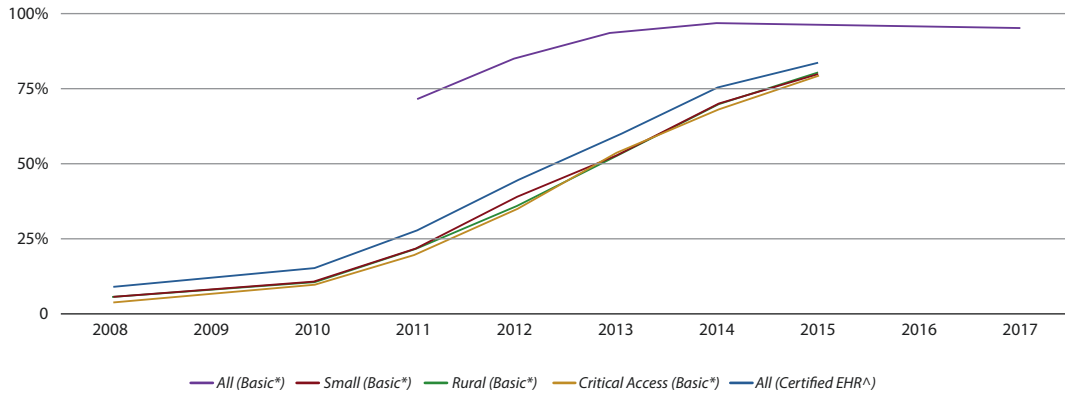


Fig. 6.3 Non-federal acute care hospital EHR adoption, from ONC (2017b)

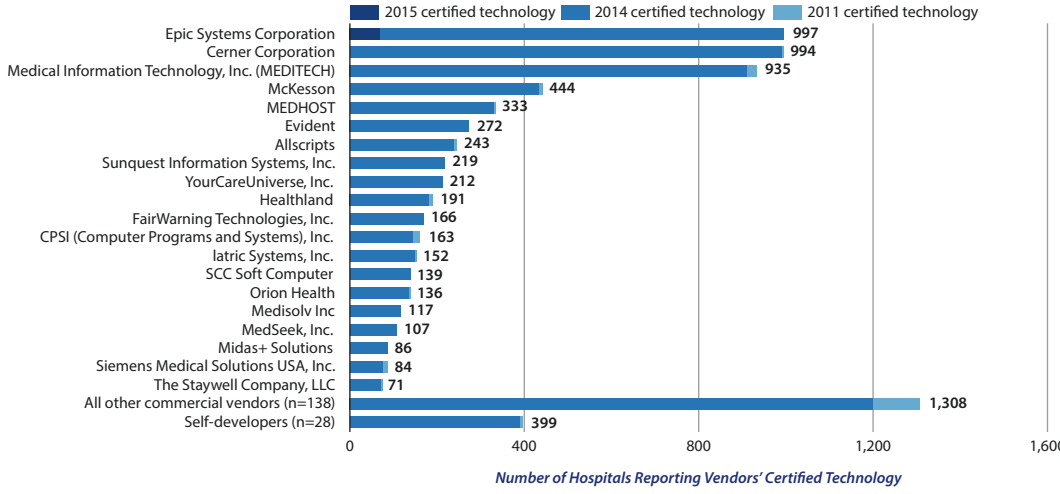


Fig. 6.4 Health IT developers product use by hospitals participating in the Medicare EHR, from ONC (2017a)

when the needs are not standard components of EHR software.

After a commercial product is selected, an organization must then choose how extensive the software will be. EHR companies typically have a core EHR system, with additional modules that have either been developed or acquired and integrated into their system. The set of modules used by each institution varies. One organization may use the core EHR system and accompanying modules for certain specialties, such as internal medicine and family practice, while choosing to purchase separate best-of breed software for other specialties, like obstetrics/gynecology and emergency medicine, even when the core EHR vendor has functional modules for those areas. Another organization may choose to purchase and implement all specialty systems offered from the core EHR vendor, and only purchase other software if a similar module is not available from the vendor. These decisions also must be made for all ancillary systems, including laboratory, pharmacy, radiology, etc. This is both a pre-implementation decision and a long-term strategy. Once the EHR is implemented, many specialties that were not included in the initial implementation plan may request software and data integration, depending on the success of the EHR implementation.

For organizations that choose components of multiple vendor offerings to any degree, they will need to address how to integrate the components together to minimize disruption to the users' workflow. There are various strategies that can be pursued to integrate modules, either at the level of user context (user authentication credentials are maintained), the level of the application view (one application is viewable as a component within another application), or at the level of data sharing (data are exchanged between the applications). If components are not integrated, a user must access each application separately, by opening the software application, logging in to each separately, and selecting the patient within each. When data are integrated at the user context, a user moves between both applications, but the user and patient context are shared. This "single sign-on" approach alleviates one of the main barriers to the user, by facilitating the login and patient selection, while retaining all the functionality of each system.

A deeper level of integration is at the application view. In this case, a primary application provides a portal that views another application; the second application shares user and patient context and is accessible through the user's main workflow system. The second application may use data from the primary application and/or other data sources. Rapidly expanding in adoption for this type of integration is the HL7 Substitutable Medical Applications and Reusable Technologies (SMART) framework, in particular in combination with HL7 Fast Healthcare Interoperability Resources (FHIR) for data interchange. This approach, known as SMART on FHIR, is enabling an ecosystem wherein applications developed by health care organizations or third-party vendors can be seamlessly integrated within the EHR (Kawamoto et al. 2019; Mandel et al. 2016).

The deepest level of integration is where l data elements from one system are also stored in the other system. With this approach, one system is determined to be the main repository, and data from the other systems are automatically stored into the repository. This approach has the advantage of the most complete use of data, e.g., decision support logic can use data from multiple systems, which can be more accurate. The disadvantage is that the integration can be expensive, requiring new interfaces for each integrated system.

Another and often overlooked consideration of EHR software modules is data analytics capabilities, usually discussed in conjunction with a data warehouse. EHR systems generally include reporting functionality, where specific reports can be configured to summarize and display data stored in the system. However, these systems often do not facilitate ad hoc data extracts that are commonly needed for more complicated data analysis. Additionally, if modules from multiple software vendors are used, the data reporting functions will be limited according to how well data are integrated. One typical approach

is to use a separate data warehouse and analysis system, with functions to create ad hoc reports, that can combine data from multiple systems. Data integration with warehouses is less expensive than with repositories, because the data do not need to be synchronized. Instead, data can be extracted in batches from source systems, transformed to the warehouse data model, and then loaded into the warehouse at periodic intervals. The greatest cost of integration is the data transformation, but this transformation is similar to what is required when receiving data through a real-time interface.

The **Meaningful Use** program, which has evolved to become the **Promoting Interoperability** program, has greatly influenced the systems that are installed by an institution. Initially, the ONC created a list of important EHR functions. They also created a requirement that hospitals and physician practices use a "certified" system – i.e., one that has demonstrated it provides those functions – to receive the incentives, and other criteria that the functions must be used in clinical care (Washington et al. 2017). As a result, health care organizations rapidly implemented EHRs that were certified and best fulfilled regulatory requirements.

6.3.3.2 Case Studies of EHR Adoption

Consider the following case studies of institutions adopting EHR systems. All examples are fictional, but reflect the complexity of the issues with EHR software.

Best-Care Medical Center had been using information systems for many years, dating back to when some researchers in the cardiology department built a small system to integrate data from the purchased laboratory and pharmacy information systems. Eventually, the infection control group for the hospital began using the system, and contributed efforts to expand its functionality. Other departments began developing decision support rules, and the system continued to grow. Eventually, the institution made a commitment to redevelop the infrastructure to support a much larger group of users and functions, and named it A-Chart. Satisfaction with the system was high where it had been initially developed, and with other related specialties. However, over time there was disproportionate development in these areas, and clinicians in other specialties complained about the rudimentary functionality, especially when compared to existing vendor systems for their specialty. As a result, the organization decided to purchase a new vendor system. This made the other specialties happy, but was a big concern to the groups that had been using A-Chart for years. These clinicians feared that they would have to reconfigure their complicated decision support rules with a new system, or worse, that functionality would no longer be supported. To alleviate concerns, representatives from each department were asked to participate in both drafting a Request for Proposals and then reviewing the proposals from four different vendors. Many clinicians liked System X, but in the end the hospital chose System Y, which seemed to have most of the same capabilities but was perceived to be more affordable than System X. However, System Y did not include a laboratory system, so the medical center purchased a separate laboratory system and built interfaces to connect it with the core EHR.

Patients' Choice is an integrated delivery system with a long history of EHR use.. Years ago, it existed as two separate systems of hospitals and clinics. Shortly before the merger of these systems, the hospitals and clinics purchased separate EHRs, InPatSys and CliniCare. At the time of the merger, the institution felt that each would be best served by a best-of-breed system, to support the different workflows, and there was no single system that both sides of the organization could agree to use. Years later, as Patients' Choice began to integrate care between the hospitals and clinics, the clinicians and administrators became increasingly frustrated at how different the InPatSys and CliniCare systems were, and that they had to use two separate systems to care for the same patients. A team was formed to evaluate the options, and the CliniCare system was eventually replaced by OutPatSys, the outpatient version of InPatSys. To prevent losing data as they moved from one system to the other, the Patients' Choice IT department prepared the OutPatSys system by loading existing laboratory results and vital sign measurements from

CliniCare. A SMART on FHIR application provided an integrated view of historical CliniCare data inside the OutPatSys system.

Hometown Community Hospital historically used various niche information systems, but no EHR. With the availability of Meaningful Use incentives, the hospital decided to acquire a commercial EHR. A leadership team visited six hospitals to investigate how various EHRs were used. Ultimately, the hospital made the decision to purchase eCompuChart, because it was highly rated and seemed best adapted to their needs. Hometown hired a new Chief Information Officer who had recently implemented eCompuChart at a community hospital in a neighboring state. They also promoted Dr. Jones, who had recently moved from another hospital that had also used eCompuChart, to Chief Medical Information Officer (CMIO). The CIO and CMIO negotiated a contract with DigiHealth, a consulting company with experience in implementing EHRs, to plan and coordinate the implementation. Among other recommendations from DigiHealth, most existing systems were replaced with modules from eCompuChart to simplify maintenance.

In practice, organizations may not adopt a complete "build" or a complete "buy" strategy. EHR vendors have advanced considerably in their ability to create systems that meet common needs in health care. Still, no system exists to date that can fully address all information needs for an organization, in part because the information needs expand as more data and new technologies become available. Additionally, EHR strategies become malleable over time, as commercial software capabilities increase and data become more consistent. As indicated through some of the examples above, organizational strategies may change over time to adapt to these capabilities and needs. Expanding options for health care organizations is the emergence of the notion of the EHR as a platform, where HL7 FHIR data interfaces can be used to read data from, and in some cases write data back to the EHR; HL7 SMART is available to integrate external applications into the native EHR user interface; and more recently, a framework known as HL7 CDS Hooks is available to interface externally developed alerts and reminders into the EHR.[3]

One consideration that is not always stated in the software selection process, but is significant in its influence over the decision, is how an organization will pay for the application. In organizations where software purchases are requested from the information technology department and budget, overall maintenance costs are considered more prominently, and software that integrates with and is a component of the overall EHR vendor offering is often selected. However, if a clinical department has direct control over their spending for the software, functionality may become a more focal concern. An additional case study illustrates this situation.

Downtown Hospital recently decided to purchase eCompuChart as the centerpiece of its overall clinical information system strategy. eCompuChart has award-winning modules for the emergency department and intensive care units. However, there are strong complaints about its capabilities for labor and delivery management and radiology. After considering capabilities of best-of-breed options and their ability to integrate with eCompuChart, Downtown Hospital eventually made a split decision. The labor and delivery module for eCompuChart was purchased because other systems with more elaborate functionality could not integrate data as well with the overall EHR. On the other hand, a separate best-of-breed system was purchased for radiology, because interfaces between the systems were seen as an acceptable solution for integrating data.

6.3.3.3 Enhancing Acquired Software

Although most institutions will choose to acquire a system rather than building it from scratch, software engineering is still required to make the systems function effectively. This involves more than just installing and configuring the software to the local environment. There is still a significant need for software development in implementing COTS, because

3 ► https://cds-hooks.org/ (last accessed June 3, 2020).

(1) applications must be integrated with existing systems, and (2) leading healthcare institutions are increasingly developing custom applications, such as SMART on FHIR applications, that supplement commercial systems.

6.3.3.4 Integration with Existing Systems

In all but the most basic health care information technology environments, multiple software applications are used for treatment, payment, and operations purposes. A partial list of applications that might be used in a hospital environment is shown in ◘ Table 6.1.

To facilitate the sharing of information among various software applications, standards have emerged for exchanging messages and defining clinical terminology (see ► Chap. 8). Message exchange between different software applications enables the following scenario:

1. A patient is admitted to the hospital. A registration clerk uses the bed management system to assign the patient's location and attending physician of record.
2. The physician orders a set of routine blood tests for the patient in the inpatient EHR computerized order entry module.
3. The request for blood work is sent electronically to the laboratory information system, where the blood specimen is matched to the patient using a barcode.
4. The results of the laboratory tests are sent to the results review module of the EHR

   Message exchange is an effective means of integrating disparate software applications in healthcare when the users rely primarily on a single "workflow system" (e.g., a physician uses the inpatient EHR and a laboratory technician uses the LIS). Because message exchange is handled by a sophisticated "interface engine" (see ► Chap. 8), little software development, in the traditional sense, is typically required. When a user accesses multiple workflow systems to perform a task, message exchange may not be sufficient and a deeper level of integration may be required. For example, consider the following addition to the previously described scenario:
5. The physician reviews the patient's blood work and notes that the patient may be suffering from renal insufficiency as evidenced by his elevated creatinine level.
6. The physician would like to review a trend of the patient's creatinine over the past 3 years. Because the hospital installed their commercial EHR less than a year ago, data from prior to that time are available in a legacy results review system that was developed locally. The physician logs into the legacy application (entering her username and password), searches for the correct patient, and reviews the patient's creatinine history.

◘ **Table 6.1** Partial list of software applications that may be used in a hospital setting

System	Primary Users
Inpatient EHR (Results Review, Order Entry, Documentation)	Physicians, nurses, allied health professionals
Pharmacy Information System	Pharmacists, pharmacy technicians
Laboratory Information System	Laboratory technicians, phlebotomists
Radiology Information System	Radiologists, radiology technicians
Pathology Information System	Pathologists
Registration/Bed Management	Registration staff
Hospital Billing System	Medical coders
Professional Services Billing System	Physicians, medical coders

While it may seem preferable in this scenario to load all data from the legacy system into the new EHR, commercial applications may not support importing such data for various reasons. To simplify and improve the user experience for reviewing information from a legacy application within a commercial EHR, one group of informaticians created the custom application shown in ◘ Fig. 6.5. The application is accessed by clicking a link

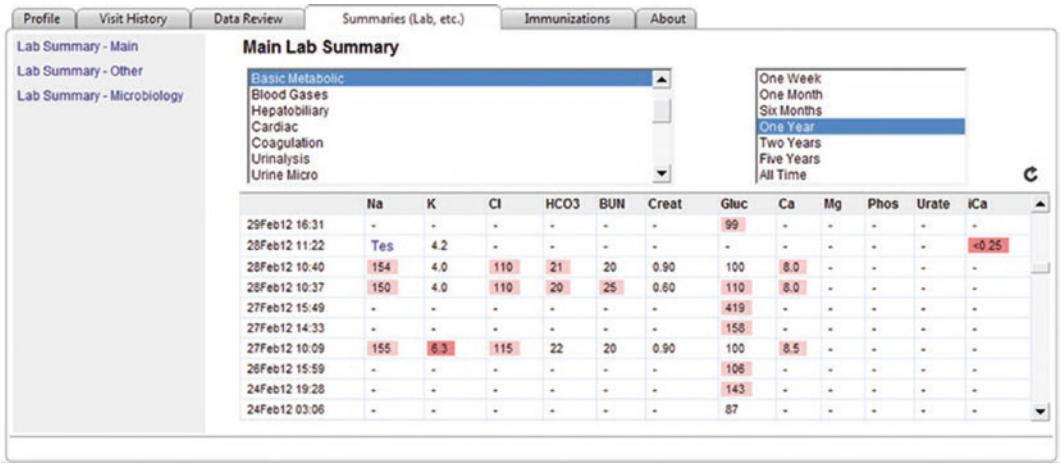


Fig. 6.5 Example screen from a custom lab summary display application integrated into a commercial EHR. The application shows a longitudinal view of laboratory results that can span multiple patient encounters

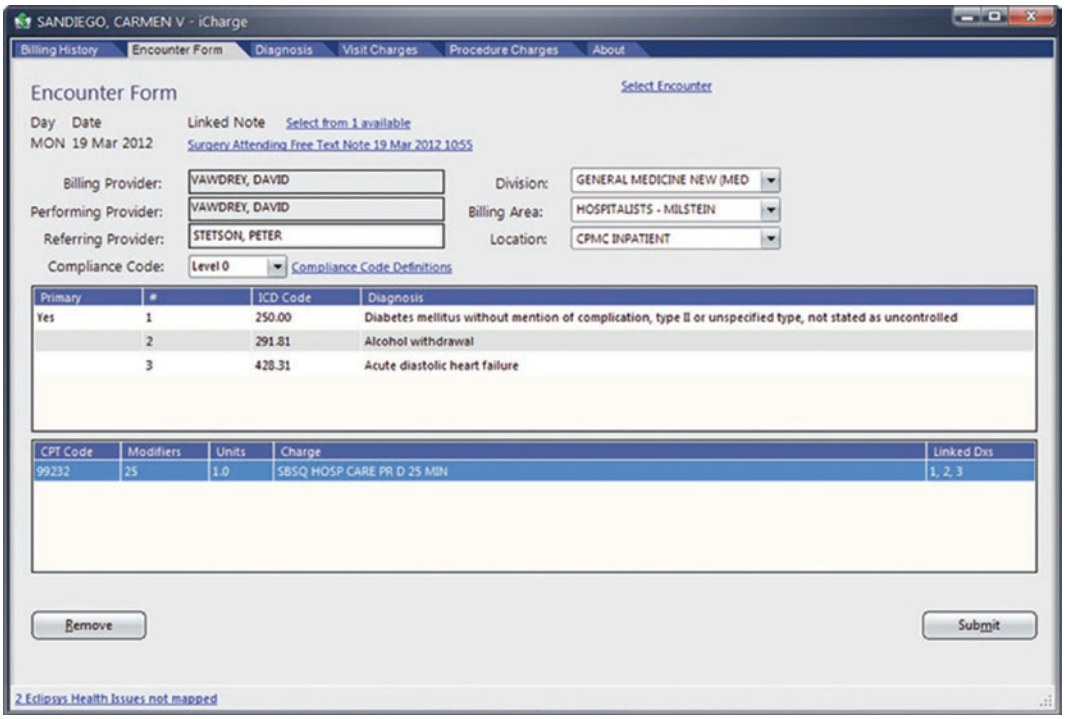


Fig. 6.6 Example screen from a custom billing application integrated into a commercial EHR. This replaced a separate application that was not integrated into the clinicians' workflow

within the commercial EHR and does not require login or patient look-up.

In an example of a more sophisticated level of "workflow integration" is shown in Fig. 6.6. In this example, informaticians developed a custom billing application within an inpatient commercial EHR. Users of the application were part of a physician practice that used a different outpatient EHR with a professional billing module with which they were already familiar. When the physicians in the practice rounded on their patients who

were admitted to the hospital, they documented their work by writing notes within the inpatient EHR, and then used their outpatient EHR to submit their professional service charges. This practice not only required a separate login to submit a bill, but also required duplicate patient lists to be maintained in each application, as well as a duplicate problem list for each patient to be managed in each application. The integrated charge application was accessed from the inpatient EHR but provided the same look-and-feel as the outpatient EHR billing module. Charges were submitted through the outpatient EHR infrastructure and would appear as normal charges in the outpatient system, with the substantial improvement of displaying the information (note name, author, and time) for the documentation that supported the charge.

6.3.3.5 Development of Custom Applications that Supplement or Enhance Commercial Systems

Commercial EHRs frequently provide customers with the ability to develop custom software modules. Some EHRs provide a flexible clinical decision support infrastructure that allows customers to develop modules that execute medical logic to generate alerts, reminders, corollary orders, and so on. Vendors may also provide customers with tools to access the EHR database, which allows development of stand-alone applications that make use of EHR data. Additionally, vendors may foster development of custom user interfaces within the EHR by providing an application programming interface through which developers can obtain information on user and patient context.

The ability to provide patient-specific clinical decision support is one of the key benefits of EHRs. Many commercial EHRs either directly support or have been influenced by the **Arden Syntax for Medical Logic Modules** (Pryor and Hripcsak 1993). The Arden Syntax is part of the HL7 family of standards. It encodes medical knowledge as **Medical Logic Modules** (MLMs), which can be triggered by various events within the EHR (e.g., the placing of a medication order) and execute serially as a sequence of instructions to access and manipulate data and generate output. MLMs have been used to generate clinical alerts and reminders, to screen for eligibility in clinical research studies, to perform quality assurance functions, and to provide administrative support (Dupuits 1994; Jenders 2008; Jenders and Shah 2001; Ohno-Machado et al. 1999). Although one goal of the Arden Syntax was to make knowledge portable, MLMs developed for one environment are not easily transferable to another. Developers of clinical decision support logic require skills in both computer programming as well as medical knowledge representation.

An example of a standalone, locally developed software application that relies on EHR data is shown in ◘ Fig. 6.7. The Web-based application, EpiPortal™, provides a comprehensive, electronic hospital epidemiology decision support system. The application can be accessed from a Web browser or directly from within the EHR. It relies on EHR data such as microbiology results, clinician orders, and bed tracking information to provide users with timely information related to infection control and prevention.

In some cases, it is desirable to develop custom applications to address specific clinical needs that are not met by a commercial EHR. For example, most commercial EHRs lack dedicated tools to support patient handoff activities. For hospitalized patients, handoffs between providers affect continuity of care and increase the risk of medical errors. Informaticians at one academic medical center developed a collaborative application supporting patient handoff that is fully integrated with a commercial EHR (Fred et al. 2009). An example screen from the application is shown in ◘ Fig. 6.8. The application creates user-customizable printed reports with automatic inclusion of patient allergies, active medications, 24-hour vital signs, recent common laboratory test results, isolation requirements, code status, and other EHR data.

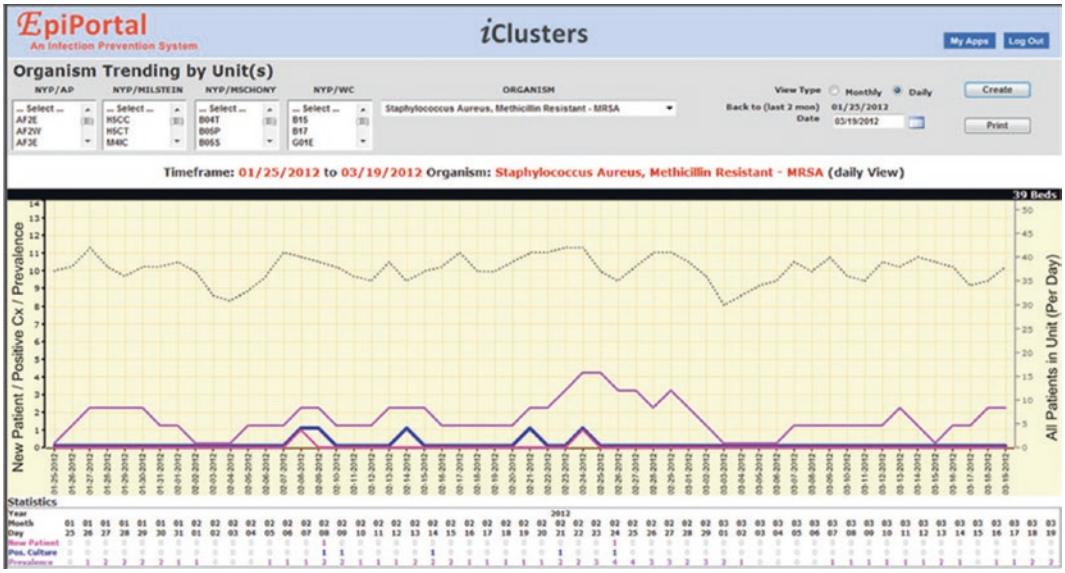
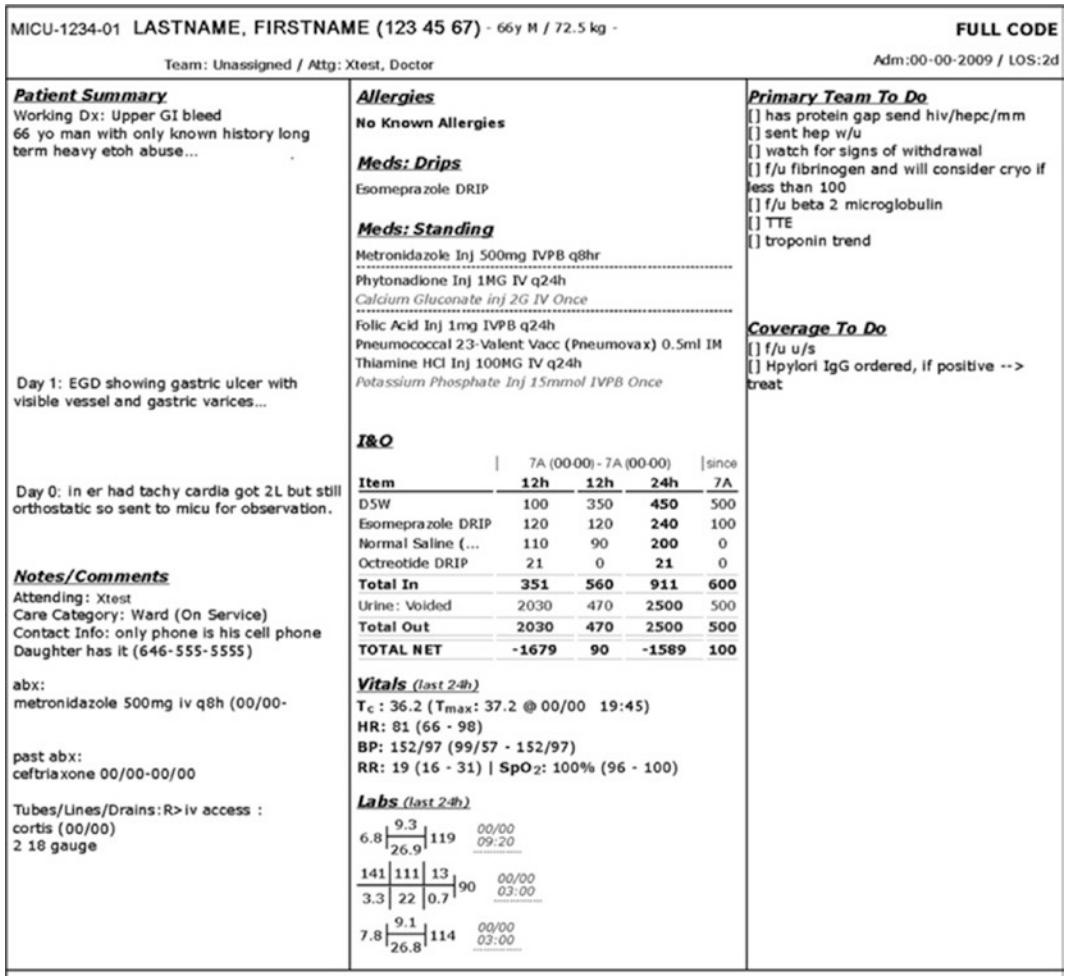


Fig. 6.7 Example screen from a standalone, software application that relies on EHR data to provide a comprehensive, electronic hospital epidemiology decision support system. (Reused with permission. Copyright 2012 The New York and Presbyterian Hospital and Columbia University – All rights reserved. EpiPortal is a trademark of The New York and Presbyterian Hospital.)

6.4 Emerging Influences and Issues

Several trends in software engineering are beginning to significantly influence biomedical information systems. While many of the trends may not be considered new to software engineering in general, they are more novel to the biomedical environment because of the less rapid and less broad adoption of information technology in this field. One area in particular that has received growing attention is **service oriented architectures** (SOA). Sometimes called "software as a service", SOA is a software design framework that allows specific processing or information functions (services) to run on an independent computing platform that can be called by simple messages from another computer application. For example, an EHR application might have native functionality to maintain a patient's medication list, but might call a drug-drug interaction program running on a third party system to check the patient's medications for potential interactions. This allows the EHR provider to off-load developing this functionality, while the drug-drug interaction service provider can concentrate efforts on this focused task, and in particular on ensuring that the drug interaction database is kept up-to-date for all users of the service. Since the service is independent of any EHR application, many different EHR providers can call the same service, as can other applications such as patient health record (PHR) applications that are focused on consumer functionality. SOA might also be grouped with the more recently computer phrase "cloud computing", which includes providing functional services to other applications, but also encompasses running entire applications and storing data in offsite or disconnected locations. A good example of SOA is the HL7 CDS Hooks standard, which specifies how EHR systems can interface with external clinical decision support services to provide point-of-care alerts and reminders to clinical end users.

Another emerging trend, discussed earlier, is the notion of the **EHR as a platform** for third-party applications and services that interface with, and add value to, the EHR. Central to this approach is HL7 FHIR



MICU-1234-01 **LASTNAME, FIRSTNAME (123 45 67)** - 66y M / 72.5 kg - **FULL CODE**

Team: Unassigned / Attg: Xtest, Doctor — Adm:00-00-2009 / LOS:2d

Patient Summary

Working Dx: Upper GI bleed

66 yo man with only known history long term heavy etoh abuse...

Day 1: EGD showing gastric ulcer with visible vessel and gastric varices...

Day 0: in er had tachy cardia got 2L but still orthostatic so sent to micu for observation.

Notes/Comments

Attending: Xtest
Care Category: Ward (On Service)
Contact Info: only phone is his cell phone
Daughter has it (646-555-5555)

abx:
metronidazole 500mg iv q8h (00/00-

past abx:
ceftriaxone 00/00-00/00

Tubes/Lines/Drains:R>iv access :
cortis (00/00)
2 18 gauge

Allergies

No Known Allergies

Meds: Drips

Esomeprazole DRIP

Meds: Standing

Metronidazole Inj 500mg IVPB q8hr
Phytonadione Inj 1MG IV q24h
Calcium Gluconate inj 2G IV Once
Folic Acid Inj 1mg IVPB q24h
Pneumococcal 23-Valent Vacc (Pneumovax) 0.5ml IM
Thiamine HCl Inj 100MG IV q24h
Potassium Phosphate Inj 15mmol IVPB Once

I&O

	7A (00-00) - 7A (00-00)			since
Item	**12h**	**12h**	**24h**	**7A**
D5W	100	350	**450**	500
Esomeprazole DRIP	120	120	**240**	100
Normal Saline (...	110	90	**200**	0
Octreotide DRIP	21	0	**21**	0
Total In	**351**	**560**	**911**	**600**
Urine: Voided	2030	470	**2500**	500
Total Out	**2030**	**470**	**2500**	**500**
TOTAL NET	**-1679**	**90**	**-1589**	**100**

***Vitals** (last 24h)*

T_c: 36.2 (T_{max}: 37.2 @ 00/00 19:45)
HR: 81 (66 - 98)
BP: 152/97 (99/57 - 152/97)
RR: 19 (16 - 31) | SpO_2: 100% (96 - 100)

***Labs** (last 24h)*

6.8 | 9.3 / 26.9 | 119 — 00/00 09:20

141 | 111 | 13 / 3.3 | 22 | 0.7 | 90 — 00/00 03:00

7.8 | 9.1 / 26.8 | 114 — 00/00 03:00

Primary Team To Do

[] has protein gap send hiv/hepc/mm
[] sent hep w/u
[] watch for signs of withdrawal
[] f/u fibrinogen and will consider cryo if less than 100
[] f/u beta 2 microglobulin
[] TTE
[] troponin trend

Coverage To Do

[] f/u u/s
[] Hpylori IgG ordered, if positive --> treat

Fig. 6.8 Example screen from a custom patient handoff application integrated into a commercial EHR. The application creates user-customizable printed reports with automatic inclusion of patient allergies, active medications, 24-hour vital signs, recent common laboratory test results, isolation requirements, code status, and other EHR data

application programming interface (API), which uses modern Internet technologies and approaches for data exchange, as well as HL7 SMART for application integration and HL7 CDS Hooks for integrating decision support services. While this notion of EHR as a platform is still in its early stages and still maturing, many EHR vendors are strongly supportive of this type of an ecosystem, and promising examples are emerging of how these technologies can be used to deliver value to health care organizations in an EHR-agnostic manner. We anticipate that this approach to health information systems, wherein core EHR systems are augmented by third-party applications and services, will play an important role in the health IT ecosystem in the years to come.

Another important consideration in clinical information systems is infrastructure to support data sharing, such as through a **health information exchange** (HIE). HIE infrastructure allows organizations to share information about patients through a common electronic framework. Robust HIE capabilities, which are now being implemented in commercial EHR systems, make it much more efficient to share patient information between organizations versus creating point-to-point interfaces between all the clinical information

systems a particular provider might need to communicate with. Effective sharing of information is predicated on the use standard message formats and terminologies (see ▶ Chap. 8), or the use of a shared EHR vendor. Where HIE functionality does not exist or is declining (Adler-Milstein et al. 2016), sharing can be coordinated directly between organizations with incentives for sharing, such as in an ambulatory care network (ACN). As health care in the United States increasingly shifts to a payment model that rewards value over volume, data sharing capabilities and the capacity for population-level analytics and health management will become increasingly critical. Moreover, there are also emerging efforts to scale health information exchange to a national scale, and to facilitate patient access to their information using APIs (ONC 2019b).

Software engineering is an ever-evolving discipline, and new ideas are emerging rapidly in this field. It is less than 30 years since the first graphical browser was used to access the World Wide Web, but today Web-based applications are the standard. Access to information through search engines has changed the way that people find and evaluate information. Social networking applications have altered our views on privacy and personal interaction. All of these developments have shaped the development of healthcare software, too. Today it is unimaginable that an EHR would not support a Web-based patient portal. Clinicians and consumers use the Web to search for health-related information in growing numbers and with growing expectations. It is not atypical for patients to discuss health issues in online forums and share intimate details on patient networking sites.

Another development that is impacting virtually all industries, including health care, is **advanced analytics**. Coupled with the rapid increase in the adoption of EHR systems, health care represents a golden opportunity for leveraging powerful computing approaches with large data sets to identify and apply new insights. For example, deep learning techniques can be applied to EHR data to predict important outcomes such as in-hospital mortality. If coupled with user-centered, workflow-integrated interventions, such insights have the potential to improve clinical decision making and enhance patient care.

6.5 Summary

The goal of software engineering in health care is to create a system that facilitates delivery of care. Much has changed in the past decade with EHRs, and today most institutions will purchase rather than build an EHR. But engineering these systems to facilitate care is still challenging, and following appropriate software development practices is increasingly important. The success of a system depends on interaction among designers of healthcare software applications and those that use the systems. Communication among the participants is very difficult when it comes to commercial applications. Informaticians have an important role to play in bridging the gaps among designers and users that result from the wide variety in background, education, experience, and styles of interaction. They can improve the process of software development by specifying accurately and realistically the need for a system and of designing workable solutions to satisfy those needs.

Suggested Reading

Carter, J. H. (2008). *Electronic health records* (2nd ed.). Philadelphia: ACP Press. Written by a clinician and for clinicians, this is a practical guide for the planning, selection, and implementation of an electronic health record. It first describes the basic infrastructure of an EHR, and then how they can be used effectively in health care. The second half of the book is written more as a workbook for someone participating in the selection and implementation of an EHR.

KLAS Reports. http://www.klasresearch.com/reports. These reports are necessary tools for a project manager who needs to know the latest industry and customer information about vendor health information technology products. The reports include information on functionality available from vendors as well as customer

opinions about how vendors are meeting the needs of organizations and whose products are the best in a particular user environment.

McConnell, S. (1996). *Rapid development: Taming wild software schedules*. Redmond: Microsoft Press. For those who would like a deeper understanding of software development and project methodologies like Agile, this is an excellent source. It is targeted to code developers, system architects, and project managers.

President's Council of Advisors on Science and Technology (December 2010). Report to the President Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: the Path Forward. http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf. This PCAST report focuses on what changes could be made in the field of electronic health records to make them more useful and transformational in the future. It gives a good summary of the current state of EHRs in general, and compares the barriers to those faced in adopting information technology in other fields. Time will tell if the suggestions really become the solution.

Stead, W. W., & Lin, H. S. (Eds.). (2009). *Computational technology for effective health care: immediate steps and strategic directions*. Washington, DC: National Academies Press. This is a recent National Research Council report about the current state of health information technology and the vision of the Institute of Medicine about how such technology could be used. It can help give a good understanding of how health IT could be used in health care, especially to technology professionals without a health care background.

Tang, P. C. (2003). *Key capabilities of an electronic health record system*. Washington, DC: National Academies Press. This is a short, letter report from an Institute of Medicine committee that briefly describes the core functionalities of an electronic health record system. Much of the report is tables that list specific capabilities of EHRs in some core functional areas, and indicate their maturity in hospitals, ambulatory care, nursing homes, and personal health records.

Wager, K. A., Lee, F. W., & Glaser, J. P. (2017). *Health care information systems: a practical approach for health care management*. John Wiley & Sons. This is a textbook giving a good overview of healthcare information systems, used in many academic courses on the subject. It reviews the different environmental factors and contexts that influence the health information landscape nationally, as well as giving guidance on implementation, management and evaluation of systems.

Questions for Discussion

1. Reread the hypothetical case study in ▶ Sect. 6.2.1.
   (a) What are three primary benefits of the software used in James's care?
   (b) How many different ways is James's information used to help manage his care?
   (c) Without the software and information, how might his care be different?
   (d) How has health care that you have experienced similar or different to this example?
2. For what types of software development projects would an agile development approach be better than a waterfall approach? For what types of development would waterfall be preferred?
3. What are reasons an institution would choose to develop software instead of purchase it from a vendor?
4. How is would various stages in the software development life cycle be different when developing software versus configuring or adding enhancements to an existing software program?
5. Reread the case studies in ▶ Sect. 6.3.3.2.
   (a) What are the benefits and advantages of the different approaches to development and acquisition among the scenarios?
   (b) What were the initial costs for each institution for the software? Where will most of the long-term costs be?
6. In what ways might new trends in software (small "apps" that accomplish focused tasks) change long-term strategies for electronic health record architectures?

References

Adler-Milstein, J., Lin, S. C., & Jha, A. K. (2016). The number of health information exchange efforts is declining, leaving the viability of broad clinical data exchange uncertain. *Health Affairs, 35*(7), 1278–1285.

Bates, D. W. (2006). Invited commentary: The road to implementation of the electronic health record. *Proceedings (Baylor University. Medical Center), 19*(4), 311–312.

Bates, D. W., Leape, L. L., et al. (1998). Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA, 280*(15), 1311–1316.

Beck, K., Beedle, M., et al. (2001). *Manifesto for agile software development*. 2012, from agilemanifesto.org.

Blumenthal, D., & Tavenner, M. (2010). The "meaningful use" regulation for electronic health records. *The New England Journal of Medicine, 363*(6), 501–504.

Campanella, P., Lovato, E., Marone, C., Fallacara, L., Mancuso, A., Ricciardi, W., & Specchia, M. L. (2016). The impact of electronic health records on healthcare quality: A systematic review and meta-analysis. *European Journal of Public Health, 26*(1), 60–64.

Chaudhry, B., Wang, J., et al. (2006). Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine, 144*(10), 742–752.

Dupuits, F. M. (1994). The use of the Arden Syntax for MLMs in HIOS+, a decision support system for general practitioners in The Netherlands. *Computers in Biology and Medicine, 24*(5), 405–410.

Evans, R. S., Pestotnik, S. L., et al. (1998). A computer-assisted management program for antibiotics and other antiinfective agents. *The New England Journal of Medicine, 338*(4), 232–238.

Fred, M., Vawdrey, D., et al. (2009). *Enhancing a commercial electronic health record with a novel patient handoff application*. Society of Hospital Medicine (SHM) Annual Meeting, Chicago, IL.

Han, Y. Y., Carcillo, J. A., et al. (2005). Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics, 116*(6), 1506–1512.

Hripcsak, G., Kuperman, G. J., et al. (1998). Extracting findings from narrative reports: Software transferability and sources of physician disagreement. *Methods of Information in Medicine, 37*(1), 1–7.

Hunt, D. L., Haynes, R. B., et al. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes: A systematic review. *JAMA, 280*(15), 1339–1346.

Jenders, R. A. (2008). Suitability of the Arden Syntax for representation of quality indicators. *American Medical Informatics Association Annual Symposium Proceedings, 991*.

Jenders, R. A., & Shah, A. (2001). Challenges in using the Arden Syntax for computer-based nosocomial infection surveillance. In *Proceedings of the AMIA Symposium* (pp. 289–293). American Medical Informatics Association.

Jones, S. S., Rudin, R. S., Perry, T., & Shekelle, P. G. (2014). Health information technology: An updated systematic review with a focus on meaningful use. *Annals of Internal Medicine, 160*(1), 48–54.

Kawamoto, K., Kukhareva, P., Shakib, J. H., Kramer, H., Rodriguez, S., Warner, P. B., Shields, D., Weir, C., Fiol, G. D., Taft, T., & Stipelman, C. H. (2019). Association of an electronic health record add-on app for neonatal bilirubin management with physician efficiency and care quality. *JAMA Network Open, 2*(11), e1915343–e1915343.

Leavitt, M., & Gallagher, L. (2006). The EHR seal of approval: CCHIT introduces product certification to spur EHR adoption. *Journal of AHIMA, 77*(5), 26–30, quiz 33–24.

Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association, 23*(5), 899–908.

McConnell, S. (1996). *Rapid development : Taming wild software schedules/Steve McConnell*. Redmond, Wash: Microsoft Press.

Office of the National Coordinator for Health Information Technology (ONC). (2017a). *Certified health IT developers and editions reported by hospitals participating in the medicare EHR incentive program, Health IT Quick-Stat #29*. dashboard.healthit.gov/quickstats/pages/FIG-Vendors-of-EHRs-to-Participating-Hospitals.php. July 2017.

Office of the National Coordinator for Health Information Technology (ONC). (2017b). *Non-federal acute care hospital electronic health record adoption, Health IT Quick-Stat #47*. dashboard.healthit.gov/quickstats/pages/FIG-Hospital-EHR-Adoption.php. September 2017.

Office of the National Coordinator for Health Information Technology (ONC). (2019a). *Office-based physician electronic health record adoption, Health IT Quick-Stat #50*. dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php. January 2019.

Office of the National Coordinator for Health Information Technology (ONC). (2019b). *Trusted exchange framework and common agreement* | HealthIT.gov. (2019). Retrieved February 2020, from https://www.healthit.gov/topic/interoperability/trusted-exchange-framework-and-common-agreement

Ohno-Machado, L., Wang, S. J., et al. (1999). Decision support for clinical trial eligibility determination in breast cancer. In *Proceedings of the AMIA symposium* (pp. 340–344). American Medical Informatics Association.

Pryor, T. A., & Hripcsak, G. (1993). The Arden syntax for medical logic modules. *International Journal of Clinical Monitoring and Computing, 10*(4), 215–224.

Shekelle, P. G., Morton, S. C., et al. (2006). Costs and benefits of health information technology. *Evid Rep Technol Assess (Full Rep), 132*(April), 1–71.

Smelcer, J., Miller-Jacobs, H., et al. (2009). Usability of electronic medical records. *Journal of Usability Studies, 4*(2), 70–84.

Sullivan, T. (2017). Coast guard seeks new EHR vendor after failed epic implementation. *Healthcare IT news*. https://www.healthcareitnews.com/news/coast-guard-seeks-new-ehr-vendor-after-failed-epic-implementation.

Thompson, C.B., Snyder-Halpern, R., & Staggers, N. (1999). Clinical informatics case studies: Analysis, processes, and techniques. *Computers in Nursing, 17*(5), 203–206.

Washington, V., DeSalvo, K., et al. (2017). The HITECH era and the path forward. *The New England Journal of Medicine, 377*, 904–906.

Standards in Biomedical Informatics

Charles Jaffe, Viet Nguyen, Wayne R. Kubick, Todd Cooper, Russell B. Leftwich, and W. Edward Hammond

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_7

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- Why are standards important in biomedical informatics?
- What data standards are necessary to be able to exchange data seamlessly among systems?
- What organizations are active in standards development?
- What aspects of biomedical information management are supported today by standards?
- What is the process for creating consensus standards?
- What factors and organizations influence the creation of standards?
- How have standards development organizations applied modern Internet technologies such as application programing interfaces (APIs) into their interoperability standards?

7.1 The Idea of Standards

Ever since Eli Whitney developed interchangeable parts for rifle assembly, standards have been created and used to make things or processes work more easily and economically—or, sometimes, to work at all. A standard can be defined in many physical forms, but essentially it comprises a set of rules and definitions that specify how to carry out a process or produce a product. Sometimes, a standard is useful because it provides a way to solve a problem that other people can use without having to start from scratch. Generally, though, a standard is useful because it permits two or more disassociated people to work in some cooperative way. Every time you screw in a light bulb or play a media file, you are taking advantage of a standard. Some standards make things work more easily. Some standards evolve over time,[1] others are developed deliberately.

The first computers were built without standards, but hardware and software standards quickly became a necessity. Although computers work with values such as 1 or 0, and with "words" such as 10101100, humans need a more readable language (see ▶ Chap. 5). Thus, standard character sets, such as ASCII and EBCDIC, were developed. The first standard computer language, COBOL, was written originally to simplify program development but was soon adopted as a way to allow sharing of code and development of software components that could be integrated. As a result, COBOL was given official standard status by the American National Standards Institute (ANSI).[2] In like manner, hardware components depend on standards for exchanging information to make them as interchangeable as were Whitney's gun barrels.

A 1987 technical report from the International Standards Organization (ISO) states that "any meaningful exchange of utterances depends upon the prior existence of an agreed upon set of semantic and syntactic rules" (International Standards Organization 1987). In biomedical informatics, where the emphasis is on collection, manipulation, and transmission of information, standards are essential and their importance is widely recognized by clinicians and policy makers. Requirements for implementation of interoperability standards have been written into laws and regulations. Over the past 2 years, the bipartisan 21st Century Cures Act[3] (Hudson and Collins 2017) has codified many standards into everyday use. At present, the standards scene is evolving so rapidly that any description is inevitably outdated within a few months. This chapter

1 The current standard for railroad-track gauge originated with Roman chariot builders, who set the axle length based on the width of two horses. This axle length became a standard as road ruts developed, requiring that the wheels of chariots—and all subsequent carriages—be the right distance apart to drive in the ruts. When carriage makers were called on to develop railway rolling stock, they continued to use the same axle standard.

2 Interestingly, medical informaticians were responsible for the second ANSI standard language: MUMPS (now known as M).

3 ▶ https://en.wikipedia.org/wiki/21st_Century_Cures_Act (accessed 12/2/19)

emphasizes the need for standards in general, standards development processes, current active areas of standards development, and key participating organizations that are making progress in the development of usable standards.

7.2 The Need for Health Informatics Standards

Standards are generally required when excessive diversity creates inefficiencies or impedes effectiveness. The healthcare environment has traditionally consisted of a set of loosely connected, organizationally independent units. Patients receive care across primary, secondary, and tertiary care settings, with little bidirectional communication and coordination among the services. Patients are cared for by one or more primary physicians, as well as by specialists. There is little coordination and sharing of data between inpatient care and outpatient care. Both the system and patients, by choice, create this diversity in care. Within the inpatient setting, the clinical environment is divided into clinical specialties that frequently treat the patient without regard to what other specialties have done.

Ancillary departments function as detached units, performing their tasks as separate service units, reporting results without follow-up about how those results are used or whether they are even seen by the ordering physician. Reimbursement requires patient information that is often derived through a totally separate process, based on the fragmented data collected in the patient's medical record and abstracted specifically for billing purposes. The resulting set of diagnosis and procedure codes often correlates poorly with the patient's original information (Jollis et al. 1993). With the transition of the US healthcare system from a fee for service payment model to a value-based care model, the need to share information between healthcare providers and their IT systems in order to coordinate care becomes even more crucial.

7.2.1 Early Standards to Support the Use of IT in Health Care

Early interest in the development of standards was driven by the need to exchange data between clinical laboratories and clinical systems, and then between independent units within a hospital. Therefore, the first standards were for data exchange and were referred to as messaging standards. Early systems were developed within independent service units, functional applications such as ADT (admission-discharge-transfer) and billing, and within primary care and specialty units. The first uses of computers in hospitals were for billing and accounting purposes and were developed on large, monolithic mainframe computers Initially the cost of computers restricted expansion into clinical areas. But in the 1960s, hospital information systems (HISs) were developed to support service operations within a hospital. These systems followed a pattern of diversity similar to that seen in the health care system itself.

As new functions were added in the 1970s, they were implemented on mainframe computers and were managed by a data processing staff that usually was independent of the clinical and even of the administrative staff. The advent of the minicomputer supported the development of departmental systems, such as those for the clinical laboratory, radiology department, or pharmacy. With the advent of minicomputers, departmental systems were introduced but connectivity to other parts of the hospital was either by paper or independent electronic systems. It was common to see two terminals sitting side-by-side with an operator typing data from one system to another. Clinical systems, as they developed, continued to focus on dedicated departmental operations and clinical-specialty systems and thus did not permit the practicing physician to see a unified view of the patient. Most HISs were either supported entirely by a single vendor or were still functionally independent and unconnected.

In the 1980s, the need to move laboratory data directly into developing electronic health

records systems (although this term was not used then), early standards were created in ASTM (formerly, the American Society for Testing and Materials, see ▶ Sect. 8.3.2) for the transfer of laboratory data from local and commercial laboratories (American Society for Testing and Materials 1999). In the late 1980s, Simborg and others developed an HIS by interfacing independent systems using a "Best of Breed" approach to create an integrated HIS (Simborg et al. 1983) Unfortunately, the cost of developing and maintaining those interfaces was prohibitive, and the need for a broader set of standards was realized. This effort resulted in the creation of the **standards developing organization** (SDO) Health Level Seven (**HL7**) in 1987. Other SDOs were created in this same time frame: EDIFACT by the United Nations and ASC X12N by ANSI to address standards for claims and billing, IEEE for device standards, ACR/NEMA (later DICOM) for imaging standards, and NCPDP for prescription standards. Internationally, the 1990s saw the creation of the European Normalization Committee (CEN) and ISO Technical Committee 215 for Health Informatics (TC251). These organizations are described in more detail in ▶ Sect. 8.3.2.

7

7.2.2 Transitioning Standards to Meet Present Needs

Early standards were usually applied within a single unit or department in which the standards addressed mainly local requirements. Even then, data acquired locally came from another source introducing the need for additional standards. These many pressures caused health care information systems to change the status quo such that data collected for a primary purpose could be reused in a multitude of ways. Newer models for health care delivery, such as integrated delivery networks, health maintenance organizations (HMOs), preferred provider organizations (PPOs), and now accountable care organizations (ACOs) have increased the need for coordinated, integrated, and consolidated information (see ▶ Chap. 15), even though the information comes from disparate departments and institutions. Various management techniques, such as continuous quality improvement and case management, require up-to-date, accurate abstracts of patient data. Post hoc analyses for clinical and outcomes research require comprehensive summaries across patient populations. Advanced tools, such as clinical workstations (▶ Chap. 5) and decision-support systems (▶ Chap. 3), require ways to translate raw patient data into generic forms for tasks as simple as summary reporting and as complex as automated medical diagnosis. All these needs must be met in the existing setting of diverse, interconnected information systems—an environment that cries out for implementation of standards.

One obvious need is for standardized identifiers for individuals, health care providers, health plans, and employers so that such participants can be recognized across systems. Choosing such an identifier is much more complicated than simply deciding how many digits the identifier should have. Ideal attributes for these sets of identifiers have been described in a publication from the ASTM (1999). The identifier must include a check digit to ensure accuracy when the identifier is entered by a human being into a system. A standardized solution must also determine mechanisms for issuing identifiers to individuals, facilities, and organizations; for maintaining databases of identifying information; and for authorizing access to such information (also see ▶ Chap. 5).

The Centers for Medicare and Medicaid Services (CMS), has defined a National Provider Identifier (NPI) as a national standard. This number is a seven-character alphanumeric base identifier plus a one-character check digit. No meaning is built into the number, each number is unique, it is never reissued, and alpha characters that might be confused with numeric characters have been eliminated (e.g., 0, 1, 2, 4, and 5 can be confused with O, I or L, Z, Y, and S). CMS was tasked to define a Payer ID for identifying health care plans. The Internal Revenue Service's employer identification number has been adopted as the Employer Identifier.

The most controversial issue is identifying each individual or patient. Many people

consider assignment and use of such a number to be an invasion of privacy and are concerned that it could be easily linked to other databases. Public Law 104–191, passed in August 1996 (see ▶ Sect. 8.3.2), required that Congress formally define suitable identifiers. Pushback by privacy advocates and negative publicity in the media resulted in Congress declaring that this issue would not be moved forward until privacy legislation was in place and implemented (see ▶ Chap. 14). The Department of Health and Human Services has recommended the identifiers discussed above, except for the person identifier. This problem has still not been resolved, although the momentum for creating such a unique personal identified seems to be increasing. A work-around non-solution is algorithmic patient matching based on electronic health record (EHR) data. The United States is one of the few developed countries without such an identifier Stead et al. 2005).

7.2.3 Settings Where Standards Are Needed

The patient care process, which can be varied and complicated, also include numerous processes that can be improved with standardization. A hospital admissions system records that a patient has the diagnosis of diabetes mellitus, a pharmacy system records that the patient has been given gentamicin, a laboratory system records that the patient had certain results on kidney function tests, and a radiology system records that a doctor has ordered an X-ray examination for the patient that requires intravenous iodine dye. Other systems need ways to store these data, to present the data to clinical users, to send warnings about possible drug-drug interactions, to recommend dosage changes, and to follow the patient's outcome. A standard for coding patient data is nontrivial when one considers the need for agreed-on definitions, use of qualifiers, differing (application-specific) levels of granularity in the data, and synonymy, not to mention the breadth and depth that such a standard would need to have.

The inclusion of medical knowledge in clinical systems is becoming increasingly important and commonplace. Sometimes, the knowledge is in the form of simple facts such as the maximum safe dose of a medication or the normal range of results for a laboratory test. Much medical knowledge is more complex, however. It is challenging to encode such knowledge in ways that computer systems can use (see ▶ Chap. 26), especially if one needs to avoid ambiguity and to express logical relations consistently. Thus the encoding of clinical knowledge using an accepted standard would allow many people and institutions to share the work done by others. One standard designed for this purpose is the Arden Syntax, discussed in ▶ Chap. 3, as well as the HL7 standard Clinical Quality Language (Odigie et al. 2019).

Because the tasks we have described require coordination of systems, methods are needed for transferring information from one system to another. Such transfers were traditionally accomplished through custom-tailored point-to-point interfaces, but this technique has become unworkable as the number of systems and the resulting permutations of necessary connections have grown. A current approach to solving the multiple-interface problem is through the development of messaging standards. Such messages must depend on the preexistence of standards for patient identification and data encoding.

Over the past decade, non-healthcare domains such as travel, package delivery and e-commerce have adopted, implemented and published standard application programing interfaces (APIs) in order to streamline their business processes and improve efficiency. The adoption of open APIs especially the HL7 Fast Healthcare Interoperability Resources (FHIR®) has increased dramatically and cited proposed regulations as an enabler of improved data sharing (Braunstein 2018).

Data sharing has become an expected functionality for any health IT system. Many of the new initiatives in health require data sharing. Data sharing is essential not only for patient care, but for aggregating data across multiple sites for research. Security must also

be addressed before such exchanges can be allowed to take place. Before a system can divulge patient information, it must ensure that requesters are who they say they are and that they are permitted access to the requested information (see ▶ Chap. 5). Standards exist for this functionality. Although each clinical system can have its own security features, system builders would rather draw on available standards and avoid reinventing the wheel. Besides, the secure exchange of information requires that interacting systems use standard technologies. Electronic Health Record systems (EHRs) are increasingly adopting standard authorization (OAuth2) and identification (OpenID) by implementing **Substitutable Medical Applications Reusable Technology** (SMART) on FHIR which allows platform independent applications to be launched within the EHR workflow and utilize EHR data via FHIR (Payne et al. 2015).

7.3 Standards Undertakings and Organizations

It is helpful to separate our discussion of the general process by which standards are created from our discussion of the specific organizations and the standards that they produce. The process is relatively constant, whereas the organizations form, evolve, merge, and are disbanded. This section will discuss how standards are created then identify the many SDOs and an overview of the types of standards they create. This section will also identify other groups and organizations that contribute or relate to standards activities.

7.3.1 The Standards Development Process

The process of creating standards is biased and highly competitive. Most standards are created by volunteers who represent multiple, disparate stakeholders. They are influenced by direct or indirect self-interest rather than judgment about what is best or required. The process is generally slow and inefficient; multiple international groups create competitive standards; and new groups continue to be formed as they become aware of the need for standards and do not look to see what standards exist. Yet, the process of creating standards largely works, and effective standards are created.

There are four ways in which a standard can be produced:

1. Ad hoc method: A group of interested people and organizations (e.g., laboratory-system and hospital-system vendors) agree on a standard specification. These specifications are informal and are accepted as standards through mutual agreement of the participating groups. A standard example produced by this method is the DICOM standard for medical imaging.
2. De facto method: A single vendor controls a large enough portion of the market to make its product the market standard. An example is Microsoft's Windows. A more recent example are the **Argonaut Implementation Guides**.[4] In this case, a collaborative of vendors and academic health systems are creating consensus standards for their requirements.
3. Government-mandate method: A government agency, such as CMS or the National Institute for Standards and Technology (NIST) creates a standard and legislates its use. An example is the HIPAA standard. Another example is the **Consolidated Clinical Data Architecture** (CCDA),[5] a standard that resulted from the US Government's creating a set of requirements and driving a standard to meet those requirements.
4. Consensus method: A group of volunteers representing interested parties works in an open process to create a standard. Most health care standards are produced by this method. An example is the Health Level 7 (HL7) standard for clinical-data interchange (◘ Fig. 7.1).

4 ▶ http://fhir.org/guides/argonaut/ (accessed 12/2/19)

5 ▶ https://www.healthit.gov/topic/standards-technology/consolidated-cda-overview (accessed 12/2/19)

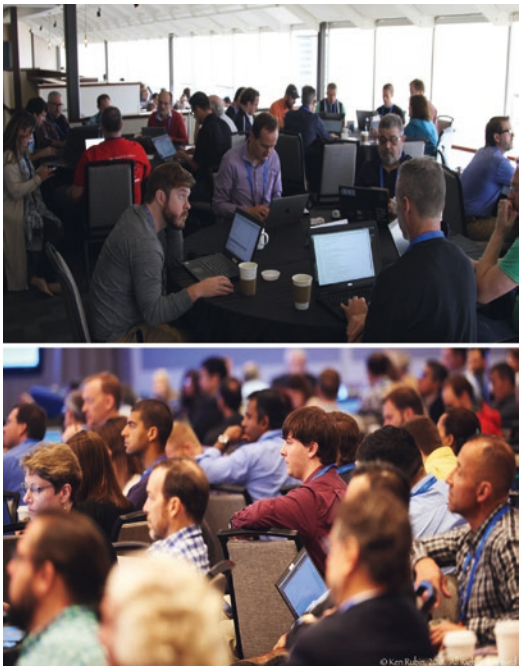


Fig. 7.1 Standards development meetings. The development of effective standards often requires the efforts of dedicated volunteers, working over many years. Work is often done in small committee meetings and then presented to a large group to achieve consensus. Here we see meetings of the HL7. Vocabulary Technical Committee (*top*) and an HL7 plenary meeting (*bottom*). See ▶ Sect. 8.5.2 for a discussion on HL7; Photo courtesy of Ken Rubin Photography

The process of creating a standard proceeds through several stages (Libicki 1995). It begins with an identification stage, during which someone becomes aware that there exists a need for a standard in some area and that technology has reached a level that can support such a standard. For example, suppose there are several laboratory systems sending data to several central hospital systems—a standard message format would allow each laboratory system to talk to all the hospital systems without specific point-to-point interface programs being developed for each possible laboratory-to-laboratory or laboratory-to-hospital combination. If the time for a standard is ripe, then several individuals can be identified and organized to help with the conceptualization stage, in which the characteristics of the standard are defined. What must the standard do? What is the scope of the standard? What will be its format? In the early years of standards development, this approach led the development of standards, and the process was supported by vendors and providers.

As those early standards have become successful, the need for "gap-standards" has arisen. These gap-standards have no champion but are necessary for completeness of an interoperable data exchange network. The need for these standards is not as obvious as for the primary standards, people are less likely to volunteer to do work, putting stress on the voluntary approach. In such cases, the need of such standards must be sold to the volunteers or developed by paid professionals.

Let us consider, for purposes of illustration, how a standard might be developed for sending laboratory data in electronic form from one computer system to another in the form of a message. The volunteers for the development might include laboratory system vendors, clinical users, and consultants. One key discussion would be on the scope of the standard. Should the standard deal only with the exchange of laboratory data, or should the scope be expanded to include other types of data exchange? Should the data elements being exchanged be sent with a XML tag identifying the data element, or should the data be defined positionally? In the ensuing discussion stage, the participants will begin to create an outline that defines content, identifies critical issues, and produces a timeline. In the discussion, the pros and cons of the various concepts are discussed. What will be the specific form for the standard? For example, will it be message based or document based? Will the data exchange be based on a query or on a trigger event? Will the standard define the message content, the message syntax, the terminology, the network protocol, or a subset of these issues? How will a data model or information model be incorporated?

The participants are generally well informed in the domain of the standard, so they appreciate the needs and problems that the standard must address. Basic concepts are usually topics for heated discussion; subsequent details may follow at an accelerated pace. Many of the participants will have experience in solving problems to be addressed

by the standard and will protect their own approaches. The meanings of words are often debated. Compromises and loosely defined terms are often accepted to permit the process to move forward. For example, the likely participants would be vendors of competing laboratory systems and vendors of competing HISs. All participants would be familiar with the general problems but would have their own proprietary approach to solving them. Definitions of basic concepts normally taken for granted, such as what constitutes a test or a result, would need to be clearly stated and agreed on.

The writing of the draft standard is usually the work of a few dedicated individuals—typically people who represent the vendors in the field. Other people then review that draft; controversial points are discussed in detail and solutions are proposed and finally accepted. Writing and refining the standard is further complicated by the introduction of people new to the process who have not been privy to the original discussions and who want to revisit points that have been resolved earlier.

The balance between moving forward and being open is a delicate one. Most standards-writing groups have adopted an **open standards development policy**: Anyone can join the process and can be heard. Most standards development organizations—certainly those by accredited groups— support an open balloting process. A draft standard is made available to all interested parties, inviting comments and recommendations. All comments are considered. Negative ballots must be addressed specifically. If the negative comments are persuasive, the standard is modified. If they are not, the issues are discussed with the submitter in an attempt to convince the person to remove the negative ballot. If neither of these efforts is successful, the comments are sent to the entire balloting group to see whether the group is persuaded to change its vote. The resulting vote then determines the content of the standard. Issues might be general, such as deciding what types of laboratory data to include (pathology? blood bank?), or specific, such as deciding the specific meanings of specific fields (do we include the time the test was ordered? specimen drawn? test performed?).

A standard will generally go through several versions on its path to maturity. The first attempts at implementation are frequently met with frustration as participating vendors interpret the standard differently and as areas not addressed by the standard are encountered. These problems may be dealt with in subsequent versions of the standard. Backward compatibility is a major concern as the standard evolves. How can the standard evolve, over time, and still be economically responsible to both vendors and users? An implementation guide is usually produced to help new vendors profit from the experience of the early implementers.

Connectathons have become increasingly important in the standards development. In the past, standards development and standards implementation have been generally separated. Today, standards are tested during one or two day connectathons where implementers bring client and server applications to test against one another. The successes reinforce the validity of the standard while failures identify gaps and errors which need to be addressed.

A critical stage in the life of a standard is early implementation, when acceptance and rate of implementation are important to success. This process is influenced by accredited standards bodies, by the federal government, by major vendors, and the marketplace. The maintenance and promulgation of the standard are also important to ensure widespread availability and continued value of the standard. Some form of conformance testing is ultimately necessary to ensure that vendors adhere to the standard and to protect its integrity.

Producing a standard is an expensive process in terms of both time and money. Vendors and users must be willing to support the many hours of work, usually on company time; the travel expense; and the costs of documentation and distribution. In the United States, the production of a consensus standard is voluntary, in contrast to in Europe and elsewhere, where most standards development

is funded by governments. In the US, a new model for funding standards development work has emerged. The **Da Vinci Project**[6] and the **CARIN Alliance**[7] are two collaboratives funded by payers and technology vendors to address interoperability needs between patients, providers and payers. These group share the cost of standards development and benefit from the accelerated pace.

An important aspect of standards is conformance, a concept that covers compliance with the standard and also usually includes specific agreements among users of the standard who affirm that specific rules will be followed. A conformance document identifies specifically what data elements will be sent, when, and in what form. Even with a perfect standard, a conformance document is necessary to define business relationships between two or more partners. Unlike past standards, recent standards have built conformance testing directly into the standard artifacts which are not only human readable documents but also machine readable.

The creation of the standard is only the first step. Ideally the first standard would be a **Standard for Trial Use** (STU), and two or more vendors would implement and test the standard to identify problems and issues. Those items would be corrected, and in a short period of time (usually 1 year), the standards would be advanced to a normative stage. A normative standard specifies to what implementers must conform. Even then, the process is only beginning. Implementation that conforms to the standard is essential if the true value of the standard is to be realized. The use of most standards is enhanced by a certification process in which a neutral body certifies that a vendor's product, in fact, does comply and conform to the standard.

There is currently no body that certifies conformance of specific standards from a vendor. There is, however, the certification of an application that uses standards. In 2010, the Office of the National Coordinator (ONC)[8] engaged with the CCHIT to certify EHR products. That certification process evolved in 2011 to include eight groups that could certify EHR products, and to date over 500 EHR products have been certified. The certification process is still undergoing change.

7.3.2 Data Standards Organizations

Sometimes, standards are developed by organizations that need the standard to carry out their principal functions; in other cases, coalitions are formed for the express purpose of developing a particular standard. The latter organizations are discussed later, when we examine the particular standards developed in this way. There are also standards organizations that exist for the sole purpose of fostering and promulgating standards. In some cases, they include a membership with expertise in the area where the standard is needed. In other cases, the organization provides the rules and framework for standard development but does not offer the expertise needed to make specific decisions for specific standards, relying instead on participation by knowledgeable experts when a new standard is being studied.

This section describes several of the best known and most influential health-related SDOs. Since most standards continue to evolve to accommodate changes in technology, policy, regulations, and requirements, links are provided to selected standards and SDO information. For a detailed understanding of an organization or the standards it has developed, you will need to refer to current primary resources. Many of the organizations maintain Web sites with excellent current information on their status.

6 ▶ http://www.hl7.org/about/davinci/ (accessed 12/2/19)

7 ▶ https://www.carinalliance.com/ (accessed 12/2/19)

8 ▶ http://www.healthit.gov/newsroom/about-onc (accessed 12/2/19)

7.3.2.1 American National Standards Institute

ANSI is a private, nonprofit membership organization founded in 1918. It originally served to coordinate the U.S. voluntary census standards systems. Today, it is responsible for approving official American National Standards. ANSI membership includes over 1100 companies; 30 government agencies; and 250 professional, technical, trade, labor, and consumer organizations.

ANSI does not write standards; rather, it assists standards developers and users from the private sector and from government to reach consensus on the need for standards. It helps them to avoid duplication of work, and it provides a forum for resolution of differences. ANSI administers the only government-recognized system for establishing American National Standards. ANSI also represents U.S. interests in international standardization. ANSI is the U.S. voting representative in the ISO and the International Electrotechnical Commission (IEC). There are three routes for a standards development body to become ANSI approved so as to produce an American National Standard: Accredited Organization; Accredited Standards Committee (ASCs); and Accredited Canvass.

An organization that has existing organizational structure and procedures for standards development may be directly accredited by ANSI to publish American National Standards, provided that it can meet the requirements for due process, openness, and consensus. HL7 (discussed in ▶ Sect. 8.5.2) is an example of an ANSI Accredited Organization.

7.3.2.2 ASC X12

ANSI may also create internal ASCs to meet a need not filled by an existing Accredited Organization. ASC X12 is an example of such a committee.

The final route, Accredited Canvass, is available when an organization does not have the formal structure required by ANSI. Through a canvass method that meets the criterion of balanced representation of all interested parties, a standard may be approved as an American National Standard. X12 develops transaction sets that transcend across a broad range of business domains.

Link: ▶ www.x12.org

Link to transaction sets: ▶ http://www.x12.org/x12-work-products/x12-transaction-sets.cfm

Link to EDI standards: ▶ http://www.x12.org/x12-work-products/x12-edi-standards.cfm

7.3.2.3 ASTM International

ASTM (formerly known as the American Society for Testing and Materials) develops standard test methods for materials, products, systems, and services. ASTM is the largest non-government of standards in the US. ASTM Committee E31 on Computerized Systems is responsible for the development of the medical information standards. The scope of this committee is the promotion of knowledge and development of standard classifications, guides, specifications, practices, and terminology for the architecture, content, storage and communication of information used within healthcare, including patient-specific information and medical knowledge. Standard also address policies for integrity and confidentiality and computer procedures that support the uses of data and healthcare decision making.

Link to E31 Standards: ▶ https://www.astm.org/COMMITTEE/E31.htm

7.3.2.4 Clinical Data Interchange Standards (CDISC)

CDISC creates standards in support of the clinical research community. Its membership includes pharma, academic researchers, vendors and others. CDISC was created in 2000 in order to facilitate electronic regulatory submission of clinical trial data. The current standards include a study data model, a data analysis model, a lab data model and an operational data model that supports audit trails and **metadata**. In 2007, CDISC began a collaborative project with HL7, the National Institutes of Health and the US FDA to link research data with data derived from clinical care. This modeling effort, BRIDG

(Biomedical Research Integrated Domain Group; Becnel et al. 2017), has created a domain analysis model of clinical research.

Link: ► www.cdisc.org

Link to standards: ► www.cdisc.org/models/sds/v2.0/index.html

Link to BRIDG: ► http://www.hl7.org/Special/committees/bridg/index.cfm

7.3.2.5 Digital Imaging and Communications in Medicine (DICOM)

DICOM was created through a joint effort by the American College of Radiology and National Electrical Manufacturers Association (NEMA) to develop standards for imaging and waveforms. The DICOM standard has been developed with an emphasis on diagnostic medical imaging as practiced in radiology, cardiology, pathology, dentistry, ophthalmology and related disciplines, and image-based therapies such as interventional radiology, radiotherapy and surgery.

Link: ► www.dicomstandard.org

Link to standard: ► www.dicomstandard.org/current

Link to history: ► www.dicomstandard.org/history/

7.3.2.6 European Committee for Standardization Technical Committee 251

The European Committee for Standardization (CEN) established, in 1991, Technical Committee 251 (TC 251—not to be confused with ISO TC 215 described below) for the development of standards for health care informatics. The major goal of TC 251 is to develop standards for communication among independent medical information systems so that clinical and management data produced by one system could be transmitted to another system. The organization of TC 251 parallels efforts in the United States through various working groups. These groups similarly deal with data interchange standard, medical record standards, code and terminology standards, imaging standards, and security, privacy and confidentiality.

CEN has made major contributions to data standards in health care. One important CEN pre-standard ENV 13606 on the electronic health record (EHR) is being advanced by CEN as well as significant input from Australia and the OpenEHR Foundation. There is an increasing cooperation among the CEN participants and several of the U.S. standards bodies.

Link: ► http://www.ehealth-standards.eu/

Link to projects: ► http://www.ehealth-standards.eu/en/projects/

7.3.2.7 GS1

GS1 (for "Global Standard 1") is a global standards organization that develops and maintains global standards for business communication. With over 1.5 million members worldwide, it has a presence in over 100 countries. Its primary standards relate to the supply chain and for assigning object identifiers and standards for barcodes. GS1 standards are designed to improve the efficiency, safety and visibility of supply chains across physical and digital channels in 25 sectors. They form a business language that identifies, captures and shares key information about products, locations, assets and more.

Link: ► www.gs1.org

7.3.2.8 Health Level Seven International (HL7)

Health Level 7 was founded as an ad hoc standards group in March 1987 to create standards for the exchange of clinical data, adopting the name "HL7" to reflect the application (seventh) level of the OSI reference model. The primary motivation was the creation of a Hospital Information System from "Best of Breed" components. The HL7 data interchange standard (version 2.n series) reduced the cost of interfacing between disparate systems to an affordable cost. Today HL7 is one of the premier SDOs in the world. It has become an international standards body with approximately 40 Affiliates, over 500 organizational members and over 2200 individual members. HL7 is ANSI accredited, and many of the HL7 standards are required by the

U.S. government as part of the certification requirements of Meaningful Use.

Link: ► www.hl7.org

Link to HL7 FHIR: ► www.hl7.org/FHIR

7.3.2.9 Institute of Electrical and Electronic Engineering (IEEE)

IEEE is an international SDO organization that is a member of both ANSI and ISO. Through IEEE, many of the world's standards in telecommunications, electronics, electrical applications, and computers have been developed.

IEEE 1073, Standard for Medical Device Communications, has produced a family of documents that defines the entire seven-layer communications requirements for the **Medical Information Bus** (MIB; Gottschalk 1991). The MIB is a robust, reliable communication service designed for bedside devices in the intensive care unit, operating room, and emergency room. These standards have been harmonized with work in CEN, and the results are released as ISO standards. IEEE and HL7 have collaborated on several key standards, including those for mobile medical devices.

Link: ► www.ieee.org

Link to standards: ► https://standards.ieee.org/standard/index.html

7.3.2.10 ISO Technical Committee 215—Health Informatics

In 1989, interests in the European Committee for Standardization (CEN) and the United States led to the creation of Technical Committee (TC) 215 for Health Information within ISO.

TC 215 meets once in a year as a TC and once as a Joint Working Group. TC 215 follows rather rigid procedures to create ISO standards. Thirty-five countries are active participants in the TC with another 23 countries acting as observers. While the actual work is done in the working groups, the balloting process is very formalized—one vote for each participating country. For most work, there are a defined series of steps, beginning with a New Work Item Proposal and getting five countries to participate; a Working Document, a Committee Document; a Draft International Standard, a Final Draft International Standard (FDIS); and finally an International Standard. This process, if fully followed, takes several years to produce an International Standard. Under certain conditions, a fast track to FDIS is permitted. Technical Reports and Technical Specifications are also permitted.

The United States has been assigned the duties of Secretariat, and that function is carried out, at this time, by ANSI. ANSI also serves as the U.S. Technical Advisory Group Administrator, which represents the U.S. position in ISO.

Link: ► https://www.iso.org/standards.html

7.3.2.11 Integrating the Healthcare Enterprise

The goal of the Integrating the Healthcare Enterprise (IHE) initiative is to stimulate integration of healthcare information resources. IHE is sponsored jointly by the Radiological Society of North America (RSNA) and the HIMSS. Using established standards and working with direction from medical and information technology professionals, industry leaders in healthcare information and imaging systems cooperate under IHE to agree upon implementation profiles for the transactions used to communicate images and patient data within the enterprise. Their incentive for participation is the opportunity to demonstrate that their systems can operate efficiently in standards-based, multi-vendor environments with the functionality of real HISs. Moreover, IHE enables vendors to direct product development resources toward building increased functionality rather than redundant interfaces.

Link: ► https://www.ihe.net/

Link to IHE domains: ► https://www.ihe.net/ihe_domains/

7.3.2.12 National Council for Prescription Drug Program (NCPDP)

NCPDP is a not-for-profit, multi-stakeholder forum for developing and promoting industry standards and business solutions that improve patient safety and health outcomes, while also decreasing costs. NCPDP is an ANSI accredited SDO and uses a consensus-building process to create national standards for real-time, electronic exchange of healthcare information. Their primary focus is on information exchange for prescribing, dispensing, monitoring, managing and paying for medications and pharmacy services crucial to quality healthcare.

Link: ► www.ncpdp.org/home

Link to standards: ► www.ncpdp.org/standards-Development

7.3.2.13 OpenEHR

OpenEHR is the name of a technology for e-health, consisting of open specifications, clinical models and software that can be used to create standards and build information and interoperability solutions for healthcare. The various artefacts of openEHR are produced by the openEHR community and managed by the openEHR Foundation, an international non-profit organization established in the year 2003.

Link: ► openehr.org/

Link to Clinical Models: ► www.openehr.org/clinicalmodels

Link to Clinical Knowledge Manager: ► www.openehr.org/ckm

Link to Software Programs: ► www.openehr.org/programs/software

Link to Specification Program: ► openehr.org/programs/specification

7.3.2.14 Personal Connected Health Alliance (PCHA)

The PCHA publishes the Continua Design Guidelines to enable a flexible implementation framework for end-to-end interoperability of personal connected health devices and systems. These Guidelines are recognized by the International Telecommunications Union (ITU) as the international standard for safe, secure and reliable exchange of data to and from personal health devices.

Link to guidelines: ► www.pchalliance.org/continua-design-guidelines

7.3.2.15 SNOMED International

SNOMED International (previously called IHTSDO) was founded in 2007 with nine charter members. Currently 19 countries, including the United States, are members. The primary purpose of IHTSDO is the continued development and maintenance of the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT; see ► Sect. 7.4.4.4, below). Member countries make SNOMED-CT freely available to its citizens. SNOMED has a number of Special Interest Groups: Anesthesia, Concept Model, Education, Implementation, International Pathology & Laboratory Medicine, Mapping, Nursing, Pharmacy, and Translation.

Link: ► http://www.snomed.org/

7.3.2.16 OHDSI

The Observational Health Data Sciences and Informatics (OHDSI) program is a multi-stakeholder, interdisciplinary open-source collaborative to leverage the value of health data through large-scale analytics. OHDSI has established an international network of researchers and observational health databases. ODHSI enables active engagement across multiple disciplines, including clinical medicine, biostatistics, computer science, epidemiology, and life sciences (◘ Fig. 7.2).

Link: ► https://www.ohdsi.org/

7.4 Detailed Clinical Models, Coded Terminologies, Nomenclatures, and Ontologies

The capture, storage, and use of clinical data in computer systems is complicated by lack of agreement on terms and meanings. In recent years there has also been a growing recognition that just standardizing the terms and codes used in medicine is not sufficient to enable interoperability. The structure or form

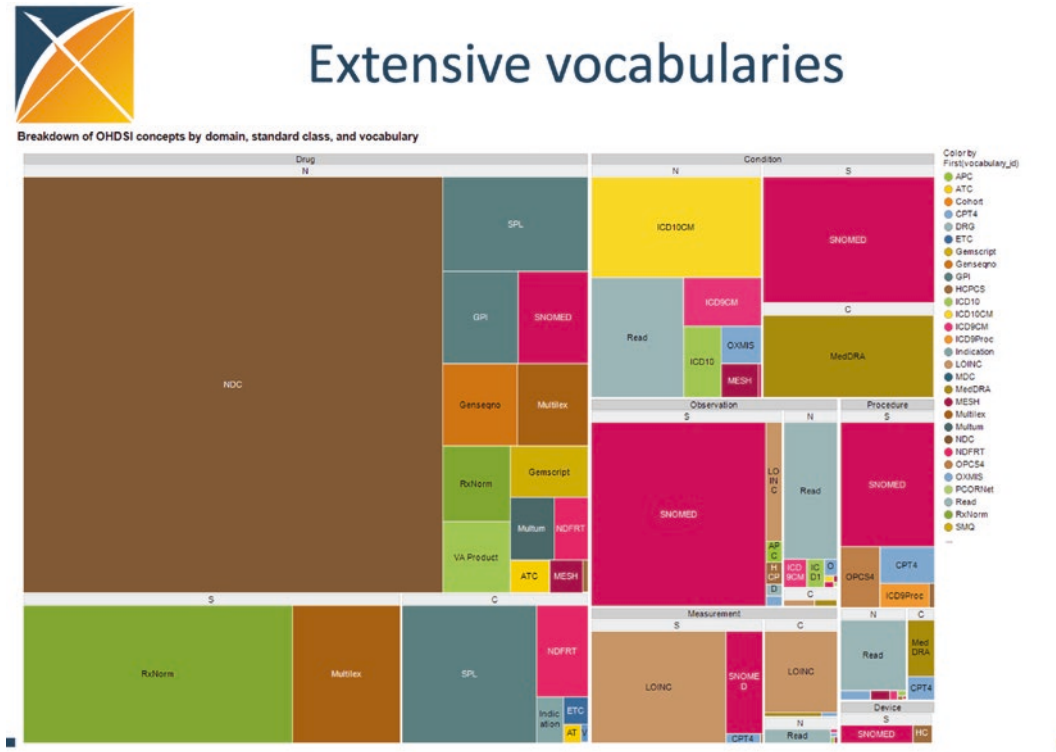


Fig. 7.2 Extensive International Vocabularies, George Hripscak, MD (by permission). A critical challenge to data analytics across international domains continues to be the requirement for integrating data sources among multiple vocabularies. (Photo courtesy of George Hripcsak, MD, with permission)

of medical data provides important context for computable understanding of the data. Terms and codes need to be interpreted in the context of clinical information models. The many terminologies and detailed clinical modeling activities discussed in this section have been developed to ease the communication of coded medical information.

7.4.1 Motivation for Structured and Coded Data

The structuring and encoding of medical information is a basic function of most clinical systems. Standards for such structuring and encoding can serve two purposes. First, they can save system developers from reinventing the wheel. For example, if an application allows caregivers to compile problem lists about their patients, using a standard structure and terminology saves developers from having to create their own. Second, using commonly accepted standards can facilitate exchange of data, applications, and clinical decision support logic among systems. For example, if a central database is accepting clinical data from many sources, the task is greatly simplified if each source is using the same logical data structure and coding scheme to represent the data. System developers often ignore available standards and continue to develop their own solutions. It is easy to believe that the developers have resisted adoption of standards because it is too much work to understand and adapt to any system that was "not invented here." The reality, however, is that the available standards are often inadequate for the needs of the users (in this case, system developers). As a result, no standard terminology enjoys the wide acceptance sufficient to facilitate the second function: exchange of coded clinical information.

The need for detailed clinical models is directly related to the second goal dis-

cussed above, that of creating interoperability between systems. The subtle relationship between terminologies and models is best understood using a couple of examples. If a physician wants to record the idea that a patient had "chest pain that radiates to the back", the following coded terms could be used from SNOMED-CT.

7.4.2 Detailed Clinical Models

The creation of unambiguous data representation is a combination of creating appropriate structures (models) for representing the form of the data and then linking or "binding" specific sets of codes to the coded elements in the structures. Several modeling languages or formalisms have been found to be useful in describing the structure of the data. They include:

- UML – the Unified Modeling Language, Object Management Group
- ADL – Archetype Definition Language, OpenEHR Foundation
- CDL – Constraint Definition Language, General Electric and Intermountain Health care
- MIF – Model Interchange Format, Health Level Seven International Inc.
- OWL – Web Ontology Language, World Wide Web Consortium

All languages used for clinical modeling need to accomplish at least two major things: they need to show the "logical" structure of the data, and they need to show how sets of codes from standard terminologies participate in the logical structure. Defining the logical structure is simply showing how the named parts of a model relate to one another. Model elements can be contained in other elements, creating hierarchies of elements. It is also important to specify which elements of the model can occur more than once (cardinality), which elements are required, and which are optional. Terminology binding is the act of creating connections between the elements in a model and concepts in a coded terminology. For each coded element in a model, the set of allowed values for the coded element are specified. The HL7 Vocabulary Working Group has created a comprehensive discussion of how value sets can be defined and used with information models.

There are many clinical information modeling activities worldwide. Some of the most important activities are briefly listed below.

- HL7 Activities
  - HL7 Detailed Clinical Models – This group has developed a method for specifying clinical models based on the HL7 Reference Information Model (RIM) that guarantees that data that conforms to the model could be sent in HL7 Version 3 messages.
  - HL7 Clinical Document Architecture (CDA) Templates – This group has defined a standard way of specifying the structure of data to be sent in XML documents that conform to the CDA standard.
  - HL7 TermInfo – This Workgroup of HL7 has specified a set of guidelines for how SNOMED-CT codes and concepts should be used in conjunction with the HL7 RIM to represent data sent in HL7 Version 3 messages.
  - HL7 Clinical Information Modeling Initiative – A formerly independent group now organized as and HL7 Workgroup is chartered to develop implementable clinical information models.
- The openEHR Foundation is developing models based on a core reference model and the Archetype Definition Language. This approach has been adopted by several national health information programs.
- EN 13606 is developing models based on the ISO/CEN 13606 standard and core reference model.
- The US Veterans Administration (VA) is creating models for integrating data across all VA facilities and for integration with military hospitals that are part of the US Department of Defense. The modeling is done primarily using Unified Modeling Language.
- US Department of Defense is creating models for integrating data across all DoD facilities and for integration with VA

facilities. The modeling is done primarily using Unified Modeling Language.
- The National Health Service in the United Kingdom (UK) is developing the Logical Record Architecture to provide models for interoperability across all health care facilities in the UK. The modeling is done primarily using UML.
- Clinical Element Models – Intermountain Health care and General Electric have created a set of detailed clinical models using a core reference model and Constraint Definition Language. The models are free-for-use and are available for download from the Internet.
- SHARE Models – CDISC is creating models to integrate data collected as part of clinical trials.
- SMART Team – This group at Boston Children's Hospital is defining standard application programming interfaces (APIs) for securely connecting with EHRs.
- Clinical Information Modeling Initiative (CIMI) – This is an international consortium that has the goal of establishing a free-for-use repository of detailed clinical models, where the models are expressed in a single common modeling language with explicit bindings to standard terminologies.
- OMOP – The OMOP Common Data Model enables the systematic analysis of disparate observational claims-based data. The initial approach was to transform data contained within claims databases into a common format (data model) as well as a common representation (terminologies, vocabularies, coding schemes). OMOP has been primarily used by the OHDSI community (see ► Sect. 7.3.2.16), but is being examined by other consortia seeking to share patient-level clinical information.

7.4.3 Vocabularies, Terminologies, and Nomenclatures

In discussing coding systems, the first step is to clarify the differences among a **terminology**, a **vocabulary**, and a **nomenclature**. These terms are often used interchangeably by creators of coding systems and by authors discussing the subject. Fortunately, although there are few accepted standard terminologies, there is a generally accepted standard about terminology: ISO Standard 1087 (Terminology—Vocabulary).

Finally, we should consider the methods by which the terminology is maintained. Every standard terminology must have an ongoing maintenance process, or it will rapidly become obsolete. The process must be timely and must not be too disruptive to people using an older version of the terminology. For example, if the creators of the terminology choose to rename a code, what happens to the data previously recorded with that code?

7.4.4 Specific Terminologies

With these considerations in mind, let us survey some of the available controlled terminologies. There are introductory descriptions of a few current and common terminologies. New terminologies appear annually, and existing proprietary terminologies often become publicly available. When reviewing the following descriptions, try to keep in mind the background motivation for a development effort. All these standards are evolving rapidly, and one should consult the Web sites or other primary sources for the most recent information.

7.4.4.1 International Classification of Diseases and Its Clinical Modifications

One of the most recognized terminologies is the International Classification of Diseases (ICD). First published in 1893, it has been revised at roughly 10-year intervals, first by the Statistical International Institute and later by the World Health Organization (WHO). The Ninth Edition (ICD-9) was published in 1977 (World Health Organization 1977) and the Tenth Edition (ICD-10) in 1992 (World Health Organization 1992). The ICD-9 coding system consists of a core classification of three-digit codes that are the minimum required for reporting mortality statistics to WHO. A fourth digit (in the first decimal

place) provides an additional level of detail; usually .0 to .7 are used for more specific forms of the core term, .8 is usually "other," and .9 is "unspecified." Codes in the ICD-10 coding system start with an alpha character and consist of three to seven characters. In both systems, terms are arranged in a strict hierarchy, based on the digits in the code. In addition to diseases, ICD includes several "families" of terms for medical-specialty diagnoses, health status, disease-related events, procedures, and reasons for contact with health care providers.

In June of 2018, the World Health Organization (WHO) replaced ICD-10 with ICD-11, with the intention to make a more comprehensive and computable version that could be maintained without the need for major version changes in the future. ICD-11 includes many new features, such as a semantic network, a polyhierarchy, a formal information model, and a set of "linearization" that constrain the terms in strict hierarchies, designed to support various functions (Chute (2018)). The US National Committee for Vital and Health Statistics (NCVHS) is currently considering the timing and method for transition to ICD-11, so close on the 2015 adoption of ICD-10.

7.4.4.2 Current Procedural Terminology

The American Medical Association developed the Current Procedural Terminology (CPT) in 1966 (American Medical Association, updated annually) to provide a pre coordinated coding scheme for diagnostic and therapeutic procedures that has since been adopted in the United States for billing and reimbursement. Like the DRG codes, CPT codes specify information that differentiates the codes based on cost. For example, there are different codes for pacemaker insertions, depending on whether the leads are "epicardial, by thoracotomy" (33200), "epicardial, by xiphoid approach" (33201), "transvenous, atrial" (33206), "transvenous, ventricular" (33207), or "transvenous, atrioventricular (AV) sequential" (33208). CPT also provides information about the reasons for a procedure. For example, there are codes for arterial punctures for "withdrawal of blood for diagnosis" (36600), "monitoring" (36620), "infusion therapy" (36640), and "occlusion therapy" (75894). Although limited in scope and depth (despite containing over 8000 terms), CPT-4 is the most widely accepted nomenclature in the United States for reporting physician procedures and services for federal and private insurance third-party reimbursement.

7.4.4.3 Diagnostic and Statistical Manual of Mental Disorders

The American Psychiatric Association published the Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) in May 2013 (American Psychiatric Association Committee on Nomenclature and Statistics (2013)). DSM-5 is the standard classification of mental disorders used by mental health professionals in the U.S. and contains a listing of diagnostic criteria for every psychiatric disorder recognized by the U.S. healthcare system.[9] The previous edition, DSM-IV, was originally published in 1994 and revised in 2000 as DSM-IV-TR. DSM-5 is coordinated with ICD-10.

7.4.4.4 SNOMED Clinical Terms and Its Predecessors

Drawing from the New York Academy of Medicine's Standard Nomenclature of Diseases and Operations (SNDO) (Plunkett 1952; Thompson and Hayden 1961; New York Academy of Medicine 1961), the College of American Pathologists (CAP) developed the Standard Nomenclature of Pathology (SNOP) as a multiaxial system for describing pathologic findings (College of American Pathologists 1971) through post-coordination of topographic (anatomic), morphologic, etiologic, and functional terms. SNOP has been used widely in pathology systems in the United States; its successor, the Systematized Nomenclature of Medicine (SNOMED) has evolved beyond an abstracting scheme to become a comprehensive coding system.

Largely the work of Roger Côté and David Rothwell, SNOMED was first published in 1975, was revised as SNOMED II in 1979, and then greatly expanded in 1993 as the

9 ► https://www.psychiatry.org/psychiatrists/practice/dsm (last accessed 12/2/2019)

Systematized Nomenclature of Human and Veterinary Medicine—SNOMED.

International (Côté and Rothwell 1993). Each of these versions was **multi-axial**; coding of patient information was accomplished through the post-coordination of terms from multiple axes to represent complex terms that did not exist as single codes in SNOMED. In 1996, SNOMED changed from a multi-axial structure to a more logic-based structure called a Reference Terminology (Campbell et al. 1998; Spackman et al. 1997a, 1997b), intended to support more sophisticated data encoding processes and resolve some of the problems with earlier versions of SNOMED (**see** Fig. 7.2). In 1999, CAP and the NHS announced an agreement to merge their products into a single terminology called SNOMED Clinical Terms (SNOMED-CT) (Spackman 2000), containing terms for over 344,000 concepts (see Fig. 7.3). SNOMED-CT is currently maintained by a not-for-profit association once called the International Health Terminology Standards Development Organization (IHTSDO), but now simply SNOMED International.

Despite the broad coverage of SNOMED-CT, it continues to allow users to create new, ad hoc terms through post-coordination of existing terms. While this increases the expressivity, users must be careful not to be too expressive because there are few rules about how the post-coordination coding should be done, the same expression might end up being represented differently by different coders. For example, "acute appendicitis" can be coded as a single disease term, as a combination of a modifier ("acute") and a disease term ("appendicitis"), or as a combination of a modifier ("acute"), a morphology term ("inflammation") and a topography term ("vermiform appendix"). Users must therefore be careful when post-coordinating terms, not to recreate a meaning that is satisfied by an already existing single code. SNOMED-CT's description logic, such as the example in Fig. 7.2, can help guide users when selecting modifiers.

Concept: Bacterial pneumonia
- Concept Status Current
- fully defined by ...
  - Is a
    - Infectious disease of lung
    - Inflammatory disorder of lower respiratory tract
    - Infective pneumonia
    - Inflammation of specifie body organs
    - Inflammation of specifie body systems
    - Bacterial infectious disease
  - Causative agent:
    - Bacterium
  - Pathological process:
    - Infections disease
  - Associated morphology:
    - Inflammation
  - Finding site:
    - Lung structure
  - Onset:
    - Subacute onset
    - Acute onset
    - Insidious onset
    - Sudden onset
  - Severity:
    - Severities
  - Episodicity:
    - Episodicities
  - Course:
    - Courses
- Descriptions:
  - Bacterial pneumonia (disorder)
  - Bacterial pneumonia
- Legacy codes:
  - SNOMED: DE-10100
  - CTV31D: X100H

Fig. 7.3 Description-logic representation of the SNOMED-CT term "Bacterial Pneumonia." The "Is a" attributes define bacterial pneumonia's position in SNOMED-CT's multiple hierarchy, while attributes such as "Causative Agent" and "Finding Site" provide definitional information. Other attributes such as "Onset" and "Severities" indicate ways in which bacterial pneumonia can be postcoordinated with other terms, such as "Acute Onset" or any of the descendants of the term "Severities." "Descriptions" refers to various text strings that serve as names for the term, while "Legacy Codes" provide back-ward compatibility to SNOMED and Read Clinical Terms (NHS Centre for Coding and Classification (1994))

7.4.4.5 GALEN

In Europe, a consortium of universities, agencies, and vendors, with funding from the Advanced Informatics in Medicine initiative (AIM), has formed the GALEN project

to develop standards for representing coded patient information (Rector et al. 1995). GALEN developed a reference model for medical concepts using a formalism called Structured Meta Knowledge (SMK) and a formal representation language called GALEN Representation and Integration Language (GRAIL). Using GRAIL, terms are defined through relationships to other terms, and grammars are provided to allow combinations of terms into sensible phrases. The reference model is intended to allow representation of patient information in a way that is independent of the language being recorded and of the data model used by an electronic medical record system. The GALEN developers are working closely with CEN TC 251 (see ▶ Sect. 8.3.2) to develop the content that will populate the reference model with actual terms.

7.4.4.6 Logical Observations, Identifiers, Names, and Codes

An independent consortium, led by Clement J. McDonald and Stanley M. Huff, has created a naming system for tests and observations. The system is called Logical Observation Identifiers Names and Codes (LOINC).[10] The coding system contains names and codes for laboratory tests, patient measurements, assessment instruments, document and section names, and radiology exams. ◘ Figure 7.4 shows some typical fully specified names for common laboratory tests. The standard specifies structured coded semantic information about each test, such as the substance measured and the analytical method used.

7.4.4.7 Nursing Terminologies

Nursing organizations and research teams have been extremely active in the development of standard coding systems for documenting and evaluating nursing care. One review counted a total of 12 separate projects active worldwide (Coenen et al. 2001), including coordination with SNOMED and LOINC. These projects have arisen because general medical terminologies fail to represent the kind of clinical concepts needed in nursing care. For example, the kinds of problems that appear in a physician's problem list (such as "myocardial infarction" and "diabetes mellitus") are relatively well represented in many of the terminologies that we have described, but the kinds of problems that appear in a nurse's assessment (such as "activity intolerance" and "knowledge deficit related to myocardial infarction") are not. Preeminent nursing terminologies include the North American Nursing Diagnosis Association (NANDA) codes, the Nursing Interventions Classification (NIC), the Nursing Outcomes Classification (NOC), the Georgetown Home Health Care Classification (HHCC), and the Omaha System (which covers problems, interventions, and outcomes).

Despite the proliferation of standards for nursing terminologies, gaps remain in the coverage of this domain (Park and Cho 2009). The International Council of Nurses and the International Medical Informatics Association Nursing Informatics Special Interest Group have worked together to produce the International Classification for Nursing Practice (ICNP®). This system uses a post-coordinated approach for describing nursing diagnoses, actions, and outcomes.

7.4.4.8 Drug Codes

A variety of public and commercial terminologies have been developed to represent terms used for prescribing, dispensing and administering drugs. The WHO Drug Dictionary is an international classification of drugs that provides proprietary drug names used in different countries, as well as all active ingredients and the chemical substances, with Chemical Abstract numbers. Drugs are classified according to the Anatomical-Therapeutic-Chemical (ATC) classification, with cross-references to manufacturers and reference sources. The current dictionary contains 25,000 proprietary drug names, 15,000 single ingredient drugs, 10,000 multiple ingredient drugs, and 7000

10 5 ▶ loinc.org (accessed 5/30/19)

Fig. 7.4 Examples of codes in SNOMED-CT, showing some of the hierarchical relationships among bacterial pneumonia terms. Tuberculosis terms and certain terms that are included in SNOMED-CT for compatibility with other terminologies are not shown. Note that some terms such as "Congenital group A hemolytic streptococcal pneumonia" appear under multiple parent terms, while other terms, such as "Congenital staphylococcal pneumonia" are not listed under all possible parent terms (e.g., it is under "Congenital pneumonia" but not under "Staphylococcal pneumonia"). Some terms, such as "Pneumonic plague" and "Mycoplasma pneumonia" are not classified under Bacterial Pneumonia, althogh the causative agents in their descriptions ("Yersinia pestis" and "Myocplasma pneumioniae", respectively) are classified under "Bacterium", the causative agent of Bacterial pneumonia

- Pneumonia
  - Bacterial pneumonia
    - Proteus pneumonia
    - Legionella pneumonia
    - Anthrax pneumonia
    - Actinomycotic pneumonia
    - Nocardial pneumonia
    - Meningocoocal pneumonia
    - Chlamydial pneumonia
      - Neonatal chlamydial pneumonia
      - Ornithosis
        - Ornithosis with complication
        - Ornithosis with pneumonia
    - Congenital bacterial pneumonia
      - Congenital staphylococcal pneumonia
      - Congenital group A hemolytic streptocoocal pneumonia
      - Congenital group B hemolytic streptocoocal pneumonia
      - Congenital *Escherichia colt* pneumonia
      - Congenital pseudomonal pneumonia
    - Chlamydial pneumonitis in all species except pig
      - Feline pneumonitis
    - Staphylocoocal pneumonia
    - Pulmonary actinobacillosis
    - Pneumonia in Q fever
    - Pneumonia due to Streptococcus
      - Group B streptococcal pneumonia
      - Congenital group A hemolytic streptococcal pneumonia
      - Congenital group B hemolytic streptococcal pneumonia
      - Pneumococcal pneumonia
      - Pneumococcal lobar pneumonia
      - AIDS with pneumococcal pneumonia
    - Pneumonia due to Pseudomonas
      - Congenital pseudomonal pneumonia
    - Pulmonary tularemia
    - Enzootic pneumonia of calves
    - Pneumonia in pertussis
    - AIDS with bacterial pneumonia
    - Enzootic pneumonia of sheep
    - Pneumonia due to *Klebstella pneumontae*
    - *Hemophilus influenzae* pneumonia
    - Porcine contagious pleuropneumonia
    - Pneumonia due to pleuropneumonia-like organism
    - Secondary bacterial pneumonia
  - Pneumonic plague
    - Primary pneumonic plague
    - Secondary pneumonic plague
  - Salmonella pneumonia
    - Pneumonia in typhoid fever
  - Infective pneumonia
    - Mycoplasma pneumonia
      - Enzootic mycoplasmal pneumonia of swine
  - Achromobacter pneumonia
  - Bovine pneumonic pasteurellosis
  - Corynebacterial pneumonia of foals
  - Pneumonia due to *Escherichia coli*
  - Pneumonia due to *Proteus mirabilis*

chemical substances. The dictionary now covers drugs from 34 countries and grows at a rate of about 2000 new entries per year.

The National Drug Codes (NDC), produced by the U.S. Food and Drug Administration (FDA), is applied to all drug packages. It is widely used in the United States, but it is not as comprehensive as the WHO codes. The FDA designates part of the code based on drug manufacturer, and each manufacturer defines the specific codes for their own products. As a result, there is no uniform class hierarchy for the codes, and codes may be reused at the manufacturer's discretion. Due in part to the inadequacies of the NDC codes, pharmacy information systems typically purchase proprietary terminologies from knowledge base vendors. These terminologies map to NDC, but provide additional information about therapeutic classes, allergies, ingredients, and forms.

The need for standards for drug terminologies led to a collaboration between the FDA, the U.S. National Library of Medicine (NLM), the Veterans Administration (VA), and the pharmacy knowledge base vendors that has produced a representational model for drug terms called RxNorm. The NLM provides RxNorm to the public as part of the Unified Medical Language System (UMLS) (see below) to support mapping between NDC codes, the VA's National Drug File (VANDF) and various proprietary drug terminologies (Nelson et al. 2002). RxNorm currently contains 14,000 terms.

7.4.4.9 Medical Subject Headings

The Medical Subject Headings (MeSH), maintained by the NLM (updated annually), is the terminology by which the world medical literature is indexed. MeSH arranges terms in a structure that breaks from the strict hierarchy used by most other coding schemes. Terms are organized into hierarchies and may appear in multiple places in the hierarchy (◘ Fig. 7.5). Although it is not generally used as a direct coding scheme for patient information, it plays a central role in the UMLS.

7.4.4.10 RadLex

RadLex is a terminology produced by the Radiology Society of North America (RSNA). With more than 30,000 terms, RadLex is intended to be a unified language of radiology terms for standardized indexing and retrieval of radiology information resources. RadLex includes the names of anatomic parts, radiology devices, imaging exams and procedure steps performed in radiology. Given the scope of the radiology domain, many RadLex terms overlap with SNOMED-CT, and LOINC.

7.4.4.11 Bioinformatics Terminologies

For the most part, the terminologies discussed above fail to represent the levels of detail needed by biomolecular researchers. This has become a more acute problem with the advent of bioinformatics and the sequencing of organism genomes (see ► Chap. 11). As in other domains, researchers have been forced to develop their own terminologies. As these researchers have begun to exchange information, they have recognized the need for standard naming conventions as well as standard ways of representing their data with terminologies. Prominent efforts to unify naming systems include the Gene Ontology (GO) (Harris et al. 2004) from the Gene Ontology Consortium and the gene naming database of the HUGO Gene Nomenclature Committee (HGNC). A related resource is the RefSeq database of the National Center for Biotechnology Information (NCBI) which contains identifiers for reference sequences.

7.4.4.12 Unified Medical Language System

In 1986, Donald Lindberg and Betsy Humphreys, at the NLM, began working with several academic centers to identify ways to construct a resource that would bring together and disseminate controlled medical terminologies. An experimental version of the UMLS was first published in 1989 (Humphreys 1990); the UMLS has been updated annually since then. Its principal component is the

Blood glucose	GLUCOSE:MCNC:PT:BLD:QN:
Plasma glucose	GLUCOSE:MCNC:PT:PLAS:QN:
Serum glucose	GLUCOSE:MCNC:PT:SER:QN:
Urine glucose concentration	GLUCOSE:MCNC:PT:UR:QN:
Urine glucose by dip slick	GLUCOSE:MCNC:PT:UR:SQ:TEST STRIP
Glucose tolerance test at 2 hours	GLUCOSE.2H POST 100 G GLUCOSE PO: MCNC:PT:PLAS:QN:
Ionized whole blood calcium	CALCIUM.FREE:SCNC:PT:BLD:QN:
Serum or plasma ionized calcium	CALCIUM.FREE:SCNC:PT:SER/PLAS:QN:
24-hour calcium excretion	CALCIUM.TOTAL:MRAT:24H:UR:QN:
Whole blood total calcium	CALCIUM.TOTAL:SCNC:PT:BLD:QN:
Serum or plasma total calcium	CALCIUM.TOTAL:SCNC:PT:SER/PLAS:QN:
Automated hematocrit	HEMATOCRIT:NFR:PT:BLD:QN: AUTOMATED COUNT
Manual spun hematocrit	HEMATOCRIT:NFR:PT:BLD:QN:SPUN
Urine erythrocyte casts	ERYTHROCYTE CASTS:ACNC:PT:URNS:SQ: MICROSCOPY.LIGHT
Erythrocyte MCHC	ERYTHROCYTE MEAN CORPUSCULAR HEMOGLOBIN CONCENTRATION:MCNC:PT:RBC:QN:AUTOMATED COUNT
Erythrocyte MCH	ERYTHROCYTE MEAN CORPUSCULAR HEMOGLOBIN:MCNC:PT:RBC:QN: AUTOMATED COUNT
Erythrocyte MCV	ERYTHROCYTE MEAN CORPUSCULAR VOLUME:ENTVOL:PT:RBC:ON:AUTOMATED COUNT
Automated Blood RBC	ERYTHROCYTES:NCNC:PT:BLD:QN: AUTOMATED COUNT
Manual blood RBC	ERYTHROCYTES:NCNC:PT:BLD:QN: MANUAL COUNT
ESR by Westergren method	ERYTHROCYTE SEDIMENTATION RATE:VEL:PT:BLD:QN:WESTERGREN
ESR by Wintrobe method	ERYTHROCYTE SEDIMENTATION RATE:VEL:PT:BLD:QN:WINTROBE

Fig. 7.5 Examples of common laboratory test terms as they are encoded in LOINC. The major components of the fully specified name are in separate columns and consist of the analyte, the property (e.g., Mcnc mass concentration, Scnc substance concentration, Acnc arbitrary concentration, Vfr volume fraction, EntMass entitic mass, EntVol entitic volume, Vel velocity, and Ncnc number concentration), the timing (Pt point in time), the system (specimen), and the method (Ord ordinal, Qn quantitative)

Metathesaurus, which contains over 8.9 million terms collected from over 160 different sources (including many of those that we have discussed), and attempts to relate synonymous and similar terms from across the different sources into over 2.6 million concepts (Fig. 7.6). Figure 7.7 lists the preferred names for many of the pneumonia concepts in the Metathesaurus; Fig. 7.8 shows how like terms are grouped into concepts and are tied to other concepts through semantic relationships. Figure 7.9 shows some of the information available in the Unified Medical Language System about selected pneumonia concepts.

7.5 Data Interchange Standards

The recognition of the need to interconnect health care applications led to the development and enforcement of **data interchange standards**. The conceptualization stage began in 1980 with discussions among individuals in an organization called the American Association for Medical Systems and Informatics (AAMSI). In 1983, an AAMSI task force was established to pursue those interests in developing standards.

The development phase was multifaceted. The AAMSI task force became subcommittee E31.11 of the ASTM and developed and pub-

Respiratory Tract Diseases
  Lung Diseases
    Pneumonia
      Bronchopneumonia
      Pneumonia, Aspiration
        Pneumonia, Lipid
      Pneumonia, Lobar
      Pneumonia, Mycoplasma
      Pneumonia, *Pneumocystis carinii*
      Pneumonia, Rickettsial
      Pneumonia, Staphylococcal
      Pneumonia, Viral
    Lung Diseases, Fungal
      Pneumonia, *Pneumocystis carinii*
  Respiratory Tract Infections
    Pneumonia
      Pneumonia, Lobar
      Pneumonia, Mycoplasma
      Pneumonia, *Pneumocystis carinii*
      Pneumonia, Rickettsial
      Pneumonia, Staphylococcal
      Pneumonia, Viral
    Lung Diseases, Fungal
      Pneumonia, *Pneumocystis carinii*

Fig. 7.6 Partial tree structure for the Medical Subject Headings showing pneumonia terms. Note that terms can appear in multiple locations, although they may not always have the same children, implying that they have somewhat different meanings in different contexts. For example, Pneumonia means "lung inflammation" in one context (line 3) and "lung infection" in another (line 16)

Fig. 7.7 Growth of the UMLS. The UMLS Metathesaurus contains 3.85 million concepts and 14.6 million unique concept names from 210 source vocabularies. The content continues to grow dynamically in response to user needs (Source: U.S. National Library of Medicine)

lished ASTM standard 1238 for the exchange of clinical-laboratory data. Two other groups were formed to develop standards, each with a slightly different emphasis: HL7 and Institute of Electrical and Electronics Engineering (IEEE) Medical Data Interchange ("Medix") Standard. The American College of Radiology (ACR) joined with the National Electronic Manufacturers Association (NEMA) to develop a standard for the transfer of image data. Two other groups developed related standards independent of the biomedical informatics community: (1) ANSI X12 for the transmission of commonly used business transactions, including health care claims and benefit data, and (2) National Council for Prescription Drug Programs (NCPDP) for the transmission of third-party drug claims.

7.5.1 General Concepts and Requirements

The purpose of a data-interchange standard is to permit one system, the **sender**, to transmit to another system, the **receiver**, all the data required to accomplish a specific communication, or **transaction set**, in a precise, unambiguous fashion. To complete this task successfully, both systems must know what format and content is being sent and must understand the words or terminology, as well as the delivery mode.

A communications model, called the Open Systems Interconnection (OSI) reference model (ISO 7498–1), has been defined by the ISO (see ▶ Chap. 5 and the discussion of software for network communications). It describes seven levels of requirements or specifications for a communications exchange: physical, data link, network, transport, session, presentation, and application (Rose 1989; Stallings 1987; Tanenbaum 1987). Level 7, the application level, deals primarily with the semantics or data-content specification of the transaction set or message. For the data-interchange standard, HL7 requires the definition of all the data elements to be sent in

Fig. 7.8 Some of the bacterial pneumonia concepts in the Unified Medical Language System Metathesaurus

C0004626: Pneumonia, Bacterial
C0023241: Legionnaires' Disease
C0032286: Pneumonia due to other specified bacteria
C0032308: Pneumonia, Staphylococcal
C0152489: Salmonella pneumonia
C0155858: Other bacterial pneumonia
C0155859: Pneumonia due to *Klebsiella pneumoniae*
C0155860: Pneumonia due to Pseudomonas
C0155862: Pneumonia due to Streptococcus
C0155865: Pneumonia in pertussis
C0155866: Pneumonia in anthrax
C0238380: PNEUMONIA, KLEBSIELLA AND OTHER GRAM NEGATIVE BACILLI
C0238381: PNEUMONIA, TULAREMIC
C0242056: PNEUMONIA, CLASSIC PNEUMOCOCCAL LOBAR
C0242057: PNEUMONIA, FRIEDLAENDER BACILLUS
C0275977: Pneumonia in typhoid fever
C0276026: Hemophilus influenzae pneumonia
C0276039: Pittsburgh pneumonia
C0276071: Achromobacter pneumonia
C0276080: Pneumonia due to *Proteus mirabilis*
C0276089: Pneumonia due to *Escherichia coli*
C0276523: AIDS with bacterial pneumonia
C0276524: AIDS with pneumococcal pneumonia
C0339946: Pneumonia with tularemia
C0339947: Pneumonia with anthrax
C0339952: Secondary bacterial pneumonia
C0339953: Pneumonia due to *Escherichia coli*
C0339954: Pneumonia due to proteus
C0339956: Typhoid pneumonia
C0339957: Meningococcal pneumonia
C0343320: Congenital pneumonia due to staphylococcus
C0343321 : Congenital pneumonia due to group A hemolytic streptococcus
C0343322: Congenital pneumonia due to group B hemolytic streptococcus
C0343323: Congenital pneumonia due to *Escherichia coli*
C0343324: Congenital pneumonia due to pseudomonas
C0348678: Pneumonia due to other aerobic Gram-negative bacteria
C0348680: Pneumonia in bacterial diseases classified elsewhere
C0348801: Pneumonia due to streptococcus, group B
C0349495: Congenital bacterial pneumonia
C0349692: Lobar (pneumococcal) pneumonia
C0375322: Pneumococcal pneumonia {*Streptococcus pneumoniae pneumonia*}
C0375323: Pneumonia due to Streptococcus, unspecified
C0375324: Pneumonia due to Streptococcus Group A
C0375326: Pneumonia due to other Streptococcus
C0375327: Pneumonia due to anaerobes
C0375328: Pneumonia due to *Escherichia coli*
C0375329: Pneumonia due to other Gram-negative bacteria
C0375330: Bacterial pneumonia, unspecified

response to a specific task, such as the admission of a patient to a hospital. In many cases, the data content requires a specific terminology that can be understood by both sender and receiver.

Presentation, the sixth level of Interoperability, addresses the syntax of the message, or how the data are formatted. There are both similarities and differences at this level across the various standards bodies. Two philosophies are used for defining syntax: one proposes a position-dependent format; the other uses a tagged-field format. In the position-dependent format, the data content is specified and defined by position.

The remaining OSI levels—session, transport, network, data link, and physical—govern the communications and networking

Bacterial *pneumonia*
Source: CSP93/PT/259S.5280; DOR27/DT/U000523; ICD91/PT/482.9; ICD91/IT/482.9
Parent: Bacterial Infections; Pneumonia; Influenza with Pneumonia
Child: Pneumonia, Mycoplasma
Narrower: Pneumonia, Lobar; Pneumonia, Rickettsial; Pneumonia, Staphylococcal; Pneumonia due to *Klebsiella pneumoniae;* Pneumonia due to Pseudomonas; Pneumonia due to *Hemophilus influenzae*
Other: *Klebsiella pneumoniae, Streptococcus pneumonlae*

Pneumonia, Lobar
Source: ICD91/IT/481 ; MSH94/PM/D011018; MSH94/MH/D011018; SNM2/RT/M-40000; ICD91/PT/481 ; SNM2/PT/D-0164; DXP92/PT/U000473; MSH94/EP/D011018; INS94/MH/D011018;1NS94/SY/D011018
Synonym: Pneumonia, diplococcal
Parent: Bacterial Infections; Influenza with Pneumonia
Broader: Bacterial Pneumonia; Inflammation
Other: *Streptococcus pneumoniae*
Semantic: inverse-is-a: *Pneumonia*
has-result: *Pneumococcal Infections*

Pneumonia, Staphylococcal
Source: ICD91/PT/482.4; ICD91/IT/482.4; MSH94/MH/D011023; MSH94/PMID011023; MSH94/EP/D011023; SNM2/PT/D-017X; INS94/MH/D011023; INS94/SY /D011023
Parent: Bacterial Infections; Influenza with Pneumonia
Broader: Bacterial Pneumonia
Semantic inverse-is-a: Pneumonia; Staphylococcal Infections

Pneumonia, Streptococcal
Source: ICD91/IT/482.3
Other: Streptococcus pneumoniae

Pneumonia due to Streptococcus
Source: ICD91/PT/482.3
ATX: Pneumonia AND Streptococcal Infections AND NOT Pneumonia, Lobar
Parent: Inftuenza with Pneumonia

Pneumonia in Anthrax
Source: ICD91/PT/484.5; ICD91/IT/022.1 ; ICD91/IT/484.5
Parent: Influenza with Pneumonia
Broader: Pneumonia in other infectious diseases classified elsewhere
Other: Pneumonia, Anthrax

Pneumonia, Anthrax
Source: ICD91/IT/022.1; ICD91/IT/484.5
Other: Pneumonia in Anthrax

Fig. 7.9 Some of the information available in the Unified Medical Language System about selected pneumonia concepts. Concept's preferred names are shown in italics. Sources are identifiers for the concept in other terminologies. Synonyms are names other than the preferred name. ATX is an associated Medical Subject Heading expression that can be used for Medline searches. The remaining fields (Parent, Child, Broader, Narrower, Other, and Semantic) show relationships among concepts in the Metathesaurus. Note that concepts may or may not have hierarchical relations to each other through Parent–Child, Broader–Narrower, and Semantic (is-a and inverse is-a) relations. Note also that Pneumonia, Streptococcal and Pneumonia due to Streptococcus are treated as separate concepts, as are Pneumonia in Anthrax and Pneumonia, Anthrax

protocols and the physical connections made to the system. Obviously, some understanding at these lower levels is necessary before a linkage between two systems can be successful. Increasingly, standards groups are defining scenarios and rules for using various protocols at these levels, such as TCP/IP. Much of the labor in making existing standards work lies in these lower levels.

Typically, a transaction set or message is defined for a particular event, called a **trigger event**. This trigger event, such as a hospital admission, then initiates an exchange of messages. The message is composed of several data segments; each data segment consists of one or more data fields. Data fields, in turn, consist of data elements that may be one of several data types. The message must identify the sender and the receiver, the message number for subsequent referral, the type of message, special rules or flags, and any security requirements. If a patient is involved, a data segment must identify the patient, the circumstances of the encounter, and additional information as required. A reply from the receiving system to the sending system is mandatory in most circumstances and completes the communications set.

It is important to understand that the sole purpose of the data-interchange standard is to allow data to be sent from the sending system to the receiving system; the standard is not intended to constrain the application system that uses those data. Application independence permits the data-interchange standard to be used for a wide variety of applications. However, the standard must ensure that it accommodates all data elements required by the complete application set.

7.5.2 Specific Data Interchange Standards

As health care increasingly depends on the connectivity within an institution, an enterprise, an integrated delivery system, a geographic system, or even a national integrated system, the ability to interchange data in a seamless manner becomes critically important. The economic benefits of data-interchange standards are immediate and obvious. Consequently, it is in this area of healthcare standards that most effort has been expended. All of the SDOs in health care have some development activity in data-interchange standards.

The following sections summarize many of the current standards for data-interchange. Examples are provided to give you a sense of the technical issues that arise in defining a data-exchange standard, but details are beyond the scope of this effort. In fact, the pace of change is so great that many of the referenced standards will have been improved at the time of publication. Rather than providing an exhaustive list of standards, links to the standards and standards platforms will provide access to the most recent technical information and its implementation.

7.5.2.1 HL7 Standards

HL7 has provided standards that have been adopted world-wide. In the United States and in many other countries, these standards are codified in legislation and in regulation. The changes in these standards most often reflect new governmental policies, new science (both clinical and pre-clinical), new care paradigms, and new models for payment models.

Compendium of HL7 Standards

- **Introduction to HL7 Standards**

HL7 provides a framework, as well as related standards, for the exchange, integration, sharing, and retrieval of electronic health information. These standards define how information is packaged and communicated from one party to another, setting the language, structure and data types required for seamless integration between systems. HL7 standards support clinical practice and the management, delivery, and evaluation of health services, and are recognized as the most commonly used in the world.

► http://www.hl7.org/implement/standards/index.cfm?ref=nav

- **HL7 Primary Standards**

Primary standards are the most widely implemented standards and are fundamental for system integrations, inter-operability and compliance. The most frequently used standards are defined in this category

► http://www.hl7.org/implement/standards/product_section.cfm?section=1&ref=nav

- **HL7 Foundational Standards**

Foundational standards define the fundamental tools and building blocks used to create the standards, and the technology infrastructure that implementers of HL7 standards must manage.

► http://www.hl7.org/implement/standards/product_section.cfm?section=2&ref=nav

- **HL7 Clinical & Administrative Domains**

Messaging and document standards for clinical specialties and groups are found in this section. These standards are usually implemented once primary standards for the organization are operational.

► http://www.hl7.org/implement/standards/product_section.cfm?section=3

- **HL7 EHR Profiles**

These standards provide *functional models* and profiles that enable the constructs for management of electronic health records. EHR System Functional Model (EHR-S FM) outlines important features and functions that should be contained in an EHR system.

► http://www.hl7.org/implement/standards/product_section.cfm?section=4

- **HL7 Implementation Guides**

Implementation guides and their supporting documents are intended to be used in conjunction with an existing standard. The supporting documents serve as supplemental material for a parent standard. Implementation guides provide the road map for transforming the technical standard into an effective working solution.

► http://www.hl7.org/implement/standards/product_section.cfm?section=5

- **HL7 Standards Rules & References**

These references provide the technical specifications, programming structures and guidelines for software and standards development. They are not stand alone solutions, but rather provide support for a standard or for a family of standards.

► http://www.hl7.org/implement/standards/product_section.cfm?section=6

- **HL7 Current Projects & Education**

This is a resource for Standards for Trial Use (STUs) and for ongoing projects and standards. The link also provides helpful resources and tools to further supplement understanding and adoption of HL7 standards.

► http://www.hl7.org/implement/standards/product_section.cfm?section=7

- **HL7 Standards Master Grid**

This is a convenient navigation tool for all HL7 standards. Because HL7 encompasses the complete life cycle of a standards specification, including the development, adoption, market recognition, utilization, and adherence. There is also an explanation of the IP Policy that provides more information about how members and non-members can use the standard

► http://www.hl7.org/implement/standards/product_matrix.cfm?ref=nav

Clinical Document Architecture

Since its initial development in 2001, the Clinical Document Architecture (CDA) standard has become globally adopted for a broad range of use (Ferranti et al. 2006). Now an ISO standard and advanced to Release 2, CDA is a document markup standard for the structure and semantics of an exchanged "clinical document." CDA is built upon the RIM and relies upon reusable templates for its ease of implementation. A CDA document is a defined and complete information object that can exist outside of a message and can include text, images, sounds, and other multimedia content. CDA supports the following features: persistence, stewardship, potential for authentication, context, wholeness, and human-readability. In the US, CDA is one of the core components of data exchange for Meaningful Use. The competing implementation processes for CCD profile development were successfully harmonized into a broadly adopted *Consolidated* Continuity of Care Document (*C*CCD).

In order to ease the path to implementation of CDA, HL7 has developed a more narrowly defined specification called greenCDA, which limits the requirements of the RIM, provides greater ease of template composition, and consumes much less bandwidth for transmission. An additional effort to promote CDA adoption was achieved with the release of the CDA Trifolia repository, which, in addition to offering a library of templates, includes tooling for template modification as well as a template-authoring language. This has enabled the adoption of native CDA for exchange of laboratory data, clinical summaries, and electronic prescriptions and well as for clinical decision support.

HL7 Fast Healthcare Interoperability Resources

FHIR (see ▶ Sect. 7.2.3) is a new highly innovative approach to standards development, first introduced by HL7 in 2011. FHIR was created in order to overcome the complexity of development based upon the HL7 Reference Information Model (RIM), without losing the successful interoperability that model-driven data interchange demands. At the same time, FHIR delivers greater ease of implementation than other high-level development processes. It is designed to be compatible with legacy systems that conform to V2 and/or V3 messaging, and it supports system-development utilizing broadly deployed Clinical Document Architecture (CDA) platforms and ubiquitous templated CDA implementations, such as *Consolidated CDA*.

Although FHIR is built upon more than a decade of the development and refinement of the RIM, FHIR utilizes unique methodologies, artifacts, tooling, and publishing approach. While FHIR is based upon the RIM, it does not require implementers to know the RIM or know the modeling language upon which it was built. FHIR defines a limited set of data models (or *resources*) as XML or JSON objects, but provides extension mechanisms for creating any elements which are incomplete or missing. The resulting structures are native XML/JSON objects which do not require knowledge of the RIM abstraction in order to be implemented. Fundamentally, each clinical concept is created as a single resource, which need not change over time. The resources remain as the smallest unit of abstraction, and the creation of each resource is based upon ***RESTful* design** principles. Base FHIR resources can be further refined by creating *profiles* which constrain existing data elements or add elements as extensions.

Inherently, development can precede around a *services* (SOA) model, which will support cloud-based applications. While a RESTful framework is enabled, it is not required. In addition, a well-defined ontology persists in the background, but knowledge of the terminology is not necessary for implementation. Fundamental to FHIR, all resources, as well as all resource attributes have a free-text expression, an encoded expression or both. Thus, FHIR supports a human-readable format, which is so valuable to the implementations supported by CDA.

Finally, FHIR is built with new *data types*, conformant with the familiar ISO 21090 format. As such, these data types are far simpler to use, with much of the complexity captured in the *extensions*. This allows mapping to other models, including those developed using *archetypes*, upon which the CEN format for electronic medical records is predicated. This allows an inherently much smaller library of resources, all mapped to the HL7 RIM, and which can be maintained in perpetuity. FHIR developers have estimated that fewer than 150 such resources will define all of health care. Other concepts can be described as extensions.

This provides a unique opportunity for creation of both new applications in mature computing environments and for low and medium resource countries without legacy implementations. Nonetheless, migrations from V2 or V3 environments to FHIR implementations are achievable through native tooling.

FHIR APIs and resources can be implemented in SMART applications and thereby extending the utility of EHR data to support externalized clinical decision support, data visualization and combining EHR data with remote monitoring devices. SMART Health IT is an open, standards based technology platform that enables innovators to create apps that seamlessly and securely run across the healthcare system (▶ https://smarthealthit.org/) SMART was created at the Boston Children's Hospital through a grant from the Office of the National Coordinator for Healthcare IT. SMART on FHIR enables cross-platform and intersystem exchange of data by enabling ISO standards-based solutions for security and authentication.

With Clinical Decision Support Hooks (CDS-Hooks; Spineth et al. 2018), triggers can be built into the EHR workflow and trigger external CDS services. One example application developed with support from the Centers

for Disease Control is an opioid medication management tool which can be automatically launched when a physician orders an opioid medication. This tool provides guidance to the provider based on the patient's history of opioid prescriptions as well as the current prescriber's intended order.

Link to CDSHooks: ► https://cds-hooks.hl7.org/

Most often, HL7 is recognized for its messaging standards, but there is a large contribution to technical specifications that support the development and implementation of these messaging standards.

7.5.2.2 American Dental Association Standards

In 1983, the American Dental Association (ADA) committee MD 156 became an ANSI-accredited committee responsible for all specifications for dental materials, instruments, and equipment. In 1992, a Task Group of the ASC MD 156 was established to initiate the development of technical reports, guidelines, and standards on electronic technologies used in dental practice. These include digital radiography, digital intraoral video cameras, digital voice-text-image transfer, periodontal probing devices, and CAD/CAM. Proposed standards include Digital Image Capture in Dentistry, Infection Control in Dental Informatics, Digital Data Formats for Dentistry, Construction and Safety for Dental Informatics, Periodontal Probe Standard Interface, Computer Oral Health Record, and Specification for the Structure and Content of electronic medical record integration.

7.5.2.3 Health Industry Business Communications Council Standards

The Health Industry Business Communications Council (HIBCC) has developed the Health Industry Bar Code (HIBC) Standard, composed of two parts. The HIBC Supplier Labeling Standard describes the data structures and bar code symbols for bar coding of health care products. The HIBCC Provider Applications Standard describes data structures and bar code symbols for bar coding of identification data in a health care provider setting. HIBCC also issues and maintains Labeler Identification Codes that identify individual manufacturers. The HIBCC administers the Health Industry Number System, which provides a unique identifier number and location information for every health care facility and provider in the United States. The HIBCC also administers the Universal Product Number Repository, which identifies specific products and is recognized internationally.

Link: ► https://www.hibcc.org/

7.5.2.4 The Electronic Data Interchange for Administration, Commerce, and Transport Standard

The EDI for Administration, Commerce, and Transport (EDIFACT) is a set of international standards, projects, and guidelines for the electronic interchange of structured data related to trade in goods and services between independent computer-based information systems (National Council for Prescription Drug Programs Data Dictionary 1994). The standard includes application-level syntax rules, message design guidelines, syntax implementation guidelines, data element dictionary, code list, composite data-elements dictionary, standard message dictionary, uniform rules of conduct for the interchange of trade data by transmission, and explanatory material.

The basic EDIFACT (ISO 9735) syntax standard was formally adopted in September 1987 and has undergone several updates. In addition to the common syntax, EDIFACT specifies standard messages (identified and structured sets of statements covering the requirements of specific transactions), segments (the groupings of functionally related data elements), data elements (the smallest items in a message that can convey data), and code sets (lists of codes for data elements). The ANSI ASC X12 standard is similar in purpose to EDIFACT, and work is underway to coordinate and merge the two standards.

EDIFACT is concerned not with the actual communications protocol but rather with the structuring of the data that are sent.

EDIFACT is independent of the machine, media, system, and application and can be used with any communications protocol or with physical magnetic tape.

Link: ► https://www.edistaffing.com/resources/unedifact-standards/

7.6 Today's Reality and Tomorrow's Directions

In the current environment, the seamless exchange of data that can be used for any purpose remains a challenge. As we move closer to sematic interoperability (► https://en.wikipedia.org/wiki/Semantic_interoperability) for healthcare and biomedical data, the challenge of true plug-and-play interoperability is still elusive. The meaning of many concepts and terms remains ambiguous, controversial, disputed, or poorly understood. Instead, the standards community has built a system of interchange that often requires mapping between terminologies and standards to overcome these issues.

7.6.1 The Interface: Standards and Systems

Historically, interchange standards evolved to support sharing of information over complex networks of distributed systems. This served a simple business model in which data was pushed from disparate repositories with inconsistent architectures and data structures. This permitted the exchange of data for both business needs and patient care.

In today's medical environment, there are several competing forces that place a burden on standards requirements. The traditional scope of data sources included business level information, principally for payment needs. These were developed utilizing coding methodologies and business architecture that did not rely upon inclusion of primary clinical data into the reimbursement decision. With the advent of statutory requirements that demand justification of insurance claims and reimbursement, additional data forms and formats became essential. This led to the development of claims attachment standards (see X12, above) that enabled more complex adjudication, comparative effectiveness, and accountable care. These standards will most certainly require structured, coded data rather than free-text and unstructured narrative.

Complexity of data requirements is constantly growing to better support evidence-based medicine, clinical decision support, personalized medicine, and accountable care. Each of these has overlapping, but fundamentally unique data streams. Moreover, the data provided at the point of care, if unfiltered, is likely to overwhelm the clinical decision making process. Elements of clinical data, such as events in pediatric years, must not compete for the attention of the caregiver. To an extent, this was solved with specifications, such as FHIRcast, which were developed to provide context aware data to that process. There are growing demands for increasing the depth and breadth of data delivered to that clinical environment. In addition, these standards must support the implicit policy decisions about the nature of this data.

To date, clinical and preclinical information populates many of the alerts that clinicians receive at the point of care. Typically, these range from information supporting complex decision trees to the selection of testing and interventions. This has been abetted by increasing knowledge of genomic data and implication for therapeutic decisions. Although this has had its greatest impact on the chemotherapy of cancer, the importance in many other clinical domains, for more common conditions (including the treatment of diabetes, hypertension, and arthritis) is now recognized. Current architectural systems are ill-prepared to manage this process. Moreover, data formats for genomic and genetic information are disparate and often incompatible.

Data privacy requirements, and the variability of these requirements among legal entities, currently pose a different set of demands for information access technologies. For example, some states permit line-item exclusion of clinical data that is transferred between providers, based on the primacy of

the information and the role of the caregiver. Other jurisdictions allow participation of health information exchanges to those individuals who agree only to dissemination of data from complete sources.

Existing data architectures enable a constant stream of data to be passed in an untended and unmonitored fashion. In evolving models, data request and acknowledgement require a more complex query and response logic. In fact, most inquiries demand the validation of the provider system and privileges that are afforded to both the caregiver and the primary data repository. This places another component of interface design between the respective systems and necessitated the development of analogous provider indices and provider repositories. Concerns of both privacy and security must be met by these specifications. The system effectively asks not only who you are but why you want the information.

Much of this process overhead has been addressed by the design and architecture of health information exchanges. Often the business case supersedes the demand for clinical knowledge. At the same time, these exchanges are designed to behave in an entirely agnostic fashion, placing no demand on either the sender or recipient for data quality, other than source identification. In fact, the metadata, so responsible for the value of the information, is often capable of specifying only its origin and value sets.

In today's clinical environment, there has been very little attention paid to the capture and validation of patient-initiated data. While so very critical to diagnosis and ongoing management, only scant standards exist for embedding patient derived information into the clinical record without intermediate human interaction and adjudication. When allowed by current systems, data provided by patients often lies within the audit trail, as a comment, rather than in the record as source data. Steps are sorely needed to define and attribute such data since it is so critical to many aspects of accountable care. Data obtained directly from patient sources is often attributed to "subjective" status, but it is no less objective that many clinician observations. Perhaps justification for that lies in the fact that this patient derived data is neither quantifiable nor codeable. This is supported by valid concerns about the patient's health care literacy, or lack thereof, but is no less required than validated decision support for caregivers.

Data obtained from clinical research and clinical data provided to inform clinical studies suffer from other concerns of failed interoperability. This is attributed, and rightfully so, to disparities of terminologies utilized for patient care and those used in clinical research. This is most dramatically highlighted in the terminology deployed by regulation for adverse event reporting (MedDRA; Medical Dictionary of Regulatory Affairs). Mapping between the MedDRA dictionary and other clinical terminologies (SNOMED-CT, LOINC, ICD, and CPT) has not proven successful. Moreover, many aspects pertaining to study subject inquiries in clinical research are often designed to elicit yes-no responses (Have you smoked in the last 5 years), rather than data that many caregivers deem relevant. Yet, today, it is more critical than ever to enable clinical research to inform patient care and care derived data to enable clinical trials. The business model of developing drugs for billon dollar markets ("blockbuster drugs") has proven itself to be unsustainable, as the cost of developing a new drug entity has now exceeded a billion dollars. From the clinical perspective, current estimates suggest that information from basic science research experiences delays of nearly 17 years before that knowledge can be incorporated into clinical care (Balas and Boren 2000).

Semantic interoperability of clinical data inherently requires *data reuse*. It is not sufficient for systems to unambiguously exchange machine readable data. Data, once required only for third party payment, must be shared by other partners in the wellness and health care delivery ecosystems. Certainly, these data must be presented to research systems, as noted above. The data must also be available for public health reporting and analysis, for comparative effectiveness research, for accountable care measurement and for enhancement of decision support systems

(including those for patients and their families). The immediate beneficiaries have been systems developed to support biosurveillance and pharmacovigilance. In practical terms, the business practices that govern our delivery systems (and the government policies and regulations that enable them) must enable these data streams to both enhance care and control costs.

7.6.2 Future Directions

The new models for health care require a very different approach. The concept of a patient-centric EHR (► Chap. 19) requires the aggregation of any and all data created from, for, and about a patient into a single real or virtual record that provides access to the required data for effective care at the place and time of care. Health information exchange (HIE; see ► Chap. 20) at regional, state, national, and potentially at a global level is now the goal. This goal can only be reached through the effective use of information technology, and that use can only be accomplished through the use of common global standards that are ubiquitously implemented across all sites of care.

Three other future trends influence the need for new and different standards. The first is secondary use of data by multiple stakeholders. This requirement can only be met through semantic interoperability – a universal ontology that covers all aspects of health, health care, clinical research, management, and evaluation. Standards for expressing what is to be exchanged and under what circumstances are important as well as standards for the exchange of data. Included in multiple uses of data is reporting to other organizations such as immunization and infectious disease reports to the **Centers for Disease Control and Prevention** (CDC), performance reports to the **National Quality Forum** (NQF) and audit reports to **The Joint Commission** (TJC). Such systems as described also enable population health studies and health surveillance for natural and bioterrorism outbreaks.

The second trend area is the expansion of the types of data that are to be included in the EHR. The new emphasis on translational informatics will require new standards for the transport, inclusion into the EHR, and use of genetic information including genes, biomarkers, and phenotypic data. Imaging, videos, waveforms, audio, and consumer-generated data will require new types of standards. Effective use of these new types of data as well as exponential increases in the volume of data will require standards for decision support, standards for creating effective filters for presentation and data exchange, and new forms of presentation including visualization. New sources of data will include geospatial coding, health environmental data, social and community data, financial data, and cultural data. Queries and navigation of very large databases will require new standards. Establishing quality measures and trust will require new standards. Ensuring integrity and trust as data is shared and used by other than the source of data will require new standards addressing provenance and responsibility.

The third trend area is the use of mobile devices, smart devices, and personal health devices. How, when and where such devices should be used is still being explored. Standards will be required for safe design, presentation, interface, integrity, and protection from interference.

True global interoperability will require a suite of standards starting with the planning of systems, the definition and packaging of the data, collection of the data including usability standards, the exchange of data, the storage and use of data, and a wealth of applications that enable the EHR for better care. IT systems must turn data collected into information for use – a process that will require the use of knowledge in real time with data to produce information for patient care. Selecting the correct knowledge from literature, clinical trials, and other forms of documentation will require standards. Knowledge representation, indexing, and linkages will require standards.

A major, and challenging, requirement to address these new types and use of data will be effective standards for privacy and security. These standards must protect, but not restrict the use of data and access to that data for

determining and giving the best care possible. The aggregation of data requires an error-free way for patient identification that will permit merger of data across disparate sources. Sharing data also requires standards for the de-identification of patient data.

The effective management of all of these resources will require additional output from the standards communities. Standards for defining the required functionality of systems and ways for certifying adherence to required functionality is essential for connecting a seamless network of heterogeneous EHRs from multiple vendors. Testing of standards, including IHE Connectathons before widespread dissemination and perhaps mandated use of standards is critical to use and acceptance. Standards for registries, standards for the rules that govern the sharing of research data, standards for patient consent, and standards for identification of people, clinical trials, collaboration, and other similar areas are necessary. Profiles for use and application from the suite of standards are a necessity. Detailed implementation guides are key to use and implementation of standards. Tools that enable content population and use of standards are mandatory for easy use of standards.

Standards for these new and evolving business and social needs must be supported by changes in standards development methodologies and harmonization. Legacy systems are not easily discarded. Recommendations for complete replacement of existing standards are neither politically expedient nor fiscally supportable. Currently, there is increasing attention to new approaches to standards development that speeds the creation process and improves the quality of standards that are developed. These evolving development platforms pay appropriate homage to existing standards and leverage previously developed models of development and analysis.

The use of the FHIR may provide a much needed solution while relying upon historically developed and refined interoperability specifications, it hides the complexity of authoring messages within the FHIR development process. This leads to more usable specifications, created in a dramatically abbreviated time frame. Other approaches to standards development, such as those focused upon services, are rapidly evolving. These services-aware architectures are governed by strict development principles that help ensure both interoperability and the ability of components to be reused.

Increasingly large data stores ("big data") have demanded some of these changes. These data have emanated from a highly diverse universe of scientific development. In fact, some of the new bio-analytic platforms for *in vitro* cellular research are generating data at a rate, which by some estimates, is faster than the data can be analyzed. Medical images, for which storage requirements are growing, must now be principally evaluated by human inspection. Newly evolving algorithms and the technologies to support them, initially developed for "star wars" type image analysis, are replacing radiologist and pathologists for the establishment of diagnoses. These machines have proven to be faster and more accurate than their human counterparts. In the very near future, such instrumentation will supplant medical scientists the same way that comparable technologies replaced human inspection in the estimation of cell differentials for blood counts. These new technologies are demanding the development of specifications and the terminologies to support them.

Tomorrow's technologies will transition from early vision through prototyping to commercial products in a more compressed life cycle. A model for this process in biomedical science was established with the emergence of the Human Genome Project (see ▶ Chap. 11). Within the next decade, routine genome determination and archiving, as well as their application to disease management, will require greatly enhanced solutions for data management and analysis. Innovative strategies for recognizing and validating biomarkers will grow exponentially from the current stable of imaging and cell surface determinants. These data streams will require adaption of existing decision support systems and comparative effectiveness paradigms. Lastly, scientific evidence supporting the diagnosis and management with the field of behavioral medicine will change the entire clinical spectrum

and approach to evaluation and care. As we emerge from the dark ages of behavioral medicine, we will certainly require new systems for recognizing, diagnosing, naming and intervening on behalf of our patients.

In some sense, the development of standards is just beginning. The immediate future years will be important to create effective organizations that include the right experts in the right setting to produce standards that are in themselves interoperable. That goal still remains in the future.

Suggested Reading

Abbey, L. M., & Zimmerman, J. (Eds.). (1991). *Dental informatics: Integrating technology into the dental environment*. New York: Springer. This text demonstrates that the issues of standards extend throughout the areas of application of biomedical informatics. The standards issues discussed in this chapter for clinical medicine are shown to be equally pertinent for dentistry.

American Psychiatric Association Committee on Nomenclature and Statistics. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: The American Psychiatric Association. Argonaut Project – HL7 FHIR, https://argonautwiki.hl7.org/Main_Page.

Benson, T. (2012). *Principles of health interoperability HL7 and SNOMED* (2nd ed.). London: Springer. This book presents a detail discussion of the HL7 version 2 messaging standard and a detail presentation of SNOMED-CT.

Braunstein, M. (2018). *Health informatics on FHIR: How HL7's new API is transforming healthcare*. Springer.

Boone, K. W. (2012). *The CDA™book*. London: Springer. This book provides an excellent presentation of the HL7 Clinical Document Architecture and related topics.

Chute, C. G. (2000). Clinical classification and terminology: some history and current observations. *Journal of the American Medical Informatics Association, 7*(3), 298–303. This article reviews the history and current status of controlled terminologies in health care.

Cimino, J. J. (1998). Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of Information in Medicine, 37*(4–5), 394–403. This article enumerates a set of desirable characteristics for controlled terminologies in health care.

Executive Office of the President; President's Council of Advisors on Science and Technology (2010). *Report to the President realizing the full potential of health information technology to Improve healthcare for Americans: the path forward*.

Henderson, M. (2003). *HL7 messaging*. Silver Spring: OTech Inc.. Description of HL7 V2 with examples. Available from HL7.

Holland, C., & Shostak, J. (2016). *Implementing CDISC using SAS* (2nd ed.). SAS Institute.

ICD-10-CM 2020: The Complete Official Codebook, AMA, 2019

Institute of Medicine. (2003). *Patient safety: Achieving a new standard for care*. Washington, DC: National Academy Press. Discusses approaches to the standardization of collection and reporting of patient data.

Jamie, G. (2019). *Starting SNOMED: A beginner's guide to the Snomed CT healthcare terminology*. SNOMED.

Kahn, A. N., et al. (2006). Standardizing laboratory data by mapping to LOINC. *JAMIA, 13*(3), 353–355.

NCPDP Standards-based Facilitator Model for PDMP White Paper, NCPDP, 2019 ISO Technical Committee 215 Healthcare Informatics, https://en.wikipedia.org/wiki/ISO/TC_215

New York Academy of Medicine. (1961). *Standard nomenclature of diseases and operations* (5th ed.). New York: McGraw-Hill.

Pianykh, O. (2011). *Digital imaging and communications in medicine (DICOM): A practical introduction and survival guide*. Springer.

Richesson, R. L., & Andrews, J. E. (2012). *Clinical research informatics*. London: Springer. This book includes a discussion of standards and applications from the clinical research perspectives.

Stallings, W. (1987). *Handbook of computer-communications standards*. New York: Macmillan. This text provides excellent details on the Open Systems Interconnection model of the International Standards Organization.

Stallings, W. (1997). *Data and computer communications*. Englewood Cliffs: Prentice-Hall. This text provides details on communications

architecture and protocols and on local and wide area networks.

Questions for Discussion

1. Who should be interested in interoperability and health data standards?
2. What are the five possible approaches to accelerating the creation of standards?
3. Define five health care standards, not mentioned in the chapter, which might also be needed?
4. What role should the government play in the creation of standards?
5. At what level might a standard interfere with a vendor's ability to produce a unique product?
6. Define a hypothetical standard for one of the areas mentioned in the text for which no current standard exists. Include the conceptualization and discussion points. Specifically state the scope of the standard.

References

American Psychiatric Association Committee on Nomenclature and Statistics. (2013). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: The American Psychiatric Association.

American Society for Testing and Materials. (1999). *Standard guide for properties of a universal healthcare identifier (UHID) (E1714-95.)*. West Conshohocken: American Society for Testing and Materials.

Balas, E., & Boren, S. (2000). Managing clinical knowledge for health care improvement. In J. H. van Bemmel & M. C. AT (Eds.), *Yearbook of medical informatics* (pp. 65–70). Stuttgart: Schattauer Verlagsgesellschaft mbH.

Becnel, L. B., Hastak, S., Ver Hoef, W., Milius, R. P., Slack, M. A., Wold, D., Glickman, M. L., Brodsky, B., Jaffe, C., Kush, R., & Helton, E. (2017). BRIDG: A domain information model for translational and clinical protocol-driven research. *Journal of the American Medical Informatics Association, 24*(5), 882–890.

Braunstein, M. (2018). *Health informatics on FHIR: How HL7's new API is transforming healthcare*. Springer.

Campbell, K. E., Tuttle, M. S., & Spackman, K. A. (1998). A "lexically-suggested logical closure" metric for medical terminology maturity. In *Proceedings of the 1998 AMIA Annual Fall Symposium* (pp. 785–789). Orlando: American Medical Informatics Association.

Chute, C. G. (2018). The rendering of human phenotype and rare diseases in ICD-11. *Journal of Inherited Metabolic Disease, 41*(3), 563–569.

Coenen, A., McNeil, B., Bakken, S., Bickford, C., Warren, J. J., & American Nurses Association Committee on Nursing Practice Information Infrastructure. (2001). Toward comparable nursing data: American Nurses Association criteria for data sets, classification systems, and nomenclatures. *Computers in Nursing, 19*(6), 240–246.

College of American Pathologists. (1971). *Systematized nomenclature of pathology*. Chicago: The College of American Pathologists.

Côté, R. A., & Rothwell, D. J. (1993). *The systematised nomenclature of human and veterinary medicine, 1993*. Northfield: College of American Pathologits.

Ferranti, J. M., Musser, R. C., Kawamoto, K., & Hammond, W. E. (2006). The clinical document architecture and the continuity of care record: A critical analysis. *Journal of the American Medical Informatics Association, 13*(3), 245–252.

Gottschalk, H. W. (1991). The medical information bus: Overview of the medical device data language. *Journal of Clinical Engineering, 16*(4), 323–329.

Harris, M. A., for the Gene Ontology Consortium, et al. (2004). The gene ontology (GO) database and informatics resource. *Nucleic Acids Research, 32*(Database issue), D258–D256.

Hudson, K. L., & Collins, F. S. (2017). The 21st century cures act — A view from the NIH. *The New England Journal of Medicine, 376*(2), 111–113.

Humphreys, B. L. (Ed.). (1990). *UMLS knowledge sources – First experimental edition documentation*. Bethesda: National Library of Medicine.

International Standards Organization. (1987). *Information processing systems-concepts and terminology for the conceptual schema and the information base.* (ISO TR 9007:1987). International Standards Organization.

Jollis, J. G., Ancukiewicz, M., DeLong, E. R., Pryor, D. B., Muhlbaier, L. H., & Mark, D. B. (1993). Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. *Annals of Internal Medicine, 119*(8), 844–850.

Libicki, M. C. (1995). *Information technology standards: Quest for the common byte*. Boston: Digital Press.

National Council for Prescription Drug Programs. (1994). *Data dictionary*.

Nelson, S. J., Brown, S. H., Erlbaum, M. S., Olson, N., Powell, T., Carlsen, B., Carter, J., Tuttle, M. S., & Hole, W. T. (2002). A semantic normal form for clinical drugs in the UMLS: Early experience with the VANDF. In *Proceedings of the AMIA Fall Symposium* (pp. 557–561). American Medical Informatics Association.

New York Academy of Medicine. (1961). *Standard nomenclature of diseases and operations* (5th ed.). New York: McGraw-Hill.

NHS Centre for Coding and Classification. (1994). *Read codes (Version 3)* (April ed.). London: NHS Management Executive, Department of Health.

Odigie, E., Lacson, R., Raja, A., Osterbur, D., Ip, I., Schneider, L., & Khorasani, R. (2019). Fast healthcare interoperability resources, clinical quality language, and systematized nomenclature of medicine—Clinical terms in representing clinical evidence logic statements for the use of imaging procedures: Descriptive study. *JMIR Medical Informatics, 7*(2), e13590.

Park, H. A., & Cho, I. (2009). Education, practice, and research in nursing terminology: Gaps, challenges, and opportunities. *Yearbook of Medical Informatics*, 103–108.

Payne, T. H., Corley, S., Cullen, T. A., Gandhi, T. K., Harrington, L., Kuperman, G. J., Mattison, J. E., McCallie, D. P., McDonald, C. J., Tang, P. C., Tierney, W. M., Weaver, C., Weir, C. R., & Zaroukian, M. H. (2015). Report of the AMIA EHR-2020 Task Force on the status and future direction of EHRs. *Journal of the American Medical Informatics Association, 22*(5), 1102–1110.

Plunkett, R. J. (1952). *Standard nomenclature of diseases and operations* (4th ed.). New York: McGraw-Hill.

Rector, A. L., Glowinski, A. J., Nowlan, W. A., & Rossi-Mori, A. (1995). Medical-concept models and medical records: An approach based on GALEN and PEN & PAD. *Journal of the American Medical Informatics Association: JAMIA, 2*(1), 19–35.

Rose, M. T. (1989). *The open book: A practical perspective on OSI*. Upper Saddle River: Prentice Hall.

Simborg, D. W., Chadwick, M., Whiting-O'Keefe, Q. E., Tolchin, S. G., Kahn, S. A., & Bergan, E. S. (1983). Local area networks and the hospital. *Computers and Biomedical Research, 16*(3), 247–259.

Spackman, K. A. (2000). SNOMED RT and SNOMEDCT. Promise of an international clinical terminology. *MD Computing, 17*(6), 29.

Spackman, K. A., Campbell, K. E., & Cote, R. A. (1997a). SNOMED RT: a reference terminology for health care. In *Proceedings of the AMIA Annual Fall Symposium* (pp. 640–644). American Medical Informatics Association.

Spackman, K. A., Campbell, K. E., et al. (1997b). SNOMED-RT: A reference terminology for health care. In D. R. Masys (Ed.), *Proceedings, AMIA annual fall symposium* (pp. 640–644). Philadelphia: Hanley and Belfus.

Spineth, M., Rappelsberger, A., & Adlassnig, K. P. (2018). Implementing CDS hooks communication in an Arden-syntax-based clinical decision support platform. *Studies in Health Technology and Informatics, 255*, 165–169.

Stallings, W. (1987). *The open systems interconnection (OSI) model and OSI-related standards* (Vol. 1). New York: Macmillian.

Stead, W. W., Kelly, B. J., & Kolodner, R. M. (2005). Achievable Steps toward Building a National Health Information Infrastructure in the United States. *Journal of the American Medical Informatics Association, 12*(2), 113–120.

Tanenbaum, A. S. (1987). *Computer networks* (2nd ed.). Englewood Cliffs: Prentice Hall.

Thompson, E. T., & Hayden, A. C. (1961). *Standard nomenclature of diseases and operations* (5th ed.). New York: McGraw-Hill.

World Health Organization. (1992). *International classification of diseases index. Tenth revision* (Vol. Tabular list, Vol. 1). Geneva: The World Health Organization.

World Health Organization. (1977). *International classification of diseases index. Ninth revision*. Geneva: The World Health Organization.

Natural Language Processing for Health-Related Texts

Dina Demner-Fushman, Noémie Elhadad, and Carol Friedman

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_8

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What are the potential uses for NLP in the biomedical, clinical, and health domains?
- What are the principal computational tasks of NLP for health-related texts?
- What are the different knowledge resources and linguistic representations that can support the development of NLP techniques?
- What are the near-future directions for health-related NLP research and applications?

8.1 Motivation

Language is the primary means of human communication, and it is no surprise clinicians, biomedical researchers, patients, and health consumers alike rely on language extensively. Clinicians document the care of their patients in the electronic health record and use patient record notes to determine next steps of care. Biomedical scientists write and read articles to keep abreast of research progress from their peers. Patients rely on online platforms to learn from and exchange informational and emotional support from their peers, and health consumers view health content online as a primary source of information to manage their health and increase their health literacy. In fact, there are continuously growing, unprecedented amounts of these biomedical and health-related texts available.

The field of health and biomedical natural language processing is concerned with the theories, principles, and computational approaches to building tools that exploit these textual data and support these stakeholders – patients and health consumers, clinicians, and biomedical researchers – in their information needs.

While there is valuable information conveyed in biomedical, clinical, and health texts, it is not in a format directly amenable to further processing. These texts are difficult to process reliably because of the inherent characteristics and variability of language. While in most automated applications, structured, standardized data are readily available for processing, there is a significant amount of manual work currently devoted to mapping textual information to coded representations in biomedicine and health care: Professional coders assign billing codes corresponding to diagnoses and procedures to hospital admissions based on discharge summaries and admission information; indexers at the National Library of Medicine assign MeSH (Medical Subject Headings) terms to represent the main topics of scientific articles; and database curators extract genomic and phenotypic information on organisms from the literature. Because of the overwhelmingly large amount of textual information in health domains, manual work is costly, time-consuming, and impossible to keep up to date. One aim of Natural Language Processing (NLP) is to facilitate these tasks by enabling use of automated methods with high validity and reliability.

Another aim of NLP is to help advance many of the fundamental aims of biomedical informatics, which are the discovery and validation of scientific knowledge, improvement in the quality and cost of health care, and support to patients and health consumers. The considerable amounts of texts amassed through clinical care, published in the scientific biomedical literature, and discussed by patients and caregivers online can help acquire new knowledge and promote discovery of new phenomena. For instance, the information in patient notes, while not originally entered for discovery purposes, but rather for the care of individual patients, can be processed, aggregated and mined to discover patterns across patients. This process of leveraging observational health data has shown much success stories when applied to health data that are highly structured (OHDSIPNAS), and there is promise that incorporating learned representations of language can help as well (Ghassemi et al. 2014).

For clinicians interacting with an electronic health record and treating a particular patient, NLP can support several points in a clinician workflow: when reviewing the patient

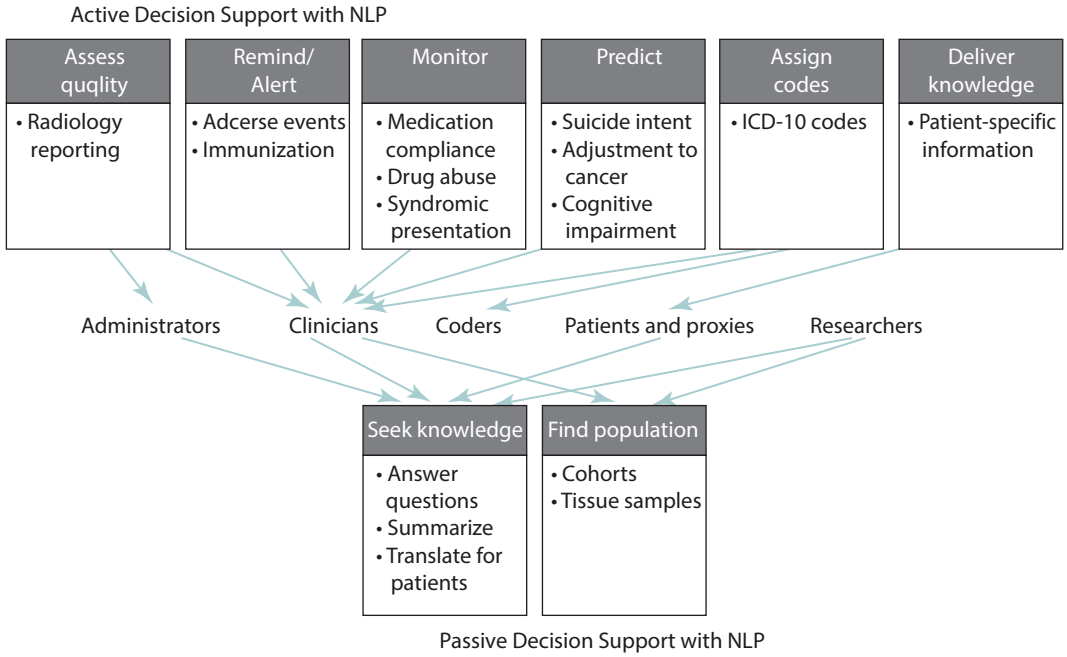


Fig. 8.1 Active (initiated by the application) and passive (initiated by the users) decision support applications to which NLP tools have contributed and have potential to contribute in the future

chart, NLP can be leveraged to aggregate and consolidate information spread across many notes and reports, and to highlight relevant facts about the patient. During the decision-making and actual care phase, information extracted through NLP from the notes can contribute to the decision support systems in the EHR as shown in Fig. 8.1 (Demner-Fushman et al. 2009). Finally, when health care professionals are documenting patient information, higher quality notes can be generated with the help of NLP-based methods.

For quality and administrative purposes, NLP can signal potential errors, conflicting information, or missing documentation in the chart. For public health administrators, EHR patient information can be monitored for syndromic surveillance through the analysis of ambulatory notes or chief complaints in the emergency room (Hripcsak et al. 2009).

The expectations and requirements for NLP support evolve and grow due to successes and demonstrated potential, such as a tool to identify tests for EGFR (epidermal growth factor receptor) mutations deployed in VA EHR clinical notes (Lynch et al. 2019) or Medical Text Indexer that supports indexers in assigning MeSH terms at the National Library of Medicine (Mork et al. 2017). New requirements arise in the domain of consumer language, due to the changes in consumers' behavior, which in turn, are changing the dynamics of healthcare interaction (▶ Chap. 11). NLP is crucial in enabling consumers to get to the right information, whether through access to clinical information or to information generated by their peers. NLP can support health consumers and patients looking for information about a particular disease or treatment, by providing better access to relevant information, targeted to their information needs, and to their health literacy levels through the analysis of the topics conveyed in a document as well as the vocabulary used in the document.

When approaching the above health-related tasks or use cases for NLP, it is beneficial to align the human tasks and end-goals with the tasks that NLP has to perform. For example, giving a clinician a full

and concise description of a patient's course relevant to the patient's current state is the task of summarization. The high-level NLP tasks or applications described later in this chapter include, but are not limited to, question answering (QA), summarization, text labeling and text generation. These tasks are bringing the field closer to the ultimate goal of text understanding, which encompasses not only direct understanding of textual information, but also the author's attitude, such as sentiment polarity and modality at the document level, and temporal reasoning at a document and document sequence levels (▶ Chap. 23).

Across all these use cases of health NLP (hNLP), the techniques of natural language processing provide a means to bridge the gap between unstructured text and data by transforming the text to data in a computable format, allowing humans to interact using familiar natural language, while enabling computer applications to process data effectively and to provide users with easy access and synthesis of the raw textual information.

This chapter is organized with two types of readers in mind: students and researchers looking for a broad introduction to health NLP prior to delving into this active field of research, and informatics practitioners looking to use hNLP technology for specific tasks or types of text. ▶ Section 8.2 presents a more in-depth description of hNLP applications and emphasizes the critical role that the context in which these applications are deployed plays when developing hNLP solutions. In ▶ Sect. 8.3 we establish the basic computational tasks involved in most hNLP applications. ▶ Section 8.4 is concerned with the different linguistic knowledge resources and types of linguistic representations that can enable and facilitate these basic NLP tasks. ▶ Section 8.5 provides further practical considerations for users of hNLP technology, while ▶ Sect. 8.6 provides further research considerations for hNLP research, including evaluation methodology. Finally, ▶ Section 8.7 briefly outlines future directions of hNLP.

8.2 NLP Applications and Their Context

In this section, we describe specific hNLP applications that have been and continue to prove being useful to biomedical, clinical, and health stakeholders in their information needs (▶ Sect. 8.2.1). We then abstract away from them and emphasize the essential role that context in which these applications exist play (▶ Sect. 8.2.2).

8.2.1 NLP Applications

Natural language processing has a wide range of potential applications in the biomedical and health domains. The following are important applications of NLP technology for biomedicine and health:

- Information extraction locates and structures specific information in text, sometimes by performing a complete linguistic analysis of the text, but more frequently, by looking for patterns in the text. This is the most common application in biomedicine. It is also one of the earliest: In the 1970s, the Linguistic String Project (LSP) under the leadership of Dr. Naomi Sager, a pioneer in NLP, developed a comprehensive computer grammar and parser of English (Grishman et al. 1973; Sager 1981), and also began work in NLP of clinical reports (Sager 1972, 1978; Sager et al. 1987) that continued into the 1990s. Other clinical and biomedical NLP systems followed, e.g., MedLee (Medical Language Extraction and Encoding System) has been used successfully primarily for clinical information extraction, but also adapted to literature processing (Friedman et al. 1994; Hripcsak et al. 1995; Friedman 2000). Other systems that have been successfully used for extraction of information from clinical notes and the literature include MetaMap (Aronson and Lang 2010), MetaMap Lite (Demner-Fushman et al. 2017) and SemRep (Kilicoglu et al. 2012). The latter three sys-

tems developed at the National Library of Medicine rely on the Unified Medical Language system (UMLS) (Lindberg et al. 1993a) described in ▶ Sect. 8.4. Other systems that rely on the UMLS include cTAKES (clinical Text Analysis and Knowledge Extraction System) (Savova et al. 2010) and CLAMP (Clinical Language Annotation, Modeling, and Processing) (Soysal et al. 2017). cTAKES combines machine learning and rule-based methods to perform clinical information extraction tasks, whereas CLAMP provides a graphic user interface to build customized NLP pipelines for clinical applications.

Once textual information is extracted and structured, it can be used for a number of different tasks. In inferring social determinants of health, for instance, one can extract social risk factors from clinical notes (Conway et al. 2019) or from social media (De Choudhury et al. 2013). The extracted data, when collected across many patients, can help understand the prevalence as well as the progression of a particular disease at the community level (Eichstaedt et al. 2015). Notably, the above system for detection of social determinants of health in clinical notes continues the lineage of clinical NLP systems at the University of Utah, which started with SPRUS (which evolved into Symtext and then MPLUS, ONYX, and TOPAZ) (Haug et al. 1990, 1994; Christensen et al. 2002; Dublin et al. 2013; Ye et al. 2014). In biology, biomolecular interactions extracted from one article or from different articles can be merged to construct biomolecular pathways. Figure 8.2 shows a pathway in the form of a graph, which was created by extracting interactions from one article published in the journal *Cell* (Maroto et al. 1997). The DARPA Big Mechanism program that aimed to assemble automatically the causal fragments found in individual scientific papers, such as in Fig. 8.2, into pathways, demonstrated the successes and the remaining challenges in the first step of the process: machine reading of the literature (Cohen 2015).

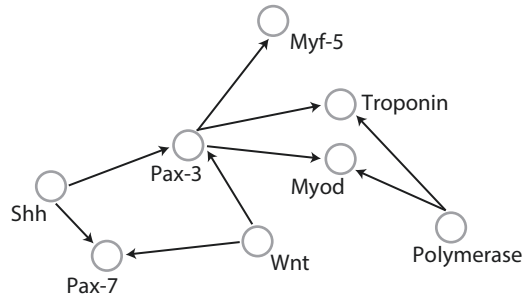


Fig. 8.2 A graph showing interactions that were extracted from an article. A vertex represents a gene or protein, and an edge represents the interaction. The arrow represents the direction of the interaction so that the agent is represented by the outgoing end of the arrow and the target by the incoming end

The techniques for information extraction may be limited to the identification of names of people or places, dates, and numerical expressions, or to certain types of terms in text (e.g. mentions of medications or proteins), which can then be mapped to canonical or standardized forms. This is referred as named-entity recognition and named-entity normalization, respectively. More sophisticated techniques identify and represent the modifiers attached to a named entity. Such advanced methods are necessary for reliable retrieval of information because the correct interpretation of a biomedical term typically depends on its relation with other terms in a given sentence. For example, the term *fever* has different interpretations in *no fever*, *high fever*, *fever lasted 2 days*, and *check for fever*. Defining the types of modifiers of interest (e.g. *no* is a negation modifier, while *lasted 2 days* is a temporal modifier), as well as techniques to recognize them in text, is an active topic of research that was in part stimulated by public release of tools and algorithms, such as NegEx (Chapman et al. 2001). Identifying relations among named entities is another important information extraction method. For example, when extracting adverse events associated with a medication, the sentences "the patient

developed a rash from amoxicillin" and "the patient came in with a rash and was given benadryl" must be distinguished. In both sentences, there is a relation between a rash and a drug, but the first sentence conveys a potential adverse drug event whereas the second sentence conveys a treatment for an adverse event. As entities are extracted within one document or across documents, one important step consists of reference resolution, that is, recognizing that two mentions in two different textual locations refer to the same entity (Kilicoglu and Demner-Fushman 2016). In some cases, resolving the references is very challenging. For instance, mentions of stroke in two different notes associated with the same patient can refer to the same stroke or two different strokes; additional contextual information and domain knowledge is often needed to resolve this problem.

- Information retrieval (IR) and NLP overlap in some of the methods that are used. IR is discussed in ▶ Chap. 23, but here we discuss the basic differences between IR and NLP. IR methods are generally geared to help users access documents in large collections, such as electronic health records, the scientific literature, or the Web. This is a crucial application in biomedicine and health, due to the explosion of information available in electronic form. The essential goal of information retrieval is to match a user's query against a document collection (usually using an index) and return a ranked list of relevant documents or the best matching snippets of text. The most basic form of indexing isolates simple words and terms, and therefore, uses minimal linguistic knowledge. More advanced approaches use NLP-based methods similar to those employed in information extraction, identifying complex named entities and determining their relationships in order to improve the accuracy of retrieval. For instance, one can search for *hypertension* and have the search operate at the concept level, returning documents that mention the phrase *high blood pressure* in addition to the ones mentioning *hypertension* only. In addition, one can search for *hypertension* in a specific context, such as in the context of *treatment* or context of *etiology*.
- Question answering (QA) involves a process whereby a user submits a natural language question, which is then automatically answered by a QA system. The availability of information in journal articles and on the Web makes this type of application increasingly important as health care consumers, health care professionals, and biomedical researchers frequently search the Web to obtain information about a disease, a medication, or a medical procedure. A QA system can be very useful for obtaining the answers to clinical questions, like "In children with an acute febrile illness, what is the efficacy of single-medication therapy with acetaminophen or ibuprofen in reducing fever?" (Demner-Fushman and Lin 2007). QA systems provide additional functionalities to an IR system. In an IR system, the user has to translate a question into a list of keywords and generate a query, but this step is carried out automatically by a QA system. Furthermore, a QA system presents the user with an actual answer (often one or several passages extracted from the source documents), rather than a list of relevant source documents. QA has focused for the most part on the literature (Demner-Fushman and Lin 2007; Cao et al. 2011), however, research on answering clinician's questions asked of EHR (Roberts and Patra 2018) and answering consumer health questions (Demner-Fushman et al. 2020; Ben Abacha et al. 2019) is burgeoning.
- Text summarization takes one or several documents as input and produces a single, coherent text, which synthesizes the main points of the input documents. Summarization helps users make sense of a large amount of data, by identifying and presenting the salient points in texts automatically. Summarization can be generic or query-focused (i.e. taking a particular information need into account when selecting important content of input documents). Query-focused summarization can

be viewed as a post-processing of IR and QA: the relevant passages corresponding to an input question are further processed into a single, coherent text. Several steps are involved in the summarization process: content selection (identifying salient pieces of information in the input document(s)), content organization (identifying redundancy and contradictions among the selected pieces of information, and ordering them so the resulting summary is coherent), and content re-generation (producing a text output from the organized pieces of information). Text summarization in the biomedical domain has focused on the literature (Elhadad et al. 2005; Zhang et al. 2011), with some forays into summarization of clinical text and Web resources (Pivovarov and Elhadad 2015; Mane et al. 2015).

- Other tasks:

  Text classification (labeling) involves categorizing text into known types. Sentences in scientific articles can be classified as key sentences indicative of the article's content (Ruch et al. 2007). Sections of a clinical note can be labeled, as, e.g., social or family history (Denny et al. 2008). At the document level, texts can be classified into different genres, etc. Related to classification is clustering, which involves grouping texts based on some intrinsic similarity without knowing a priori what these similar properties are.

  Text generation formulates natural language sentences from a given source of information, which is not directly readable by humans. Generation can be used to create a text from a structured database, such as summarizing trends and patterns in laboratory data (Hüske-Kraus 2003).

  Machine translation converts text in one language (e.g. English) into another (e.g. Spanish). These applications are important in multilingual environments in which human translation is too expensive or time consuming (Deléger et al. 2009a).

  Text readability assessment and simplification is becoming relevant to the health domain, as patients and health consumers access more and more medical information on the Web, but need support because their health literacy levels do not match the ones of the documents they read (Elhadad 2006; Keselman et al. 2007).

Finally, sentiment analysis and emotion detection belong to the general task of automated content analysis (Zunic et al. 2020).

8.2.2 Context for NLP Applications

Understanding the context and intent of the speaker on meaning (pragmatics) is crucial to performing many NLP tasks correctly. Although the context of health-related NLP applications is broad and varied, it is bounded by specific tasks and environments that can be relatively easily enumerated and therefore taken into account during processing. It is defined by those who produce the text, by the purposes for which the text is produced, and by the intended readers. For example, clinicians can write a patient status report for their colleagues, or a simplified summary for the patient. Clinical researchers can describe the arms of a clinical trial in a scientific publication or describe the inclusion criteria of the same trial in a simpler language for recruitment and patient education purposes. The style of communication is often determined by cultural conventions and ecosystems in which these texts are written and read, and inference is needed for correct interpretation and generation of language. Powerful context models are missing in the open domain (Bunt 2017) but can be approximated through the semantic lexicon and rules about the discourse of a text in the biomedical domain. Biomedical sublanguages are easier to interpret than general languages because they exhibit more restrictive semantic patterns that can be represented more easily (Harris et al. 1989; Harris 1991; Sager et al. 1987). Sublanguages tend to have a relatively small number of well-defined semantic types (e.g. medication, gene, disease, body part, or organism) and a small number of semantic patterns (e.g. medication-treats-disease, gene-interacts with-gene).

The fact that texts belong to a particular domain, be it clinical, biological or related

to health-consumers, allows us to capture domain-specific characteristics in the lexicon, the grammar, and the discourse structure. Thus, the more specific the domain of a text, the more knowledge can be encoded to help its processing, but then the NLP system would be extremely limited and specialized. For instance, in the domain of online patient discourse, patients discussing breast cancer among their peers online rely on a very different set of terms than caregivers of children on the autism spectrum. One can develop a lexicon for each subdomain, online breast cancer patients and online autism caregivers. But maintaining separate lexicons can be inefficient and error prone, since there can be a significant amount of overlap among terms across subdomains. Conversely, if a single lexicon is developed for all subdomains, ambiguity can increase as terms can have different meanings in different subdomains. For example, in the emergency medicine domain shock will more likely refer to a procedure used for resuscitating a patient, or to a critical condition brought about by a drop in blood flow, whereas in psychiatry notes it will more likely denote an emotional condition or occasionally electric shock therapy. Deciding on whether to model a domain as a whole or to focus on its subdomains independently of each other is a tradeoff. Careful determination of the use cases of a system can help determine the best choice for the system.

8.3 Basic Computational NLP Tasks

The different applications we reviewed in the previous sections have in common the need to process text, but they differ widely in the types of input and output they produce. Tasked with an application that entails text processing, one can think in these terms: is the goal to cluster, label, extract, generate or a combination of these? And what is my input: a corpus, a document, a fragment of text, or a sequence of words? The step of casting an application into one or a set of NLP tasks is important in determining the choice and design of the approaches. An evaluation of the application then helps determine which design was the more appropriate. There is not always a single solution to the application casting step. For instance, take the task of identifying within a large pool of patient records the notes with documented heart attack in the past year. We can cast this application as a note labeling task: given a note, label it according to a binary label: documentation present or absent. In this case, the application will retrieve notes of patients that are labeled as a whole as likely to contain documentation of heart attack within the past year or not. To achieve the task, we will then need to compile a training set of notes and their gold-standard labels and train a document-level classifier. An alternative approach is to cast the same application as an information extraction or template filling task. There, the template would contain for instance the location in the note that mentions the condition heart attack, along with the temporal expression documenting the time at which the heart attacks occurred in the patient. In this case, all heart attacks and temporal aspects will be extracted in our pool of patient notes, and only the ones that satisfy our temporal constraint will be kept. Both tasks will achieve the same goal (identify the patients who have a documented heart attack in the past year), but will do so in different ways and with slightly different outputs: the labeling approach will not be able to provide data provenance, but might be more feasible because it is easier to label a document as a whole than it is to annotate templates and extract information as shown in Fig. 8.3.

In this section, we review the basic tasks that are most common when processing health-related texts and in biomedical and health applications. These tasks do not constitute a pipeline, but rather a portfolio of the tasks one can rely on depending on their application and goals. They each take specific inputs (sometimes a collection of documents as in topic modeling, or a sequence of words as in sequence labeling) and produce specific outputs (groups of words in topic modeling, one label in document labeling, triples in relation extraction, or complex frames in event extraction, as shown in Fig. 8.3).

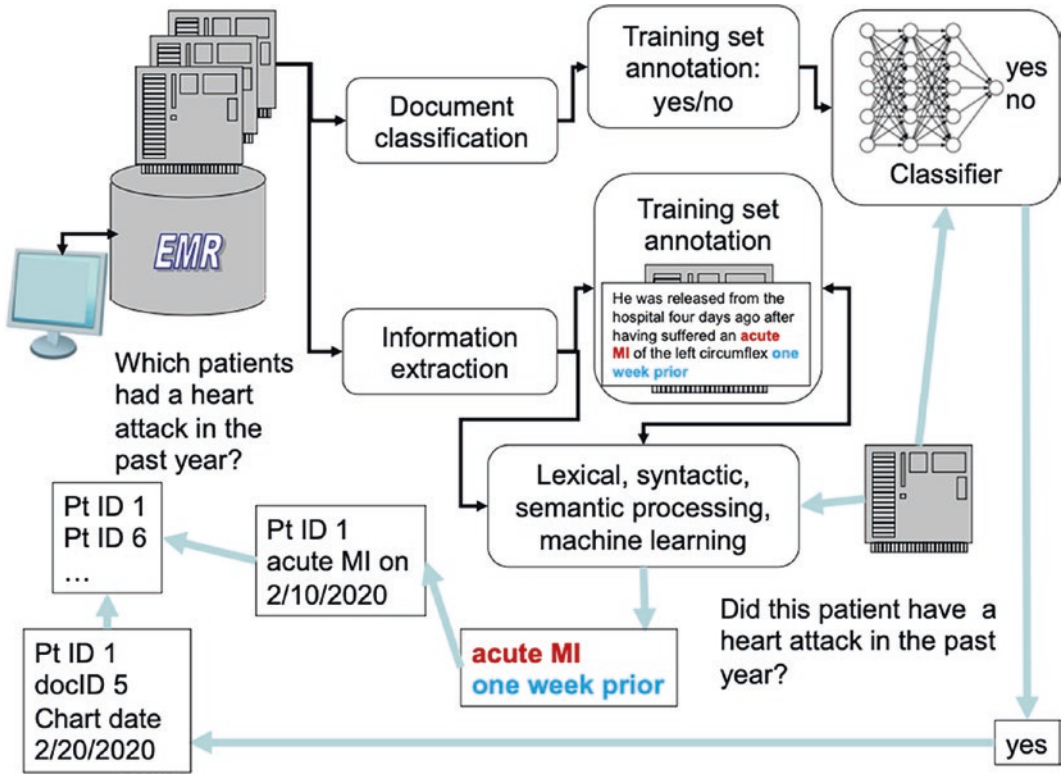


Fig. 8.3 Two different approaches to identifying patients who suffered a heart attack in the past year. The document labeling/classification approach and the template filling/information extraction approach

8.3.1 Topic Modeling

Topic modeling takes as input a collection of texts and identifies the topics discussed in the documents that comprise the collection. It is unsupervised in nature; that is, it does not require any guidance, document labels, or dictionary to identify topics. The discovered topics are expressed as clusters of words. As such, topic modeling is a particularly useful task for exploratory data analysis. A variety of methods have been proposed for the task of topic modeling, including latent semantic analysis (Deerwester et al. 1990), probabilistic latent semantic analysis (Hofmann 1999), and Latent Dirichlet Allocation (LDA) (Blei et al. 2003). All topic modeling methods take a simple representation of the documents, namely a "bag of words" approach. That is, the text is split into words using a tokenizer, and the order between the words is not taken into account during processing.

LDA belongs to the family of probabilistic generative models, and several extensions to LDA, also in the family of probabilistic generative models, were proposed in the literature and can facilitate exploratory analysis of large corpora further (Blei 2012). The generative process of LDA and its variant is what enables topic modeling to go further than a simple exploratory device within a corpus. In addition to discovering the clusters of words that determine the topics of a corpus, it is able to infer for any new document the topics which best represent it.

For instance, Fig. 8.4 depicts the result of applying LDA for 20 topics on a corpus of 1700 documents. Each document in the corpus contains the title and abstract of a scientific publication from PubMed Central. Each topic consists of a ranked list of words, where the ten most likely words are presented. Through examining the topics, one can get a quick overview of the content of the corpus.

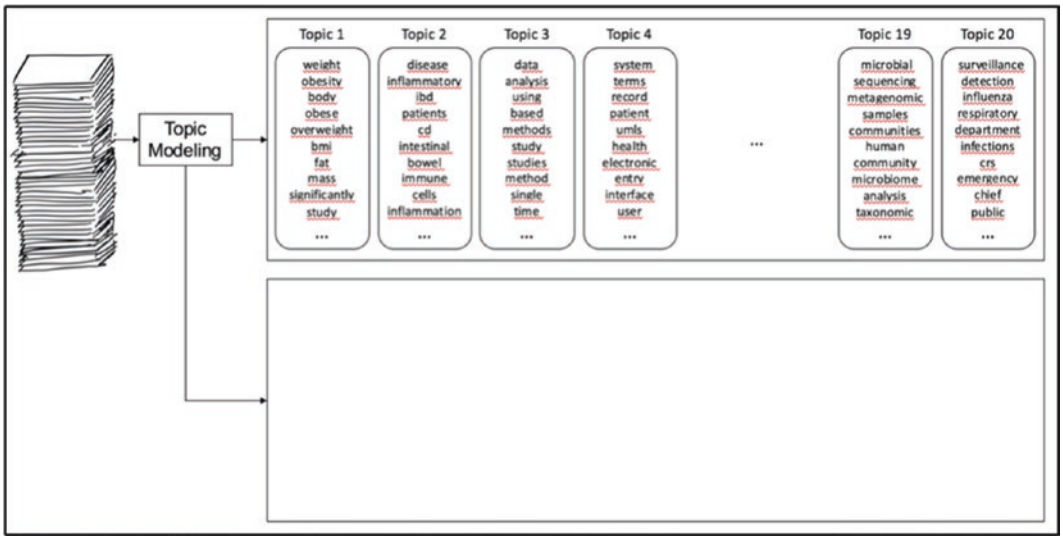


Fig. 8.4 Topic modeling is an unsupervised basic computational tasks in hNLP. Given a corpus, two types of output are generated: topics defined as distributions over words in the corpus, and an inference mechanism to assign topic assignments to any new document

In our example, we see that our documents are a mixture of articles in the fields of clinical, public health, informatics, and biological domains. Themes that emerge from the analysis include obesity (topic 1), microbiology (topic 19), and electronic health records (topic 4). Because the topic modeling task is unsupervised, the discovered topics do not always depict actual themes, but can sometimes elicit groups of words representative of the genre of texts analyzed. Topic 3 in our example is such a topic, where the most highly ranked words are common to the genre of biomedical studies.

Beyond utility of topic modeling as exploratory technique for large corpora, it is a clustering technique which can be found useful in many other applications. For instance, in a machine learning application that leverages both text and other types of data.

8.3.2 Text Labeling

In Text Labeling tasks, the input is text, either in its entirety or a fragment, and the output is a label or a set of labels. Examples of applications in healthcare, consumer health, and biomedicine that use text labeling abound and some are already discussed in ▶ Sect. 8.2, but we list here a few examples:

- Automated coding of discharge summary according to diagnostic codes. This is an example of text labeling task where the input is an entire document – a discharge summary, the clinical note authored at the end of a hospital admission by a physician – and the output is a set of diagnostic codes, where the codes are chosen from a taxonomy, e.g., ICD-10. This task is typically carried out in hospital for billing and administrative goals, and the tasks are done in a semi-automated fashion with coders, professionals trained to select the appropriate codes (Resnik et al. 2006).
- Sentiment analysis of hospital reviews written by health consumers is another example of text labeling task (Greaves et al. 2013). There, a short text written in lay language is labeled according to a single label – its polarity or sentiment. In fact, it might even be a fragment of the text at a time that gets labeled according to its sentiment.

Most approaches to this task use classification techniques from the field of machine learning and thus require a training set of text inputs

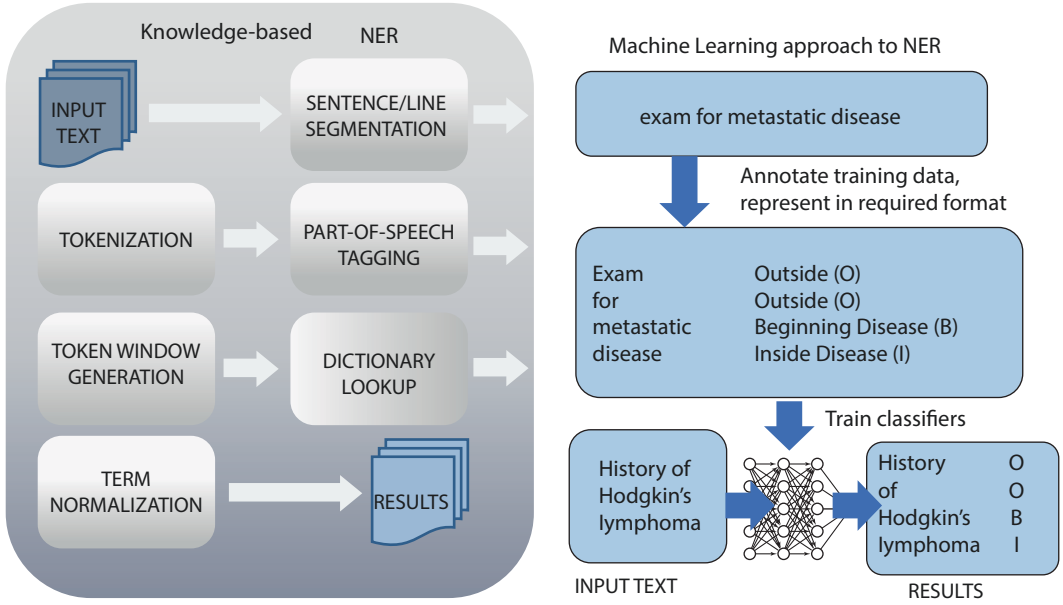


Fig. 8.5 Two approaches to named entity recognition. The approach on the left is a schematic representation of MetaMap – a knowledge-based tool for NER. The right side depicts an abstraction of a machine learning approach to NER. Provided with the following note as input: "Indication: Ca history, exam for metastatic disease, Impression: 1.5 cm nodule in the left mid-lung zone may contain calcium." MetaMap will output the following:

- C0582103: exam: Medical Examination : [hlca]: 0 : 4 : 0.0
- C0027627: metastatic disease: Neoplasm Metastasis : [neop]: 9 : 18 : 0.0
- C4284036: exam: Exam : [ftcn]: 0 : 4 : 0.0
- C2939420: metastatic disease: Metastatic Neoplasm : [neop]: 9 : 18 : 0.0

and their associated labels, although rule-based approaches are also a possibility, e.g., in labeling sections of clinical notes (Denny et al. 2008). Which features to include and how to represent them depends on the task at hand. For instance, in automated diagnosis coding, word-level representations described in ▶ Sect. 8.4 have been found helpful, as well as simple bag-of-words approaches.

8.3.3 Sequence Labeling

The task of sequence labeling is a specific case of text labeling. In sequence labeling, we are keeping track of the order in which textual units occur (be it words or text fragments), and the approaches work better at labeling the entire sequence jointly. Here are two examples of sequence labeling tasks in health-related texts.

We focus on a traditional example for sequence labeling, namely, **named entity recognition** (NER). NER is a classic task of general NLP and in the health domain. The task consists of identifying within a text the different mentions of different types of entities as shown in Fig. 8.5. The task is difficult because we don't know in advance how many words comprise the entity, e.g., *lymphoma* is a single word, *Hodgkin's lymphoma* is a two-word phrase, but many other terms include these entities:

- non-Hodgkin's lymphoma of lung,
- Hodgkin's lymphoma of lung (without the non), and
- non-Hodgkin's lymphoma.

Each one of these phrases is a different term, and sometimes nested recognition might be needed, e.g., if a clinician is looking for all patients with lymphoma grouped by the type and location of the lesion, we will need to identify the full names, as well as the nested mentions of lymphoma. Sequence labeling is

also difficult because terms are often ambiguous with respect to the type of entity they instantiate. The context in which the term occurs might help identify the correct entity type and link the entity to its class or an appropriate identifier in a terminology. For instance, the term "ca" in itself in the clinical text is ambiguous towards two frequent entity types: (as in calcium) measurement or the neoplasm UMLS semantic type (as in cancer).

The output provides UMLS identifiers, positional information of the terms in the text, and their semantic types. Note that in addition to ambiguity of **Ca** that will not be easy to disambiguate to cancer because the note also contains the term calcium, **Exam** is also ambiguous due to the different semantic types assigned to physical exam. Also note the fine-grained differences in normalizing metastatic disease. The output of the machine learning model trained to label disease name sequences will identify Hodgkin's lymphoma as disease.

Like in text labeling, sequence labeling rely on supervised approaches. Training data with sequences fully annotated are required as training examples, see Fig. 8.5. While probabilistic approaches like Hidden Markov Models and Conditional Random Fields were historically used for this task, neural architectures have shown impressive results in sequence learning. Recurrent neural networks (LSTM, bi-LSTM, GRUs) in particular encode in their architecture the ability to handle almost arbitrarily long sequences and keep track of the relevant words within them (whether long-term dependency or short-term dependency) (Habibi et al. 2017).

8.3.4 Relation Extraction

Similarly to NER, relation extraction can be decomposed into relation detection and determination of the relation type. Until recently, research in biomedical relation extraction was limited to protein-protein interactions and gene binding, primarily due to availability of resources created for BioCreative evaluations (Krallinger et al. 2008). In the clinical domain, relations of interest are those between medical problems and treatments, problems and tests, problems and genes, and genes and treatments (the latter two becoming increasingly important due to interest in translational research and precision medicine). The resources for extraction of some of these relations are available due to the i2b2 relation extraction challenge (Uzuner et al. 2011). Another source of literature partially annotated for entities and relations is the pharmacogenomics knowledge base (PharmGKB) (Barbarino et al. 2018). Among other activities, PharmGKB provides literature annotated for genetic variants and gene-drug-disease relationships and annotates associations between genetic variants and drugs, and drug pathways.

Initially, the existence of relations between entities was assumed if the entities co-occurred more frequently than by chance in a unit of text, such as a MEDLINE abstract or a sentence. Mutual information, chi-square and log-likelihood ratio were often used to extract co-occurrence-based relations (Hakenberg et al. 2012). This approach has two drawbacks: (1) the results are often noisy and (2) the nature of the relations is undefined.

Both the knowledge-based and statistical approaches are used to extract specific relations. Knowledge-based approaches often involve lexical-semantic or syntactic-semantic patterns. For example, to extract a relationship between a complication of a patient's health condition and its cause, we can define a "Complications of" relation. The expressions: "status post", "secondary to", and others indicate the presence of the "Complications of" relation. These expressions can be combined with semantic categories to form patterns for a rule-based system. For example, the "[concept Problem][s/p][concept any|word noun]" pattern, in which "s/p" represent the set of the indicator expressions. A rule-based system that extracts "treats" relations can include the following rule:

If dependency path contains a treatment indicator, Procedure and Problem ⇒ (treats, Procedure, Problem)

where treatment indicators are stored in a **gazetteer** and Procedure and Problem are identified by one of the NER tools presented above. As common for the rule-based methods, this approach has relatively low recall and is potentially brittle. The currently better-performing systems use supervised machine learning. Most supervised machine learning methods assume that the entities are identified by a NER tool and require positive examples in which the relations and entities are annotated and negative examples in which there are no relations between annotated entities. Given these examples, classifiers (Support Vector Machines in the near past, and currently Deep Learning frameworks (Peng et al. 2018)) are trained to determine if a specific relationship between the candidate entities exists. If several relations are possible between the two entities, for example, drug causing a problem or drug treating a problem, a two-step approach can be applied: first determining if a relation exists; and then determining the type of the relation. As with other classification tasks, feature selection is one of the factors determining the success of the method. In addition to the standard "bag of words" in the windows preceding, following, and between the two concepts, features often used in relation extraction are: the semantic types of the concepts; the distance between the concepts, parts of speech, paths to the root of the parse tree and dependency relations between the concepts.

8.3.5 Template Filling

Often, n-ary relations, e.g., the size, the location and the borders of a lesion, are of interest. Capturing such information requires event (frame) extraction and is viewed as template filling task.

Biomedical events involve a change in the state of biomedical objects. Examples of the events are gene expression, protein binding and regulation in the biological domain and medication or phenotype events in the clinical domain. Events usually involve multiple entities and relations between the entities and can be nested. Medication events, as shown in Fig. 8.6, exemplify relatively simple events: a medication event involves a drug, its form, dosage, administration route, duration, and indications.

The complex biomolecular events are usually more involved. For example, consider the phrase "SYK-TLR4 binding increases upon TLR4 dimerization and phosphorylation" (PMID: 22776094) that introduces a complex positive regulation event in which binding of spleen tyrosine kinase (SYK) to the toll-like receptor-4 (TLR4) is regulated by two simple events: TLR4 dimerization and phosphorylation. Secondary arguments of the events may provide additional information about the event, such as the specific domain or region of the theme of the event. For example, in "binding of SYK to the cytoplasmic domain of toll- like receptor-4 (TLR4)", cytoplasmic domain is the secondary argument associated with the TLR4 theme of the binding event.

Many systems for event recognition are built as pipelines that start with recognizing protein names. For example, one approach to extracting biomedical events on PubMed scale (Björne et al. 2010) is to use publicly available NER tools as the first step in extraction of protein events from PubMed abstracts. In the subsequent steps, the extraction pipelines rely on the graph representations of sentence syntax and semantics, which we will present in the next section. The syntactic graph generated by a dependency parser and the identified named entities are used to generate a semantic graph for each sentence independently. The system uses the graphs as the source of features to supervised machine learning for event trigger and edge detection, which are followed by the rule-based event construction step.

8.4 Linguistic Knowledge and Representations

It is important to know the principles of linguistics and computational linguistics that, when incorporated in computational tasks, help achieve better results. Processing of language is not as simple as applying a pipeline of independent modules: one to determine

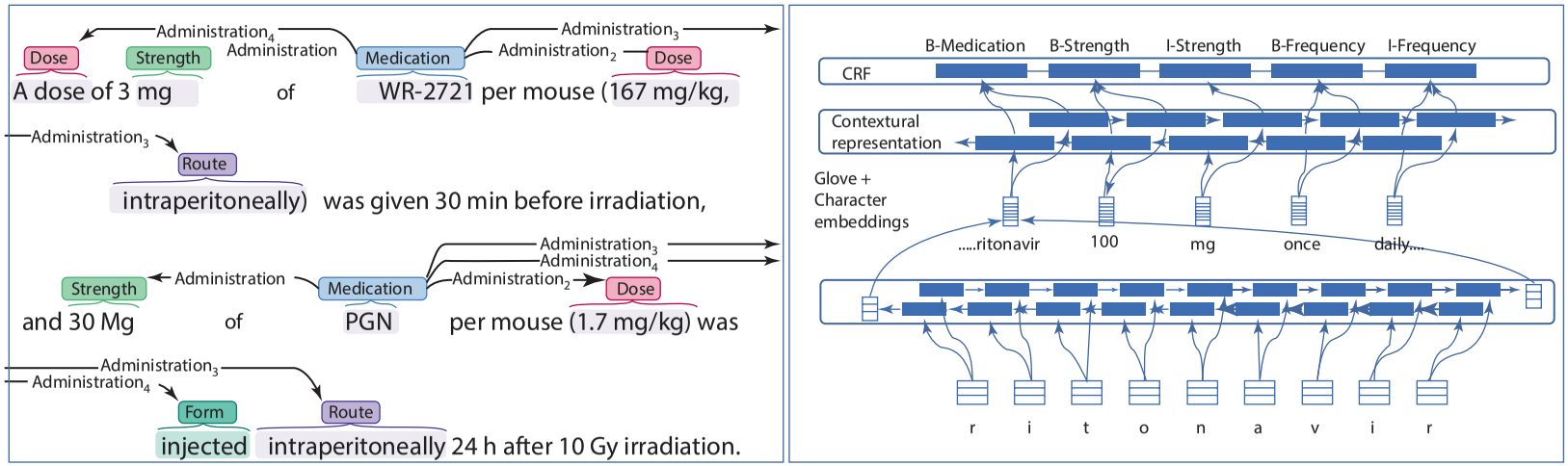


Fig. 8.6 Annotation of a medication event on the left using brat annotation tool (Stenetorp et al. 2012) and on the right, a Long Short-Term Memory (LSTM) neural network with a Conditional Random Field (CRF) layer using character embeddings trained to extract medication frames. Character level word representations are extracted using Bi-directional LSTM and concatenated with word embeddings. These word representations feed into another Bi-LSTM layer to extract contextual representations of the words. The contextual representations are fed to the CRF layer to decode the best label sequence

tokens, one to assign part-of-speech tags to tokens and to parse the syntax, one to interpret the meaning of a sentence, and one to resolve the discourse-level characteristics of the text. In reality, all linguistic levels influence each other. Low-level decisions about how to tokenize a string impact named-entity recognition; determining which sense to attribute to a named entity depends on its place in the syntactic tree, the pragmatics of the text, and its place in the discourse structure. How to model these interactions is one of the primary open research questions of natural language processing, which is currently addressed by modeling the tasks jointly, e.g., using deep learning approaches (James et al. 2013). Although language processing is not a simple pipeline, the practical applications still often approximate the process as one, and linguistic knowledge contributes to performing the basic tasks and modeling the interactions.

Linguistic levels consist of word-level representations (tokens and morphology; sentence-level representations (syntax and semantics); and document-level representations (pragmatics and discourse). Linguistic knowledge is captured in lexicons, domain knowledge, e.g., lists of diseases and drugs can be found in terminologies, and domain semantics, such as relations among diseases and drugs comprises ontologies.

8.4.1 Terminological and Ontological Knowledge

In the biomedical and clinical domains, interpreting text might require extensive background knowledge, as many facts are implied, e.g., inferring that a patient is likely hypertensive or has an edema if loop diuretics are prescribed, even though there are no explicit mentions of either high blood pressure or swelling in the text. Ontologies contain some of the background knowledge, and some can be learned from text.

The Unified Medical Language System (UMLS), including the Metathesaurus, Semantic Network, and the Specialist Lexicon) – can be used as a knowledge base and resource for NLP tasks. MedDRA and RxNorm, two of the over 130 sources contributing to the UMLS, are examples of terminologies specific to adverse events and medications, respectively. They are particularly helpful in the clinical domain, in biosurveillance, pharmacovigilance and in pharmacogenomics. The UMLS Metathesaurus is organized by concept. It preserves the meaning and structure of the contributing sources and links alternative surface representations of a concept in many languages. It also establishes relationships between concepts. All concepts in the Metathesaurus are assigned to at least one Semantic Type from the Semantic Network. Most English strings in the Metathesaurus also appear in the SPECIALIST Lexicon. The Specialist lexicon provides detailed syntactic knowledge for words and phrases and includes a comprehensive medical vocabulary. It also provides a set of tools to assist in NLP, e.g., a lexical variant generator.

Although ontologies and lexicons are maintained and regularly updated, the language changes, new concepts enter the language, and terms fall out of use and become obsolete (although data represented with the terms may persist). The biomedical and health domains are highly dynamic in the influx of new terms (e.g. new drug names, but also sometimes new disease names, like COVID-19, SARS and H1N1). Modern corpus-based approaches compensate for that lag leveraging the existing knowledge. For example, all diseases in the UMLS can be marked in PubMed and the representation of disease can be learned as described in the section on word embeddings. A new disease such as COVID-19 can then be recognized as such using its context and the learned representation of *Disease*.

8.4.2 Word-Level Representations

- **Tokens**

Tokens are basic language units defined based on their utility for solving a specific language

processing task. The units include morphemes, words (often morpheme sequences), numbers, symbols (e.g. mathematical operators), and punctuation. The notion of what constitutes a token is far from trivial. The primary indication of a token in general English is the occurrence of white space before and after it; however there are many exceptions: a token may be followed by certain punctuation marks without an intervening space, such as by a period, comma, semicolon, or question mark, or may have a "–" in the middle, as shown in ► Example 8.1:

► Example 8.1

… a decoy receptor for IL-2 in the T cell …
… IL 2-regulated genes …
… Il2 and Csf2 are increased as T-cells … ◄

The three snippets in ► Example 8.1 contain three different spellings for IL2. Ideally, we would like to link all three surface forms to interleukin-2. To achieve this normalization, we need a tokenizer that will treat the dash and the white space equally and segment IL2 into IL and 2 (or merge IL and 2 in the first two example snippets.)

In biomedicine, periods and other punctuation marks can be part of words (e.g., *p.o.* means *per os (by mouth; orally)* in the clinical domain, and *M03F4.2A*, is a gene name that includes a period.) Moreover, punctuation marks are used inconsistently, thereby complicating the tokenization process: In clinical text, it is common to abbreviate "discontinue" as d/c, without as w/o, but it is also common to write s/p for status post, and, finally, use slashes in measurements and units. In addition, chemical and biological names often include parentheses, commas and hyphens, for example *(w)adh-2*, which also complicate the tokenization process. For example, replacing non-alphanumeric characters with spaces will prevent us from correctly identifying entities in the following sentence: "PBMC of HLA-DR3(+) but not HLA-DR3(−) cured TB patients." For agglutinative languages, tokenization needs to be augmented by segmentation as shown in ► Example 8.2.

► Example 8.2 *Rückschmerzen nach Brustwirbelbruch*

Ich habe oft Schmerzen im Rücken und in der Schulter. [Back pain after thoracic fracture: I often have pain in the back and in the shoulder]. ◄

In this example blog post, *Rückschmerzen* in the title needs to be split into *Rück[en]* (back) and *Schmerzen* (pains). In addition, to identify the reason for back pain, information about the fracture (*bruch*) has to be separated from its location (thoracic spine – *Brust – wirbel*.)

Some thought should be given to upper and lower case in the original text. In some situations, it makes sense to keep all tokens in lowercase, as it reduces variations in vocabulary. But in others, it might hinder further NLP: in both the biology and clinical domains, there are many acronyms, which when lowercased, might be confused with regular words, such as *CAT* (*computerized axial tomography*) and *cat* or *FISH* (*fluorescent in situ hybridization*) and *fish*.

Morphology concerns the combination of morphemes (roots, prefixes, suffixes) to produce words or lexemes, where a lexeme generally constitutes several forms of the same word (e.g. activate, activates, activating, activated, activation). There has been little work concerning morphology in the field of NLP in the biomedicine and health domains, especially for the English language. In other languages that are morphologically rich (e.g., Turkish, German, and Hebrew), encoding morphological knowledge is necessary. For example, morphological proximity can identify important terminological relations (Claveau and L'Homme 2005) or generate definitions of medical terms (Deléger et al. 2009b).

■ Word Embeddings

Word embedding is a feature learning techniques in NLP in which words or phrases from the vocabulary are numeric vectors such that similar words will have similar vectors. Word embeddings have two functions: they capture the meaning of a word using its context and, at the same time, condense the representation

of the word into a vector, e.g., a vocabulary of thousands of words can be compressed into a 300-dimensional vector. The idea of "recognizing a word by the company it keeps" originated from Firth (1957). One of the earliest corpus-based implementations, Brown clustering, aggregated words into classes using hierarchical clustering (Turian et al. 2010), and now deep learning provides more robust approaches to pre-computing representations of words using large corpora, e.g., 2.5 billion Wiki words. The growing family of text embeddings and pre-trained language models started with Word2Vec (Mikolov et al. 2013), and now includes GloVe (Pennington et al. 2014), fastText (Bojanowski et al. 2016), ELMo (Peters et al. 2018), BERT (Devlin et al. 2018), GPT (Radford et al. 2019), and BART (Lewis et al. 2019), to name a few. The language models, such as BERT, have been shown to be open to fine-tuning for specific NLP tasks. Domain-specific embeddings and models trained on PubMed and clinical text also exist, and have been shown superior to those pre-trained on the open-domain text. BioBert, for example, significantly advanced the state-of-the-art for biomedical named entity recognition, relation extraction, and question answering (Lee et al. 2019).

Spelling Variants and Errors

In addition to common American English and British **spelling variants**, such as -or/-our (e.g., color/colour), e/ae, e/oe, er/-re (e.g., liter/litre), and -ize/-ise, generic drug names may also differ (e.g., adrenaline (British) vs. epinephrine). The complex origins and spelling of the biomedical terms often lead to **misspellings**. Misspellings in the published literature are relatively rare, but the queries submitted to the search engines are often misspelled. Clinical notes and informal communications also often contain misspelled terms. Clinicians, when typing free text in the EHR, do so under time pressure and generally do not have the time to proofread their notes carefully. In addition, they frequently use abbreviations (e.g. *HF* for *Hispanic female* or *heart failure, 2/2* for *secondary to* or a date), many of which are non-standard and ambiguous. For patients and health consumers, when posting content online, misspellings, typos, and non-standard abbreviations are pervasive like in the rest of the social Web. Ignoring these variations may cause an NLP system to lose or misinterpret information. At the same time, errors can be introduced when correcting the typos automatically. For instance, it is not trivial to correct *hypetension* automatically without additional knowledge because it may refer to *hypertension* or *hypotension*. This type of error is troublesome not only for automated systems, but also for clinicians when reading a note, as this phenomenon is aggravated by the large amount of short, misspelled words in notes. In the clinical domain, misspellings can be found even in the definitions of clinical variables.

8.4.3 Sentence-Level Representations

Sentence Boundary Detection

Detecting the beginning and end of a sentence may seem like an easy task, but it is highly domain dependent. Not all sentences end with a punctuation mark (this is especially true in texts with minimal editing, such as online patient posts and clinical notes entered by physicians). Sentences in scientific publications are usually well-formed and delimited by final punctuation, primarily a period. Some care has to be taken to avoid breaking up sentences on periods used in abbreviations (e.g., vs.) and in honorifics, chemical names, and decimal numbers (as discussed in tokenization). In most cases an off-the-shelf sentence tokenizer is expected to be highly accurate.

The informal biomedical text is much harder to split into meaningful utterances. Clinical notes often contain table-like structures and lists. Moreover, some electronic health records might enforce a certain line length, which could be violated by, for example, a de-identification tool. Therefore, custom solutions might be needed to detect end of sentences in these texts.

Syntax concerns the categorization of the words in the language, and the structure of the phrases and sentences. Each word belongs to

one or more parts of speech in the language, such as noun (e.g. chest), adjective (e.g. mild), or tensed verb (e.g. improves) Lexemes can consist of more than one word as in foreign phrases (ad hoc), prepositions (along with), and idioms (follow up, on and off). Lexemes combine in well-defined ways and according to their parts of speech, to form sequences of words or phrases, such as noun phrases (e.g. severe chest pain), adjectival phrases (e.g. painful to touch), or verb phrases (e.g. has increased). Each phrase generally consists of a main part of speech and modifiers, e.g. nouns are frequently modified by adjectives, while verbs are frequently modified by adverbs. The phrases then combine in well-defined ways to form sentences (e.g. "he complained of severe chest pain" is a well-formed sentence, but "pneumococcal vaccine how often?" is not).

General English imposes many restrictions on the formation of sentences, e.g. every sentence requires a verb, and count nouns (like cough) require an article (e.g., a or the). Clinical language, in contrast, is often telegraphic, relaxing many of these restrictions of the general language to achieve a highly compact form. For example, clinical language allows all of the following as sentences: *the cough worsened*, *cough worsening*, and *cough*. Because the community widely uses and accepts these alternate forms, they are not considered ungrammatical, but constitute a sublanguage (Friedman et al. 2004).

Representation of Syntactic Knowledge

Phrases and sentences can be represented as a sequence where each word is coupled with its corresponding part of speech as shown in Fig. 8.7. For example, *Severe joint pain* can be represented as Severe/*adjective* joint/*noun* pain/*noun*. Formalisms that can be used to represent syntactic linguistic knowledge include probabilistic context free grammars (Jurafsky and Martin 2019), which, along with dependency formalisms are widely used in language processing.

Dependency is a binary asymmetrical relation between a head and its dependents or modifiers. The head of a sentence is usually a tensed verb. Thus dependency structures are

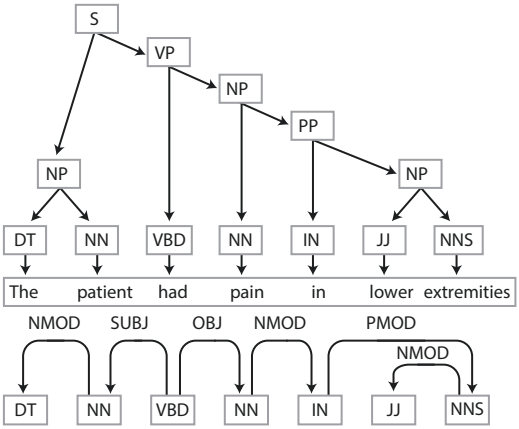


Fig. 8.7 A syntactic parse tree for the sentence "The patient had pain in lower extremities" according to the context-free grammar is shown above the sentence. Notice that the terminal nodes in the tree correspond to the syntactic categories of the words in the sentence. The parse tree below the sentence, is in a dependency grammar framework

basically directed relations between words. For example, in the sentence, "The patient had pain in lower extremities", the head of the sentence is the verb "had", which has two arguments, a subject noun "patient" and an object noun "extremities", that modifies or is dependent on "patient", while "in" is dependent on "pain", "extremities" is dependent on "in", and "lower" is dependent on "extremities". As such, in a dependency grammar, the relations among words and the concept of head in particular (e.g. "extremities" is the head of "lower") is closer to the semantics of a sentence.

Semantics

Semantics concerns the meaning or interpretation of words, phrases and sentences, generally associated with real-world applications. There are many different theories for representation of meaning, such as logic-based (e.g., first order logic and lambda calculus), frame-based, conceptual graph formalisms, and distributional semantics, i.e., vector representations learned from data (Jurafsky and Martin 2019).

Each word has one or more meanings – word senses (e.g. resistance, as in psychological resistance, social resistance, multidrug

resistance, and capillary resistance), and other terms may modify the senses (e.g. no, as in no fever, or last week as in fever last week). Recognizing polysemy or ambiguity of the word is important. Complementary to the phenomenon of polysemy, there are often terms that are different variations of the same concepts (synonymy). For instance, the term blood sugar is often used by health consumers to refer to a glucose measurement, but it is used rarely if ever in the clinical literature or in clinical notes.

Additionally, the meanings of the words combine to form a meaningful sentence, as in "there was thickening in the renal capsule"). Representation of the semantics of general language is extremely important, but the underlying concepts are not as clear or uniform as those concerning syntax. Interpreting the meaning of words and text in general is very challenging. In biomedical informatics, interpreting the meaning of text focuses, largely, on entity linking (i.e., representing a word or group of words with a unique semantic concept from a relatively small number of well-defined semantic types, e.g. medication, gene, disease, body part, or organism.) The semantics of phrases and sentences is also restricted to a smaller set of patterns than in general language (e.g. medication-treats-disease, gene-interacts-with-gene).

There are often several ways to express a particular medical concept as well as numerous ways to express modifiers for that concept. For example, ways to express severity include faint, mild, borderline, 1+, 3rd degree, severe, extensive, and moderate. Often, to complicate matters, modifiers can be composed or nested. For instance, in the phrase "no improvement in pneumonia," improvement is a change modifier that modifies the concept pneumonia, and no is a negation marker that modifies improvement (not pneumonia). Complex semantic structures containing nesting can be represented using a semantic grammar, which is a context free grammar based on semantic categories. An alternative representation would facilitate processing by flattening the nesting. In this case, some information may be lost but ideally only information that is not critical. For example, *slightly improved* may not be clinically different from *improved*. Since this type of information is fuzzy and imprecise, the loss of information may not be significant. However, the loss of a negation modifier would be significant. Another such example concerns hedging, which frequently occurs in radiology reports as well as in the scientific articles. Implementing a semantic grammar would require a large corpus that has been annotated with both syntactic and semantic information. Since a semantic grammar is domain and/or application specific, annotation involving the phrase structure would be costly and not portable, and therefore is not generally done.

8.4.4 Document-Level Representations

As is the case for text in general, documents within the biomedical domain are expected to have a certain structure. Text or discourse structure refers to the way in which authors organize information within documents. The organization of biomedical text is reflective of both the type of information being conveyed as well as its intended audience. The structure of biomedical text aids in its comprehension by the reader, and can be utilized in performing various natural language processing tasks. The structure of biomedical text can be examined at a local or global level. The local level primarily concerns coherence and cohesion, aspects of structure that connect text and give it meaning, whereas the global level concerns aspects of the overall organization and rhetorical structure of a document, such as its sectioning.

The cohesive devices responsible for the local structure of text play an important role in comprehension of biomedical documents. The recognition of phenomena such as coordinating constructions, anaphora, as well as ellipsis is needed to represent biomedical text at the document level. In this section, we exemplify discourse processing with automated resolution of referential expressions and then discuss pragmatics, which concerns everything

extraneous to the text that contributes to its meaning and context. In clinical and health-consumers language processing, the context is sometimes readily accessible and sometimes easily extractable from datasets, e.g., medication orders, who wrote the text, at what time, etc. Interaction patterns can be inferred from the datasets as well.

8.4.4.1 Automated Resolution of Referential Expressions

Determining which words or phrases in a text referring to the same entity, called **coreference resolution** can draw on both syntactic and semantic information in the text.

Syntactic information for resolving referential expressions includes:

- Agreement of syntactic features between the referential phrase and potential referents
- Recency of potential referents (nearness to referential phrase)
- Syntactic position of potential referents (e.g. subject, direct object, object of preposition)
- The pattern of transitions of topics across the sentences

Syntactic features that aid the resolution include such distinctions as singular/plural, animate/inanimate, and subjective/objective/possessive. For instance, the inanimate pronoun "it" usually refers to things, but sometimes does not refer to anything when it occurs in *cleftconstructions*, such as "it was noted", "it was decided to" and "it seemed likely that".

Referential expressions are usually very close to their referents in the text. The syntactic position of a potential referent is an important factor. For example, a referent in the subject position is a more likely candidate than the direct object, which in turn is more likely than an object of a preposition. Centering theory accounts for reference by noting how the center (focus of attention) of each sentence changes across the discourse (Grosz et al. 1995). In this approach, resolution rules attempt to minimize the number of changes in centers. Semantic information for resolving referential expressions involves consideration of the semantic type of the expression and the way it relates to potential referents (Hahn et al. 1999; Kilicoglu 2016).

8.4.5 Pragmatics

Pragmatics concerns how the intent of the author of the text, or, more generally, the context in which the text is written, influences the meaning of a sentence or a text. For example, in a mammography report "mass" generally denotes breast mass, whereas a radiological report of the chest denotes mass in lung. In yet a different genre of texts, like a religious journal, it is likely to denote a ceremony. Similarly, in a health care setting "he drinks heavily" is assumed to be referring to alcohol and not water. In these two examples, pragmatics influences the meaning of individual words. It can also influence the meaning of larger linguistic units. For instance, when physicians document the chief-complaint section of a note, they list symptoms and signs, as reported by the patient. The presence of a particular symptom, however, does not imply that the patient actually has the symptom. Rather, it is understood implicitly by both the author of the note and its reader that this is the patient's impression rather than the truth. Thus, the meaning of the chief-complaint section of a note is quite different from the assessment and plan, for instance.

Another pragmatic consideration is the interpretation of pronouns and other referential expressions (there, tomorrow). For example, in the two following sentences "An infiltrate was noted in right upper lobe. It was patchy", the pronoun "it" refers to "infiltrate" and not "lobe". In a sentence containing the term "tomorrow", it would be necessary to know when the note was written in order to interpret the actual date denoted by "tomorrow". As mentioned above, in the biomedical domain, pragmatics can be encoded through the semantic lexicon and rules about the discourse of a text.

8.5 Practical Considerations

The recent years see a steadily growing demand for biomedical language processing. The traditionally manual tasks, such as assigning medical billing codes for reimbursement, indexing biomedical literature, populating biological knowledge bases, and providing evidence for clinical decision support are assisted by information extraction and classification tools. With few exceptions, such tools are not yet widely used, and the need for them exceeds their supply.

When embarking on adding natural language processing to the workflow, practitioners and users often ask where to start and if there are any tools, corpora, and resources available. These excellent questions should be asked, but with the exception of few well-established resources, we refrain from pointing to specific collections because the field is extremely active and fast moving. The questions, therefore, should be answered at the time of need using a literature search, including searching PubMed, a widely used and growing collection of citations in biomedical literature and PubmedCentral, a growing collection of open access full text biomedical articles.

Independently of the specific task, corpora and tools, any NLP endeavor starts with dealing with raw data, which entails dealing with file formats, character sets and machine settings. Bird et al. (2019) discuss the practical considerations of working with unstructured text in Python.

Specific biomedical software toolkits for many tasks, such as named entity recognition, introduced in ▶ Sect. 8.2.1, or modality detection, are freely available and widely used. Many of the existing approaches are built using the open domain tools, such as NLTK (Bird et al.), OpenNLP (openNLP) and UIMA (UIMA). Many solutions to specific problems leverage the existing tools to build pipelines that include the existing tools, e.g. MetaMap for NER, combined with the local implementation of the task-specific algorithms.

8.5.1 Patient Privacy and Ethical Concerns

As an NLP system deals with patient information, its designers must remain cognizant of the privacy and ethical concerns entailed in handling protected health information. In the clinical domain for instance, the Health Insurance Portability and Accountability Act (HIPAA) regulates the protection of patient-sensitive information (see ▶ Chap. 12 for a detailed description of privacy matters in the clinical domain). Online, patients provide much information about their own health in blogs and online communities. While there are no regulations in place concerning online patient-provided information, researchers have established guidelines for the ethical study and processing of patient-generated speech (Eysenbach and Till 2001).

The somewhat opposing needs for large amounts of data for NLP processing and protecting patients' privacy, led to development of de-identification and anonymization tools (Meystre et al. 2010). Many researchers are exploring transfer learning (Ruder 2019), where the tools are trained on openly available data, e.g., general domain or veterinary, or synthetically generated life-like data and then fine-tuned on the small amounts of task and domain specific data (Wu et al. 2019).

8.5.2 Good System Performance

If the output of an NLP system is to be used to help manage and improve the quality of healthcare and to facilitate research, it must have high enough performance for the intended application. Evidently, different applications require varying levels of performance, and the desired level of performance needs to be discussed with the intended users of the system. While discussing a system's performance, it is important to make sure the users understand the benefits and the limitations of the approach and have reasonable expectations with respect to the results, prob-

ability and nature of errors, and potential requirements for curation. An example application is finding and ranking patients with respect to eligibility for a cohort. Depending on the nature of the cohort, the users might want to see only the patients for which information was extracted with high accuracy, e.g., for a retrospective study on a large dataset, or all patients that might fit the criteria, e.g., identifying patients at risk for disease exacerbation. These pragmatic questions are external to NLP processing, but still need to be answered to optimize the models and approaches as needed.

8.5.3 System Interoperability

NLP-based systems are often part of larger applications. There must be seamless integration of the NLP component into its parent application. This is equally important in the clinical domain, where the system must follow standards for interoperability among different health information technology systems, such as Health Level 7 (HL7) and the Clinical Document Architecture (CDA; see ▶ Chap. 7), and in processing biomedical literature and social media, where the system should be able to communicate with the downstream applications. For example, information extracted automatically to support database curation, e.g., model organism databases, should be provided to curators within their workflow, along with an easy access to the context that suggested these terms.

8.6 Research Considerations

Biomedical natural language processing has a wide range of practical applications. It facilitates clinical and biomedical research, quality assurance of clinical care and delivery of information to patients. This wide range of tasks stimulates an ongoing and constantly growing research of foundational principles of biomedical language processing. Due to the growing interest in literature-based discovery, management of big data, secondary use of clinical data, clinical decision support, and population studies through social media, the demand for biomedical language processing has increased significantly and will continue growing.

Being primarily motivated by the needs of the domain, biomedical and clinical NLP always was data driven. All NLP research starts with exploratory data analysis that takes into account the context in which the text was created and the context in which it is used. Some of the text always needs to be annotated to create gold standards for evaluation, and, depending on the approach that is researched, e.g. supervised machine learning, large amounts of annotated text might be needed to train the models. Annotation takes time, effort and money, so leveraging the existing annotated collections and approaches that allow adding minimal amounts of task-specific annotations to improve the results is growing in popularity.

Before an NLP-based system can be used for a practical task, it must be evaluated carefully, both **intrinsically** and **extrinsically**, in a setting where the system will be used. A variety of techniques exists for the evaluation and testing of natural language processing programs. They vary with respect to cost, repeatability, and the kind of information that is obtainable from them. In this section, we first discuss data annotation and annotation guidelines, and then present evaluation principles and approaches.

8.6.1 Data Annotation

During the initial task definition and data exploration for gold standard construction, an annotation schema is established to capture the minimal amount of annotations sufficient to perform the task. At the same time, annotation guidelines are created to describe the task, the schema and the annotation rules. There are ongoing community efforts, starting with The Canon group (Evans et al. 1994), to create established representations for certain

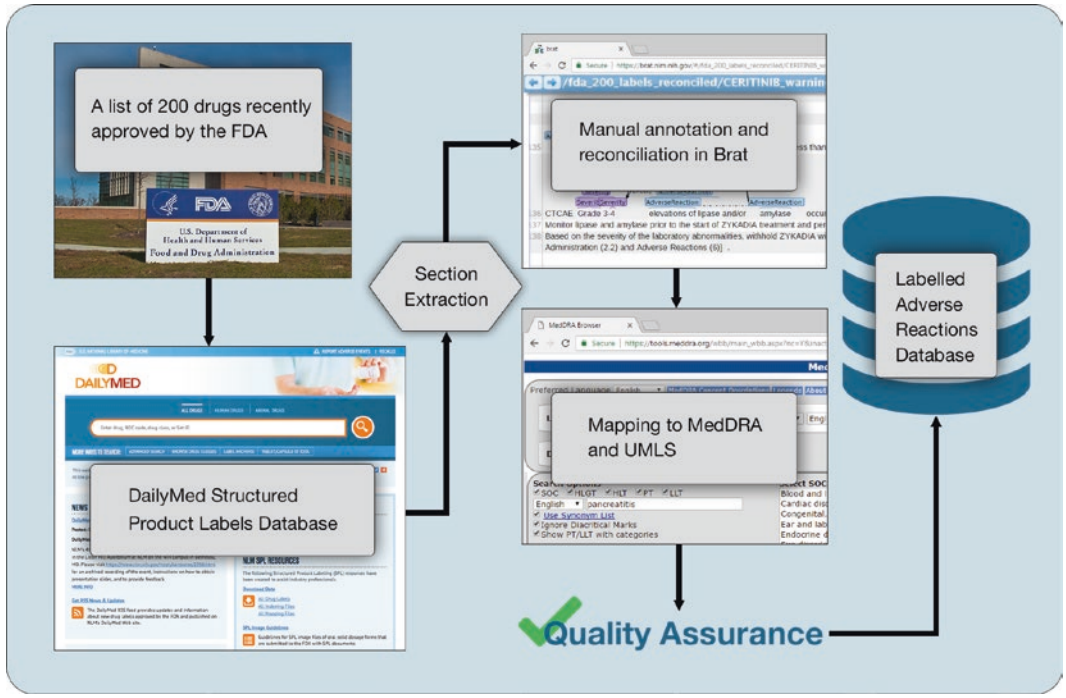


Fig. 8.8 Document annotation process for detection of Adverse Drug Reactions

aspects of information, such as the different modifiers of concepts and the relations among concepts that can occur in texts. Although in most cases the specific datasets are still using local representations, the de-facto standard is standoff annotation that provides offsets of the strings and meta-description to represent concepts. More variety exists in representing relations and events.

Once the initial schema is established, a small number of documents is annotated by a group of annotators to determine if the schema allows annotating all required entities and relations and if the guidelines are clear. The next step is to finalize and freeze the guidelines, and then annotate the required number of documents, ensuring some overlap to measure inter-annotator agreement. The guidelines can be modified during annotation for one purpose only: to add information about a new case that was not covered by the existing rules. Figure 8.8 illustrates the process of creating a specific corpus of drug labels annotated with Adverse Drug Reactions and all steps of this annotation effort (Demner-Fushman et al. 2018).

8.6.2 Evaluation

Evaluating the performance of an NLP system is crucial whether the NLP system targets the end-users directly or as a part of a larger application. In the biomedical NLP domain, evaluation brings together two traditions: evaluation in biology and clinical research and evaluation of software, both with respect to its output and usability in the eyes of the intended end-users.

Biomedical and clinical researchers expect health technology assessment to include "properties of a medical technology used in health care, such as safety, efficacy, feasibility, and indications for use, cost, and cost effectiveness, as well as social, economic, and ethical consequences, whether intended or unintended" (IOM 1985). Measuring the social, economic, and ethical consequences of NLP systems in biomedical domain has not been systematically researched. Several studies looked into social consequences of delivering information to clinicians. For example, over 500 clinicians interviewed by Lindberg et al. used MEDLINE searches to choose the most

appropriate test, make the diagnosis, develop and implement a treatment plan, maintain an effective physician-patient relationship, and modify patients' health behaviors. In 8 cases, MEDLINE was credited with saving a patient's life, and in another 17 with increasing the length of life (Lindberg et al. 1993b).

The efficacy, feasibility and cost of NLP systems and tools, on the other hand, are relatively easy to measure and these evaluations follow the principles of the software evaluation tradition. The metrics described below were developed to evaluate the software performance using sets of benchmarks independent of the tasks for which the tools might be used. These intrinsic evaluations measure changes in the system's output caused by changes in the system's parameters, as well as the differences between systems that implement different algorithms. For example, we can compare different parsers against an established reference standard, such as the Penn Treebank (Taylor et al. 2003). Alternatively, in an extrinsic evaluation that measures a method's performance in a given task, we can ask what parser will improve the overall performance in a relation extraction task. Most of the large-scale evaluations (shared tasks) provide venues and generate collections that allow evaluating systems' performance in a specific task, e.g., the adverse drug reaction collection in Fig. 8.8 was used in a Text Analysis Conference evaluation (Roberts et al. 2017).

8.6.2.1 Evaluation Metrics

The most commonly used metrics for evaluations conducted by computational linguists are precision, recall, and F-measure. The clinical informatics community prefers referring to recall as sensitivity, and pairs it with specificity and the area under the ROC (Receiver Operating Characteristic) curve, if the task allows applying these metrics. The above metrics are based on the **confusion matrix** (or error matrix) and are often defined in terms of the four cells of the matrix:

- true positive (tp) – outputs correctly labeled as having the characteristics of interest, for example, tagging a string as gene name
- false positive (fp) – outputs incorrectly labeled as having the characteristics of interest, for example, tagging a string that is not a gene name as gene name
- true negative (tn) – outputs correctly labeled as not having the characteristics of interest, for example, tagging a string that is not a gene name as such
- false negative (fn) – outputs incorrectly labeled as not having the characteristics of interest, for example, failing to tag a string that is a gene name as gene name

In many NLP tasks, using metrics based on the true negative values is problematic because the number of true negatives is not countable. Even if we arbitrarily specify what constitutes a true negative for this task, the annotation effort for the reference set will become even more daunting and expensive than the efforts described above, and our solution will not solve the problem in general. Another reason to avoid measures based on true negatives, is the prevalence of negative results, for example, gene names will constitute a small percentage of an article, even in an article describing major pathways.

The three basic quantitative measures used to assess performance in an extrinsic or intrinsic evaluation are calculated as follows: Recall is the percentage of results that should have been obtained according to the test set that actually were obtained by the system:

$$\textbf{Recall} = \text{Number of correct results obtained by system}\,(\text{TP}) \,/\, \text{Number of results specified in gold standard}\,(\text{TP}+\text{FN})$$

Precision is the percent of results that the system obtained that were actually correct according to the test set:

$$\textbf{Precision} = \text{Number of correct results obtained by system}\,(\text{TP}) \,/\, \text{Total number of results obtained by system}\,(\text{TP}+\text{FP})$$

There is usually a tradeoff between recall and precision, with higher precision usually being attainable at the expense of recall, and

vice versa. The F-measure is a combination of both measures and can be used to weigh the importance of one measure over the other by giving more weight to one. If both measures are equally important, the F measure is the harmonic mean of the two measures. When reporting the results, an error analysis provides insights into ways to improve a system. This process involves determining reasons for errors in recall and in precision. In an extrinsic evaluation, some errors can be due to the NLP system and other errors can be due to the subsequent application component. Some NLP errors in recall (i.e. false negatives) can be due to failure of the NLP system to tokenize the text correctly, to recognize a word, to detect a relevant pattern, or to interpret the meaning of a word or a structure correctly. Some errors in precision can be due to errors in interpreting the meaning of a word or structure or to loss of important information. Errors caused by the application component can be due to failure to access the extracted information properly or failure of the reasoning component.

Understanding the errors is the first step in bringing the NLP applications closer to being incorporated in a wider range of biomedical and clinical text processing tasks.

8.7 Recapitulation and Future Directions

This chapter targets both the students and researchers looking for a broad introduction to health NLP prior to delving into this active field of research, and the informatics practitioners looking to use NLP for specific tasks or types of text. The chapter introduced NLP applications and emphasized the critical role that the context in which these applications are deployed plays when developing NLP solutions. It presented the basic computational tasks involved in most NLP applications and the different linguistic knowledge resources and types of linguistic representations that can enable and facilitate these basic NLP tasks. The chapter listed the practical considerations for users of NLP technology and the research considerations for moving the field forward.

Although NLP continues to advance towards practical applications and more NLP methods are used in large-scale real-life health information applications, more needs to be done to make NLP use in biomedical and clinical applications a routine widespread reality. Some of the applications described in this chapter, are already used in practice, e.g., named entity recognition and text labeling are used to support MEDLINE indexers. Some research approaches are already outperforming humans on research datasets that approximate real-life tasks, for example, on reading comprehension tests (SQUAD). This does not mean, however, that NLP in general and health NLP are solved. In addition to improvements in the existing applications, new areas are emerging, and some of the well-known impediments still need to be addressed. The impediments include the data access challenges, which are partially addressed by synthetic data and transfer learning; the lack of interoperability and standards, particularly in the evaluation of tools included in the clinical workflows. The emerging areas and areas of active research include, but are not limited to: multi-modal data integration, interpretability of machine learning results, understanding machine learning models bias, and distributed large-scale computational models.

Suggested Reading

NLP is a very active field of research in the open domain. Many of the applications and techniques described in this chapter are investigated in other domains. For a review of NLP methods in the general domain, we refer the reader to the following textbooks:

Jurafsky, D., & Martin, J. H. (2019). *Speech and language processing. An introduction to natural language processing, computational linguistics and speech recognition*. Upper Saddle River: Prentice Hall. See a draft of the 3-rd edition at https://web.stanford.edu/~jurafsky/slp3/.

Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

This chapter provides a comprehensive and concise overview of health NLP. For additional details and examples see

Cohen, K. B., & Demner-Fushman, D. (2014). *Biomedical natural language processing*. Amsterdam: John Benjamins Publishing.

Questions for Discussion

1. Develop a regular expression to regularize the tokens in lines four to nine of the following cardiac catheterization report (Complications through Heart Rate):

> Procedures performed: Right Heart Catheterization Pericardiocentesis
> Complications: None
> Medications given during procedure: None
> Hemodynamic data
> Height (cm): 180
> Weight (kg): 74.0
> Body surface area (sq. m): 1.93
> Heart rate: 102
> Pressure (mmHg)
> Sys Dias Mean Sat
> RA 14 13 8
> RV 36 9 12
> PA 44 23 33 62% PCW253021
> Hemoglobin (g/dL):
> Conclusions: Post Operative Cardiac Transplant Abnormal Hemodynamics Pericardial Effusion
> Successful Pericardiocentesis
> General Comments:
> 1600 cc of serosanguinous fluid were drained from the pericardial sac with improvement in hemodynamics.

2. Create a lexicon for the last seven lines of the cardiac catheterization report above (Conclusions through the last sentence). For each word, determine all the parts of speech that apply. Which words have more than one part of speech? Choose eight clinically relevant words in that section of the report, and suggest appropriate semantic categories for them that would be consistent with the SNOMED-CT terminology and with the UMLS semantic network.
3. Draw a parse tree for the last sentence of cardiac catheterization report above.
4. Draw parse trees for the following sentences: no increase in temperature; low grade fever; marked improvement in pain; not breathing. (Hint: some lexemes have more than one word.)
5. Identify all the referential expressions in the text below and determine the correct referent for each. Assume that the compute attempts to identify referents by finding the most recent noun phrase. How well does this resolution rule work? Suggest a more effective rule.

> The patient went to receive the AV fistula on December 4. However, he refuses transfusion. In the operating room it was determined upon initial incision that there was too much edema to successfully complete the operation and the incision was closed with staples. It was well tolerated by the patient.

6. In the two following scenarios, an off-the-shelf NLP system that identifies terms and normalizes them against UMLS concepts is applied to a large corpus of texts. In the first scenario, the corpus consists of patient notes. Looking at the frequency of different concepts, you notice that there is a large number of patients with the concept C0019682 (HIV) present, much larger than the regular incidence of HIV in the population reported in the literature. In the second scenario, the corpus consists of full-text biology articles. Looking at the frequency of different concepts, you notice that the failed axon connection (fax) gene is one of the most frequently mentioned genes in your corpus. Describe how you would check the validity of these results. For both cases, discuss what can explain the high frequency counts.

7. The following is an excerpt from a de-identified clinical discharge summary (as shown in Uzuner et al. (2008)).

> HISTORY OF PRESENT ILLNESS: The patient is a 77-year-old woman with long standing hypertension who presented as a Walk-in to me at the [REMOVED] Health Center on [REMOVED]. Recently had been started q.o.d. on Clonidine since [REMOVED] to taper off of the drug. Was told to start Zestril 20 mg. q.d. again. The patient was sent to the [REMOVED] Unit for direct admission for cardioversion and anticoagulation, with the Cardiologist, Dr. [REMOVED] to follow. SOCIAL HISTORY: Lives alone, has one daughter living in [REMOVED]. Is a non-smoker, and does not drink alcohol. HOSPITAL COURSE AND TREATMENT: During admission, the patient was seen by Cardiology, Dr. [REMOVED], was started on IV Heparin, Sotalol 40 mg PO b.i.d. increased to 80 mg b.i.d., and had an echocardiogram. By [REMOVED] the patient had better rate control and blood pressure control but remained in atrial fibrillation. On [REMOVED], the patient was felt to be medically stable.

   (a) Annotate all elliptical constructions and anaphoric references.
   (b) Develop an algorithm to identify section headings.

8. The following is the abstract of the article entitled "Tissue-specific distributions of alternatively spliced human PECAM-1 isoforms" by Wang et al. (as cited by Agarwal and Yu (2009)). Annotate each sentence according to the four categories: Introduction, Methods, Results, and Discussion.

> PECAM-1 plays an important role in endothelial cell-cell and cell-matrix interactions, which are essential during vasculogenesis and/or angiogenesis. Here, we examined expression of PECAM-1 mRNA in vascular beds of various human tissues and compared it with expression of PECAM-1 in human endothelial and hematopoietic cells. A short exposure of the blot probed with GAPDH is shown, because poly(A)+ RNA from the cell lines gives a strong signal within several hours compared with the total RNA from human tissue. Therefore, total RNA from various tissues required a much longer exposure to reveal GAPDH mRNA. Human tissue and cell lines expressed multiple RNA bands for PECAM-1, which may represent alter- natively spliced PECAM-1 isoforms, the identity of which required further analysis.

9. Develop a regular expression that is capable of differentiating in-text parenthetical citations of the form "(Author, Year)" from other parentheticals.
10. Manually or programmatically, repeat Swansonian literature-based discovery ((Swanson 1986), see some implementation details in (Ganiz et al. 2005)):

    (a) Pick a topic of interest (Raynaud's Disease)
    (b) Search to find literature C = {Raynaud's}
    (c) Guess that B (e.g., blood factors) should be studied in relation to Raynaud's
    (d) Search literature C, = C ∩ blood
    (e) Notice two common descriptors: blood viscosity, red blood cell rigidity
    (f) Search literature A = {blood viscosity} ∪ {red blood cell rigidity}
    (g) Notice the term "Fish Oil"
    (h) Search literature A = {Fish Oil}
    (i) Show {Fish Oil} ∩ {Raynaud's} = ∅
    (j) Show plausible connection between Raynaud's and Fish Oil

References

Agarwal, S., & Yu, H. (2009). Automatically classifying sentences in full-text biomedical articles into introduction, methods, results and discussion. *Bioinformatics, 25*(23), 3174–3180.

Aronson, A. R., & Lang, F. M. (2010). An overview of MetaMap: Historical perspective and recent advances. *Journal of the American Medical Informatics Association, 17*(3), 229–236.

Barbarino, J. M., Whirl-Carrillo, M., Altman, R. B., & Klein, T. E. (2018). PharmGKB: A worldwide resource for pharmacogenomic information. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine, 10*(4), e1417.

Ben Abacha, A., Shivade, C., & Demner-Fushman, D. (2019). Overview of the MEDIQA 2019 shared task on textual inference, question entailment and question answering. *Proceedings of the BioNLP 2019 Workshop*.

Bird, S., Klein, E., & Loper, E.. https://www.nltk.org/book/.

Björne, J., Ginter, F., Pyysalo, S., Tsujii, J. I., & Salakoski, T. (2010). Complex event extraction at PubMed scale. *Bioinformatics, 26*(12), i382–i390.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching word vectors with subword information. arXiv preprint arXiv:1607.01759.

Bunt, H. (2017). Computational pragmatics. In Y. Huang (Ed.), *The Oxford handbook of pragmatics* (pp. 326–345). Oxford: Oxford University Press.

Cao, Y., Liu, F., Simpson, P., Antieau, L., Bennett, A., Cimino, J. J., et al. (2011). AskHERMES: An online question answering system for complex clinical questions. *Journal of Biomedical Informatics, 44*(2), 277–288.

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics, 34*(5), 301–310.

Christensen, L., Haug, P., & Fiszman, P. (2002). MPLUS: A probabilistic medical language understanding system. *Proceedings of the ACL BioNLP*, 29–36.

Claveau, V., & L'Homme, M.-C. (2005). Structuring terminology using analogy-based machine learning. *Proceedings of the 7th International Conference on Terminology and Knowledge Engineering, TKE*.

Cohen, P. R. (2015). DARPA's Big Mechanism program. *Physical Biology, 12*(4), 045008. IOP Publishing Ltd. https://iopscience.iop.org/article/10.1088/1478-3975/12/4/045008/meta.

Conway, M., Keyhani, S., Christensen, L., South, B. R., Vali, M., Walter, L. C., et al. (2019). Moonstone: A novel natural language processing system for inferring social risk from clinical narratives. *Journal of Biomedical Semantics, 10*(1), 6. https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-019-0198-0.

De Choudhury, M. D., Counts, S., & Horvitz, E.. (2013). Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference (WebSci '13)*. Association for Computing Machinery, New York, NY, USA, 47–56. https://dl.acm.org/doi/abs/10.1145/2464464.2464480.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391–407.

Deléger, L., Merkel, M., & Zweigenbaum, P. (2009a). Translating medical terminologies through word alignment in parallel text corpora. *Journal of Biomedical Informatics, 42*(4), 692–701.

Deléger, L., Namer, F., & Zweigenbaum, P. (2009b). Morphosemantic parsing of medical compound words: Transferring a French analyzer to English. *International Journal of Medical Informatics, 78*, S48–S55.

Demner-Fushman, D., & Lin, J. (2007). Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics, 33*(1), 63–103.

Demner-Fushman, D., Chapman, W. W., & McDonald, C. J. (2009). What can natural language processing do for clinical decision support? *Journal of Biomedical Informatics, 42*(5), 760–772.

Demner-Fushman, D., Rogers, W. J., & Aronson, A. R. (2017). MetaMap Lite: An evaluation of a new Java implementation of MetaMap. *Journal of the American Medical Informatics Association, 24*(4), 841–844.

Demner-Fushman, D., Shooshan, S. E., Rodriguez, L., Aronson, A. R., Lang, F., Rogers, W., et al. (2018). A dataset of 200 structured product labels annotated for adverse drug reactions. *Scientific Data, 5*, 180001.

Demner-Fushman, D., Mrabet, Y., & Ben Abacha, A. (2020). Consumer health information and question answering: Helping consumers find answers to their health-related information needs. *Journal of the American Medical Informatics Association, 27*(2), 194–201.

Denny, J. C., Miller, R. A., Johnson, K. B., & Spickard, A. III. (2008). Development and evaluation of a clinical note section header terminology. In *AMIA annual symposium proceedings 2008* (Vol. 2008, pp. 156–160). Bethesda: American Medical Informatics Association.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018, Oct 11). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dublin, S., Baldwin, E., Walker, R. L., Christensen, L. M., Haug, P. J., Jackson, M. L., et al. (2013).

Natural language processing to identify pneumonia from radiology reports. *Pharmacoepidemiology and Drug Safety, 8*(22), 834–841.

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., et al. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science, 26*(2), 159–169. https://doi.org/10.1177/0956797614557867.

Elhadad, N. (2006). Comprehending technical texts: Predicting and defining unfamiliar terms. *Proceedings AMIA Symposium*, 239–243.

Elhadad, N., Kan, M. Y., Klavans, J. L., & McKeown, K. R. (2005). Customization in a unified framework for summarizing medical literature. *Artificial Intelligence in Medicine, 33*(2), 179–198.

Evans, D. A., Cimino, J. J., Hersh, J. J., Huff, S. M., & Bell, D. S. (1994). Toward a medical-concept representation language. The Canon Group. *Journal of the American Medical Informatics Association: JAMIA, 1*(3), 207–217.

Eysenbach, G., & Till, J. E. (2001). Ethical issues in qualitative research on internet communities. *BMJ, 323*(7321), 1103–1105.

Firth, J. R. (1957). A synopsis of linguistic theory. In *Studies in linguistic analysis*. Oxford: Blackwell.

Friedman, C. (2000). A broad-coverage natural language processing system. *American Medical Informatics Association Annual Symposium Proceedings, 2000*, 270–274.

Friedman, C., Alderson, P. O., Austin, J., Cimino, J. J., & Johnson, S. B. (1994). A general natural language text processor for clinical radiology. *Journal of the American Medical Informatics Association: JAMIA, 1*(2), 161–174.

Friedman, C., Shagina, L., Lussier, Y., & Hripcsak, G. (2004). Automated encoding of clinical documents based on natural language processing. *Journal of the American Medical Informatics Association, 11*(5), 392–402.

Ganiz, M. C., Pottenger, W. M., & Janneck, C. D. (2005). Recent advances in literature based discovery. *Journal of the American Society for Information Science and Technology: JASIST* (Submitted).

Ghassemi, M., Naumann, T., Doshi-Velez, F., Brimmer, N., Joshi, R., Rumshisky, A., & Szolovits, P. (2014). Unfolding physiological state: Mortality modelling in intensive care units. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 75–84). New York: ACM.

Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2013). Use of sentiment analysis for capturing patient experience from free-text comments posted online. *Journal of Medical Internet Research, 15*(11), e239.

Grishman, R., Sager, N., Raze, C., & Bookchin, B. (1973). The linguistic string parser. *Proceedings of the National Computer Conference, 42*, 427–434.

Grosz, B., Joshi, A., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics, 2*(21), 203–225.

Habibi, M., Weber, L., Neves, M., Wiegandt, D. L., & Leser, U. (2017). Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics, 33*(14), i37–i48.

Hahn, U., Romacker, M., & Schulz, S. (1999). Discourse structures in medical reports – watch out! The generation of referentially coherent and valid text knowledge bases in the MEDSYNDIKATE system. *International Journal of Medical Informatics, 53*(1), 1–28.

Hakenberg, J., Voronov, D., Nguyên, V. H., Liang, S., Anwar, S., Lumpkin, B., et al. (2012). A SNPshot of PubMed to associate genetic variants with drugs, diseases, and adverse reactions. *Journal of Biomedical Informatics, 45*(5), 842–850.

Harris, Z. (1991). *A theory of language and information – a mathematical approach*. New York: Oxford University Press.

Harris, Z., Gottfried, M., Ryckman, T., Mattick, P., Daladier, A., Harris, T., & Harris, S. (1989). *The form of information in science – analysis of an immunology sublanguage*. Dordrecht: Kluwer Academic.

Haug, P. J., Ranum, D. L., & Frederick, P. R. (1990). Computerized extraction of coded findings from free-text radiology reports. *Radiology, 174*, 543–548.

Haug, P., Koehler, S., Lau, L. M., Wang, P., Rocha, R., & Huff, S. (1994). A natural language understanding system combining syntactic and semantic techniques. *Proceedings of the Annual Symposium on Computer Applications in Medical Care*, 247–251.

Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the Twenty-Second Annual International SIGIR Conference*.

Hripcsak, G., Friedman, C., Alderson, P. O., DuMouchel, W., Johnson, S. B., & Clayton, P. D. (1995). Unlocking data from narrative reports: A study of natural language processing. *Annals of Internal Medicine, 122*(9), 681–688.

Hripcsak, G., Soulakis, N. D., Li, L., Morrison, F. P., Lai, A. M., Friedman, C., et al. (2009). Syndromic surveillance using ambulatory electronic health records. *Journal of the American Medical Informatics Association, 16*(3), 354–361.

Hüske-Kraus, D. (2003). Text generation in clinical medicine – a review. *Methods of Information in Medicine, 42*(1), 51–60.

Institute of Medicine (US) Committee for Evaluating Medical Technologies in Clinical Use. (1985). *Assessing medical technologies*. Washington, DC: National Academies Press.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R* (Springer texts in statistics). New York: Springer Science+Business Media.

Keselman, A., Tse, T., Crowell, J., Browne, A., Ngo, L., & Zeng, Q. (2007). Assessing consumer health vocabulary familiarity: An exploratory study. *Journal of Medical Internet Research, 9*(1), e5.

Kilicoglu, H., & Demner-Fushman, D. (2016). BioSCoRes: A smorgasbord architecture for coreference resolution in biomedical text. *PLoS One, 11*(3), e0148538.

Kilicoglu, H., Shin, D., Fiszman, M., Rosemblat, G., & Rindflesch, T. C. (2012). SemMedDB: A PubMed-scale repository of biomedical semantic predications. *Bioinformatics, 28*(23), 3158–3160.

Krallinger, M., Morgan, A., Smith, L., Leitner, F., Tanabe, L., Wilbur, J., et al. (2008). Evaluation of text-mining systems for biology: Overview of the Second BioCreative community challenge. *Genome Biology, 9*(2), S1.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2019, Jan 25). BioBERT: Pre-trained biomedical language representation model for biomedical text mining. arXiv preprint arXiv:1901.08746.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., et al. (2019, Oct 29). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.

Lindberg, D. A. B., Humphreys, B. L., & McCray, A. T. (1993a). The unified medical language system. *Methods of Information in Medicine, 32*, 281–291.

Lindberg, D. A., Siegel, E. R., Rapp, B. A., Wallingford, K. T., & Wilson, S. R. (1993b). Use of MEDLINE by physicians for clinical problem solving. *Journal of the American Medical Association, 269*(24), 3124–3129.

Lynch, J. A., Kelley, M. J., Lee, K. M., Hung, A., Li, Y., Hintze, B. J., et al. (2019). An NLP tool to identify molecular diagnostic testing in veterans with stage IV NSCLC. *Journal of Clinical Oncology, 37*(27_suppl), 318. https://ascopubs.org/doi/abs/10.1200/JCO.2019.37.27_suppl.318.

Mane, V. L., Panicker, S. S., & Patil, V. B. (2015, Jan 8). Summarization and sentiment analysis from user health posts. In *2015 International Conference on Pervasive Computing (ICPC)* (pp. 1–4). IEEE.

Maroto, M., Reshef, R., Munsterberg, A. E., Koester, S., Goulding, M., & Lassar, A. B. (1997). Ectopic Pax-3 activates MyoD and Myf-5 expression in embryonic mesoderm and neural tissue. *Cell, 89*, 139–148.

Meystre, S. M., Friedlin, F. J., South, B. R., Shen, S., & Samore, M. H. (2010). Automatic de-identification of textual documents in the electronic health record: A review of recent research. *BMC Medical Research Methodology, 10*(1), 70.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013, Jan 16). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Mork, J., Aronson, A., & Demner-Fushman, D. (2017). 12 years on – is the NLM medical text indexer still useful and relevant? *Journal of Biomedical Semantics, 8*(1), 8.

OHDSIPNAS, Lancet – Noemie?

openNLP. http://opennlp.apache.org/index.html.

Peng, Y., Rios, A., Kavuluru, R., & Lu, Z. (2018). Extracting chemical–protein relations with ensembles of SVM and deep learning models. *Database, 2018*, bay073.

Pennington, J., Socher, R., & Manning, C. D. (2014, Oct). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Peters, M. P., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv:1802.05365.

Pivovarov, R., & Elhadad, N. (2015, Sep). Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association, 22*(5), 938–947. https://doi.org/10.1093/jamia/ocv032.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog, 1*(8), 9.

Resnik, P., Niv, M., Nossal, M., Schnitzer, G., Stoner, J., Kapit, A., & Toren, R. (2006). Using intrinsic and extrinsic metrics to evaluate accuracy and facilitation in computer-assisted coding. *Perspectives in Health Information Management Computer Assisted Coding Conference Proceedings*, Fall, 6 Sept 2006.

Roberts, K., & Patra, B. G. (2018). A semantic parsing method for mapping clinical questions to logical forms. *American Medical Informatics Association Annual Symposium Proceedings, 2017*, 1478–1487.

Roberts, K., Demner-Fushman, D., & Tonning, J. M. (2017). Overview of the TAC 2017 adverse reaction extraction from drug labels track. *Proceedings of the 2017 Text Analysis Conference*, 13 Nov 2017.

Ruch, P., Boyer, C., Chichester, C., Tbahriti, I., Geissbühler, A., Fabry, P., et al. (2007). Using argumentation to extract key sentences from biomedical abstracts. *International Journal of Medical Informatics, 76*(2–3), 195–200.

Ruder, S. (2019). *Neural transfer learning for natural language processing* (Diss). NUI Galway.

Sager, N. (1972). Syntactic formatting of science information. *Proceedings of the AFIPS* (pp. 791–800). In R. Kittredge & J. Lehrberger (Eds.). (1982). *Reprinted in Sublanguage: Studies of language in restricted semantic domains* (pp. 9–26). Berlin: Walter de Gruyter.

Sager, N. (1978). Natural language information formatting: The automatic conversion of texts to a structured data base. In M. C. Yovits (Ed.), *Advances in computers* (Vol. 17, pp. 89–162). New York: Academic Press.

Sager, N. (1981). *Natural language information processing: A computer grammer of English and its applications*. Reading: Addison-Wesley.

Sager, N., Friedman, C., & Lyman, M. (1987). *Medical language processing – computer management of narrative data*. Reading: Addison-Wesley.

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, compo-

nent evaluation and applications. *Journal of the American Medical Informatics Association, 17*(5), 507–513.

Soysal, E., Wang, J., Jiang, M., Wu, Y., Pakhomov, S., Liu, H., & Xu, H. (2017). CLAMP – a toolkit for efficiently building customized clinical natural language processing pipelines. *Journal of the American Medical Informatics Association, 25*(3), 331–336.

SQUAD. Stanford Question Answering Dataset Leaderboard https://rajpurkar.github.io/SQuAD-explorer/.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., & Tsujii, J. I. (2012). BRAT: A web-based tool for NLP-assisted text annotation. In *Proceedings of the demonstrations at the 13th conference of the European chapter of the Association for Computational Linguistics 2012 Apr 23* (pp. 102–107). Stroudsburg: Association for Computational Linguistics.

Swanson, D. R. (1986). Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine, 30*, 7–18.

Taylor, A., Marcus, M., & Santorini, B. (2003). The Penn treebank: An overview. In *Treebanks* (pp. 5–22). Dordrecht: Springer.

Turian, J., Ratinov, L., & Bengio, Y. (2010). Word representations: A simple and general method for semi-supervised learning. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*.

UIMA. https://uima.apache.org/.

Uzuner, O., Goldstein, I., Luo, Y., & Kohane, I. (2008). Identifying patient smoking status from medical discharge records. *Journal of the American Medical Informatics Association: JAMIA, 15*(1), 14–24.

Uzuner, Ö., South, B. R., Shen, S., & DuVall, S. L. (2011). 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association, 18*(5), 552–556.

Wu, S., Roberts, K., Datta, S., Du, J., Ji, Z., Si, Y., et al. (2019). Deep learning in clinical natural language processing: A methodical review. *Journal of the American Medical Informatics Association, 27*, 457–470. https://doi.org/10.1093/jamia/ocz200.

Ye, Y., Tsui, F. R., Wagner, M., Espino, J. U., & Li, Q. (2014). Influenza detection from emergency department reports using natural language processing and Bayesian network classifiers. *Journal of the American Medical Informatics Association, 5*(21), 815–823.

Zhang, H., Fiszman, M., Shin, D., Miller, C. M., Rosemblat, G., & Rindflesch, T. C. (2011). Degree centrality for semantic abstraction summarization of therapeutic studies. *Journal of Biomedical Informatics, 44*(5), 830–838.

Zunic, A., Corcoran, P., & Spasic, I. (2020). Sentiment analysis in health and well-being: Systematic review. *JMIR Medical Informatics, 8*(1), e16023. https://doi.org/10.2196/16023.

Bioinformatics

Sean D. Mooney, Jessica D. Tenenbaum, and Russ B. Altman

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_9

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- Why are sequence, structure, and biological pathway information relevant to medicine?
- Where on the Internet should you look for a DNA sequence, a protein sequence, or a protein structure?
- What are two problems encountered in analyzing biological sequence, structure, and function?
- How has the age of genomics changed the landscape of bioinformatics?
- What are two computational challenges in bioinformatics for the future?

9.1 The Problem of Handling Biological Information

Bioinformatics is the study of how information is represented and analyzed in biological systems, especially information derived at the molecular level. Whereas clinical informatics deals with the management of information related to the delivery of health care, bioinformatics focuses on the management of information related to the underlying basic biological sciences. As such, the two disciplines are closely related—more so than generally appreciated (see ▶ Chap. 1). Bioinformatics and clinical informatics share a concentration on systems that are inherently uncertain, difficult to measure, and the result of complicated interactions among multiple complex components. Both deal with living systems that generally lack straight edges and right angles. Although **reductionist approaches** to studying these systems can provide valuable lessons, it is often necessary to analyze those systems using **integrative models** that are not based solely on first principles. Nonetheless, the two disciplines approach the patient from opposite directions. Whereas applications within clinical informatics usually are concerned with the social systems of medicine, the cognitive processes of medicine, and the technologies required to understand human physiology, bioinformatics is concerned with understanding how basic biological systems conspire to create molecules, organelles, living cells, organs, and entire organisms. Remarkably, however, the two disciplines share significant methodological elements, so an understanding of the issues in bioinformatics can be valuable for the student of clinical informatics and vice versa.

The discipline of bioinformatics continues to be in a period of rapid growth, because the needs for information storage, retrieval, and analysis in biology—particularly in molecular biology and **genomics**—have increased dramatically over the past two decades. History has shown that scientific developments within the basic sciences tend to have a delayed effect on clinical care and there is typically a lag of a decade before the influence of basic research on clinical medicine is realized. It cannot be understated the impact that genomics and bioinformatic approaches are having in the clinic and the point of care. Indeed, chapters focusing on "Translational Bioinformatics" and "Precision Medicine and Informatics" (▶ Chaps. 28 and 30) describe how these foundational advances are leading toward impacts on human health and improved approaches to clinical care.

9.1.1 Many Sources of Biological Data

There are many sources of information that are revolutionizing our understanding of human biology and that are creating significant challenges for computational processing. New technologies are enabling the miniaturization of laboratory experiments, increased automation of experiments and through advanced computer processing, and the interpretation of data quickly. These technologies are producing data at a staggering rate. The data produced can interrogate different views into the **Central Dogma of Biology**, the **metabolome**, the **metagenome** and ancillary molecular processes.

The most dominant new type of information is the **sequence information** produced by genetic studies. This was enabled by the **Human**

Genome Project, an international undertaking intended to determine the complete sequence of human DNA as it is encoded in each of the 23 human chromosomes. The first draft of the sequence was published in 2001 (Lander et al. 2001) and a final version was announced in 2003 coincident with the 50th anniversary of the solving of the Watson and Crick structure of the DNA double helix. The sequence continues to be revised and refined and now the sequence the genomes of many different individuals have been realized. Initially, the 1000 genomes consortium provided >1000 genomes of healthy individuals (1000 Genomes Consortium, 2010), and now datasets exist with >100,000 genomes of individuals with a variety of conditions.[1] Essentially, the entire set of genetically driven events from conception through embryonic development, childhood, adulthood, and aging are encoded by the DNA blueprints within most human cells. Given a complete knowledge of these DNA sequences, we are in a position to understand these processes at a fundamental level and to consider the possible use of DNA sequences for diagnosing and treating disease. This has led to the application of bioinformatics (and other foundational domains) as Translational Bioinformatics and Precision Medicine Informatics (► Chaps. 28 and 30).

Additionally, large-scale experimental methodologies are used to collect data on thousands or millions or more molecules simultaneously. Scientists apply these methodologies longitudinally over time and across a wide variety of organisms or within an organism to observe the development of various physiological phenomena. Technologies give us the ability to follow the production and degradation of molecules, such as the expression (transcription) of large numbers of genes simultaneously, the presence of proteins or metabolites in a biosample, or the populations of microorganisms in a sample.

The first high throughput experiments measured the expression of genes on **gene expression microarrays** (Lashkari et al. 1997). This enabled the study of the expression of large numbers of genes with one another (Bai and Elledge 1997) and to study multiple variations on a genome to explore the implications of changes in genome function on human disease. This work has led to the field of **genomics**, the study of the molecular state of a cell, tissue or organism through the state and activity of its genome. With technology advancements, gene expression can now be measured by directly sequencing messenger RNA molecules in a cell and counting the number of copies of that RNA molecule that is observed.

While some scientists are studying the human genome, other researchers are studying the functions of the genomes of numerous other biological organisms, including important model organisms (such as mouse, rat, fruit fly and yeast) as well as important human pathogens (such as *Mycobacterium tuberculosis* or *Haemophilus influenzae*). The genomes of these organisms have been determined, and efforts are underway to characterize them. These allow two important types of analysis: the analysis of mechanisms of pathogenicity and the analysis of animal models for human disease. In both cases, the functions encoded by genomes can be studied, classified, and categorized, allowing us to decipher how genomes affect human health and disease.

These ambitious scientific projects are not only proceeding at a furious pace, but also are often accompanied by another approach to biology, which produces another source of biomedical information: **proteomics**, the study of the protein gene products of the genome—the proteome. Proteomics enables researchers to discover the state (quantity and configuration) of proteins within an organism. These protein states can be correlated with different physiological conditions, including disease states. Some of these protein states can be used as identifying markers of human disease. Similar approaches are being applied to understanding the diversity, concentration levels and functions of non-DNA, RNA or protein molecules such as metabolites through the study of the small molecules in the **metabolome**.

Using these technologies together, we can now study the **epigenome**, the non-genetic

1 ► https://www.nhlbiwgs.org/ (accessed December 1, 2018).

effects that influence genome function. These include molecules that directly alter the structure of DNA but not its sequence (such as **DNA methylation**) or proteins that bind to DNA and affect how that DNA expresses genes. Epigenomics gives us a more complete picture of how biology functions and what its implications are for human health.

All these technologies, along with the genome-sequencing projects, are conspiring to produce a volume of biological information that at once contains secrets to age-old questions about health and disease and threatens to overwhelm our current capabilities of data analysis. Thus, bioinformatics is becoming critical for medicine in the twenty-first century.

9.1.2 Implications for Clinical Informatics

The effects of this new biological information on clinical medicine and clinical informatics are still evolving. It is already clear, however, that some major changes to medicine will have to be accommodated. These efforts have emerged as important areas of biomedical informatics that have become their own domains, Translational Bioinformatics (▶ Chap. 26) and Precision Medicine and Informatics (▶ Chap. 28) and use of biotechnology data is now common in Clinical Research Informatics (▶ Chap. 27).

1. *Genetic information in the medical record*. With the first set of human genomes now available and prices for gene sequencing rapidly decreasing, it is now cost-effective to consider sequencing every patient genome or at least genotyping key sections of the genomes and integrating that with the medical record.
2. *New diagnostic and prognostic information sources*. One of the main contributions of the genome-sequencing projects (and of the associated biological innovations) is that we are likely to have unprecedented access to new diagnostic and prognostic tools. Diagnostically, the genetic markers from a patient with an autoimmune disease, or of an infectious pathogen within a patient, will be highly specific and sensitive indicators of the subtype of disease and of that subtype's probable responsiveness to different therapeutic agents. Several **genotype**-based databases have been developed to identify markers that are associated with specific **phenotypes** and identify how genotype affects a patient's response to therapeutics. ClinVar[2] and The Human Gene Mutation Database (HGMD)[3] both annotate mutations with disease phenotype. This resource has become invaluable for genetic counselors, basic researchers, and clinicians. Additionally, the Pharmacogenomics Knowledge Base (PharmGKB) collects genetic information that is known to affect a patient's response to a drug (more on PharmGKB is described in Translational Bioinformatics, ▶ Chap. 26).[4]
3. *Ethical considerations*. One of the critical questions facing the genome-sequencing and other related projects is "Can genetic or other molecular information be misused?" The answer is certainly yes. With knowledge of a complete genome for an individual, it may be possible in the future to predict the types of disease for which that individual is at risk years before the disease actually develops. If this information fell into the hands of unscrupulous employers or insurance companies, the individual might be denied employment or coverage due to the likelihood of future disease, however distant. There is even debate about whether such information should be released to a patient even if it could be kept confidential. Should a patient be informed that he or she is likely to get a disease for which there is no treatment? What about that patient's relatives, who share genetic information with the patient? This is a matter of intense debate, and such questions have significant implications for what information is collected and for how and to whom that information

2 ▶ https://www.ncbi.nlm.nih.gov/clinvar/ (accessed November 1, 2018).
3 ▶ http://www.hgmd.org/ (accessed November 1, 2018).
4 ▶ http://www.pharmgkb.org/ (accessed November 1, 2018).

is disclosed (Durfy 1993). Passage of the Genetic Information Nondiscrimination Act in 2008 set initial federal guidelines on use of genetic information.[5] Additionally, the Personal Genome Project (PGP) has been working to define **open consent models** for releasing genetic information.[6] The Clinical Sequencing and Exploratory Research Consortium (CSER) has been tackling the difficult issues in translation of genomic data to the clinic broadly.[7]

9.2 The Rise of Bioinformatics

A brief review of the biological basis of medicine will bring into focus the magnitude of the revolution in molecular biology and the tasks that are created for the discipline of bioinformatics. The genetic material that we inherit from our parents, that we use for the structures and processes of life, and that we pass to our children is contained in a sequence of chemicals known as **deoxyribonucleic acid (DNA)**.[8] The total collection of DNA for a single person or organism is referred to as the **genome**. DNA is a long polymer chemical made of four basic subunits. The sequence in which these subunits occur in the polymer distinguishes one DNA molecule from another and directs a cell's production of proteins and all other basic cellular processes. **Genes** are discreet units encoded in DNA and they are transcribed into **ribonucleic acid (RNA)**, which has a composition very similar to DNA. Genes are transcribed into messenger RNA (mRNA) and a majority of mRNA sequences are translated by complex macromolecular machines, called ribosomes, into protein. Not all RNAs are messengers for the translation of proteins. Ribosomal RNA, for example, is used in the construction of the ribosome, the huge molecular engine that translates mRNA sequences into protein sequences. Additionally, mRNAs can be modified through alternative splicing, degradation, and formation of secondary structures that influence transcriptions. Once expressed, proteins are frequently modified (e.g. phosphorylated), and these modifications can change the function of the protein. This process of DNA being transcribed to RNA and RNA being translated to protein is commonly referred to as the **Central Dogma of Biology**.

Understanding the basic building blocks of life requires understanding the function of genomic sequences, genes, and proteins. When are genes expressed? Once genes are transcribed and translated into proteins, into what cellular compartment are the proteins directed? How do the proteins function once there? Do the proteins need to be modified in order for them to become active? How are the proteins turned off? Experimentation and bioinformatics have divided the research into several areas, and the largest are: (1) DNA and protein sequence analysis, (2) macromolecular structure–function analysis, (3) gene expression analysis, (4) proteomics, (5) metabolomics, (6) metagenomics, and (5) systems biology.

9.2.1 Roots of Modern Bioinformatics

Practitioners of bioinformatics have come from many backgrounds, including medicine, molecular biology, chemistry, physics, statistics, mathematics, engineering, and computer science. It is difficult to define precisely the ways in which this discipline emerged. There are, however, two main developments that have created opportunities for the use of information technologies in biology. The first is the progress in our understanding of how biological molecules are constructed and how they perform their functions. This dates back as far as the 1930s with the invention of electrophoresis, and then in the 1950s with the elucidation of the structure of DNA and the subsequent sequence of discoveries in the relationships

5 ► http://www.genome.gov/10002328 (accessed November 1, 2018).

6 ► http://www.personalgenomes.org/ (accessed November 1, 2018).

7 ► https://cser-consortium.org/ (accessed November 1, 2018).

8 If you are not familiar with the basic terminology of molecular biology and genetics, reference to an introductory textbook in the area would be helpful before you read the rest of this chapter.

among DNA, RNA, and protein structure. The second development has been the parallel increase in the availability of computing power. Starting with mainframe computer applications in the 1950s and moving to modern workstations, and 'the Cloud', there have been hosts of biological problems addressed with computational methods.

9.2.2 The Genomics Explosion

The benefit of the human genome sequence to medicine is both in the short and in the long term. The short-term benefits lie principally in diagnosis; the availability of sequences of normal and variant human genes will allow for the rapid identification of these genes in any patient (e.g., Babior and Matzner 1997). The long-term benefits will include a greater understanding of the proteins produced from the genome: how the proteins interact with drugs; how they malfunction in disease states; and how they participate in the control of development, aging, and responses to disease.

The effects of genomics on biology and medicine cannot be overstated. We now have the ability to measure the activity and function of genes within living cells. Genomics data and experiments have changed the way biologists think about questions fundamental to life. Whereas in the past, reductionist experiments probed the detailed workings of specific genes, we can now assemble those data together to build an accurate understanding of how cells work.

9.3 Biology Is Now Data-Driven

Nearly 30 years ago, the use of computers was proving to be useful to the laboratory researcher. Today, computers are an essential component of modern research. This has led to a change in thinking about the role of computers in biology. Before, they were optional tools that could help provide insight to experienced and dedicated enthusiasts. Today, they are required by most investigators, and experimental approaches rely on them as critical elements. This is because advances in research methods such as **genetic sequencing**, **experimental robotics and microfluidics**, **X-ray crystallography**, **nuclear magnetic resonance spectroscopy**, **cryoelectron microscopy**, **proteomic mass spectrometry** and other high throughput experiments have resulted in experiments that generate massive amounts of data. These data pose new problems for basic researchers on how the data are properly stored, analyzed, and disseminated.

The volume of data being produced by genomics projects is staggering. There are now more than 211 million sequences in **GenBank** comprising more than 285 billion digits. Since 2008, sequencing has bested Moore's law (see ► Chap. 1).[9] But these data do not stop with sequence data: PubMed contains over 28 million literature citations, the **Protein Data Bank** (**PDB**) contains three-dimensional structural data for over 45,538 distinct protein structures, and the **Gene Expression Omnibus** (**GEO**) contains over 2.8 million arrayed samples. These data are of incredible importance to biology, and in the following sections we introduce and summarize the importance of sequences, structures, gene expression experiments, systems biology, and their computational components to medicine.

9.3.1 Sequences in Biology

Sequence information (including DNA sequences, RNA sequences, and protein sequences) is critical in biology: DNA, RNA, and protein can be represented as a set of sequences of basic building blocks (bases for DNA and RNA, amino acids for proteins). Computer systems within bioinformatics thus must be able to handle biological sequence information effectively and efficiently. To that end, the bioinformatics community has developed central databases to store sequence information, data models to represent that information and software analysis tools to process sequence data.

9 ► http://www.genome.gov/sequencingcosts/ (accessed November 1, 2018).

9.3.2 Structures in Biology

The sequence information mentioned in ▶ Sect. 9.3.1 is rapidly becoming inexpensive to obtain and easy to store. On the other hand, the **three-dimensional structure information** about the proteins, DNA, and RNA is much more difficult and expensive to obtain, and presents a separate set of analysis challenges. Currently, only about 45,000 distinct three-dimensional structures of biological macromolecules are known.[10] These models are incredibly valuable resources, however, because an understanding of structure often yields detailed insights about biological function. As an example, the structure of the ribosome has been determined for several species and contains more atoms than any other structure to date. This structure, because of its size, took two decades to solve, and presents a formidable challenge for functional annotation (Cech 2000). Yet, the functional information for a single structure is dwarfed by the potential for comparative genomics analysis between the structures from several organisms and from varied forms of the functional complex. Since the ribosome is ubiquitously required for all forms of life these types of comparisons are possible. Thus, a wealth of information comes from relatively few structures. To address the problem of limited structure information, the publicly funded structural genomics initiative aims to identify all of the common structural scaffolds found in nature and to increase the number of known structures considerably. In the end, it is the physical interactions between molecules that determine what happens within a cell; thus the more complete the picture, the better the functional understanding. In particular, understanding the physical properties of therapeutic agents is the key to understanding how agents interact with their targets within the cell (or within an invading organism). These are the key questions for structural biology within bioinformatics:

1. How can we analyze the structures of molecules to learn their associated function? Approaches range from detailed molecular simulations (Levitt 1983) to statistical analyses of the structural features that may be important for function (Wei and Altman 1998).
2. How can we extend the limited structural data by using information in the sequence databases about closely related proteins from different organisms (or within the same organism, but performing a slightly different function)? There are significant unanswered questions about how to extract maximal value from a relatively small set of examples.
3. How should structures be grouped for the purposes of classification? The choices range from purely functional criteria ("these proteins all digest proteins") to purely structural criteria ("these proteins all have a toroidal shape"), with mixed criteria in between. One interesting resource available today is the Structural Classification of Proteins (SCOP),[11] which classifies proteins based on shape and function.

9.3.3 Genome Sequencing Data in Biology

Advances in sequencing technology are pivotal in enabling the practice of genomic medicine. Whereas the first human genome sequence was carried out over approximately 13 years at a cost of $2.7 billion (Davies 2010), whole human genomes can now be sequenced in a matter of days at a cost that is growing ever-closer to the magic, if somewhat arbitrary, $1000 price tag. This amount is commonly seen as the price at which it becomes feasible to sequence a patient in the course of clinical care, justifiable both clinically and financially. In 2004, and again in 2011, the National Human Genome Research Institute (part of the National Institutes of Health) funded a number of efforts specifically aimed

10 For more information see ▶ http://www.rcsb.org/ (accessed November 1, 2018).

11 ▶ http://scop2.mrc-lmb.cam.ac.uk/ (accessed December 1, 2018).

at increasing speed and decreasing the cost of genome scale sequencing.

Traditional sequencing involves a method referred to as Sanger sequencing. This method typically is applied to sequences ranging from 300 to 1000 nucleotides in a non-high throughput manner.[12] In the early to mid 2000s, several technologies were introduced to sequence large amounts of DNA in parallel. These **high throughput sequencing methods** (of which there are many including sequencing by synthesis, single molecule sequencing, combinatorial probe anchor synthesis, and others) typically involve shorter sequences than Sanger based approaches, but can generate gigabases of sequence in short fragments at low cost (<$0.05 per megabase sequenced). These methods are being used for many applications, including identification of genetic variants in clinical studies, characterizing genome function with specific experiments and sequencing novel species genomes. These studies have already discovered the genetic basis of rare genetic disorders by sequencing entire families (Ng et al. 2010), and we have seen a glimpse of the future of genome sequencing for routine health care in the analysis of a single genome of a healthy man (Ashley et al. 2010). As will be described in detail in the Translational Bioinformatics chapter (► Chap. 26), these sequencing approaches have been put to practice clinically. One emergent area of research is **metagenomics**, the study of microorganism ecosystems using DNA sequencing, including the association of human gut flora populations to disease phenotypes in humans (Qin et al. 2010).

9.3.4 Expression Data in Biology

The development of DNA microarrays led to a wealth of data and unprecedented insight into the fundamental biological machine. The traditional premise is relatively simple; tens of thousands of gene sequences derived from genomic data are fixed onto a glass slide or filter. The sequences for each spot are derived from a single gene sequence and the sequences are attached at only one end, creating a forest of sequences in each spot that are all identical. An experiment is performed where two samples (e.g. groups of cells that are grown in different conditions or for comparisons of normal and cancer tissue), one group is a control group and the other is the experimental group. The control group is grown normally, while the experimental group is grown under experimental conditions. For example, a researcher may be trying to understand how a cell compensates for a lack of sugar. The experimental cells will be grown with limited amounts of sugar. As the sugar depletes, some of the cells are removed at specific intervals of time. When the cells are removed, all of the mRNA from the cells is separated from the cells and converted back to DNA, using reverse transcriptase (a special enzyme that can create a DNA copy from an RNA template). This leaves a pool of cDNA molecules (DNA derived from mRNA is called complementary DNA or cDNA) that represent the genes that were expressed (turned on) in that group of cells. In the development of genomics experimentation, these cDNA molecules would be tagged with florescence and hybridized to slides containing single stranded DNA "probes" that are arrayed in a grid. These microarray "chips" can then be analyzed for color differences between grid points that correspond to specific gene regions. Today, with the advent of high throughput sequencing the RNA/cDNA can be sequenced directly to measure expression levels and using DNA barcoding technology and microfluidics, individual cells can be sequenced alone instead of in pooled samples where all cells' contributions to mRNA is in the same analysis. High throughput single cell sequencing is an exciting advancement which adds orders of complexity to the required computational analysis (Shapiro et al. 2013).

Computers become critical for analyzing these data because it is impossible for a researcher to measure and analyze all of the datasets by hand. Currently scientists are using gene expression experiments to study how cells from different organisms compen-

12 ► http://en.wikipedia.org/wiki/DNA_sequencing (accessed November 1, 2018).

sate for environmental changes, how pathogens fight antibiotics, and how cells grow uncontrollably (as is found in cancer). A challenge for biological computing is to develop methods to analyze these data, tools to store these data, and computer systems to collect the data automatically.

9.3.5 Metabolomics Data in Biology

Genomics and proteomics study the function of the genome and the proteome, while **metabolomics** studies the diversity and function of small molecules in a biosample. These include metabolites such as lipids, carbohydrates, metal ions, hormones, signaling molecules, etc. Interest in the metabolome has increased significantly with the development of separation and mass spectrometry technologies that can identify small molecule molecular mass and identities in a high throughput fashion. Bioinformatics is a key component of both the identification of specific molecules by matching mass spectrometry "fingerprints" with a database of known molecules as well as in the analysis the resulting data. For example, researchers have characterized the metabolome of human colorectal cancers and stool and identified disease enriched metabolites as a possible detectable markers of disease or treatment outcomes (Brown et al. 2016).

9.3.6 Epigenetics Data in Biology

Epigenetics consists of heritable changes that are not encoded in the primary DNA sequence. Several types of epigenetic effects can now be studied in the laboratory, and they have been associated to disease and risks of disease (Goldberg et al. 2007). First, the regional structure of chromosomes affects which regions of the genome can be transcribed, i.e. which regions can be *expressed*. Large proteins, called histones, coordinate the structure of chromosomes and their structure and positions are regulated with protein posttranslational modifications to the histones bound to the DNA. These changes have been associated with spontaneous mutations in cancer, complex genetic diseases, and Mendelian inherited genetic diseases. Second, cytosine bases in the DNA can be methylated and this can affect gene expression. DNA methylation patterns can be passed on when DNA is replicated. Like chromosome structure, these modifications have been associated with human disease (Bird 2002).

9.3.7 Systems Biology

Recent advances in high throughput technologies have enabled a new, dynamic approach to studying biology, that of **systems biology**. In contrast to the historically reductionist approach to biology, studying one molecule at a time, systems biology looks at the entirety of a system including dynamic relationships between the different components. With that said, systems biology is still maturing. As an analogy, consider an airplane. Having a "parts list" for a Boeing 747 does not enable us to understand how those parts work together to make the airplane operate. If the airplane breaks, the parts list alone does not tell us how to remedy the situation. Rather, we need to understand how the parts interact, how one affects another, and how perturbations to one part of the system affect the rest of the system. Similarly, systems biology involves understanding not only the "parts list", i.e. the list of all genes, proteins, metabolites, etc., but also the dynamic networks of interactions among these parts. An integrated simulation of an entire bacterial cell has shown the feasibility of accurate computational simulations of cell physiology (Karr et al. 2012).

Current research in **-omics technologies** have both enabled and catalyzed the advancement of systems biology. However, a systems biology approach goes beyond simply performing these high bandwidth methods for the purpose of biological discovery. Rather, systems biology implies a systematic, hypothesis-driven approach based on omic-scale (very large) hypotheses. Once the interactions in a biological network are understood, one can model that network to make predictions

regarding the system's behavior, particularly in light of specific perturbations. Understanding how the system has evolved to work can also help us understand what goes wrong when the system breaks down, and how to intervene in order to restore the system to normal.

9.4 Key Bioinformatics Algorithms

There are a number of common computations that are performed in many contexts within bioinformatics. In general, these computations can be classified as sequence alignment, structure alignment, pattern analysis of sequence/structure, gene expression analysis, and pattern analysis of biochemical function.

9.4.1 Early Work in Sequence and Structure Analysis

As it became clear that the information from DNA and protein sequences would be voluminous and difficult to analyze manually, algorithms began to appear for automating the analysis of sequence information. The first requirement was to have a reliable way to align sequences so that their detailed similarities and distances could be examined directly. Needleman and Wunsch (1970) published an elegant method for using **dynamic programming** techniques to align sequences in time related to the cube of the number of elements in the sequences. Smith and Waterman (1981) published refinements of these algorithms that allowed for searching both the best global alignment of two sequences (aligning all the elements of the two sequences) and the best local alignment (searching for areas in which there are segments of high similarity surrounded by regions of low similarity). A key input for these algorithms is a matrix that encodes the similarity or substitutability of sequence elements: When there is an inexact match between two elements in an alignment of sequences, it specifies how much "partial credit" we should give to the overall alignment based on the similarity of the elements, even though they may not be identical. Looking at a set of evolutionarily related proteins, Dayhoff (1974) published one of the first matrices derived from a detailed analysis of which amino acids (elements) tend to substitute for others.

Within structural biology, the vast computational requirements of the experimental methods (such as X-ray crystallography and nuclear magnetic resonance) for determining the structure of biological molecules drove the development of powerful structural analysis tools. In addition to software for analyzing experimental data, graphical display algorithms allowed biologists to visualize these molecules in great detail and facilitated the manual analysis of structural principles (Langridge 1974; Richardson 1981). At the same time, methods were developed for simulating the forces within these molecules as they rotate and vibrate (Gibson and Scheraga 1967; Karplus and Weaver 1976; Levitt 1983).

The most important development to support the emergence of bioinformatics, however, has been the creation of databases with biological information. In the 1970s, structural biologists, using the techniques of X-ray crystallography, set up the Protein Data Bank (PDB) specifying the Cartesian coordinates of the structures that they elucidated (as well as associated experimental details) and made PDB publicly available. The first release, in 1977, contained 77 structures. The growth of the database is chronicled on the Web: the PDB now has over 75,000 detailed atomic structures and is the primary source of information about the relationship between protein sequence and protein structure.[13] Similarly, as the ability to obtain the sequence of DNA molecules became widespread, the need for a database of these sequences arose. In the mid-1980s, the GENBANK database was formed as a repository of sequence information. Starting with 606 sequences and 680,000 bases in 1982, the GENBANK has grown by much more than 135 million sequences and 125 billion bases.[14] The GENBANK database of DNA sequence information supports the experimental reconstruction of genomes and acts as a focal point

13 See ► http://www.rcsb.org/ (accessed December 1, 2018).

14 ► http://www.ncbi.nlm.nih.gov/genbank/ (accessed December 1, 2018).

for experimental groups. Numerous other databases store the sequences of protein molecules[15] and information about human genetic diseases.[16]

Included among the databases that have accelerated the development of bioinformatics is the Medline database of the biomedical literature and its paper-based companion Index Medicus (see ▶ Chap. 23).[17] Including articles as far back as 1809 and brought online free on the Web in 1997, Medline provides the glue that relates many high-level biomedical concepts to the low-level molecule, disease, and experimental methods. In fact, this "glue" role was the basis for creating the NCBI suite of databases and software and PubMed systems (see ▶ Sect. 9.5) for integrating access to literature references and the associated databases.

9.4.2 Sequence Alignment and Genome Analysis

Perhaps the most basic activity in computational biology is comparing two biological sequences to determine (1) whether they are similar and (2) how to align them. The problem of alignment is not trivial but is based on a simple idea. Sequences that perform a similar function should, in general, be descendants of a common ancestral sequence, with mutations over time. These mutations can be replacements of one amino acid with another, deletions of amino acids, or insertions of amino acids. The goal of **sequence alignment** is to align two sequences so that the evolutionary relationship between the sequences becomes clear. If two sequences are descended from the same ancestor and have not mutated too much, then it is often possible to find corresponding locations in each sequence that play the same role in the evolved proteins. The problem of solving correct biological alignments is difficult because it requires knowledge about the evolution of the molecules that we typically do not have. There are now, however, well-established algorithms for finding the mathematically optimal alignment of two sequences. These algorithms require the two sequences and a scoring system based on (1) exact matches between amino acids that have not mutated in the two sequences and can be aligned perfectly; (2) partial matches between amino acids that have mutated in ways that have preserved their overall biophysical properties; and (3) gaps in the alignment signifying places where one sequence or the other has undergone a deletion or insertion of amino acids. The algorithms for determining optimal sequence alignments are based on a technique in computer science known as **dynamic programming** and are at the heart of many computational biology applications (Gusfield 1997). ◘ Figure 9.1 shows an example of a Smith-Waterman matrix, the first described local alignment algorithm that utilizes a dynamic programming approach. The algorithm works by calculating a similarity matrix between two sequences, then finding optimal paths through the matrix that maximize a similarity score between the two sequences.

Unfortunately, the dynamic programming algorithms are too computationally expensive to apply to large numbers of sequences, so a number of faster, more heuristic methods have been developed. The most popular algorithm is the **Basic Local Alignment Search Tool (BLAST)** (Altschul et al. 1990). BLAST is based on the observation that sections of proteins are often conserved without gaps (so the gaps can be ignored—a critical simplification for speed) and that there are statistical analyses of the occurrence of small subsequences within larger sequences that can be used to prune the search for matching sequences in a large database. These tools work well for both protein and nucleic acid sequences. Other tools have been developed that are better suited for nucleic acid sequence assembly and mapping of short read high throughput sequencing data including BLAT (Kent 2003), SOAP (Li et al. 2008), and others.

Protein 3D structures can be aligned, visualized and compared in a similar way to linear protein sequences (◘ Fig. 9.2). Tools such

15 ▶ http://www.uniprot.org/ (accessed December 1, 2018).
16 ▶ http://www.ncbi.nlm.nih.gov/omim (accessed December 1, 2018).
17 ▶ http://www.ncbi.nlm.nih.gov/pubmed (accessed December 1, 2018).

a) Pairwise alignment between human chymotrypsin and human trypsin.

```
CTRB_HUMAN    MAFLWLLSCWALLGTTFGCGVPAIHPVLSGLSRIVNGEDAVPGSWPWQVSLQDKTGFHFC
TRY1_HUMAN    MNPLLILTFVA- - - - - - - - - - - - AALAAPFDDDDKIVGGYNCEENSVPYQVSLN- - SGFHFC

CTRB_HUMAN    GGSLISEDWVVTAAHCGVRTSDDVVVAGEFDQGSDEENIQVLKIAKVFKNPKFSILTVNND
TRY1_HUMAN    GGSLINEQWVVSAGHC- YKSRIQVRLGEHNIEVLEGNEQFINAAKIIRHPQYDRKTLNND

CTRB_HUMAN    ITLLKLATPARFSQTVSAVCLPSADDDFPAGTLCATTGWGKTKYNANKTPDKLQQAALPL
TRY1_HUMAN    IMLIKLSSRAVINARVSTISLPTAPP - - ATGTKCLISGWGNTASSGADYPDYPDELQCLDAPV

CTRB_HUMAN    LSNAECKKSWGRRITDVMICAG - - ASGVSSCMGDSGGPLVCQKDGAWTLVGIVSWGSDTC
TRY1_HUMAN    LSQAKCEASYPGKITSNMFCVGFLEGGKDSCQGDSGGPVVCNG - - - - QLQGVVSWGDGCA

CTRB_HUMAN    STSSPGVYARVTKLIPWVQKILLAN -
TRY1_HUMAN    QKNKPGVYTKVYNYVKWIKNTIAANS
```

b) Smith Waterman matrix illustrating the aligned region in A, using the BLOSUM62 mutation matrix (Henikff and Henikoff, 1994).

	G	F	L	E	G	G	K	D	S	C	Q	G	D	S	G	G	P	V	V	C	N	G	Q	L	Q
G	6	-3	-4	-2	6	6	-2	-1	0	-3	-2	6	1	0	6	6	-2	-3	-3	-3	0	6	-2	-4	-2
A	0	-2	-1	-1	0	0	-1	-2	1	0	-1	0	-2	1	0	0	-1	0	0	0	-2	0	-1	-1	-1
S	0	-2	-2	0	0	0	0	0	4	-1	0	0	0	4	0	0	-1	-2	-2	-1	1	0	0	-2	0
G	6	-3	-4	-2	6	6	-2	-1	0	-3	-2	6	1	0	6	6	-2	-3	-3	-3	0	6	-2	-4	-2
V	-3	-1	1	-2	-3	-3	-2	-3	-2	-1	-2	-3	-3	-2	-3	-3	-2	4	4	-1	-3	-3	-2	1	-2
S	0	-2	-2	0	0	0	0	0	4	-1	0	0	0	4	0	0	-1	-2	-2	-1	1	0	0	-2	0
S	0	-2	-2	0	0	0	0	0	4	-1	0	0	0	4	0	0	-1	-2	-2	-1	1	0	0	-2	0
C	-3	-2	-1	-4	-3	-3	-3	-3	-1	9	-3	-3	-3	-1	-3	-3	-3	-1	-1	9	-3	-3	-3	-1	-3
M	-3	0	2	-2	-3	-3	-1	-3	-1	-1	0	-3	-3	-1	-3	-3	-2	1	1	-1	-2	-3	0	2	0
G	6	-3	-4	-2	6	6	-2	-1	0	-3	-2	6	1	0	6	6	-2	-3	-3	-3	0	6	-2	-4	-2
D	-1	-3	-4	2	-1	-1	-1	-6	0	-3	0	-1	6	0	-1	-1	-1	-3	-3	-3	1	-1	0	-4	0
S	0	-2	-2	0	0	0	0	0	4	-1	0	0	0	4	0	0	-1	-2	-2	-1	1	0	0	-2	0
G	6	-3	-4	-2	6	6	-2	-1	0	-3	-2	6	1	0	6	6	-2	-3	-3	-3	0	6	-2	-4	-2
G	6	-3	-4	-2	6	6	-2	-1	0	-3	-2	6	1	0	6	6	-2	-3	-3	-3	0	6	-2	-4	-2
P	-2	-4	-3	-1	-2	-2	-1	-1	-1	-3	-1	-2	-1	-1	-2	-2	7	-2	-2	-3	-2	-2	-1	-3	-1
L	-4	0	4	-3	-4	-4	-2	-4	-2	-1	-2	-4	-4	-2	-4	-4	-3	1	1	-1	-3	-4	-2	4	-2
V	-3	-1	1	-2	-3	-3	-2	-3	-2	-1	-2	-3	-3	-2	-3	-3	-2	4	-4	-1	-3	-3	-2	1	-2
C	-3	-2	-1	-4	-3	-3	-3	-3	-1	9	-3	-3	-3	-1	-3	-3	-3	-1	-1	9	-3	-3	-3	-1	-3
Q	-2	-3	-2	-2	-2	-2	1	0	0	-3	5	-2	0	0	-2	-2	-1	-2	-2	-3	0	-2	5	-2	5
K	-2	-3	-2	1	-2	-2	5	-1	0	-3	1	-2	-1	0	-2	-2	-1	-2	-2	-3	0	-2	-1	-1	1
D	-1	-3	-4	2	-1	-1	-1	6	0	-3	0	-1	6	0	-1	-1	-1	-3	-3	-3	1	-1	1	-4	1
G	6	-3	-4	-2	6	6	-2	-1	0	-3	-2	6	1	0	6	6	-2	-3	-3	-3	0	6	-2	-4	-2
A	0	-2	-1	-1	0	0	-1	-2	1	0	-1	0	-2	1	0	0	-1	0	0	0	-2	0	-1	-1	-1
W	-2	1	-2	-3	-2	-2	-3	-4	-3	-2	-2	-2	-4	-3	-2	-2	-4	-3	-3	-2	-4	-2	-2	-2	-2
T	-2	-2	-1	-1	-2	-2	-1	-1	1	-1	-1	-2	-1	1	-2	-2	-1	0	0	-1	0	-2	-1	-1	-1
L	-4	0	4	-3	-4	-4	-2	-4	-2	-1	-2	-4	-4	-2	-4	-4	-3	1	1	-1	-3	-4	-2	-4	-2
V	-3	-1	1	-2	-3	-3	-2	-3	-2	-1	-2	-3	-3	-2	-3	-3	-2	4	4	-1	-3	-3	-2	1	-2

Fig. 9.1 Example of sequence alignment using the Smith Waterman algorithm

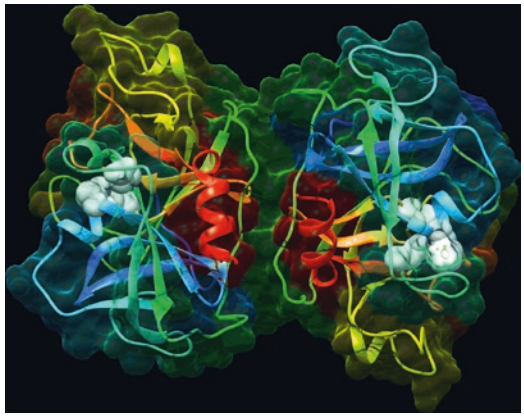


Fig. 9.2 Example of structural visualization and comparison. Comparison of the serine protease protein structures and catalytic amino acids using Chimera (► http://www.cgl.ucsf.edu/chimera; accessed December 15, 2018)

as PyMol[18] and UCSF Chimera[19] provide sophisticated and extensible applications for relatively easy visualization of 3D structures. Tools for 3D alignment of the structures are provided with these applications.

9.4.3 Prediction of Structure and Function from Sequence

One of the primary challenges in bioinformatics is taking a newly determined DNA sequence (as well as its translation into a protein sequence) and predicting the structure of the associated molecules, as well as their function. Both problems are difficult, being fraught with all the dangers associated with making predictions without hard experimental data. Nonetheless, the available sequence data are starting to be sufficient to allow good predictions in a few cases. For example, there is a Web site devoted to the assessment of biological macromolecular structure prediction methods.[20] Results suggest that when two protein molecules have a high degree (more than 40%) of sequence identity and one of the structures is known, a reliable model of the other can be built by analogy. In the case that sequence similarity is less than 25%, however, performance of these methods is much less reliable.

With the advent of deep learning, there has been an acceleration of progress in many machine learning tasks, including structure prediction. Recently, the use of convolutional neural networks by DeepMind Inc. called AlphaFold (Senior, et al. 2020) has lead to a quantum leap in the quality of predicted structures—so much so that some experts in protein structure prediction have said that parts of this challenge can now be considered "solved[21]." They make this claim because on multiple prediction tasks, the accuracy of the predicted structure is similar to those determined experimentally. Of course, it is likely that there are classes of proteins that may not perform as well, but for a large fraction of protein sequences, the structure seems to be predictable by these methods. An important caveat is that these methods must be carefully reviewed by the community, reproduced and made generally available before they will have their full impact

When scientists investigate biological structure, they commonly perform a task analogous to sequence alignment, called **structural alignment**. Given two sets of three-dimensional coordinates for a set of atoms, what is the best way to superimpose them so that the similarities and differences between the two structures are clear? Such computations are useful for determining whether two structures share a common ancestry and for understanding how the structures' functions have subsequently been refined during evolution. There are numerous published algorithms for finding good structural alignments. We can apply these algorithms in an automated fashion whenever a new structure is determined, thereby classifying the new structure into one of the protein families.

There are also algorithms for using the structure of a large biomolecule and the structure of a small organic molecule (such as a

18 ► https://pymol.org/ (accessed December 1, 2018).
19 ► http://www.cgl.ucsf.edu/chimera/ (accessed December 1, 2018).
20 ► http://predictioncenter.org/ (accessed December 1, 2018).
21 ► https://www.nature.com/articles/d41586-020-03348-4.

drug or cofactor) to try to predict the ways in which the molecules will interact. An understanding of the structural interaction between a drug and its target molecule often provides critical insight into the drug's mechanism of action. The most reliable way to assess this interaction is to use experimental methods to solve the structure of a drug–target complex. Once again, these experimental approaches are expensive, so computational methods play an important role. Typically, we can assess the physical and chemical features of the drug molecule and can use them to find complementary regions of the target. For example, a highly electronegative drug molecule will be most likely to bind in a pocket of the target that has electropositive features.

Prediction of function often relies on use of sequential or structural similarity metrics and subsequent assignment of function based on similarities to molecules of known function. These methods can guess at general function for roughly 60–80% of all genes, but leave considerable uncertainty about the precise functional details even for those genes for which there are predictions, and have little to say about the remaining genes.

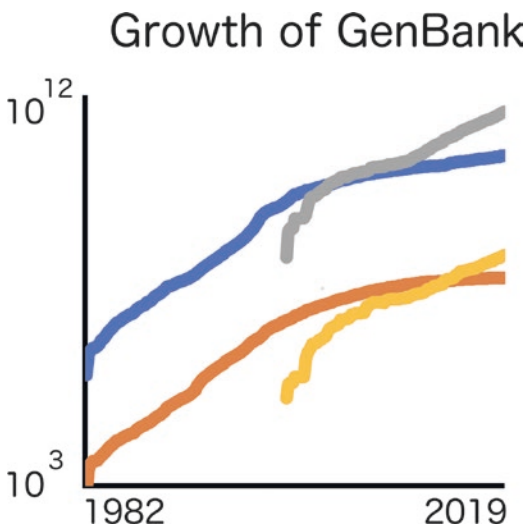
9.4.4 Clustering of Gene Expression Data

Analysis of gene expression data often begins by clustering the expression data. A typical experiment is represented as a large table, where the rows are the genes on each chip and the columns represent the different experiments, whether they be time points or different experimental conditions. Each row is then a vector of values that represent the results of the experiment with respect to a specific gene. Clustering can then be performed to determine which genes are being expressed similarly. Genes that are associated with similar expression profiles are often functionally associated. For example, when a cell is subjected to starvation (fasting), ribosomal genes are often downregulated in anticipation of lower protein production by the cell. It has similarly been shown that genes associated with neoplastic progression could be identified relatively easily with this method, making gene expression experiments a powerful assay in cancer research (see Yan and Gu 2009, for a review). In order to cluster expression data, a distance metric must be determined to compare a gene's profile with another gene's profile. If the vector data are a list of values, Euclidian distance or correlation distances can be used. If the data are more complicated, more sophisticated distance metrics may be employed. These methods fall into two categories: supervised and unsupervised. Supervised learning methods require some preconceived knowledge of the data at hand (discussed below). Usually, the method begins by selecting profiles that represent the different groups of data, e.g., genes that represent certain pathways, and then the clustering method associates each of the genes with the representative profile to which they are most similar. Unsupervised methods are more commonly applied because these methods require no knowledge of the data, and can be performed automatically.

Two such unsupervised learning methods are the hierarchical and K-means clustering methods. Hierarchical methods build a dendrogram, or a tree, of the genes based on their expression profiles. These methods are agglomerative and work by iteratively joining close neighbors into a cluster. The first step often involves connecting the closest profiles, building an average profile of the joined profiles, and repeating until the entire tree is built. K-means clustering builds k clusters or groups automatically. The algorithm begins by picking k representative profiles randomly. Then each gene is associated with the representative to which it is closest, as defined by the distance metric being employed. Then the center of mass of each cluster is determined using all of the member gene's profiles. Depending on the implementation, either the center of mass or the nearest member to it becomes the new representative for that cluster. The algorithm then iterates until the new center of mass and the previous center of mass are within some threshold. The result is k groups of genes that are regulated similarly. One drawback of K-means is that one must chose the value for k. If k is too large, logical "true" clusters may be split into pieces and if k is too small, there

will be clusters that are merged. One way to determine whether the chosen k is correct is to estimate the average distance from any member profile to the center of mass. By varying k, it is best to choose the lowest k where this average is minimized for each cluster. Another drawback of K-means is that different initial conditions can give different results, therefore it is often prudent to test the robustness of the results by running multiple runs with different starting configurations (◘ Fig. 9.3).

The future clinical usefulness of these algorithms cannot be overstated. In 2002, van't Veer et al. (2002) found that a gene expression profile could predict the clinical outcome of breast cancer. The global analysis of gene expression showed that some cancers were associated with different prognosis, not detectable using traditional means. Another exciting advancement in this field is the potential use of microarray expression data to profile the molecular effects of known and potential therapeutic agents. This molecular understanding of a disease and its treatment will soon help clinicians make more informed and accurate treatment choices (for more, see ► Chap. 26).



◘ **Fig. 9.3** The exponential growth of GENBANK. This plot shows that since 1982 the number of bases in GENBANK has grown by five full orders of magnitude and continues to grow by a factor of 10 every 4 years

9.4.4.1 Classification and Prediction

A high level description of some common approaches to classification or supervised learning are described below, but note that entire courses could be, and are, taught on each of these methods. For further details we refer readers to the suggested texts at the end of this chapter.

One of the simplest methods for classification is that of *k*-nearest-neighbor, or KNN. Essentially, KNN uses the classification of the *k* closest instances to a given input as a set of votes regarding how that instance should be classified. Unfortunately, KNN tends not to be useful for omics-based classification because it tends to break down in high-dimensional space. For high-dimensional data, KNN has difficulty in finding enough neighbors to make prediction, which will lead to large variation in the classification. This breakdown is one aspect of the "curse of dimensionality," described in more detail below (Hastie et al. 2009).

A more general statistical approach to supervised learning, and one which encompasses a number of popular methods, is that of function approximation. In this approach, one attempts to find a useful approximation of the function $f(x)$ that underlies the actual relation between the inputs and outputs. In this case, one chooses a metric by which to judge the accuracy of the approximation, for example the residual sum of squares, and uses this metric to optimize the model to fit the training data. Bayesian modeling, logistic regression, and Support Vector Machines all use variations on this approach.

Finally, there is the class of rule-based classifiers. This type of classifier may be thought of as a series of rules, each of which splits the set of instances based on a given characteristic. Details such as what criteria are used to choose the feature on which to base a rule, and whether the algorithm uses enhancements such ensemble learning (i.e., multiple models together) determine the specifics of the classifier type, for example decision trees, random forests, or covering rules.

Which approach to use depends both on the nature of the data and the question being

asked. The question might prioritize sensitivity over specificity or vice versa. For example, for a test to detect a life-threatening infection that is easily treatable by readily available antibiotics, one might want to err on the side of sensitivity. In addition, data may be numeric or categorical or have differing degrees of noise, missing values, correlated features or non-linear interactions among features. These different qualities are better handled by different methods. In many cases the best approach is actually to try a number of different methods and to compare the results. Such comparative analysis is facilitated through freely available software packages such as R/Bioconductor[22] and Weka.[23]

9.4.5 The Curse of Dimensionality

In the post-genomic era, there is no shortage of data to analyze. Rather, many researchers have more data than they know what to do with. However this overabundance tends to be a factor of the dimensionality of the data, rather than the number of subjects. This mismatch can lead to challenges for experimental design and statistical analysis. Type 1 error, or the tendency to incorrectly reject the null hypothesis and say that indeed there is statistical significance to a pattern (see ▶ Chap. 13), is amplified by looking at high-dimensional data. This is one aspect of what is known as the "curse of dimensionality" (Hastie et al. 2009). Consider analysis of gene expression data for 20,000 genes, trying to detect a pattern that can predict outcome. In a sample of, say, 30 subjects—a reasonable number when testing a single hypothesis—by random chance, some number of genes will correlate with outcome. Essentially one is testing not one but 20,000 hypotheses simultaneously. One must therefore correct for multiple hypothesis testing. The Bonferroni method is a common and straightforward approach to correct for multiple hypothesis testing.[24] It entails dividing the threshold p-value one would use, traditionally 0.05, by the number of hypotheses. So, for a test of 20,000 genes, one would require a p-value of 2.5×10^{-6} to call a gene significant. Typically, analyses using high dimensional data such as gene expression are not sufficiently powered to pass this stringent test. One would need thousands of samples to be sufficiently powered. Another approach is to use q-value, or false discovery rate (Storey and Tibshirani 2003), rather than p-value. This approach relies on empirical permutation to determine the expected number of false positives if indeed the null hypothesis is correct, which enables approximation of the proportion of false positives among all reported positives. Consider again the microarray experiment above in which each array includes 20,000 genes. We want to know whether gene X was differentially expressed in cases versus controls. Choosing a threshold p-value, or false *positive* rate, of 0.05 means that 1 time in 20 we will erroneously reject the null hypothesis and predict a false positive. If a statistical test returns 2000 positives, i.e. 2000 genes appear to be significantly differentially expressed, we expect 1 in 20 of the genes being analyzed ($20{,}000 \times (1/20) = 1000$) or approximately half of them to be false positives. A false *discovery* rate of 0.05, on the other hand, would mean that 5% *of those called positive*, in this case 100 out of 2000, are false positives. Q-value is thus less stringent than p-value, but may be of greater utility in a high-dimensional omics context than a traditional p-value or correction for multiple hypotheses.

Another approach to analysis of high dimensional data sets is to use dimensionality reduction methods such as feature selection or feature extraction. Feature selection entails extracting only a subset of the features at hand, in this case genes. This may be done in a number of ways, based on which genes vary the most, or on which genes seem to best predict the categorization at hand. In contrast,

22 ▶ http://bioconductor.org/ (accessed December 1, 2018).

23 ▶ http://www.cs.waikato.ac.nz/ml/weka/ (accessed December 1, 2018).

24 ▶ http://en.wikipedia.org/wiki/Bonferroni_correction (accessed December 1, 2018).

feature extraction creates a new smaller set of features that captures the essence of the original variation. As an example, imagine a plane flight from Seattle, WA to Key West, FL. One could use a 3-dimensional vector consisting of latitude and longitude to describe the plane's position at any given point along the way. In this case, one value would describe how far the plane had gone in the north/south direction, and one would indicate how far the plane had gone in the east/west direction. However, if we change the axis along which we are measuring to instead be the direct route along which the plane is flying, then we only need 1 dimension to describe where the plane is located. The distance flown tells us where the plane is located at any given time. This approach of changing the axes is the basis for principle components analysis (PCA), a common method for feature extraction. Instead of going from two dimensions to one, PCA on gene expression data typically goes from tens of thousands of features to just a few. Both for feature selection and feature extraction, it is important to replicate the findings in an independently generated data set in order to be sure the model is not over fitting the data on which it was trained.

9.5 Current Application Successes from Bioinformatics

Biologists have embraced the Internet in a remarkable way and have made access to data a normal and expected mode for doing business. Hundreds of databases curated by individual biologists create a valuable resource for the developers of computational methods who can use these data to test and refine their analysis algorithms. With standard Internet search engines, most biological databases can be found and accessed within moments. The large number of databases has led to the development of meta-databases that combine information from individual databases to shield the user from the complex array that exists. There are various approaches to this task.

The National Center for Biotechnology Information (NCBI) suite of databases and software (previously known as the 'Entrez' gives integrated access to the biomedical literature, protein, and nucleic acid sequences, macromolecular and small molecular structures, and genome project links (including both the Human Genome Project and sequencing projects that are attempting to determine the genome sequences for organisms that are either human pathogens or important experimental model organisms) in a manner that takes advantages of either explicit or computed links between these data resources.[25] Newer technologies are being developed that will allow multiple heterogeneous databases to be accessed by search engines that can combine information automatically, thereby processing even more intricate queries requiring knowledge from numerous data sources. One example is the Bioconductor project, a toolbox for bioinformatics in the R programming language.[26]

9.5.1 Data Sharing

In 1996, the First International Strategy Meeting on Human Genome Sequencing was held in Bermuda. In this meeting, a set of principles was agreed upon regarding sharing of human genome sequencing data. These principles came to be known as the Bermuda principles. They stipulated that (1) all sequence assemblies larger than 1 kb should be released as soon as possible, ideally within 24 h; (2) finished annotated sequences should be published immediately to public databases; and (3) that all human sequence data generated in large-scale sequencing centers should be made available in the public domain.[27]

Increasingly, journals and funders require that researchers deposit all types of research data in publicly available repositories (Fischer and Zigmond 2010). In 2009, President Obama announced an Open Government

25 ► https://www.ncbi.nlm.nih.gov/search/ (accessed December 7th, 2020).

26 ► http://bioconductor.org/ (accessed December 1, 2018).

27 ► http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml (accessed December 1, 2018).

Directive that included plans to make federally funded research data available to the public.[28] This announcement describes the NIH's policy regarding published manuscripts in particular, but also notes that the results of vgovernment-funded research can take many forms, including data sets. Currently the NIH requires that proposals for funding of over $500,000 include a data sharing plan.[29]

To that end, a significant advancement in bioinformatics is in making research datasets more available and reusable. From the community of researchers who are enabling this effort the concept of FAIR data has emerged. FAIR datasets are Findable, Accessible, Interoperable and Reusable. FAIR data principles lay out a framework to encourage increased sharing and use of scientific datasets. Findable data includes the use of global persistent identifiers and metadata standards. Accessible data is available on the Internet and searchable through metadata usage. Interoperable data use a "formal, accessible, shared and broadly applicable language for knowledge representation". Finally, reusable data have clear attribution and license that enables reuse. The webportal FAIRsharing provides curated resources on datasets, standards and collections that are more FAIR.[30] Resources such as BioCaddie DataMed enable discovery of datasets through a Data Discovery Index.[31]

9.5.2 Data Standards, Metadata and Biomedical Ontologies

▶ Chapter 7 on standards in biomedical informatics addresses standardized terminologies as well as standards for data exchange, and terminologies for translational research are discussed in ▶ Chap. 27. The development of such schemes necessitates the creation of terminology standards, just as in clinical informatics. There are now many controlled vocabularies (or ontologies) and metadata standards for annotation of genomic or proteomic data. Metadata standards help define information which should be collected and annotated upon various types of datasets. Furthermore, a great many tools have been developed to help researchers access and analyze this data. For example, the previously mentioned Bioconductor project provides bioinformatic tools in the R language for solving common problems. Other commonly used tools include BioPerl, BioPython and MATLAB.[32]

Biomedical ontologies have become a key component in the development of metadata standards for the management and exchange of bioinformatic datasets and in making data more FAIR (see ▶ Sect. 9.5.1). The open biomedical ontologies consortium (OBO) has developed a number of reference ontologies that are in wide use in bioinformatics including Gene Ontology, Human Phenotype Ontology and the UBERON anatomy ontology (Smith et al. 2007). For example, **Gene Ontology (GO)** is an ontology used for annotation of gene function, and arguably the most widely used ontology in basic research. Ontologies enable indexing, exchange and computing with biomedical datasets and metadata.

Metadata standards for bioinformatics datasets are an intellectual challenge for researchers to enable the sharing and interoperability of data and to make data more FAIR. There are a number of tools and web portals such as the Center for Expanded Data Annotation and Retrieval (CEDAR) provide tools for creation and sharing of metadata about datasets.[33] Metadata can include information about an experiment such as the protocol, the time the experiment was performed, who performed the experiment and technology used to generate or analyze the experiment, but

28 ▶ http://edocket.access.gpo.gov/2009/E9-29322.htm (accessed December 1, 2018).
29 ▶ http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html (accessed December 1, 2018).
30 ▶ https://fairsharing.org/ (accessed December 1, 2018).
31 ▶ https://datamed.org/ (accessed April 20, 2019).
32 ▶ http://www.open-bio.org/ (accessed December 1, 2018).
33 ▶ https://metadatacenter.org/ (accessed December 1, 2018).

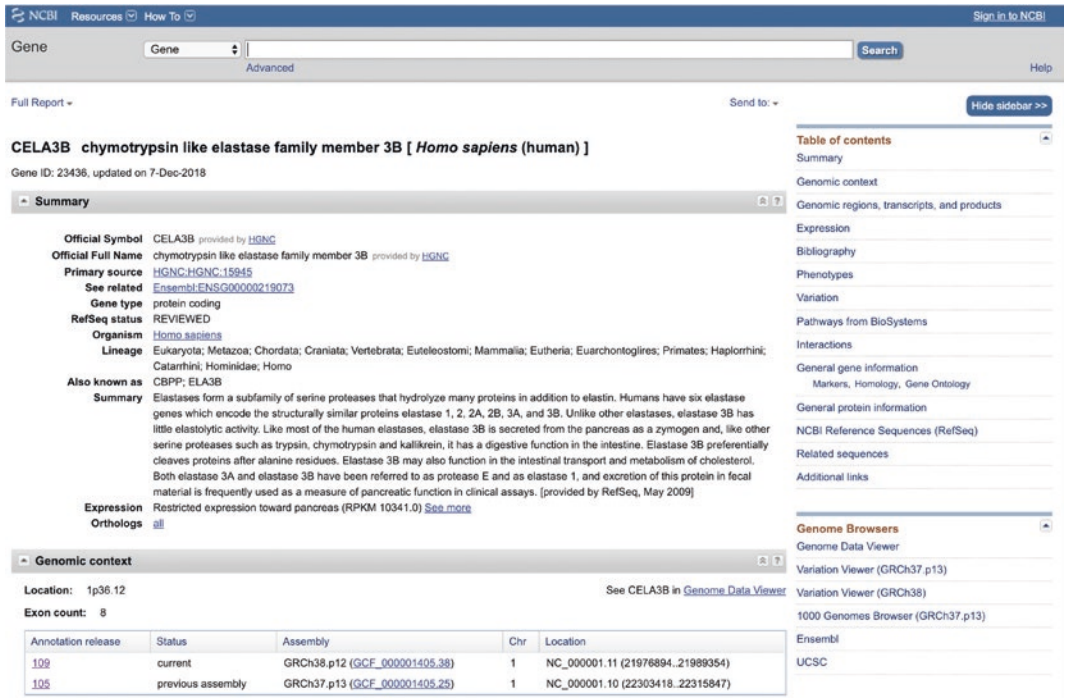


Fig. 9.4 The NCBI Gene entry for the digestive enzyme chymotrypsin. Basic information about the original report is provided, as well as some annotations of the key regions in the sequence and the complete sequence of DNA bases (a, g, t, and c) is provided as a link. (Courtesy of NCBI)

can also include information such as organism, disease model, tissue, conditions, etc.

9.5.2.1 Sequence and Genome Databases

The main types of sequence information that must be stored are DNA and protein. One of the largest DNA **sequence databases** is GENBANK, which is managed by the NCBI.[23] GENBANK is growing rapidly as genome-sequencing projects feed their data (often in an automated procedure) directly into the database. Figure 9.3 shows the logarithmic growth of data in GENBANK since 1982. NCBI Gene curates some of the many genes within GENBANK and presents the data in a way that is easy for the researcher to use (Fig. 9.4).

In addition to GENBANK, there are numerous special-purpose DNA databases for which the curators have taken special care to clean, validate, and annotate the data. The work required of such curators indicates the degree to which raw sequence data must be interpreted cautiously. GENBANK can be searched efficiently with a number of algorithms and is usually the first stop for a scientist with a new sequence who wonders "Has a sequence like this ever been observed before? If one has, what is known about it?" There are increasing numbers of stories about scientists using GENBANK to discover unanticipated relationships between DNA sequences, allowing their research programs to leap ahead while taking advantage of information collected on similar sequences.

A database that has become very useful recently is the University of California Santa Cruz Genome Browser[34] (Fig. 9.5). This data set allows users to search for specific sequences in the UCSC version of the human genome. Powered by the similarity search tool BLAT, users can quickly find annotations on the human genome that contain their sequence of interest. These annotations include known

34 ▶ http://genome.ucsc.edu/ (accessed December 1, 2018).

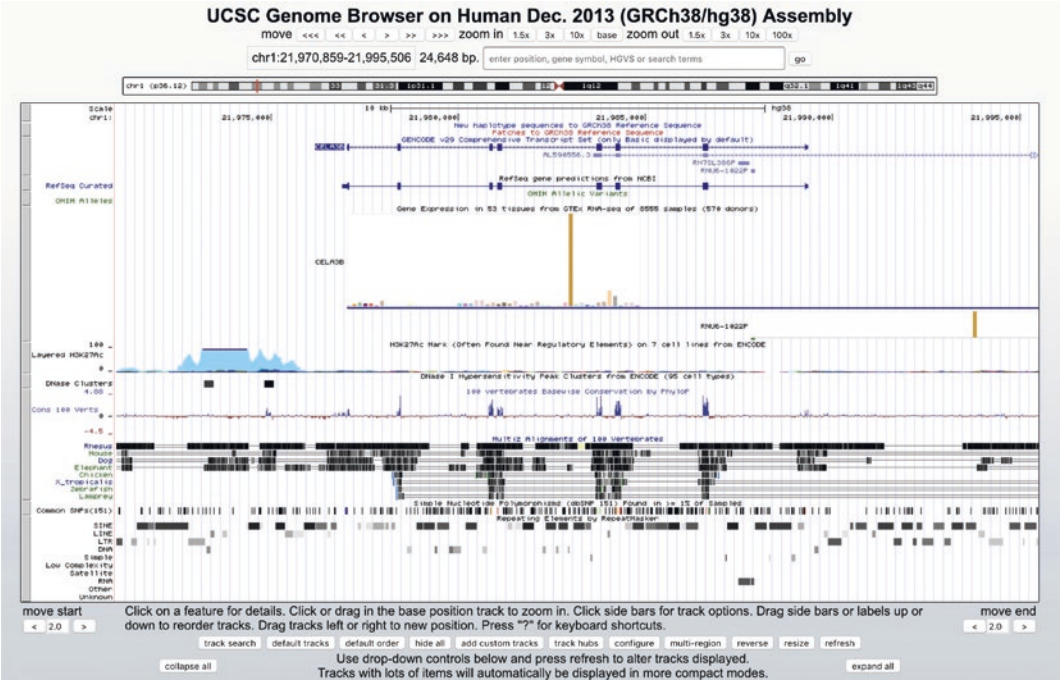


Fig. 9.5 Screen from the UC Santa Cruz genome browser showing the chymotrypsin C gene. The rows in the browser show annotations on the gene sequence. The browser window here shows a small segment of human chromosome 15, as if the sequence of a, g, c and t are represented from left to right (5–3). The annotations include gene predictions and annotations as well as an alignment of the similarity of this region of the genome when compared with the mouse genome

variations (mutations and SNPs), genes, comparative maps with other organisms, and many other important data.

9.5.3 Structure Databases

Although sequence information is obtained relatively easily, structural information remains expensive on a per-entry basis. The experimental protocols used to determine precise molecular structural coordinates are expensive in time, materials, and human power. Therefore, we have only a small number of structures for all the molecules characterized in the sequence databases. The two main sources of structural information are the Cambridge Structural Database[35] for small molecules (usually less than 100 atoms) and the PDB[36] for macromolecules (see ▶ Sect. 9.3.2), including proteins and nucleic acids, and combinations of these macromolecules with small molecules (such as drugs, cofactors, and vitamins). The PDB has approximately 75,000 high-resolution structures, but this number is misleading because many of them are small variants on the same structural architecture. There are approximately 100,000 proteins in humans; therefore, many structures remain unsolved (e.g., Burley and Bonanno 2002). In the PDB, each structure is reported with its biological source, reference information, manual annotations of interesting features, and the Cartesian coordinates of each atom within the molecule. Given knowledge of the three-dimensional structure of

35 ▶ https://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd/ (accessed December 15, 2018).

36 ▶ https://www.rcsb.org/ (accessed December 18, 2018).

molecules, the function sometimes becomes clear. For example, the ways in which the medication methotrexate interacts with its biological target have been studied in detail for two decades. Methotrexate is used to treat cancer and rheumatologic diseases, and it is an inhibitor of the protein dihydrofolate reductase, an important molecule for cellular reproduction. The three-dimensional structure of dihydrofolate reductase has been known for many years and has thus allowed detailed studies of the ways in which small molecules, such as methotrexate, interact at an atomic level. As the PDB increases in size, it becomes important to have organizing principles for thinking about biological structure. SCOP2[37] provides a classification based on the overall structural features of proteins. It is a useful method for accessing the entries of the PDB.

9.5.4 Analysis of Biological Pathways and Understanding of Disease Processes

The ECOCYC project is an example of a computational resource that has comprehensive information about biochemical pathways. ECOCYC is a knowledge base of the metabolic capabilities of *E. coli*; it has a representation of all the enzymes in the *E. coli* genome and of the chemical compounds those enzymes transform.[38] It also links these enzymes to their genes, and genes are mapped to the genome sequence.

EcoCyc also encodes the genetic regulatory network of *E. coli*, describing all protein and RNA regulators of *E. coli* genes. The network of pathways within ECOCYC provides an excellent substrate on which useful applications can be built. For example, they provide: (1) the ability to guess the function of a new protein by assessing its similarity to *E. coli* genes with a similar sequence, (2) the ability to ask what the effect on an organism would be if a critical component of a pathway were removed (would other pathways be used to create the desired function, or would the organism lose a vital function and die?), and (3) the ability to provide a rich user interface to the literature on *E. coli* metabolism. Similarly, the Kyoto Encyclopedia of Genes and Genomes (KEGG) provides pathway datasets for organism genomes.[39]

9.5.5 Integrative Databases

A **integrative database** is a postgenomic database that bridges the gap between molecular biological databases with those of clinical importance. One excellent example of a postgenomic database is the Online Mendelian Inheritance in Man (OMIM) database, which is a compilation of known human genes and genetic diseases, along with manual annotations describing the state of our understanding of individual genetic disorders.[40] Each entry contains links to special-purpose databases and thus provides links between clinical syndromes and basic molecular mechanisms (Fig. 9.6).

9.6 Future Challenges as Bioinformatics and Clinical Informatics Converge

Bioinformatics didn't solve all of its problems with the sequencing of the human genome. There is a series of challenges for which the completion of the first human genome sequence is only the beginning.

9.6.1 Linkage of Molecular Information with Symptoms, Signs, and Patients

There is currently a gap in our understanding of disease processes. Although we have a good understanding of the principles by

37 ► http://scop2.mrc-lmb.cam.ac.uk/ (accessed December 15, 2018).

38 ► http://ecocyc.org/ (accessed December 15, 2018).

39 ► http://www.genome.jp/kegg/pathway.html (accessed December 1, 2018).

40 ► http://www.ncbi.nlm.nih.gov/omim/ (accessed December 1, 2018).

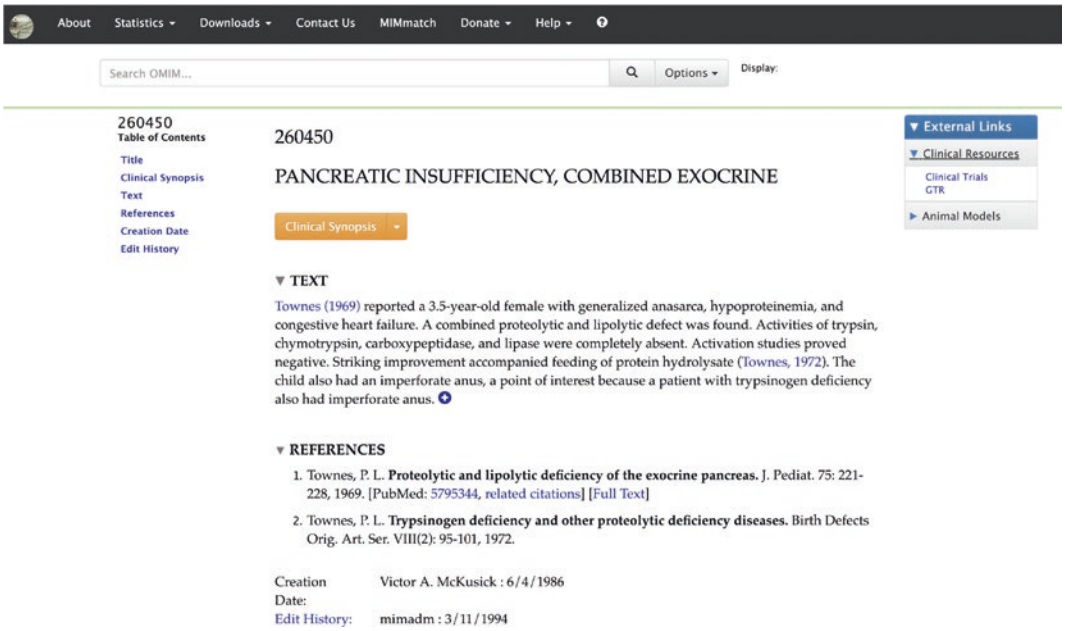


Fig. 9.6 Screen from the Online Mendelian Inheritance in Man (*OMIM*) database showing an entry for pancreatic insufficiency, an autosomal recessive disease in which chymotrypsin (NCBI Gene entry shown in Fig. 9.2) is totally absent (as are some other key digestive enzymes). (Courtesy of NCBI)

which small groups of molecules interact, we are not able to explain fully how thousands of molecules interact within a cell to create both normal and abnormal physiological states. As the databases continue to accumulate information ranging from patient-specific data to fundamental genetic information, a major challenge is creating the conceptual links among these databases to create an audit trail from molecular-level information to macroscopic phenomena, as manifested in disease. The availability of these links will facilitate the identification of important targets for future research and will provide a scaffold for biomedical knowledge, ensuring that important literature is not lost within the increasing volume of published data.

9.6.2 Computational Representations of the Biomedical Literature

An important opportunity within bioinformatics is the linkage of biological experimental data with the published papers that report them. Electronic publication of the biological literature provides exciting opportunities for making data easily available to scientists. Already, certain types of simple data that are produced in large volumes are expected to be included in manuscripts submitted for publication, including new sequences that are required to be deposited in GENBANK and new structure coordinates that are deposited in the PDB. However, there are many other experimental data sources that are currently difficult to provide in a standardized way, either because the data are more intricate than those stored in GENBANK or PDB or they are not produced in a volume sufficient to fill a database devoted entirely to the relevant area. Knowledge base technology can be used, however, to represent multiple types of highly interrelated data.

Knowledge bases can be defined in many ways (see ▶ Chap. 24); for our purposes, we can think of them as databases in which (1) the ratio of the number of tables to the number of entries per table is high compared with usual databases, (2) the individual entries (or records) have unique names, and (3) the values of many fields for one record in the database are the names of other records, thus creating

a highly interlinked network of concepts. The structure of knowledge bases often leads to unique strategies for storage and retrieval of their content. To build a knowledge base for storing information from biological experiments, there are some requirements. First, the set of experiments to be modeled must be defined. Second, the key attributes of each experiment that should be recorded in the knowledge base must be specified. Third, the set of legal values for each attribute must be specified, usually by creating a controlled terminology for basic data or by specifying the types of knowledge-based entries that can serve as values within the knowledge base.

9.6.3 Computational Challenges with an Increasing Deluge of Biomedical Data

An increasing challenge in biomedicine is storing, interpreting and integrating the massive amount of datasets the biomedical community is generating, largely from modern technologies in high throughput experimentation. The amount of DNA sequence data being generated over time has dwarfed Moore's Law, for example. This issue is important for all areas of biomedical informatics, and is discussed in more detail in the on Translational Bioinformatics (► Chap. 26).

9.7 Conclusion

Bioinformatics is closely allied to translational and clinical informatics. It differs in its emphasis on a reductionist view of biological systems, starting with sequence information and moving to structural and functional information. The emergence of the genome sequencing projects and the new technologies for measuring metabolic processes within cells is beginning to allow bioinformaticians to construct a more synthetic view of biological processes, which will complement the whole-organism, top-down approach of clinical informatics. More importantly, there are technologies that can be shared between bioinformatics and clinical informatics because they both focus on representing, storing, and analyzing biological or biomedical data. These technologies include the creation and management of standard terminologies and data representations, the integration of heterogeneous databases, the organization and searching of the biomedical literature, the use of machine learning techniques to extract new knowledge, the simulation of biological processes, and the creation of knowledge-based systems to support advanced practitioners in the two fields.

Suggested Readings

Altman, R. B., Dunker, A. K., Hunter, L., & Klein, T. E. (2003). Pacific symposium on Biocomputing'03. Singapore: World Scientific Publishing. The proceedings of one of the principal meetings in bioinformatics, this is an excellent source for up-to-date research reports. Other important meetings include those sponsored by the International Society for Computational Biology (ISCB, http://www.iscb.org/), Intelligent Systems for Molecular Biology (ISMB, http://iscb.org/conferences.shtml.35), and the RECOMB meetings on computational biology (http://www.ctw-congress.de/recomb/). ISMB and PSB have their proceedings indexed in PubMed.

Baldi, P., & Brunak, S. (2001). Bioinformatics: The machine learning approach. Cambridge, MA: MIT Press. This introduction to the field of bioinformatics focuses on the use of statistical and artificial intelligence techniques in machine learning.

Baldi, P., & Hatfield, G. W. (2002). DNA microarrays and gene expression. Cambridge: Cambridge University Press. Introduces the different microarray technologies and how they are analyzed.

Berg, J. M., Tymoczko, J. L., & Stryer, L. (2010). Biochemistry. New York: W.H. Freeman. The textbook by Stryer and colleagues is well written, and is illustrated and updated on a regular basis. It provides an excellent introduction to basic molecular biology and biochemistry.

Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). Biological sequence analysis: Probabilistic models of proteins and nucleic acids. Cambridge: Cambridge University Press.

This edited volume provides an excellent introduction to the use of probabilistic representations of sequences for the purposes of alignment, multiple alignment, and analysis.

Gusfield, D. (1997). Algorithms on strings, trees and sequences: Computer science and computational biology. Cambridge: Cambridge University Press. Gusfield's text provides an excellent introduction to the algorithmics of sequence and string analysis, with special attention paid to biological sequence analysis problems.

Malcolm, S., & Goodship, J. (Eds.). (2007). Genotype to phenotype (2nd ed.). Oxford: BIOS Scientific Publishers. This volume illustrates the different efforts to understand how diseases are linked to genes.

Pevsner, P. (2009). Bioinformatics and functional genomics. Hoboken: Wiley. A widely used excellent introduction to bioinformatics algorithms.

Questions for Discussion

1. How are DNA and protein sequence information changing the way that medical records are managed? Which types of systems are or will be most affected (laboratory, radiology, admission and discharge, financial, order entry)?
2. It has been postulated that clinical informatics and bioinformatics are working on the same problems, but in some areas one field has made more progress than the other. Identify three common themes. Describe how the issues are approached by each subdiscipline.
3. Why should an awareness of bioinformatics be expected of clinical informatics professionals? Should a chapter on bioinformatics appear in a clinical informatics textbook? Explain your answers.
4. Why should an awareness of clinical informatics be expected of bioinformatics professionals? Should a chapter on clinical informatics appear in a bioinformatics textbook? Explain your answers.
5. One major problem with introducing computers into clinical medicine is the extreme time and resource pressure placed on physicians and other health care workers. Do you think that the same problems are arising in basic biomedical research?
6. Why have biologists and bioinformaticians embraced the Web as a vehicle for disseminating data so quickly, whereas clinicians and clinical informaticians have been more hesitant to put their primary data online?
7. If a patient's entire genome were present in their medical record how would one go about interpreting it clinically? Similarly, if we had an entire electronic health record database that included human genomes, how would a researcher go about finding new or novel genetic associations?
8. With the many high throughput experiments that are used in biomedical research, how are some ways to integrate those datasets using systems biology? For example, if you had a microarray dataset that annotated gene expression levels and a proteomics dataset that identified protein interactions, how could you jointly use both datasets to identify markers for a disease?

References

Altschul, S. F., Gish, W., Mille, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403–410.

Ashley, E. A., Butte, A. J., Wheeler, M. T., Chen, R., Klein, T. E., Dewey, F. E., et al. (2010). Clinical assessment incorporating a personal genome. *The Lancet, 375*, 1525–1535.

Babior, B. M., & Matzner, Y. (1997). The familial Mediterranean fever gene—cloned at last. *The New England Journal of Medicine, 337*(21), 1548–1549.

Bai, C., & Elledge, S. J. (1997). Gene identification using the yeast two-hybrid system. *Methods in Enzymology, 283*, 141–156.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development, 16*(1), 6–21.

Brown, D. G., Rao, S., Weir, T. L., O'Malia, J., Bazan, M., Brown, R. J., & Ryan, E. P. (2016). Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. *Cancer and Metabolism, 4*, 11.

Burley, S. K., & Bonanno, J. B. (2002). Structuring the universe of proteins. *Annual Review of Genomics and Human Genetics, 3*, 243–262.

Cech, T. R. (2000). Structural biology. The ribosome is a ribozyme. *Science, 289*(5481), 878–879.

Davies, K. (2010). Physicians and their use of information: A survey comparison between the United States, Canada, and the United Kingdom. *Journal of the Medical Library Association, 99*, 88–91.

Dayhoff, M. O. (1974). Computer analysis of protein sequences. *Federation Proceedings, 33*(12), 2314–2316.

Durfy, S. J. (1993). Ethics and the human genome project. *Archives of Pathology & Laboratory Medicine, 117*(5), 466–469.

Fischer, B. A., & Zigmond, M. J. (2010). The essential nature of sharing in science. *Science and Engineering Ethics, 16*(4), 783–799.

Gibson, K., & Scheraga, H. (1967). Minimization of polypeptide energy. I. Preliminary structures of bovine pancreatic ribonuclease S-peptide. *Proceedings of the National Academy of Sciences, 58*(2), 420–427.

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A landscape takes shape. *Cell, 128*(4), 635–638.

Gusfield, D. (1997). *Algorithms on strings, trees and sequences: Computer science and computational biology*. Cambridge: Cambridge University Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Karplus, M., & Weaver, D. L. (1976). Protein-folding dynamics. *Nature, 260*(5550), 404–406.

Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Jr., et al. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell, 150*(2), 389–401.

Kent, W. J. (2003). BLAT—the BLAST-like alignment tool. *Genome Research, 12*(4), 656–664.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*, 860–921.

Langridge, R. (1974). Interactive three-dimensional computer graphics in molecular biology. *Federation Proceedings, 33*(12), 2332–2335.

Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., et al. (1997). Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proceedings of the National Academy of Sciences, 94*(24), 13057–13062.

Levitt, M. (1983). Molecular dynamics of native protein. I. Computer simulation of trajectories. *Journal of Molecular Biology, 168*(3), 595–617.

Li, R., Li, Y., Kristiansen, K., & Wang, J. (2008). SOAP: Short oligonucleotide alignment program. *Bioinformatics, 24*(5), 713–714.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology, 48*(3), 443–453.

Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics, 42*, 30–35.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature, 464*(7285), 59–65.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advances in Protein Chemistry, 34*, 167–339.

Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2020) Improved protein structure prediction using potentials from deep learning. Nature, 577, 706–710.

Shapiro, E., Biezunner, T., & Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics, 14*, 618–630.

Smith, T., & Waterman, M. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology, 147*(1), 195–197.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., et al. (2007). The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology, 25*(11), 1251–1255.

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America, 100*(16), 9440–9445.

Van't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature, 415*(6871), 484–485.

Wei, L., & Altman, R. B. (1998). Recognizing protein binding sites using statistical descriptions of their 3D environments. In *Proceedings of the pacific symposium on Biocomputing '98* (pp. 497–508), Singapore.

Yan, J., & Gu, W. (2009). Gene expression microarrays. In Y. Lu & R. I. Mahato (Eds.), *Cancer research pharmaceutical perspectives of cancer therapeutics* (pp. 645–672). New York: Springer.

Biomedical Imaging Informatics

Daniel L. Rubin, Hayit Greenspan, and Assaf Hoogi

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_10

Learning Objectives

After reading this chapter, you should know the answers to these questions:

1. What makes images a challenging type of data to be processed by computers when compared to non-image clinical data?
2. Why are there many different imaging modalities, and by what major two characteristics do they differ?
3. How are visual and knowledge content in images represented computationally? How are these techniques similar to representation of non-image biomedical data?
4. What sort of applications can be developed to make use of the semantic image content made accessible using the Annotation and Image Markup model?
5. What are four different types of image processing methods? Why are such methods assembled into a pipeline when creating imaging applications?
6. What is an imaging modality with high spatial resolution? What is a modality that provides functional information? Why are most imaging modalities not capable of providing both?
7. What is the goal in performing segmentation in image analysis? Why is there more than one segmentation method?
8. What are the main segmentation methods and what are their limitations? Should deep learning be always used as a first choice as its performance is relatively high?
9. What are two types of quantitative information in images? What are two types of semantic information in images? How might this information be used in medical applications?
10. What is the difference between image registration and image fusion? What are examples of each?
11. Can medical image analysis methods replace physicians who interpret images, or should their role to serve as adjunct tools to assist their image interpretations?

10.1 Introduction

Imaging plays a central role in the healthcare process. The field is crucial not only to health care, but also to medical communication and education, as well as in research. In fact much of our recent progress, particularly in diagnosis, can be traced to the availability of increasingly sophisticated imaging techniques that not only show the structure of the body in incredible detail, but also show the function of the tissues within the body.

Although there are many types (or modalities) of imaging equipment, the images the modalities produce are nearly always acquired in or converted to digital form. The evolution of imaging from analog, film-based acquisition to digital format has been driven by the necessities of cost reduction, efficient throughput, and workflow in managing and viewing an increasing proliferation in the number of images produced per imaging procedure (currently hundreds or even thousands of images). At the same time, having images in digital format makes them amenable to image processing methodologies for enhancement, analysis, display, storage, and even enhanced interpretation.

Because of the ubiquity of images in biomedicine, the increasing availability of images in digital form, the rise of high-powered computer hardware and networks, and the commonality of image processing solutions, digital images have become a core data type that must be considered in many biomedical informatics applications. Therefore, this chapter is devoted to a basic understanding of the unique aspects of images as a core data type and the unique aspects of imaging from an informatics perspective. ▶ Chapter 22, on the other hand, describes the use of images and image processing in various applications, particularly those in radiology since that field places the greatest demands on imaging methods.

The topics covered by this chapter and ▶ Chap. 22 comprise the growing discipline of biomedical imaging informatics (Kulikowski 1997), a subfield of biomedical informatics (see ▶ Chap. 1) that has arisen in recognition

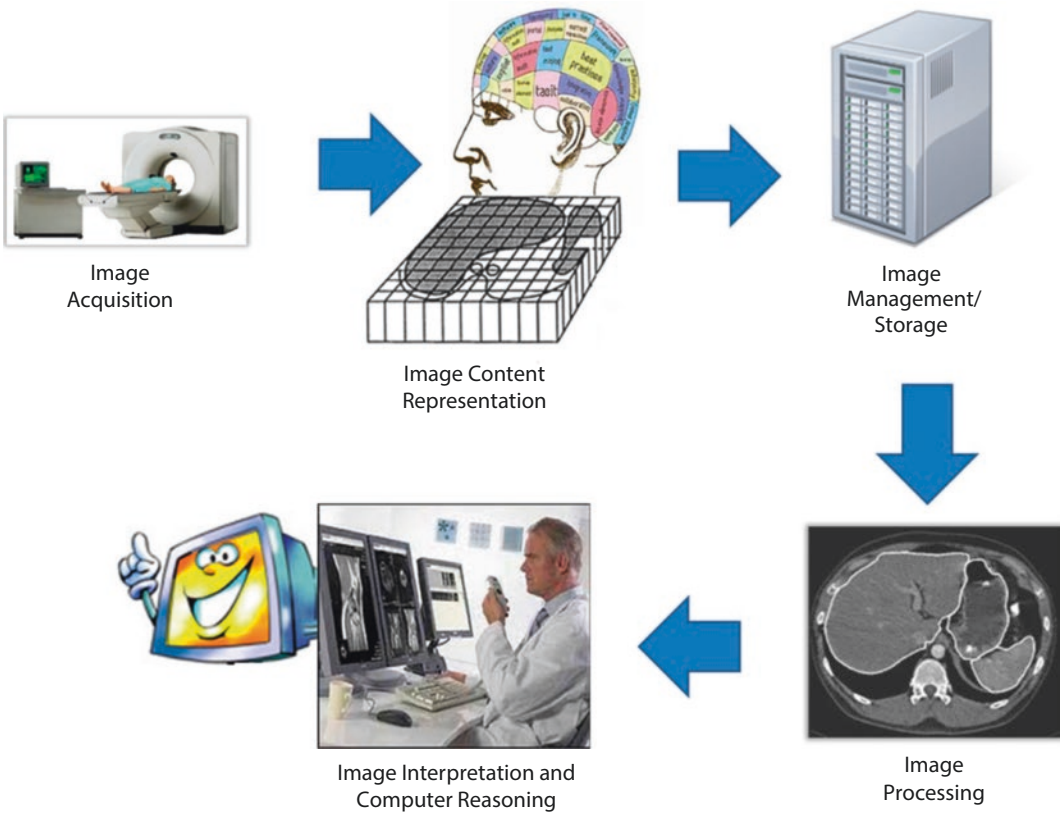


Fig. 10.1 The major topics in biomedical imaging informatics follow a workflow of activities and tasks commencing with include image acquisition, followed by image content representation, management/storage of images, image processing, and image interpretation/computer reasoning

of the common issues that pertain to all image modalities and applications once the images are converted to digital form.

Biomedical imaging informatics is a dynamic field, recently evolving from primarily focusing on image processing to broader informatics topics such as representing and processing the semantic contents (Rubin and Napel 2010) and integrating image data with other types of data (Scheckenbach et al. 2017; Pujara et al. 2018; Valdora et al. 2018; Weaver and Leung 2018). At the same time, imaging informatics shares common methodologies and challenges with other domains in biomedical informatics. By trying to understand these common issues, we can develop general solutions that can be applied to all images, regardless of the source.

The major topics in biomedical imaging informatics include image acquisition, image content representation, management/storage of images, image processing, and image interpretation/computer reasoning (Fig. 10.1). **Image acquisition** is the process of generating images from the modality and converting them to digital form if they are not intrinsically digital. **Image content representation** makes the information in images accessible to machines for processing. **Image management/storage** includes methods for storing, transmitting, displaying, retrieving, and organizing images. **Image processing** comprises methods to enhance, segment, visualize, fuse, or analyze the images. **Image interpretation/computer reasoning** is the process by which the individual viewing the image renders an impression of the medical significance of the results of imaging study, potentially aided by computer methods. ▶ Chapter 22 is primarily concerned with information systems for

image management and storage, whereas this chapter concentrates on these other core topics in biomedical imaging informatics.

An important concept when thinking about imaging from an informatics perspective is that images are an *unstructured data type*; though they are readily understood and interpreted by knowledgeable human experts, their contents are not readily machine understanding except at the granular pixel level. As such, while machines can readily manage the raw image data in terms of storage/retrieval, they cannot easily access image contents (recognize the type of image, annotations made on the image, or anatomy or abnormalities within the image), except for newer deep learning methods (▶ Sect. 10.4.5). In this regard, biomedical imaging informatics shares much in common with natural language processing (NLP; ▶ Chap. 8). In fact, as the methods of computationally representing and processing images is presented in this chapter, parallels to NLP should be considered, since there is synergy from an informatics perspective.

As in NLP, a major purpose of the methods of imaging informatics is to extract particular information; in biomedical informatics the goal is often to extract information about the structure of the body and to collect features that will be useful for characterizing abnormalities based on morphological alterations. In fact, imaging provides detailed and diverse information very useful for characterizing disease, providing an "imaging phenotype" useful for characterizing disease, since "a picture is worth a thousand words,[1]" and the informatics methods for capturing imaging phenotypes complement the informatics methods that are now being applied to electronic medical records data to capture "electronic phenotypes" of diseased patients. However, to overcome the challenges posed by the unstructured image data type, recent work is applying semantic methods from biomedical informatics to images to make their content explicit for machine processing (Rubin and Napel 2010), as well as processing entire images to learn certain semantic image content (Hosny et al. 2018; Yamashita et al. 2018). Many of the topics in this chapter therefore involve how to represent, extract and characterize the information that is present in images, such as anatomy and abnormalities. Once that task is completed, useful applications that process the image contents can be developed, such as image search and decision support to assist with image interpretation.

While we seek generality in discussing biomedical imaging informatics, many examples in this chapter are taken from a few selected domains such as brain imaging, which is part of the growing field of **neuroinformatics** (Koslow and Huerta 1997). Though our examples are specific, we attempt to describe the topics in generic terms so that the reader can recognize parallels to other imaging domains and applications.

10.2 Image Acquisition

In general, there are two different strategies in imaging the body: (1) delineate *anatomic structure* (anatomic/structural imaging), and (2) determine *tissue composition or function* (functional imaging) (◘ Fig. 10.2). In reality, one does not choose between anatomic and functional imaging; many modalities provide information about both morphology and function. However, in general, each imaging modality is characterized primarily as being able to render high-resolution images with good contrast resolution (anatomic imaging) or to render images that depict tissue function (functional imaging).

10.2.1 Anatomic (Structural) Imaging

Imaging the structure of the body has been and continues to be the major application of medical imaging, although, as described in ▶ Sect. 10.2.2, functional imaging is a very active area of research. The goal of anatomic imaging is to accurately depict the structure

1 Frederick Barnard, "One look is worth a thousand words," Printers' Ink, December, 1921.

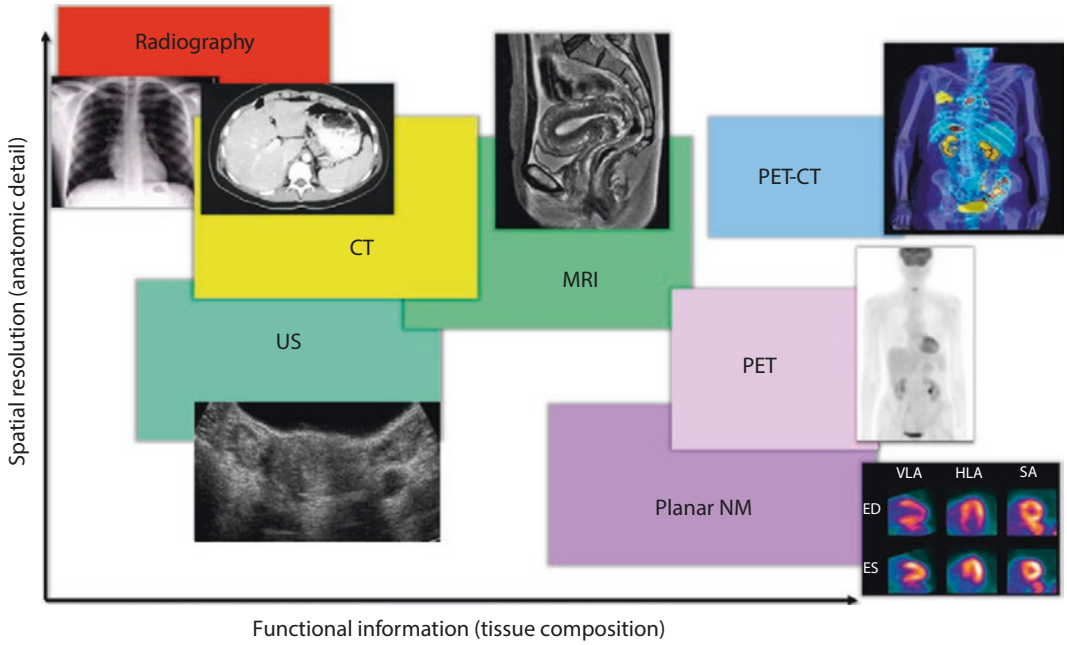


Fig. 10.2 The various radiology imaging methods differ according to two major axes of information of images, spatial resolution (anatomic detail) and functional information depicted (which represents the tissue composition—e.g., normal or abnormal). A sample of the more common imaging modalities is shown

of the body—the size and shape of organs—and to visualize abnormalities clearly. Since the goal in anatomic imaging is to depict and understand the structure of anatomic entities accurately, high spatial resolution is an important requirement of the imaging method (Fig. 10.2). Conversely, in anatomic imaging, recognizing tissue function (e.g., tissue ischemia, neoplasm, inflammation, etc.) is not the goal, though this is crucial to functional imaging and to patient diagnosis. In most cases, imaging will be done using a combination of methods or modalities to derive both structural/anatomic information as well as functional information.

10.2.2 Functional Imaging

Many imaging techniques not only show the structure of the body, but also the function, where for imaging purposes function can be inferred by observing changes of structure over time. In recent years this ability to image function has greatly accelerated. For example, ultrasound and angiography are widely used to show the functioning of the heart by depicting wall motion, and ultrasound doppler can image both normal and disturbed blood flow (Mehta et al. 2000). Molecular imaging (▶ Sect. 10.2.3) is increasingly able to depict the expression of particular genes superimposed on structural images, and thus can also be seen as a form of functional imaging.

A particularly important application of functional imaging is for understanding the cognitive activity in the brain. It is now routinely possible to put a normal subject in a scanner, to give the person a cognitive task, such as counting or object recognition, and to observe which parts of the brain light up. This unprecedented ability to observe the functioning of the living brain opens up entirely new avenues for exploring how the brain works.

Functional brain imaging modalities can be classified as *image-based* or *non-image based*. In both cases it is taken as axiomatic that the functional data must be mapped to the individual subject's anatomy, where the anatomy is extracted from structural images using techniques described in the previous sections.

Once mapped to anatomy, the functional data can be integrated with other functional data from the same subject, and with functional data from other subjects whose anatomy has been related to a template or probabilistic atlas. Techniques for generating, mapping and integrating functional data are part of the field of Functional Brain Mapping, which has become very active in the past few years, with several conferences (Organization for Human Brain Mapping 2001) and journals (Fox 2001; Toga et al. 2001) devoted to the subject.

Image-Based Functional Brain Imaging

Image-based functional data generally come from scanners that generate relatively low-resolution volume arrays depicting spatially-localized activation. For example, **positron emission tomography** (PET) (Heiss and Phelps 1983; Aine 1995; Alberini et al. 2011) and **magnetic resonance spectroscopy** (MRS) (Ross and Bluml 2001) reveal the uptake of various metabolic products by the functioning brain; and **functional magnetic resonance imaging** (fMRI) reveals changes in blood oxygenation that occur following neural activity (Aine 1995). The raw intensity values generated by these techniques must be processed by sophisticated statistical algorithms to sort out how much of the observed intensity is due to cognitive activity and how much is due to background noise.

As an example, one approach to fMRI imaging is language mapping (Corina et al. 2000). The subject is placed in the **magnetic resonance imaging** (MRI) scanner and told to silently name objects shown at 3-second intervals on a head-mounted display. The actual objects ("on" state) are alternated with nonsense objects ("off" state), and the fMRI signal is measured during both the on and the off states. Essentially the **voxel** values at the off (or control) state are subtracted from those at the on state. The difference values are tested for significant difference from non-activated areas, then expressed as t-values. The voxel array of t-values can be displayed as an image.

A large number of alternative methods have been and are being developed for acquiring and analyzing functional data (Frackowiak et al. 1997). The output of most of these techniques is a low-resolution 3-D image volume in which each voxel value is a measure of the amount of activation for a given task. The low-resolution volume is then mapped to anatomy guided by a high-resolution structural MR dataset, using one of registration techniques described in ▶ Sect. 10.4.7.

Many of these and other techniques are implemented in the SPM program (Friston et al. 1995), the AFNI program (Cox 1996), the Lyngby toolkit (Hansen et al. 1999), and several commercial programs such as Medex (Sensor Systems Inc. 2001) and BrainVoyager (Brain Innovation B.V. 2001). The FisWidgets project at the University of Pittsburgh is developing an approach that allows customized creation of graphical user interfaces in an integrated desktop environment (Cohen 2001). A similar effort (VoxBox) is underway at the University of Pennsylvania (Kimborg and Aguirre 2002).

The ultimate goal of functional neuroimaging is to observe the actual electrical activity of the neurons as they perform various cognitive tasks. fMRI, MRS and PET do not directly record electrical activity. Rather, they record the results of electrical activity, such as (in the case of fMRI) the oxygenation of blood supplying the active neurons. Thus, there is a delay from the time of activity to the measured response. In other words these techniques have relatively poor temporal resolution (▶ Sect. 10.2.4). **Electro-encephalography** (EEG) or **magnetoencephalography** (MEG), on the other hand, are more direct measures of electrical activity since they measure the electromagnetic fields generated by the electrical activity of the neurons. Current EEG and MEG methods involve the use of large arrays of scalp sensors, the output of which are processed in a similar way to CT in order to localize the source of the electrical activity inside the brain. In general this "source localization problem" is under-constrained, so information about brain anatomy obtained from MRI is used to provide further constraints (George et al. 1995).

10.2.3 Imaging Modalities

There are many different approaches that have been developed to acquire images of the body. A proliferation in imaging modalities reflects the fact that there is no single imaging technique that satisfies all the desiderata for depicting the broad variety of types of pathology. Some abnormalities are better seen on some modalities than on others. The primary difference among the imaging modalities is the type of energy source used to generate the images. In radiology, nearly every type of energy in the electromagnetic spectrum has been used, in addition to other physical phenomena such as sound and heat. We describe the more common methods according to the type of energy used to create the image.

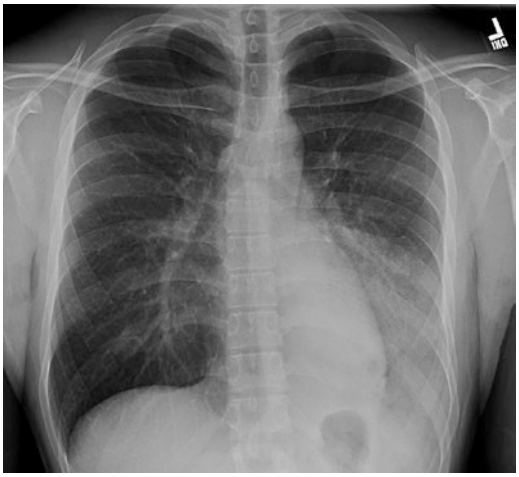


Fig. 10.3 A radiograph of the chest (Chest X-ray) taken in the frontal projection. The image is shown as if the patient is facing the viewer. This patient has abnormal density in the left lower lobe

■ Light

The earliest medical images used visible light to create photographs, either of gross anatomic structures and skin lesions or, if a microscope was used, of histological specimens. Light is still an important source for creation of images, and in fact optical imaging has seen a resurgence of interest and application for areas such as molecular imaging (Weissleder and Mahmood 2001; Ray 2011) and imaging of brain activity on the exposed surface of the cerebral cortex (Pouratian et al. 2003). Visible light is the basis for dermatological imaging (Katragadda, Finnane et al. 2016), retinal imaging (Panwar et al. 2016), and a newer modality called "optical imaging" that has promising applications such as cancer imaging (Solomon et al. 2011). Visible light, however, does not allow us to see more than a short distance beneath the surface of the body; thus other modalities are used for imaging structures deep inside the body.

■ Radiography

X-rays were first discovered in 1895 by Wilhelm Conrad Roentgen, who was awarded the 1901 Nobel Prize in Physics for this achievement. The discovery caused worldwide excitement, especially in the field of medicine; by 1900, there already were several medical radiological societies. Thus, the foundation was laid for a new branch of medicine devoted to imaging the structure and function of the body (Kevles 1997).

Radiography (colloquially called "X-ray") is still the primary modality used in radiology departments today, both to record a static image (Fig. 10.3) as well as to produce a real-time view of the patient (fluoroscopy) or a movie (cine). Both film and fluoroscopic analog screens were used initially for recording radiology images, but the fluoroscopic images very faint and required dark adaption (radiologists wore red goggles during the daytime to maximally sensitize their vision). By the 1940s, however, television and image-intensifier technologies were developed to produce clear real-time fluorescent images. Fluoroscopic examinations commonly combine real-time video monitoring of fluoroscopic images with the creation of selected higher resolution images.

Radiography is a projection technique; an X-ray beam—one form of ionizing radiation—is projected from an X-ray source through a patient's body (or other object) onto an X-ray array detector (a specially coated cassette that is scanned by a computer to capture the image in digital form), or film (to produce an non-digital image). Because an X-ray beam is differentially absorbed by

the various body tissues based on the thickness and atomic number of the tissues, the X-rays produce varying degrees of brightness and darkness on the radiographic image. The differential amounts of brightness and darkness on the image are referred to as "image contrast;" differential contrast among structures on the image is the basis for recognizing anatomic structures. Since the image in radiography is a projection, radiographs show a superposition of all the structures traversed by the X-ray beam. Much of the art (and difficulty) in interpretation of radiographs is understanding the imaging patterns that result from these superimposed structured and how to differentiate pathologies from normal structures or artifacts.

Radiographic images have very high spatial resolution because a high photon flux is used to produce the images, and a high resolution detector (film or digital image array) that captures many line pairs per unit area is used. On the other hand, since the contrast in images is due to differences in tissue density and atomic number, the amount of functional information that can be derived from radiographic images is limited (Fig. 10.2). Radiography is also limited by relatively poor contrast resolution (compared with other modalities such as computed tomography (CT) or magnetic resonance imaging (MRI) described below), their use of ionizing radiation, the challenge of spatial localization due to projection ambiguity, and their limited ability to depict physiological function. As described below, newer imaging modalities have been developed to increase contrast resolution, to eliminate the need for X-rays, and to improve spatial localization. A benefit of radiographic images is that they can be generated in real time (fluoroscopy) and can be produced using portable devices.

Digital radiography (DR) is an imaging technique that directly creates digital radiographs from the imaging procedure. Storage phosphor replaces film by substituting a reusable phosphor plate in a standard film cassette. The exposed plate is processed by a reader system that scans the image into digital form, erases the plate, and packages the cassette for reuse. An important advantage of CR systems is that the cassettes are of standard size, so they can be used in any equipment that holds film-based cassettes (Horii 1996). More recently, digital radiography uses charge-coupled device (CCD) arrays to capture the image directly. Currently, nearly all radiology departments no longer acquire radiographic images on film (analog images) and instead use digital radiography (Korner et al. 2007) to acquire digital images. This evolution was driven by the cost of film and technological advances in digital image acquisition detectors and monitors whose resolution approached that of film. At the same time, digitization of radiology drove the evolution of methods of imaging informatics we describe in this chapter.

Computed Tomography (CT) is an important imaging method that uses X-rays to produce cross sectional and volumetric images of the body (Lee 2006). Similar to radiography, X-rays are projected through the body onto an array of detectors; however, the beam and detectors rotate around the patient, making numerous views at different angles of rotation. Using computer reconstruction algorithms, an estimate of absolute density at each point (volume element or **voxel**) in the body is computed. Thus, the CT image is a computed image (Fig. 10.4); CT did not become practical for generating high quality images until the advent of powerful computers and development of computer-based reconstruction techniques, which represent one of the most

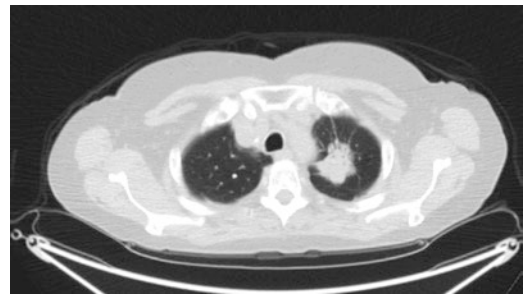


Fig. 10.4 A CT image of the upper chest. CT images are slices of a body plane; in this case, a cross sectional (axial) image of the chest. Axial images are viewed from below the patient, so that the patient's left is on viewer's right. This image shows a cancer mass in the left upper lobe of the lung

spectacular applications of computers in all of medicine (Buxton 2009). The spatial resolution of images is not as high in CT as it is in radiography, due to the computed nature of the images; however, the contrast resolution and ability to derive functional information of tissues in the body are superior for CT than for radiography (Fig. 10.2).

Ultrasound

A common energy source used to produce images is **ultrasound**, which developed from research performed by the Navy during World War II in which sonar was used to locate objects of interest in the ocean. Ultrasonography uses pulses of high-frequency sound waves rather than ionizing radiation to image body structures (Kremkau 2006). The basis of image generation is due to a property of all objects called acoustical impedance. As sound waves encounter different types of tissues in a patient's body (particularly interfaces where there is a chance in acoustical impedance), a portion of the wave is reflected and a portion of the sound beam (which is now attenuated) continues to traverse into deeper tissues. The time required for the echo to return is proportional to the distance into the body at which it is reflected; the amplitude (intensity) of a returning echo depends on the acoustical properties of the tissues encountered and is represented in the image as brightness (more echoes returning to the source is shown as image brightness). The system constructs two-dimensional images (B-scans) by displaying the echoes from pulses of multiple adjacent one-dimensional paths (A-scans). Current ultrasound machines are essentially specialized computers with attached peripherals, with active development of three-dimensional imaging. The ultrasound transducer now often sweeps out a 3-D volume rather than a 2-D plane, and the data are written directly into a three-dimensional array memory, which is displayed using volume or surface-based rendering techniques (Ritchie et al. 1996).

Ultrasound images are acquired as digital images from the outset. They may also be recorded as frames in rapid succession (cine loops) for real-time imaging. Ultrasound imaging captures not only structural information but also functional information. Doppler methods in ultrasound are used to measure and characterize the blood flow in blood vessels in the body (Fig. 10.5). More recently, ultrasound techniques called elastography have been developed to measure tissue stiffness, which is improving the ability of ultrasound to diagnose a variety of pathology conditions such as liver fibrosis (Pawlus et al. 2015; Zaleska-Dorobisz et al. 2015). The low cost of ultrasound and the fact it doesn't use ionizing radiation makes it very attractive as a primary modality for imaging worldwide, particularly for obstetrical and pediatric imaging.

Since the image contrast in ultrasound is based on differences in the acoustic impedance of tissue, ultrasound provides functional information (e.g., tissue composition and

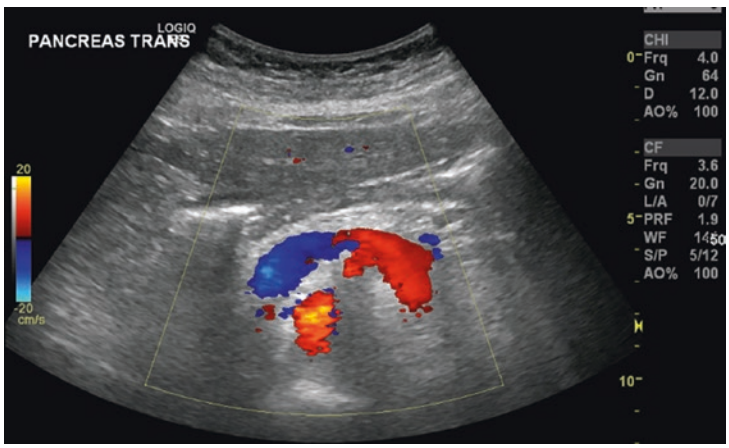


Fig. 10.5 An ultrasound image of abdomen. Like CT and MRI, ultrasound images are slices of a body, but because a user creates the images by holding a probe, any arbitrary plane can be imaged (so long as the probe can be oriented to produce that plane). This image shows an axial slice through the pancreas, and flow in nearby blood vessels (in color) is seen due to Doppler effects incorporated into the imaging method

blood flow). On the other hand, the flux of sound waves is not as dense as the photon flux used to produce images in radiography; thus ultrasound images are generally lower resolution images than other imaging modalities (Fig. 10.2).

Magnetic Resonance Imaging (MRI)

Creation of images from the resonance phenomena of unpaired spinning charges in a magnetic field grew out of **nuclear magnetic resonance** (NMR) **spectroscopy**, a technique that has long been used in chemistry to characterize chemical compounds. Many atomic nuclei within the body have a net magnetic moment, so they act like tiny magnets. When a small chemical sample is placed in an intense, uniform magnetic field, these nuclei line up in the direction of the field, spinning around the axis of the field with a frequency dependent on the type of nucleus, on the surrounding environment, and on the strength of the magnetic field.

If a radio pulse of a particular frequency is then applied at right angles to the stationary magnetic field, those nuclei with rotation frequency equal to that of the radiofrequency pulse resonate with the pulse and absorb energy. The higher energy state causes the nuclei to change their orientation with respect to the fixed magnetic field. When the radiofrequency pulse is removed, the nuclei return to their original aligned state (a process called "relaxation"), emitting a detectable radiofrequency signal as they do so. Characteristic parameters of this signal—such as intensity, duration, and frequency shift away from the original pulse—are dependent on the density and environment of the nuclei. In the case of traditional NMR spectroscopy, different molecular environments cause different frequency shifts (called chemical shifts), which we can use to identify the particular compounds in a sample. In the original NMR method, however, the signal is not localized to a specific region of the sample, so it is not possible to create an image.

Creation of medical images from NMR signals, known as Magnetic Resonance Imaging (MRI), came about shortly after fast computer-based reconstruction techniques, similar to CT, were developed. The basis of image formation in MRI is based on proton relaxation (referred to as T1 and T2 relaxation); differences in T1 and T2 are inherent properties of tissue and they vary among tissues. Thus, MRI provides detailed functional information about tissue and can be valuable in clinical diagnosis (Fig. 10.6). At the same time, the flux of radiofrequency waves used to produce the images is high, and MRI thus has high spatial resolution (Fig. 10.2).

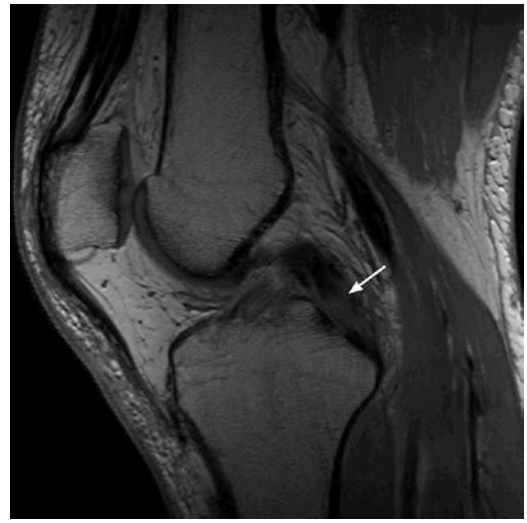


Fig. 10.6 An MRI image of the knee. Like CT, MRI images are slices of a body. This image is in the saggital plane through the mid knee, showing in a tear in the posterior cruciate ligament (arrow)

Many new modalities are being developed based on magnetic resonance. For example, magnetic resonance arteriography (MRA) and venography (MRV) are used to image blood flow (Lee 2003) and diffusion tensor imaging (DTI) is increasingly being used to image white matter fiber tracts in the brain (Le Bihan et al. 2001; Hasan et al. 2010; de Figueiredo et al. 2011; Gerstner and Sorensen 2011). More recently, a technique called MRI elastography has been developed to measure tissue stiffness (Venkatesh and Ehman 2015).

Nuclear Medicine Imaging

In **nuclear medicine imaging**, the imaging approach is a reverse of the radiographic imaging: instead of the imaging beam being outside the subject and projecting into the

subject, the imaging source is inside the subject and projects out. Specifically, a radioactive isotope is chemically attached to a biologically active compound (such as an analogue of glucose) and then is injected into the patient's peripheral circulation. The compound collects in the specific body compartments or organs (such as metabolically-active tissues), where it is stored or processed by the body. The isotope emits radiation locally, and the radiation is measured using a special detector. The resultant nuclear-medicine image depicts the level of radioactivity that was measured at each spatial location of the patient. Because the counts are inherently quantized, digital images are produced. Multiple images also can be processed to obtain temporal dynamic information, such as the rate of arrival or of disappearance of isotope at particular body sites.

Nuclear medicine images, like radiographic images, are usually acquired as projections—a large planar detector is positioned outside the patient and it collects a projected image of all the radioactivity emitted from the patient. The images are similar in appearance to radiographic projection images. However, since the photon flux is extremely low (to minimize the radiation dose to the patient), the spatial resolution of nuclear medicine images is low. On the other hand, since the only places where radioisotope accumulates will be places in the body that are targeted by the injected agent, nearly all the information in nuclear medicine images is functional information; thus nuclear imaging methods have high functional information and low spatial resolution (Fig. 10.2). Nuclear medicine techniques have recently attracted much attention because of an explosion in novel imaging probes and targeting mechanisms to localize the imaging agent (Drude et al. 2017).

In addition to projection images, a computed tomography-like method called **single-photon emission computed tomography** (SPECT) (Alberini et al. 2011) is often used. A camera rotates around the patient similar to CT, producing a computed volumetric image that may be viewed and navigated in multiple planes. A technique called Positron Emission Tomography (PET) uses a special

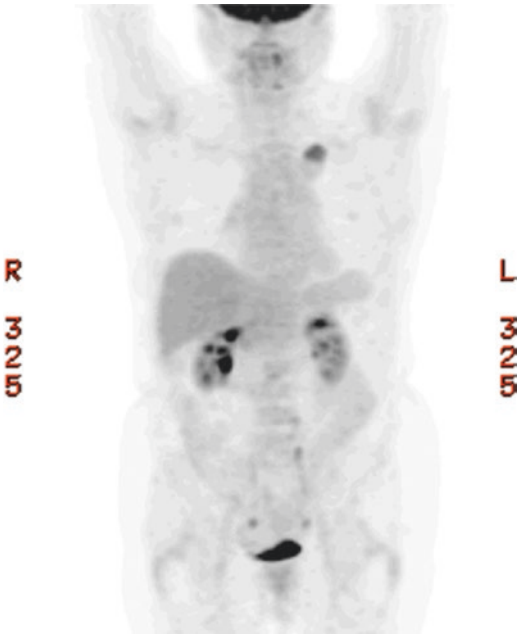


Fig. 10.7 A PET image of the body in a patient with cancer in the left lung (same patient as in Fig. 10.4). This is a projection image taken in the frontal plane after injection of a radioactive isotope that accumulates in cancers. A small black spot in the left upper lobe is abnormal and indicates the cancer mass in the upper lobe of the left lung

type of radioactive isotope that emits positrons, which, upon encountering an electron, produces an annihilation event that sends out two gamma rays in opposite directions that are simultaneously detected on an annular detector array and used to compute a cross sectional slice through the patient, similar to CT and SPECT (Fig. 10.7). These volumetric nuclear medicine imaging methods, like the projection methods, have high functional information and low spatial resolution. However, recently a newer modality called PET/CT has been developed that integrates a PET scanner and CT with image fusion (discussed below) to get the best of both worlds—functional information about lesions in the PET image plus spatial localization of the abnormality on the CT image (Figs. 10.2 and 10.8).

A subdomain of nuclear imaging called **molecular imaging** has emerged that embodies this work on molecularly-targeted imaging (and therapeutic) agents (Weissleder and

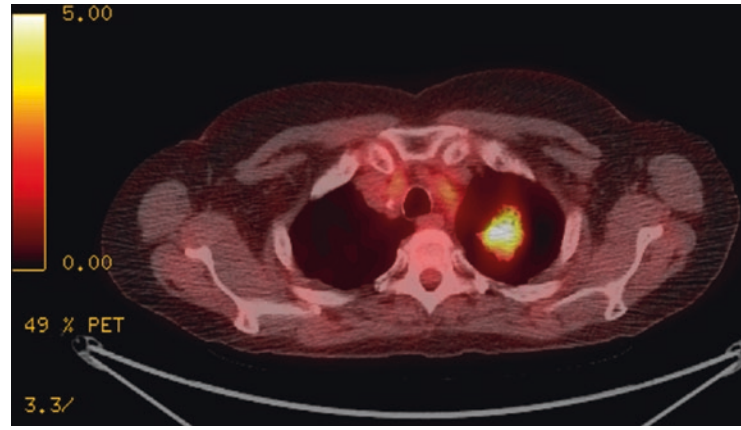


Fig. 10.8 A PET/CT fused image. The axial slice from the PET study (Fig. 10.7) and the corresponding axial slice from the CT study (Fig. 10.4) are combined into a single image that has both good spatial resolution and functional information, showing that the lung mass has abnormal uptake of isotope, indicating it is metabolically active

Mahmood 2001; Massoud and Gambhir 2003; Biswal et al. 2007; Hoffman and Gambhir 2007; Margolis et al. 2007; Ray and Gambhir 2007; Willmann et al. 2008; Pysz et al. 2010). Molecularly-tagged molecules are increasingly being introduced into the living organism, and imaged with optical, radioactive, or magnetic energy sources, often using reconstruction techniques and often in 3-D. It is becoming possible to combine gene sequence information, gene expression array data, and molecular imaging to determine not only which genes are expressed, but where they are expressed in the organism (Kang and Chung 2008; Min and Gambhir 2008; Singh et al. 2008; Lexe et al. 2009; Smith et al. 2009; Harney and Meade 2010). These capabilities will become increasingly important in the post-genomic era for determining exactly how genes generate both the structure and function of the organism.

10.2.4 Image Quality

Characteristics of Image Quality

The imaging modalities described above are complex devices with many parameters that need to be specified in generating the image, and most of the parameters can have substantial impact on the following key characteristics of the final image appearance: spatial resolution, contrast resolution, and temporal resolution, all of which have substantial impact on image quality and diagnostic value of the image. These characteristics provide an objective means for comparing images formed by digital imaging modalities.

- **Spatial resolution** is related to the sharpness of the image; it is a measure of how well the imaging modality can distinguish points on the object that are close together. For a digital image, spatial resolution is generally related to the number of pixels per image area. Spatial resolution is critical for detecting abnormalities in very small structures, such as microcalcifications on mammograms or diffuse lung diseases on chest radiographs.
- **Contrast resolution** is a measure of the ability to distinguish small differences in intensity in different regions of the image, which in turn are related to differences in measurable parameters, such as X-ray attenuation. For digital images, the number of bits per pixel is related to the contrast resolution of an image. Contrast resolution is critical to image interpretation, since differences in contrast are the basis for an object (of sufficient size) to be appreciated by the human eye or by a computerized image detection algorithm.
- **Temporal resolution** is a measure of the time needed to create an image. We consider an imaging procedure to be a real-time application if it can generate images concurrent with the physical process it is imaging. At a rate of at least 30 images per second, it is possible to produce unblurred images of the beating heart.

Other parameters that are specifically relevant to medical imaging are the degree of invasiveness, the dosage of ionizing radiation, the degree of patient discomfort, the size (portability) of the instrument, the ability to depict physiologic function as well as anatomic structure, and the availability and cost of the procedure at a specific location.

A perfect imaging modality would produce images with high spatial, contrast, and temporal resolution; it would be available, low in cost, portable, free of risk, painless, and noninvasive; it would use nonionizing radiation; and it would depict physiological function as well as anatomic structure. As seen above, the different modalities differ in these characteristics and none is uniformly strong across all the parameters (Fig. 10.2).

Contrast Agents

One of the major motivators for development of new imaging modalities is the desire to increase contrast resolution. A contrast agent is a substance introduced into the body to enhance the imaging contrast of structures or fluids in medical imaging. Contrast agents can be introduced in various ways, such as by injection, inspiration, ingestion, or enema. The chemical composition of contrast agents vary with modality so as to be optimally visible based on the physical basis of image formation. For example, iodinated contrast agents are used in radiography and CT because iodine has high atomic number, greatly attenuating X-rays, and thus greatly enhancing image contrast in any tissues that accumulate the contrast agent. Contrast agents for radiography are referred to as "radiopaque" since they absorb X-rays and obscure the beam. Contrast agents in radiography are used to highlight the anatomic structures of interest (e.g., stomach, colon, urinary tract). In an imaging technique called angiography, a contrast agent is injected into the blood vessels to opacify them on the images. In pathology, histological staining agents such as haematoxylin and eosin (H&E) have been used for years to enhance contrast in tissue sections, and magnetic contrast agents such as gadolinium have been introduced to enhance contrast in MR images.

Recently, contrast agents have been developed for ultrasound to greatly enhance image contrast (Durot et al. 2018). Ultrasound contrast agents generally comprise microbubbles—bubbles in the blood that are too small to cause damage to tissues, but that in aggregate alter the impedance mismatch between blood and tissue to enhance image contrast.

Although contrast agents have been very successful and they are commonly used, their enhancement tends to be non-specific in that any vascularized structure will be enhanced. In recent years, advances in molecular biology have led to the ability to design contrast agents that are highly specific for individual molecules. In addition to radioactively tagged molecules used in nuclear medicine, molecules are tagged for imaging by magnetic resonance and optical energy sources. Tagged molecules are imaged in 2-D or 3-D, often by application of reconstruction techniques developed for clinical imaging (Pysz et al. 2010; Jokerst and Gambhir 2011; Weissleder et al. 2016). Tagged molecules have been used for several years *in vitro* by such techniques as immunocytochemistry (binding of tagged antibodies to antigen) (Van Noorden 2002) and *in situ* hybridization (binding of tagged nucleotide sequences to DNA or RNA) (King et al. 2000). More recently, methods have been developed to image these molecules in the living organism, thereby opening up entirely new avenues for understanding the functioning of the body at the molecular level (Biswal et al. 2007; Hoffman and Gambhir 2007; Margolis et al. 2007; Ray and Gambhir 2007; Willmann et al. 2008; Pysz et al. 2010). Recent work in altering microbubbles of ultrasound contrast agents to target them to particular tissues and types of disease raises the exciting prospects for ultrasound imaging to provide even greater functional information about tissues in a minimally invasive and cost effective manner (Deshpande et al. 2010; Abou-Elkacem et al. 2015; Zhang et al. 2017).

10.2.5 Imaging Methods in Other Medical Domains

Though radiology is a core domain and driver of many clinical problems and applications of medical imaging, several other medical domains are increasingly relying on imaging to provide key information for biomedical discovery and clinical insight. The methods of biomedical informatics presented in this chapter, while focusing on radiology in our examples, are generalizable and applicable to these other domains. We briefly highlight these other domains and the role of imaging in them.

Microscopic/cellular imaging

At the microscopic level, there is a rapid growth in **cellular imaging** (Larabell and Nugent 2010; Toomre and Bewersdorf 2010; Wessels et al. 2010), including use of computational methods to evaluate the features in cells (Carpenter et al. 2006). The confocal microscope uses electronic focusing to move a two-dimensional slice plane through a three-dimensional tissue slice placed in a microscope. The result is a three-dimensional voxel array of a microscopic, or even submicroscopic, specimen (Wilson 1990; Paddock 1994). Confocal endomicroscopy, in which high resolution microscopic imaging technology is integrated into endoscopes, is opening up exciting opportunities for real-time histopatological evaluation in disease (Neumann et al. 2010). At the electron microscopic level electron tomography generates 3-D images from thick electron-microscopic sections using techniques similar to those used in CT (Perkins et al. 1997).

Pathology/tissue imaging

The radiology department was revolutionized by the introduction of digital imaging and **Picture Archiving and Communication Systems** (PACS). Pathology has likewise begun to shift from an analog to a digital workflow (Leong and Leong 2003; Gombas et al. 2004). Pathology informatics is a rapidly emerging field (Becich 2000; Gabril and Yousef 2010), with goals and research problems similar to those in radiology, such as managing huge images, improving efficiency of workflow, learning new knowledge by mining historical cases, identifying novel imaging features through correlative quantitative imaging analysis, and decision support. A particularly promising area is deriving novel quantitative image features from pathology images to improve characterization and clinical decision making (Giger and MacMahon 1996; Nielsen et al. 2008; Armstrong 2010) or to improve detection of disease within the specimen (Nagarkar et al. 2016; Ehteshami Bejnordi et al. 2017). Given that pathology and radiology produce images that characterize phenotype of disease, there is tremendous opportunity for information integration and linkage among pathology, radiology, and molecular data for discovery (Permuth et al. 2016).

Ophthalmologic imaging

Visualization of the retina is a core task of ophthalmology to diagnose disease and to monitor treatment response (Bennett and Barry 2009). Imaging modalities include retinal photography, autofluorescence, and fluorescein angiography. Recently, tomographic-based imaging has been introduced through a technique called **optical coherence tomography** (OCT; Fig. 10.9) (Figurska et al. 2010). This modality is showing great progress in evaluating a variety of retinal diseases (Freton and Finger 2011; Schimel et al. 2011; Sohrab et al. 2011; de Sisternes et al. 2014). As with radiological imaging, a number of quantitative and automated segmentation methods are being created to evaluate disease objectively (Cabrera Fernandez et al. 2005; Baumann et al. 2010; Hu et al. 2010a, b; Niu et al. 2016; de Sisternes et al. 2017a, b). Likewise, image processing methods for image visualization and fusion are being developed, similar to those used in radiology.

Dermatologic imaging

Imaging is becoming an important component of dermatology in the management of patients with skin lesions. Dermatologists

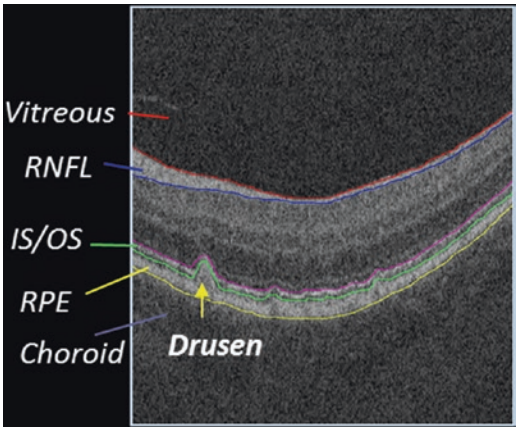


Fig. 10.9 An OCT image of the retina. Like ultrasound, OCT produces an image slice at any arbitrary angle (depending on how the light beam can be oriented), but it is limited to visualizing superficial structures due to poor penetration by light. In this image, the layered structure of the retina can be seen, as well as abnormalities (drusen)

frequently take photographs of patients with skin abnormalities, and while initially this was done for clinical documentation, increasingly this is done to leverage imaging informatics methods for training, to improve clinical care, for consultation, for monitoring progression or change in skin disease, and for image retrieval (Bittorf et al. 1997; Diepgen and Eysenbach 1998; Eysenbach et al. 1998; Lowe et al. 1998; Ribaric et al. 2001). Like radiology and pathology, recent work is being done to analyze the image content to enable decision support (Seidenari et al. 2003; Esteva et al. 2017).

10.3 Image Content Representation

The image contents comprise two components of information, the *visual content* and the *knowledge content*. The visual content is the raw values of the image itself, the information that a computer can access in a digital image directly. The knowledge content arises as the observer, who has biomedical knowledge about the image content, views the visual information in the image. For example, a radiologist viewing a CT image of the upper abdomen immediately recognizes that the image contains the liver, spleen, and stomach (anatomic entities), as well as image abnormalities such as a mass in the liver with rim enhancement (imaging observation entities). Unlike the visual content, the knowledge content of images is not directly accessible to computers from the image itself. However, semantic methods are being developed to make this content machine-accessible (▶ Sect. 10.3.2). In this section we describe imaging informatics methods for representing the visual and knowledge content of images.

10.3.1 Representing Visual Content in Digital Images

The visual content of digital images typically is represented in a computer by a two-dimensional array of numbers (a bit map). Each element of the array represents the intensity of a small square area of the picture, called a picture element (or pixel). Each **pixel** element corresponds to a volume element (or **voxel**) in the imaged subject that produced the pixel. If we consider the image of a volume, then a three-dimensional array of numbers is required. Another way of thinking of a volume is that it is a stack of two-dimensional images. However, it is also important to be aware of the voxel dimensions that correspond to the pixels when doing this. In many 2-D imaging applications, the in-plane resolution (the size of the voxels in the x,y plane) is higher than the resolution in the z-axis (i.e., the slice thickness); this is often referred to as "non-isotropic voxels." Non-isotropic voxels creates a problem when re-sampling the volume data to create other projections, such as coronal or saggital from primary axial image data. If the dimensions of the voxels (and pixels) are uniform in all dimensions, they are referred to as "isotropic." Fortunately, nearly all modern volumetric imaging methods (e.g., CT and MRI) currently produce images with isotropic voxels.

We can store any image in a computer as a matrix of integers (or real-valued numbers), either by converting it from an analog

to a digital representation or by generating it directly in digital form. Once an image is in digital form, it can be handled just like all other data. It can be transmitted over communications networks, stored compactly in databases on magnetic or optical media, and displayed on graphics monitors. In addition, the use of computers has created an entirely new realm of capabilities for image generation and analysis: images can be computed rather than measured directly. Furthermore, digital images can be manipulated for display or analysis in ways not possible with film-based images.

In addition to the 2D (slice) and 3D (volume) representation for image data, there can be additional dimensions to representing the visual content of images. It is often the case that multi-modality data are required for the diagnosis; this can be a combination of varying modalities, (e.g., CT and PET, CT and MRI) and can be a combination of imaging sequences *within* a modality (e.g., T1, T2, or other sequences in MRI) (Fig. 10.10). Pixel (or voxel) content, from each of the respective acquisition modalities, are combined in what is known as a "feature-vector" in the multi-dimensional space. For example, a 3-dimensional intensity-based feature vector, based on 3 MRI pulse sequences, can be defined as a set of three values for each pixel in the image, where the intensity of each pixel in each of the three MRI images is extracted and recorded (e.g., [Intensity(Sequence 1), Intensity(Sequence 2), Intensity(Sequence 3)]. Any imaging performed over time (e.g., cardiac echo videos) can be represented by the set of values at each time point, thus the time is added as an additional dimension to the representation.

Finally, in addition to representing the visual content, medical images also need to represent certain information about that visual content (referred to as **image metadata**). Image metadata include such things as the name of the patient, date the image was acquired, the slice thickness, the modality that was used to acquire the image, etc. All image metadata are usually stored in the header of the image file. Given that there are many different types of equipment and software that produce and consume images, standards are crucial. For images, the Digital Imaging and Communications in Medicine (DICOM) standard is for distributing and viewing any kind of medical image regardless of the origin (Bidgood Jr. and Horii 1992). DICOM has become pervasive throughout radiology and is becoming a standard in other domains such as pathology, ophthalmology, and dermatol-

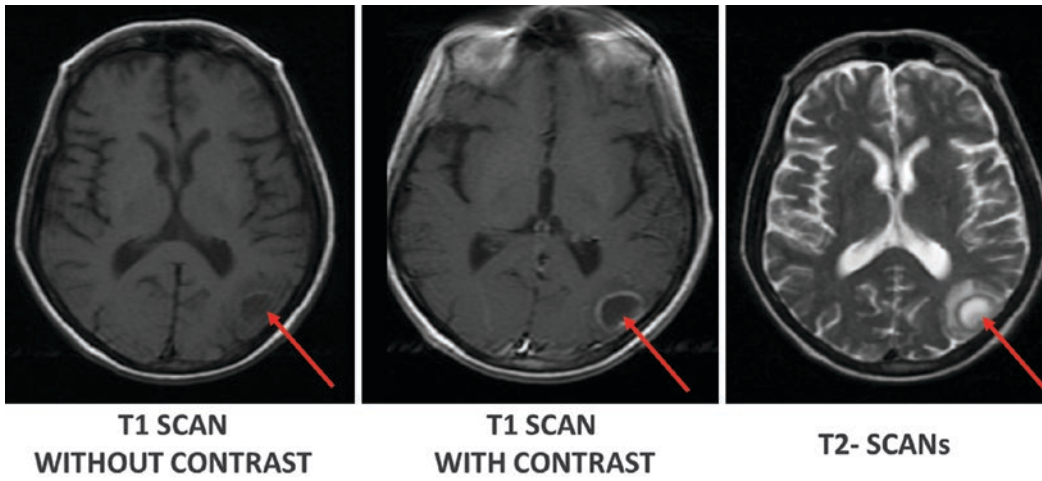


Fig. 10.10 Multi-modality imaging. Images of the brain from three modalities (T1 without contrast, T1 with contrast, and T2) are shown. The patient has a lesion in the left occipital lobe that has distinctive image features on each of these modalities, and the combination of these different features on different modalities establishes characteristic patterns useful in diagnosis

ogy. In addition to specifying a standard file syntax and metadata structure, DICOM specifies a standard protocol for communicating images among imaging devices.

10.3.2 Representing Knowledge Content in Digital Images

As noted above, the knowledge content related to images is not directly encoded in the images, but it is recognized by the observer of the images. This knowledge includes recognition of the anatomic entities in the image, imaging observations and characteristics of the observations (sometimes called "findings"), and interpretations (probable diseases). Representing this knowledge in the imaging domain is similar to knowledge representation in other domains of biomedical informatics (see ▶ Chap. 24). Specifically, for representing the entities in the domain of discourse, we adopt terminologies or ontologies. To make specific statements about individuals (images), we use "information models" (described below) that reference ontological entities as necessary. As described below, different aspects of the knowledge content of images is stored in different ways (which is one of the challenges of leveraging this information).

■ Knowledge Representation of Anatomy

Given segmented anatomical structures, whether at the macroscopic or microscopic level, and whether represented as 3-D surface meshes or extracted 3-D regions, it is often desirable to attach labels (names) to the structures in images. If the names are drawn from a controlled terminology or ontology, they can be used as an index into a database of segmented structures, thereby providing a qualitative means for comparing structures from multiple subjects or retrieving images containing particular structures.

If the terms in the vocabulary are organized so as to assert relationships that are true for all instances (the case in "ontologies"), they can support systems that manipulate and retrieve image contents in intelligent ways. If anatomical ontologies are linked to other ontologies of physiology and pathology they can provide increasingly sophisticated knowledge about the meaning of the various images and other data that are increasingly becoming available in online databases. This kind of knowledge (by the computer, as opposed to the scientist) will be required in order to achieve the seamless integration of all forms of imaging and non-imaging data.

At the most fundamental level, Nomina Anatomica (International Anatomical Nomenclature Committee 1989) and its successor, Terminologia Anatomica (Federative Committee on Anatomical Terminology 1998) provide a classification of officially sanctioned terms that are associated with macroscopic and microscopic anatomical structures. This canonical term list, however, has been substantially expanded by synonyms that are current in various fields, and has also been augmented by a large number of new terms that designate structures omitted from Terminologia Anatomica. Many of these additions are present in various controlled terminologies (e.g., MeSH (National Library of Medicine 1999), SNOMED (Spackman et al. 1997), Read Codes (Schultz et al. 1997), GALEN (Rector et al. 1993)). Unlike Terminologia these vocabularies are entirely computer-based, and therefore lend themselves for incorporation in computer-based applications.

Classification and ontology projects to date have focused primarily on arranging the terms of a particular domain in hierarchies. As noted with respect to the evaluation of Terminologia Anatomica (Rosse 2000), insufficient attention has been paid to the relationships among these terms. These relationships are named (e.g., "*is-a*" and "*part-of*") to indicate how the entities connected by them are related (e.g., Left Lobe of Liver *part-of* Liver). Linking entities with relations encodes knowledge and is used by computer reasoning applications in making inferences. Terminologia, as well as anatomy sections of the controlled medical terminologies, mix -*is a*- and -*part of*- relationships in the anatomy segments of their hierarchies. Although such heterogeneity does not interfere with using these term lists for keyword-based retrieval, these programs will

fail to support higher level knowledge (reasoning) required for knowledge-based applications. To meet this gap, the Foundational Model of Anatomy (FMA) was developed to define a comprehensive symbolic description of the structural organization of the body, including anatomical concepts, their preferred names and synonyms, definitions, attributes and relationships (Rosse et al. 1998a, b; Rosse and Mejino 2003).

In the FMA, anatomical entities are arranged in class-subclass hierarchies, with inheritance of defining attributes along the *is-a* link, and other relationships (e.g., parts, branches, spatial adjacencies) represented as additional descriptors associated with the concept. The FMA currently consists of over 75,000 concepts, represented by about 120,000 terms, and arranged in over 2.1 million links using 168 types of relationships. These concepts represent structures at all levels: macroscopic (to 1 mm resolution), cellular and macromolecular. Brain structures have been added by integrating NeuroNames with the FMA as a Foundational Model of Neuroanatomy (FMNA) (Martin et al. 2001).

The FMA can be useful for symbolically organizing and integrating biomedical information, particularly that obtained from images. But in order to answer non-trivial queries in neuroscience and other basic science areas, and to develop "smart tools" that rely on deep knowledge, additional ontologies must also be developed (e.g., for physiological functions mediated by neurotransmitters, and pathological processes and their clinical manifestations, as well for the radiological appearances with which they correlate). The relationships that exist among these concepts and anatomical parts of the body must also be explicitly modeled. Next-generation informatics efforts that link the FMA and other anatomical ontologies with separately developed functional ontologies will be needed in order to accomplish this type of integration.

Knowledge Representation of Radiology Imaging Features

While FMA provides a comprehensive knowledge representation for anatomy, it does not cover other portions of the radiology domain. As is discussed in ▶ Chap. 7, there are controlled terminologies in other domains, such as MeSH, SNOMED, and related terminologies in the UMLS (Cimino 1996; Bodenreider 2008); however, these lack terminology specific to radiology for describing the features seen in imaging. The Radiological Society of North America (RSNA) recently developed RadLex, a controlled terminology for radiology (Langlotz 2006; Rubin 2008). The primary goal of RadLex is to provide a means for radiologists to communicate clear, concise, and orderly descriptions of imaging findings in understandable, unambiguous language. Another goal is to promote an orderly thought process and logical assessments and recommendations based on observed imaging features based on terminology-based description of radiology images and to enable decision support (Baker et al. 1995; Burnside et al. 2009). Another goal of RadLex is to enable radiology research; data mining is facilitated by the use of standard terms to code large collections of reports and images (Channin et al. 2009a, b).

RadLex includes thousands of descriptors of visual observations and characteristics for describing imaging abnormalities, as well as terms for naming anatomic structures, radiology imaging procedures, and diseases (◘ Fig. 10.11). Each term in RadLex contains a unique identifier as well as a variety of attributes such as definition, synonyms, and foreign language equivalents. In addition to a lexicon of standard terms, the RadLex ontology includes term relationships—links between terms to relate them in various ways to encode radiological knowledge. For example, the *is-a* relationship records subsumption. Other relationships include part-of, connectivity, and blood supply. These relationships are enabling computer-reasoning applications to process image-related data annotated with RadLex.

RadLex has been used in several imaging informatics applications, such as to improve search for radiology information. RadLex-based indexing of radiology journal figure captions achieved very high precision and recall, and significantly improved

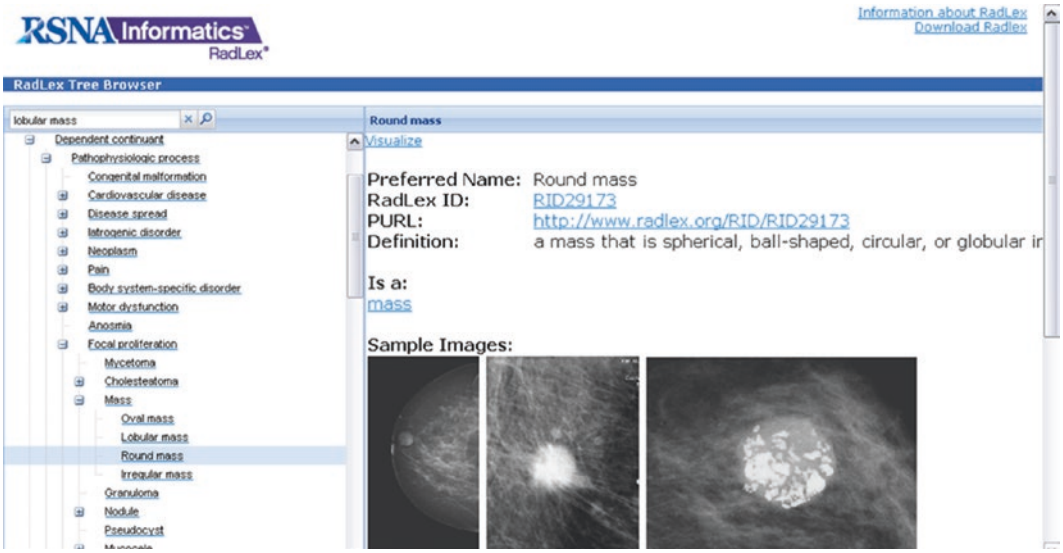


Fig. 10.11 RadLex controlled terminology (► http://radlex.org). RadLex includes term hierarchies for describing anatomy ("anatomical entity"), imaging observations ("imaging observation") and characteristics ("imaging observation characteristic"), imaging procedures and procedure steps ("procedure step"), diseases ("pathophysiologic process"), treatments ("treatment"), and components of radiology reports ("report"). Each term includes definitions, preferred name, image exemplars, and other term metadata and relationships such as subsumption

image retrieval over keyword-based search (Kahn and Rubin 2009). RadLex has been used to index radiology reports (Marwede et al. 2008). Work is underway to introduce RadLex controlled terms into radiology reports to reduce radiologist variation in use of terms for describing images (Kahn Jr. et al. 2009). Tools are beginning to appear enabling radiologists to annotate and query image databases using RadLex and other controlled terminologies (Rubin et al. 2008b; Channin et al. 2009a, b).

In addition to RadLex, there are other important controlled terminologies for radiology. The Breast Imaging Reporting and Data System (BI-RADS) is a lexicon of descriptors and a reporting structure comprising assessment categories and management recommendations created by the American College of Radiology (D'Orsi and Newell 2007). Terminologies are also being created in other radiology imaging domains, including the Fleischner Society Glossary of terms for thoracic imaging (Hansell et al. 2008), the Nomenclature of Lumbar Disc Pathology (Appel 2001), terminologies for image guided tumor ablation (Goldberg et al. 2009) and transcatheter therapy for hepatic malignancy (Brown et al. 2009), and the CT Colonography Reporting and Data System (Zalis et al. 2005).

Knowledge Representation of Radiology Procedures

A very important type of knowledge representation for images is the type of imaging procedure that produced it. While RadLex contains atomic terms for the various modalities, such as CT and MRI, it lacks the full spectrum of types of procedures that can be performed on a patient. The RadLex Playbook (Wang et al. 2017) is a project of the Radiological Society of North America that provides a standard system for naming radiology procedures, based on atomic terms (usually from RadLex) that define an imaging procedure, such as "CT Head." By providing standard names and codes for radiologic studies, the RadLex Playbook can facilitate a variety of operational and quality improvement efforts, including workflow optimization, chargemaster management, radiation dose tracking, enterprise integration and image exchange (Wang et al. 2017).

The RadLex Playbook grammar describes how to create the pre-coordinated Playbook terms across the defining name attributes. Each such term is comprised of several RadLex atomic terms. The unique combination of RadLex clinical terms defines a unique Playbook term, which is given a unique identifier (the RadLex Playbook ID, or RPID). Thus, for each RPID there is a corresponding set of RadLex IDs that link to the associated RadLex clinical terms. This knowledge representation can be very useful for retrieving particular types of images from systems that support Playbook, such as "retrieve all CT of the head," which would include *CT Head w/wo* (with and without contrast agent), *CT Head Angio w/wo*, *CT Orbits wo*, *CT Temporal Bone w/wo*, etc.

The Logical Observation Identifiers, Names, and Codes (LOINC®) terminology includes radiology terms (Vreeman and McDonald 2005), and recently a unified model LOINC/Playbook model and terminology for radiology procedure names has been created that represents the attributes of term names with an extensible set of values and provides LOINC codes and display name for each procedure (Vreeman et al. 2018). There is also a single integrated governance process for managing the unified terminology.

Semantic Representation of Image Contents

While ontologies and controlled terminologies are useful for representing knowledge related to images, they do not provide a means to directly encode assertions for recording the semantic content in images. For example, we may wish to record the fact that "there is a mass 4x5 cm in size in the right lobe of the liver." The representation of this semantic image content certainly will use ontologies and terminologies to record the entities to which such assertions refer; however, an **information model** is required to provide the required grammar and syntax for recording such assertions. There are two approaches to recording these assertions, no formal information model (narrative text) and a formal information model.

Narrative text

In the current workflow, nearly all semantic image content is recorded in narrative text (radiology reports). The advantage of text reports is that they are simple, quick to produce (the radiologist speaks freely into a microphone), and they can be expressive, capturing the subtle nuances (and ambiguities) that the English language provides. There are several downsides, however. First, text reports are unstructured; there is no adherence to controlled terminology and not consistent structure that would permit reliable information extraction. Second, the reports may be incomplete, vague, or contradictory. Further, free text is challenging for computers (see ► Chap. 8), which makes it difficult to leverage free text in applications. Finally, radiology images and the corresponding radiologist report are currently disconnected; e.g., the report may describe a mass in an organ, and the image may contain a **region of interest** (ROI) measuring the lesion, but there is no information directly linking the description of the lesion in the report with the ROI in the image. Such linkage could enable applications such as content-based image retrieval, as described below.

Structured reporting of radiology results has recently become increasingly the standard process for generating reports, usually using macros and templates (Weiss and Langlotz 2008; Langlotz 2009; Schwartz et al. 2011). In structured reports, a variety of fields is provided, such as a list of organs visualized on the image and a list of radiologist observations about them. Though structured reports improve on the ability to recognize and extract particular types of information in reports, they generally do not use controlled terminologies, and their content is usually recorded in narrative free text, so this method of recording semantic information about images is not much more computer-accessible than a fully narrative radiology report.

Information model

An information model provides an explicit specification of the types of data to be collected and the syntax by which it will be saved.

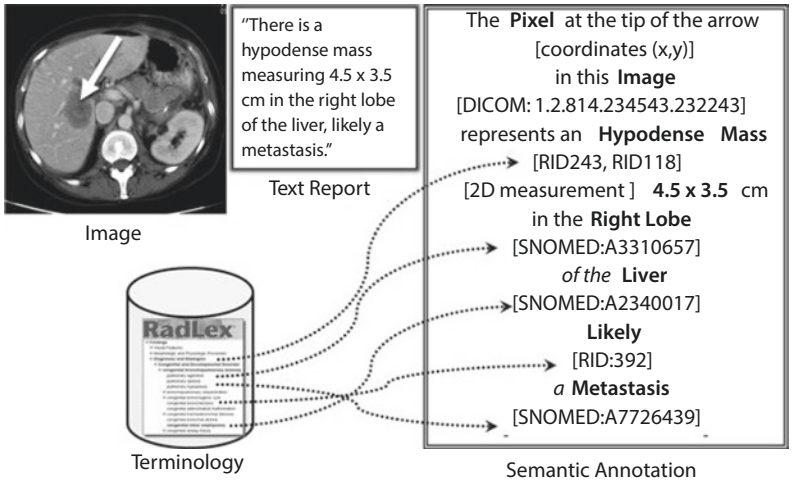


Fig. 10.12 Semantic annotation of images. The radiologist's image annotation (left) and interpretation (middle) associated with the annotation are not represented in a form such that the detailed content is directly accessible. The same information can be put into a structured representation as a semantic annotation (right), comprising terms from controlled terminologies (Systematized Nomenclature of Medicine (SNOMED) and RadLex) as well as numeric values (coordinates and measurements). (Figure reprinted with permission from (Rubin and Napel 2010). © Thieme)

So-called "semantic annotation" methods are being developed to adapt the semantic content about images that would have been put into narrative text so that it can instead be put in structured annotations compliant with the information model. The information model conveys the pertinent image information explicitly and in human-readable and machine accessible format. For example, a semantic annotation might record the coordinates of the tip of an arrow and indicate the organ (anatomic location) and imaging observations (e.g., mass) in that organ. These annotations can be recorded in a standard, searchable format, such as the Annotation and Image Markup (AIM) schema, developed by the National Cancer Imaging Program of NCI for storing and sharing image metadata (caBIG In-vivo Imaging Workspace 2008; Rubin et al. 2008a, 2009a). AIM captures a variety of information about image annotations, e.g., regions of interest, a lesion identification label, lesion location, measurements of image regions, method of measurement, radiologist observations, anatomic locations of abnormalities, calculations, inferences, and other qualitative and quantitative image features (Channin et al. 2009a, b). The image metadata also include information about the image, such as the name of imaging procedure and how or when the image was acquired. AIM supports controlled terminologies, enabling semantic interoperability. AIM has recently been incorporated into the DICOM Structured Report (DICOM SR) standard (DICOM Standards Committee - Working Group 8 – Structured Reporting 2017), with specifications for saving AIM in DICOM-SR (DICOM Standards Committee 2017). Given that DICOM is the international standard for specifying image data, it is hoped that there will be widespread adoption of the AIM/DICOM-SR format to enable interoperability of image annotations across systems.

The AIM information model includes use of controlled terms as semantic descriptors of lesions (e.g., RadLex). It also provides a syntax associating an ROI in an image with the aforementioned information, enabling raw image data to be linked with semantic information, and thus bridges the current disconnect between semantic terms and the lesions in images being described. In conjunction with RadLex, the AIM information model provides a standard syntax (in XML schema) to create a structured representation of the semantic contents of images (Fig. 10.12). Once the semantic contents are recorded in

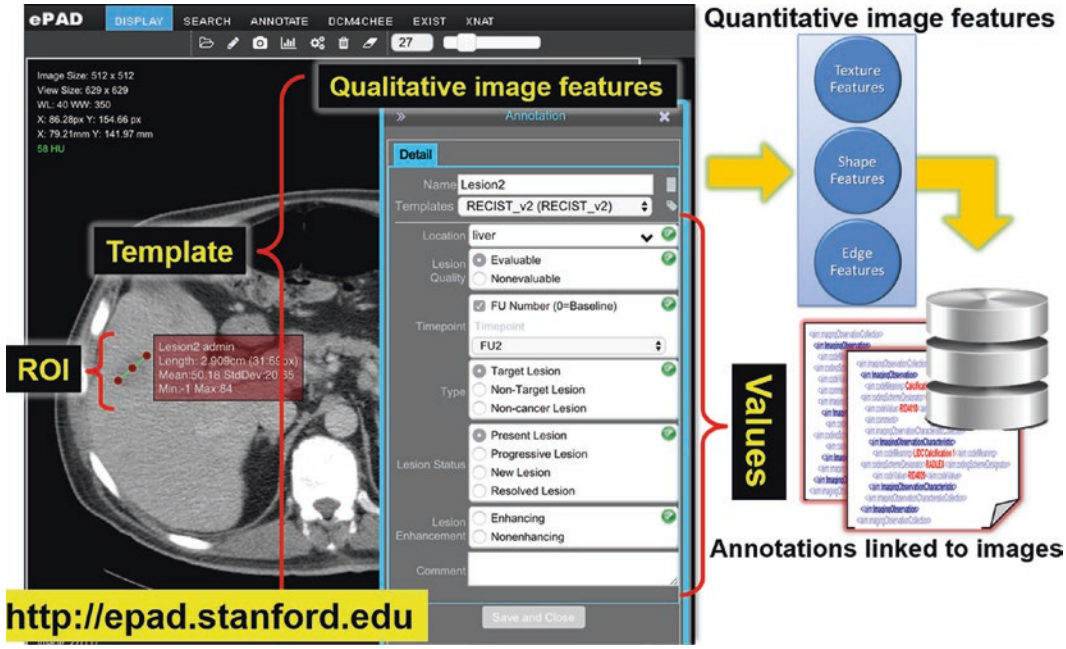


Fig. 10.13 The electronic Imaging Physician Annotation Device (ePAD). This tool creates structured semantic annotations on images using a graphical interface to minimize impact on image viewing workflow. The user views the image in and draws a region of interest (left). ePAD incorporates ontologies so that users can specify controlled terms as values in making their annotations (pull down panel on right). As they make their annotation, they receive feedback to ensure data entries are complete and that there are no violations of pre-specified annotation logic (panel on lower right). The ePAD tool saves image annotations in the AIM information model XML format

10

AIM (as XML instances of the AIM XML schema), applications can be developed for image query and analysis.

AIM has been gaining traction in the research community. A number of diverse research projects have embraced and have been enabled by AIM (Levy et al. 2009, Napel et al. 2010, Gevaert et al. 2011, Gimenez et al. 2011a, b, Hoang et al. 2011, Levy and Rubin 2011a, b, Napel et al. 2011, Plevritis et al. 2011, Gevaert et al. 2012a, b; Levy et al. 2012). An increasing number of tools are supporting AIM to facilitate creating semantic annotations on images as part of the image viewing workflow are being developed, including open source projects such as Osirix (Rubin and Snyder 2011), ClearCanvas (Klinger 2010; National Cancer Institute 2012), Slicer (Pieper et al. 2004; Fedorov 2012; Fedorov et al. 2012), and ePAD (Rubin and Snyder 2011). There are also several commercial applications using AIM that are in development (Rubin et al. 2010; Zimmerman et al. 2011). Automated semantic image annotation methods are also being pursued (Carneiro et al. 2007; Mechouche et al. 2008; Yu and Ip 2008) that will ultimately make the process of generating this structured information efficient.

The electronic Imaging Physician Annotation Device (ePAD (Rubin and Snyder 2011)) is a freely available web-based image viewing and AIM-compliant annotation tool. ePAD permits the user to draw image annotations in a manner in which they are accustomed while viewing images, while simultaneously collecting semantic information about the image and the image region directly from the image itself as well as from the user using a structured reporting template (Fig. 10.13). The tool also features a panel to provide feedback so as to ensure complete and valid annotations. Image annotations are saved in the AIM XML format.

By making the semantic content of images explicit and machine-accessible, these structured annotations of images will help radiologists analyze data in large databases of images.

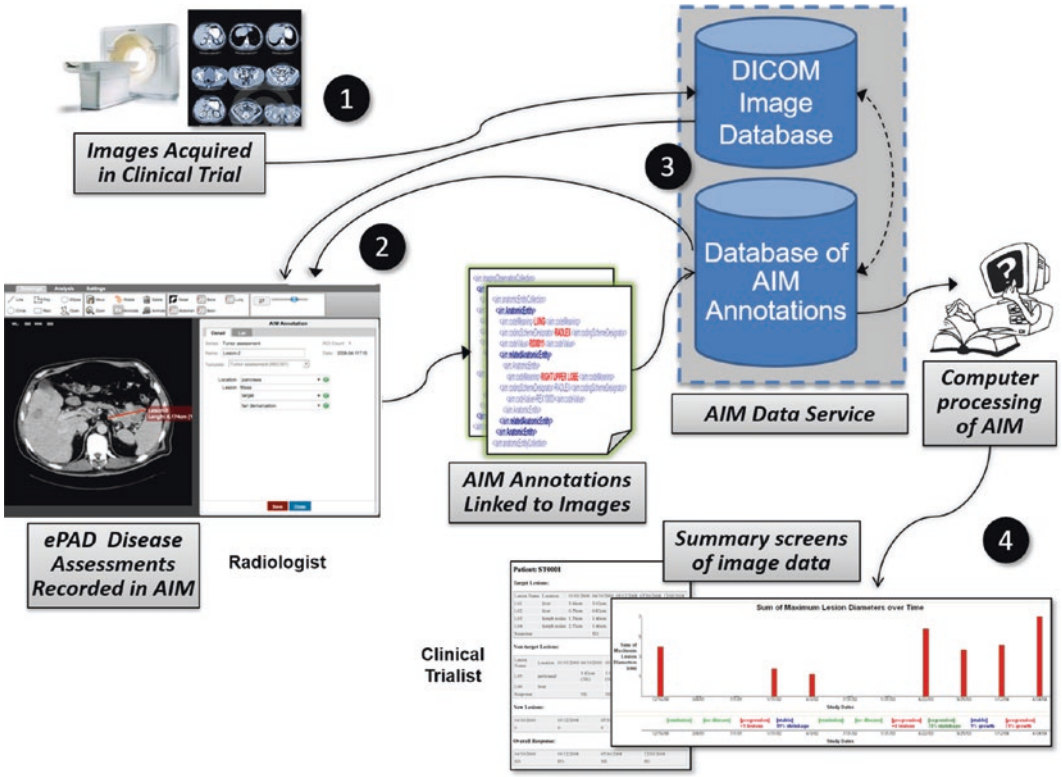


Fig. 10.14 ePAD and AIM within the clinical/research environment. (1) Images are acquired and stored in the hospital PACS, (2) the Radiologist uses ePAD to review the images and to make measurements on cancer lesions, (3) The measurements (saved as AIM XML in ePAD) with links to the images are stored by the AIM Data Service, and (4) a variety of software applications can use the AIM Data Service to access the image metadata for different purposes, such as listing the lesion measurements or generating a summary of patient response assessment

Figure 10.14 shows how image annotations in AIM can be integrated into routine research and clinical workflows. Images acquired from imaging devices flow into the PACS and can be viewed on an imaging workstation. If the imaging workstation supports AIM (e.g., ePAD as shown in the figure), then image annotations and radiologist observations are saved in the AIM format and stored in database of AIM annotation files (an XML database in the case of ePAD). The AIM annotations have a pointer to their corresponding images and queries and analyses can be done on the image annotations for clinical applications, such as summarizing the change in cancer lesion sizes over time for assessing response assessment to cancer (Fig. 10.14). Cancer patients often have many serial imaging studies in which a set of lesions is evaluated at each time point. Automated tools such as ePAD can use semantic image annotations to identify the measurable lesions at each time point and produce a summary of, and automatically reason about, the total tumor burden over time, helping physicians to determine how well patients are responding to treatment (Levy and Rubin 2008).

Atlases

Spatial representations of anatomy, in the form of segmented regions on 2-D or 3-D images, or 3-D surfaces extracted from image volumes, are often combined with symbolic representations to form digital atlases. A digital atlas (which for this chapter refers to an atlas created from 3-D image data taken from real subjects, as opposed to artists' illustrations) is generally created from a single indi-

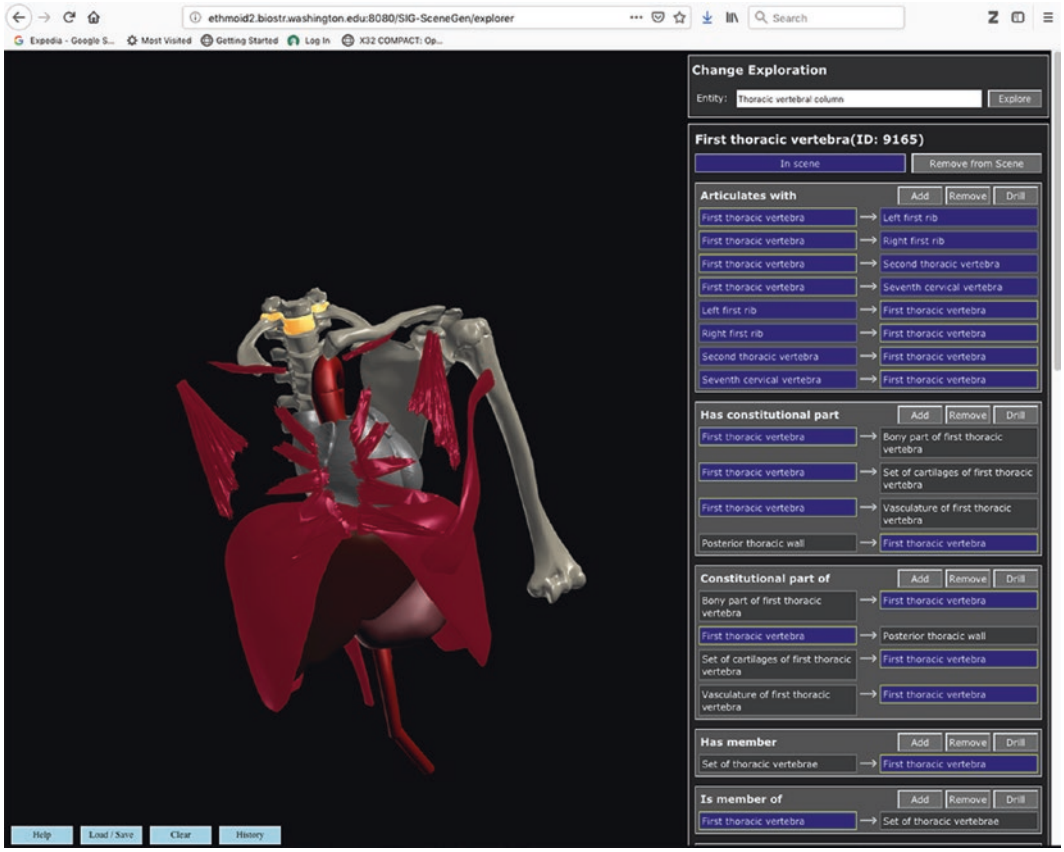


Fig. 10.15 The Digital Anatomist Dynamic Scene Generator. The user can select a set of 3-D anatomical models, in which each model is associated with an entity from the Foundational Model of Anatomy (FMA) to select a starting structure, load relations from the FMA between that structure and related structures, and, if 3-D models are available, to add those models to the scene. The evolving scene can be manipulated in real-time on the web and can be saved as a standalone scene that can be saved locally or accessed via URL, thus permitting it to be embedded in other web apps. The scene shown here was created by first requesting one of the muscles of the thorax, then exploring FMA to find the heart, aorta and branches, and thoracic vertebral column. In the scene on the left, the first thoracic vertebra was clicked, which caused it to be highlighted, and the relations between it and other structures are then shown in the panel on the right. The top pane of the right panel = shows the structures with which the first thoracic vertebra articulates. (Figure used with permission from Jim Brinkley)

vidual, which therefore serves as a "canonical" instance of the species. Traditionally, atlases have been primarily used for education, and most digital atlases are used the same way.

As an example in 2-D, the Digital Anatomist Interactive Atlases (Sundsten et al. 2000) were created by outlining ROIs on 2-D images (many of which are snapshots of 3-D scenes generated by reconstruction from serial sections) and labeling the regions with terminology from the FMA. The atlases, which are available on the web, permit interactive browsing, where the names of structures are given in response to mouse clicks; dynamic creation of "pin diagrams", in which selected labels are attached to regions on the images; and dynamically-generated quizzes, in which the user is asked to point to structures on the image (Brinkley et al. 1997).

As an example 3-D, the Digital Anatomist Dynamic Scene Generator (DSG, Fig. 10.15) creates interactive 3-D atlases "on-the-fly" for viewing and manipulation over the web (Brinkley et al. 1999; Wong et al.

1999). An example of a 3-D brain atlas created from the Visible Human is Voxelman (Hohne et al. 1995), in which each voxel in the Visible Human head is labeled with the name of an anatomic structure in a "generalized voxel model" (Hohne et al. 1990), and highly-detailed 3-D scenes are dynamically generated. Several other brain atlases have also been developed, primarily for educational use (Johnson and Becker 2001; Stensaas and Millhouse 2001).

Atlases have also been developed for integrating functional data from multiple studies (Bloom and Young 1993; Toga et al. 1994, 1995; Swanson 1999; Fougerousse et al. 2000; Rosen et al. 2000; Martin and Bowden 2001). In their original published form these atlases permit manual drawing of functional data, such as neurotransmitter distributions, onto hardcopy printouts of brain sections. Many of these atlases have been or are in the process of being converted to digital form. The Laboratory of Neuroimaging (LONI) at the University of California Los Angeles has been particularly active in the development and analysis of digital atlases (Toga 2001), and the California Institute of Technology Human Brain Project has released a web-accessible 3-D mouse atlas acquired with micro-MR imaging (Dhenain et al. 2001).

The most widely used human brain atlas is the Talairach atlas, based on post mortem sections from a 60-year-old woman (Talairach and Tournoux 1988). This atlas introduced a proportional coordinate system (often called "Talairach space") which consists of 12 rectangular regions of the target brain that are piecewise affine transformed to corresponding regions in the atlas. Using these transforms (or a simplified single affine transform based on the anterior and posterior commissures) a point in the target brain can be expressed in Talairach coordinates, and thereby related to similarly transformed points from other brains. Other human brain atlases have also been developed (Schaltenbrand and Warren 1977; Hohne et al. 1992; Caviness et al. 1996; Drury and Van Essen 1997; Van Essen and Drury 1997).

10.4 Image Processing

Image processing is a form of **signal processing** in which computational methods are applied to an input image to produce an output image or a set of characteristics or parameters related to the image. Most image processing techniques involve treating the image as a two-dimensional signal and analyzing it using signal-processing techniques or a variety of other transformations or computations. There are a broad variety of image processing methods, including transformations to enhance visualization, computations to extract features, and systems to automate detection or diagnose abnormalities in the images. The latter two methods, referred to as computer-assisted detection and diagnosis (CAD) is discussed in ▶ Sect. 10.5.2. In this section we discuss the former methods, which are more elemental and generic processing methods.

The rapidly increasing number and types of digital images has created many opportunities for image processing, since one of the great advantages of digital images is that they can be manipulated just like any other kind of data. This advantage was evident from the early days of computers, and success in processing satellite and spacecraft images generated considerable interest in biomedical image processing, including automated image analysis to improve radiological interpretation. Beginning in the 1960s, researchers devoted a large amount of work to this end, with the hope that eventually much of radiographic image analysis could be improved. One of the first areas to receive attention was automated interpretation of chest X-ray images, because, previously, most patients admitted to a hospital were subjected to routine chest X-ray examinations. (This practice is no longer considered cost effective except for selected subgroups of patients.) Interestingly, recent research in deep learning (discussed below), however, has raised enthusiasm and hopes for automating certain tasks of radiographic image interpretation, such as detection of pneumonia (Rajpurkar et al. 2017). While most of the emphasis of image processing

continues to be on systems that aid the user in viewing and manipulating images, there is a quickly growing body of work completely automated analysis of images, particularly for lesion detection, image segmentation, and image classification (diagnosis).

Medical image processing utilizes tools similar to general image processing. But there are unique features to the medical imagery that present different, and often more difficult, challenges from those that exist in general image processing tasks. To begin with, the images analyzed all represent the 3D body; thus, the information extracted (be it in 2D or 3D) is based on a 3D volumetric object. The images themselves are often taken from multi-modalities (CT, MRI, PET), where each modality has its own unique physical characteristics, leading to unique noise, contrast and other issues that need to be addressed. The fusion of information across several modalities is a challenge that needs to be addressed as well.

When analyzing the data, it is often desirable to segment and characterize specific organs. The human body organs, or various tissue of interest within them, cannot be described with simple geometrical rules, as opposed to objects and scenes in non-medical images that usually can be described with such representations. This is mainly because the objects and free-form surfaces in the body cannot easily be decomposed into simple geometric primitives. There is thus very little use of geometric shape models that can be defined from *a-priori* knowledge. Moreover, when trying to model the shape of an organ or a region, one needs to keep in mind that there are large inter-person variations (e.g., in the shape and size of the heart, liver and so on), and, as we are frequently analyzing images of patients, there is a large spectrum of abnormal states that can greatly modify tissue properties or deform structures. Finally, especially in regions of interest that are close to the heart, complex motion patterns need to be accounted for as well. These issues make medical image processing a very challenging domain.

The widespread availability of digital images, combined with image management systems such as PACS (▶ Chap. 22) and powerful workstations, has led to many applications of image processing techniques. In general, routine techniques are available on the manufacturer's workstations (e.g., a vendor-provided console for an MR machine or an ultrasound machine), whereas more advanced image-processing algorithms are available as software packages that run on independent workstations.

The primary uses of image processing in the clinical environment are for image enhancement, screening, and quantitation. Software for such image processing is primarily developed for use on independent workstations. Several journals are devoted to medical image processing (e.g., *IEEE Transactions on Medical Imaging*, *Journal of Digital Imaging*, *Neuroimage*), and the number of journal articles is rapidly increasing as digital images become more widely available. Several books are devoted to the spectrum of digital imaging processing methods (Yoo 2004; Gonzalez et al. 2009), and the reader is referred to these for more detailed reading on these topics. We describe a few examples of image-processing techniques in the remainder of this section.

10.4.1 Types of Image-Processing Methods

Image processing methods are applied to representations of image content (▶ Sect. 10.3). One may use the very low-level, pixel representation. The computational effort is minimal in the representation stage, with substantial effort (computational cost) in further analysis stages such as segmentation of the image, matching between images, registration of images, etc. A second option is to use a very high-level image content representation, in which each image is labeled according to its semantic content (medical image categories such as "abdomen vs chest", "healthy vs pathology"). In this scenario, a substantial computational effort is needed in the representation stage, including the use of automated image segmentation methods to recognize ROIs as well as advanced learning techniques to clas-

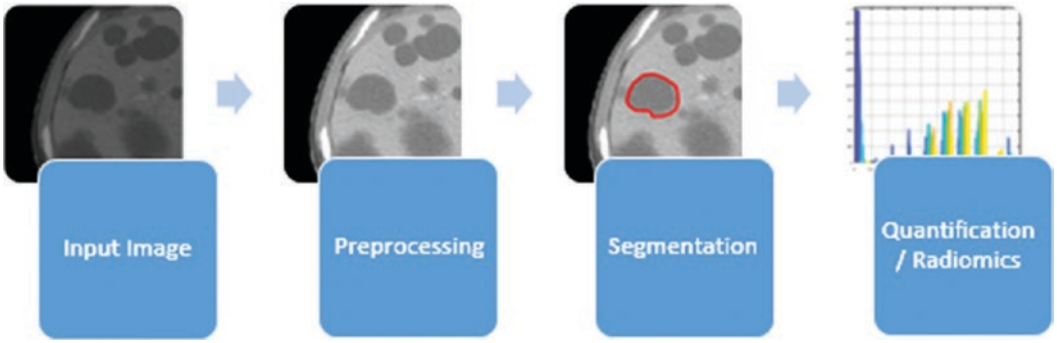


Fig. 10.16 Diagram of a typical image processing pipeline

sify the regions of image content. Further analysis can utilize knowledge resources such as ontologies, linked to the images using category labels. A mid-level representation exists, that balances the above two options, in which a transition is made from pixels to semantic features. Feature vectors are used to represent the spectrum of image content compactly and subsequent analysis is done on the feature vector representation.

Image processing is the foundation for creating image-based applications, such as image enhancement to facilitate human viewing, to show views not present in the original images, to flag suspicious areas for closer examination by the clinician, to quantify the size and shape of an organ, and to prepare the images for integration with other information. To create such applications, several types of image processing are generally performed sequentially in an *image processing pipeline*, although some processing steps may feed back to earlier ones, and the specific methods used in a pipeline varies with the application. Most image processing pipelines and applications generalize from two-dimensional to three-dimensional images, though three-dimensional images pose unique image processing opportunities and challenges. Image processing pipelines are generally built using one or more of the following fundamental image processing methods: **global processing**, **image enhancement**, **image rendering/visualization**, **image quantitation**, image **segmentation**, image **registration**, and **image reasoning** (e.g., classification). Those steps are shown in Fig. 10.16**.** In the remainder of this section we describe these methods, except for image reasoning which is discussed in ▶ Sect. 10.5.

10.4.2 Global Processing

Global processing involves computations on the entire image, without regard to specific regional content. The purpose is generally to enhance an image for human visualization or for further analysis by the computer ("preprocessing"). A simple but important example of global image processing is gray-scale windowing of CT images. The CT scanner generates pixel values (**Hounsfield numbers**, or CT numbers) in the range of −1000 to +3000. Humans, however, cannot distinguish more than about 100 shades of gray. To appreciate the full precision available with a CT image, the operator can adjust the midpoint and range of the displayed CT values. By changing the level and width (i.e., intercept and slope of the mapping between pixel value and displayed gray scale or, roughly, the brightness and contrast) of the display, radiologists enhance their ability to perceive small changes in contrast resolution within a subregion of interest.

Other types of global processing change the pixel values to produce an overall enhancement or desired effect on the image: *histogram equalization*, *convolution*, and *filtering*. In histogram equalization, the pixel values are changed, spreading out the most frequent intensity values to increase the global contrast of the image. It is most effective when the usable data of the image are represented by a

narrow range of contrast values. Through this adjustment, the intensities can be better distributed on the histogram, improving image contrast by allowing for areas of lower local contrast to gain a higher contrast. In convolution and filtering, mathematical functions are applied to the entire image for a variety of purposes, such as de-noising, edge enhancement, and contrast enhancement.

10.4.3 Image Enhancement

Image enhancement uses global processing to improve the appearance of the image either for human use or for subsequent processing by computer. The consoles of all vendor image viewing platforms and independent image-processing workstations provide some form of image enhancement. We have already mentioned CT windowing. Another technique is unsharp masking, in which a blurred, or "unsharp," positive is created to be used as a "mask" that is combined with the original image, creating the illusion that the resulting image is sharper than the original. The technique increases local contrast and enhances the visibility of fine-detail (high-frequency) structures. Histogram equalization spreads the image gray levels throughout the visible range to maximize the visibility of those gray levels that are used frequently. Temporal subtraction subtracts a reference image from later images that are registered to the first. A common use of temporal subtraction is **digital-subtraction angiography** (DSA) in which a background image is subtracted from an image taken after the injection of contrast material.

10.4.4 Image Rendering/ Visualization

Image rendering and visualization refer to a variety of techniques for creating image displays, diagrams, or animations to display images more in a different perspective from the raw images. Image volumes are comprised of a stack of 2-D images. If the voxels in each image are isotropic, then a variety of arbitrary projections can be derived from the volume, such as a sagittal or coronal view, or even curved planes. A technique called maximum intensity projection (MIP) and minimum intensity projection (MinIP) can also be created in which imaginary rays are cast through the volume, recording the maximum or minimum intensity encountered along the ray path, respectively, and displaying the result as a 2-D image.

In addition to these planar visualizations, the volume can be visualized directly in its entirety using **volume rendering** techniques (Lichtenbelt et al. 1998) (◘ Fig. 10.17) which project a two-dimensional image directly from a three-dimensional voxel array by casting rays from the eye of the observer through the volume array to the image plane. Because each ray passes through many voxels, some form of segmentation (usually simple thresholding) often is used to remove obscuring structures. As workstation memory and processing power have advanced, volume rendering has become widely used to display all sorts of three-dimensional voxel data—ranging from cell images produced by confocal microscopy, to three-dimensional ultrasound images, to brain images created from MRI or PET.

Volume images can also be given as input to image-based techniques for warping the image volume of one structure to other. However, more commonly the image volume is processed in order to extract an explicit *spatial* (or quantitative) representation of anatomy (▶ Sect. 10.4.5). Such an explicit representation permits improved visualization, quantitative analysis of structure, comparison of anatomy across a population, and mapping of functional data. It is thus a component of most research involving 3-D image processing.

10.4.5 Image Quantitation

Image quantitation is the process of extracting useful numerical parameters or deriving calculations from the image or from ROIs in the image (often as part of "radiomics" analyses, described below) (Scheckenbach et al. 2017;

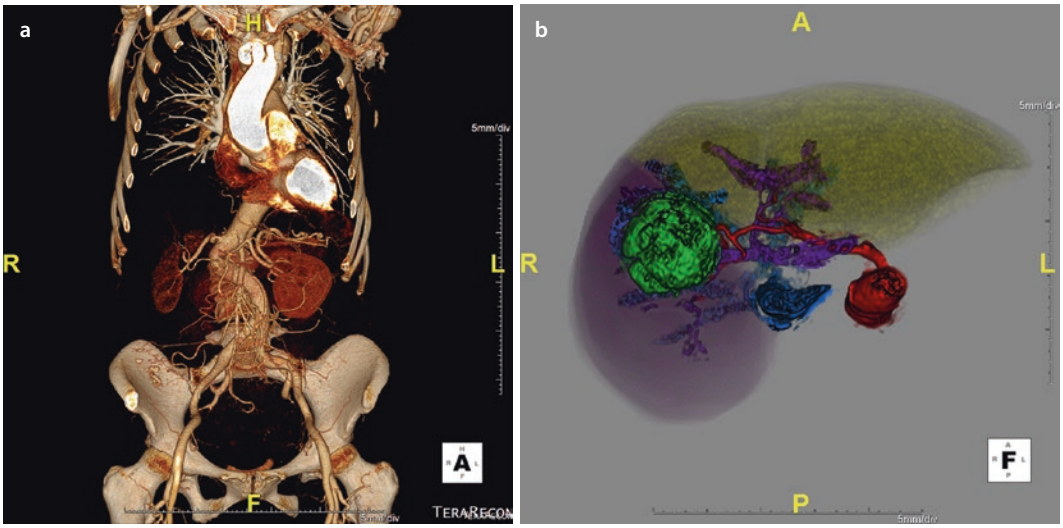


Fig. 10.17 Three-dimensional CT scan of the full abdomen and pelvis **a** and of the liver **b**. The volumetric CT scan comprises a set of contiguous slices containing isotropic voxels (the dimension of pixels in the x,y plane is the same as in the z dimension). The volume is rendered directly using volume-rendering techniques

Lohmann et al. 2018; Valdora et al. 2018). These values are also referred to as "quantitative imaging features." These parameters may themselves be informative—for example, the volume of the heart or the size of the fetus. They also may be used as input into an automated classification procedure, which determines the type of object found. For example, small round regions on chest X-ray images might be classified as tumors, depending on such features as intensity, perimeter, and area.

Mathematical models often are used in conjunction with image quantitation. In classic pattern-recognition applications, the mathematical model is a classifier (learned using some type of supervised machine learning) that assigns a label to the image; e.g., to indicate if the image contains an abnormality, or indicates the diagnosis underlying an abnormality.

Quantitative Image Features

Quantitation of images uses global processing and segmentation to characterize regions of interest in the image with numerical values. There are two kinds of *quantitative image features*, pre-defined image features and *learned features*. Pre-defined image features encode domain knowledge, since they are designed to capture specific characteristics of the image or image region, such as texture, shapes, lesion margin characteristics, and image noise. Pre-defined image features are generally designed to capture characteristics of the image that reflect underlying biology, tissue function, or disease. For example, heart size, shape, and motion are subtle indicators of heart function and of the response of the heart to therapy (Clarysse et al. 1997). Similarly, fetal head size and femur length, as measured on ultrasound images, are valuable indicators of fetal well-being (Brinkley 1993).

Pre-defined image features are quantitative representations of visual signals contained in an image. Two types of pre-defined image features are *photometric* features, which exploit color and texture cues, derived directly from raw pixel intensities, and *geometric* features, which use shape-based cues. While color is one of the visual cues often used for content description (Hersh et al. 2009), most medical images are grayscale. Texture features encode spatial organization of pixel values of an image region. Shape features describe in quantitative terms the contour of a lesion and complement the information captured by

color or texture (Depeursinge et al. 2014). In addition, the histogram of pixel values within an ROI or transforms on those values is commonly performed to compute quantitative image features.

Pre-defined image features are commonly represented by feature-vectors in a N-dimensional space, where each dimension of the feature vector describes an aspect of the individual pixel (e.g., color, texture, etc.) (Haralick and Shapiro 1992) Image analysis tasks that use the quantitative features, such as segmentation and classification are then approached in terms of distance measurements between points (samples) in the chosen N-dimensional feature space.

In contrast to pre-defined image features, *learned features* are derived by computational analysis of the image itself without incorporating any domain knowledge. Deep learning methods (described below) are a common and popular approach for deriving learned features. In deep learning, the goal is to learn a task directly from a large collection of images; the parameters of deep learning models reflect image features that are extracted during the learning of these models (Milletari et al. 2016; Yasaka et al. 2018).

Image Patches

In the last several years, "patch-based" representations and "bag-of-features" classification techniques have been proposed and used as an approach to processing image contents (Jurie and Triggs 2005; Nowak et al. 2006; Avni 2009). An overview of the methodology is shown in Fig. 10.18, and represents one of the types of "feature learning" being used for automated computer analysis of images (the other type of feature learning, deep learning, is discussed below). In image patch approaches, a shift is made from the pixel as being the atomic entity of analysis to a "patch" – a small window centered on the pixel, thus region-based information is included. A very large set of patches is extracted from an image. Each small patch shows a localized "glimpse" at the image content; the collection of thousands and more such patches, randomly selected, have the capability to identify the entire image content (similar to a puzzle being formed from its pieces).

Patch extraction approaches include using a regular sampling grid, a random selection of points, or the selection of points with high information content using salient point detectors, such as SIFT (Lowe 1999). Once patches are selected, the information content within a patch is extracted. It is possible to take the patch information as a collection of pixel values, or to shift the representation to a different set of features based on the pixels, such as SIFT features. A final step in the process is to learn a dictionary of words over a large collection of patches, extracted from a large set of images. The vector represented patches are converted into "visual words" which form a representative "dictionary". A visual word can be considered as a representative of several similar patches. A frequently-used method is to perform K-means clustering (Bishop 1995) over the vectors of the initial collection, and then cluster them into K groups in the feature space. The resultant cluster centers serve as a vocabulary of K visual words, with K often in the hundreds and thousands.

Once a global dictionary is learned, each image is represented as a collection of words (also known as a "bag of words", or "bag of features"), using an indexed histogram over the defined words. Various image processing tasks can then be undertaken, ranging from the categorization of the image content, giving the image a "high-level," more semantic label, the matching between images, or between an image and an image class, using patches for image segmentation and region-of-interest detection within an image. For these various tasks, images are compared using a distance measure between the representative histograms. In categorizing an image as belonging to a certain image class, well-known classifiers, such as the k- nearest neighbor and support-vector machines (SVM) (Vapnik 2000), can be used.

Using patches in a bag-of-visual-words (BoW) representation was shown to be successful in general scene and object recognition

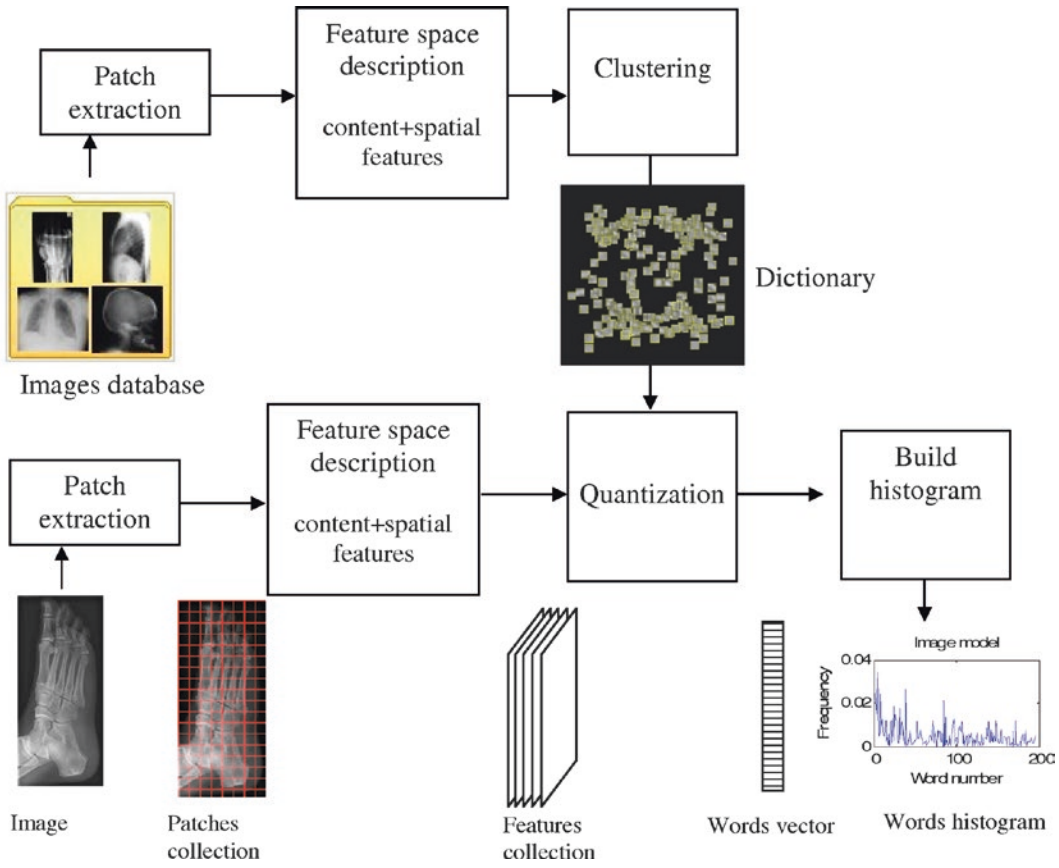


Fig. 10.18 A block diagram of the patch-based image representation. A radiographic image is shown with a set of patches indicated for processing the image data. Subsequent image processing is performed on each patch, and on the entire set of patches, rather than on individual pixels in the image. A dictionary of visual words is learned from a large set of images, and their respective patches. Further analysis of the image content can then be pursued based on a histogram across the dictionary words

tasks (Fei-Fei and Perona 2003, Varma and Zisserman 2003, Sivic and Zisserman 2003, Nowak et al. 2006, Jiang et al. 2007). A few research studies were conducted in the medical domain as well. For example, in (André et al. 2009) BoW was used as the representation of endomicroscopic images and achieved high accuracy in the tasks of classifying the images into neoplastic (pathological) and benign. In (Bosch et al. 2006) an application to texture representation for mammography tissue classification and segmentation was presented. The use of BoW techniques for large scale radiograph archive categorization can be found in the ImageCLEF competition, in a task to classify over 12,000 X-ray images to 196 different (organ-level) categories (Tommasi et al. 2010). This competition provides an important benchmarking tool to assess different feature sets as well as classification schemes on large archives of Radiographs. For several years, approaches based on local patch representation achieved the highest scores for categorization accuracy (Deselaers et al. 2006; Caputo et al. 2008; Greenspan et al. 2011).

Radiomics, Machine Learning and Deep Learning

Radiomics describes a broad set of computational methods that extract quantitative features from radiology images (though similar approaches can be applied to other image types like histopathology or ophthalmology images) (Kumar et al. 2012; Lambin et al. 2012; Grossmann et al. 2017). The term "radiomics" has been used to mean a variety

of concepts, being wide in scope to include several fields, including clinical radiology (eg, imaging interpretation), computer vision (eg, quantitative feature extraction), and machine learning (eg, classifier evaluation). In recent years, radiomics is commonly used to refer to the quantitation of image features in large collections of images, akin to the use of "omics" for large scale collection and analysis of other types of biomedical data. It includes all the quantification techniques described in this Chap.. A more distinctive and unique formalism for the term is to view the quantitation that it represents as one that focuses on the identification of quantitative imaging indicators that predict important clinical outcomes, e.g. prognosis and response or resistance to a specific cancer treatment (Zhou et al. 2018).

One motivation of radiomics is information integration—to merge image features with other known quantitative descriptors, including patient information and genomic data to generate a unique patient signature (or electronic phenotype). Initially, *pre-defined image features* were extracted, including for example a large number of quantifications defined from texture features, SIFT features, etc. (Napel et al. 2010) Once machine learning tools, specifically *deep learning* tools emerged, the latter have begun to take over several of the stages within the radiomics processing cycle – specifically, the generation of large sets of automatically extracted features for the quantification of the data, since deep learning models learn image features as part of the training process (Kontos et al. 2017; Giger 2018).

Radiomics relies on computational techniques in computer vision to extract many quantitative features from radiologic images. The extracted quantitative features are typically within a defined ROI that could include the whole tumor or specific regions within it. Computational image descriptors quantify visual characteristics at different scales from ROIs. For example, the scale-invariant feature transform (SIFT) (Lowe 1999) is computed through key point detection using a difference of Gaussian function and local image gradient measurement with radius and scale selections. This permits a quantitative measurement of the tumor shape so that subtle variations during treatment can be observed and quantified. Local-level feature extraction provides an image descriptor used to compare a pixel being tested with its immediate pixel neighborhood. This allows identification of a small area within an otherwise homogeneous, larger tumor region. This can be achieved, for example, with local binary patterns (LBP). These are local image descriptors sensitive to small monotonic gray-level differences (Ojala et al. 2002). Texture descriptors, such as the LBP are very common in ROI descriptions, including gray-level co-occurrence matrices (Haralick et al. 1973) that examine the spatial relationships of pixels through a series of statistical measures, and histogram of oriented gradients (HOG) (Dalal and Triggs 2005) features to quantify image-gradient statistics with multiple directions not obvious to radiologists.

Machine Learning is commonly used for discovering predictive radiomics features. In machine learning, the parameter space is searched for an imaging feature statistically associated with clinical outcome (Zhou et al. 2018). Before one evaluates machine-learning models, a specification for the medical diagnostic task is needed so that models can be appropriately trained. For example, supervised, unsupervised, and semisupervised learning models are fundamental learning strategies used in accordance with the different levels of available clinical outcome labels.

In *supervised learning*, the goal is to learn from a certain portion of trained samples with known class labels and to predict classes or numeric values for unknown patterns from large and noisy datasets. Conversely, *unsupervised learning* finds the natural structure from data without having any prior labels. As a hybrid setting, *semisupervised learning* needs only a small portion of labeled training data. The unlabeled data samples,

instead of being discarded, are also used in the learning process.

Deep Learning as a new frontier in machine learning that is quickly rising as a primary

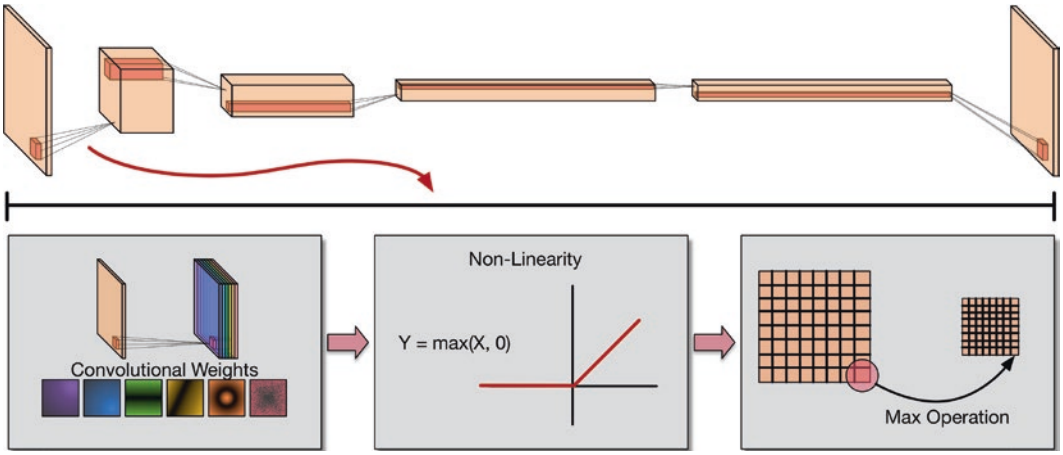


Fig. 10.19 Above is a simply 5-layer fully convolutional network. The first convolutional block is separated out into three components. The first is the actual 2-dimensional convolution. Different convolutional weights act on the input tensor to create the output tensor, each weight acting as a feature extractor. This is followed by a nonlinearity. Common nonlinearities include the sigmoid, hyperbolic tangent, and the rectified linear unit, or ReLU. Finally, should we need to conserve some memory, we can use pooling operations, such as a MaxPool layer to decrease the size of our tensors. This fundamental combination of convolutions, nonlinearities, and pooling operations is used in every convolutional neural network since its rise to prominence in 2012. (Figure courtesy of Darvin Yi)

approach to many image analysis problems, and it has advanced large-scale medical image analysis. (Greenspan et al. 2016; Shen et al. 2017). Excitement for using deep learning for attacking many image analysis problems in medical imaging has grown quickly because these methods were the first to be top performing methods in the ImageNet classification challenge (Krizhevsky et al. 2012), and much of medical imaging analysis is an image classification problem. The development of deep learning, as part of the machine-learning field, provided a new approach in which the input data is automatically quantified while being analyzed.

Deep learning has been termed one of the 10 breakthrough technologies as of 2013 (MIT Technology Review 2013). It is an improvement of artificial neural networks, architectures of computational units ("neurons"), which are designed in several (all the way to thousands) of layers ("deep") – where it was found that more layers permit higher levels of abstraction and improved predictions from data (LeCun et al. 2015). To date, it is emerging as the leading machine-learning tool in the general imaging and computer vision domains.

Deep learning methods have achieved record-breaking performances for numerous computer vision applications when the number of available training samples is sufficiently large (Deng et al. 2009).

Among the network architectures and models, convolutional neural networks (CNNs) have proven to be powerful tools for a broad range of computer vision tasks. The typical CNN architecture for image processing consists of a series of layers of convolution filters, followed by or interspersed with a series of data reduction or pooling layers (Fig. 10.19). The convolution filters are applied to small patches of the input image. Like the low-level vision processing in the human brain, the convolution filters detect increasingly more relevant image features, for example lines or circles that may represent straight edges (such as for organ detection) or circles (such as for round objects like colonic polyps), and then higher order features like local and global shape and texture. The output of the CNN is typically one or more probabilities or class labels. The convolution filters are learned from training data. This is desirable because it reduces the necessity of the time-consuming hand-crafting of features

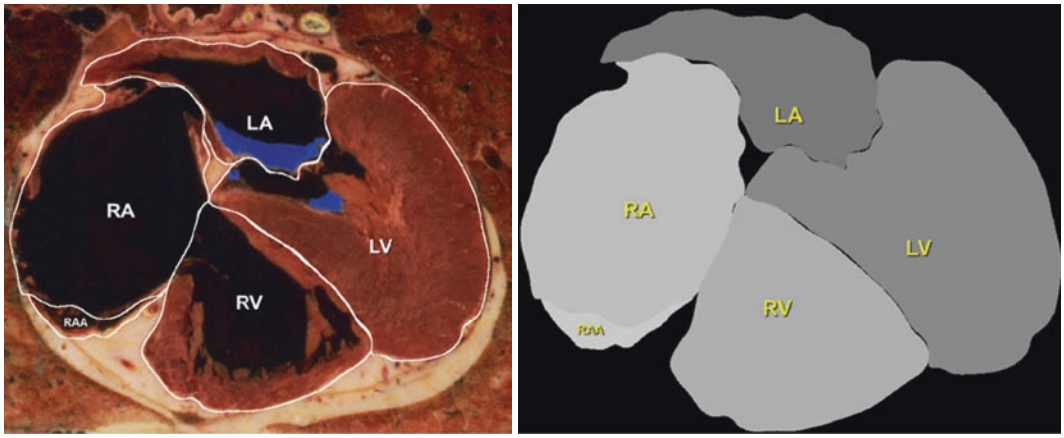


Fig. 10.20 Image segmentation. This figure illustrates the process of segmenting and labeling the chambers of the heart. On the left, a cross sectional atlas image of the heart has been segmented by hand and each chamber was labeled (RAA right atrial appendage, RA right atrium, LA left atrium, RV right ventricle, LV left ventricle). The boundary of each circumscribed anatomic region can be converted into a digital mask (right) which can be used in different applications where labeling anatomic structures in the image is needed

that would otherwise be required to pre-process the images with application-specific filters or by calculating computable features.

Deep CNNs automatically learn mid-level and high-level abstractions obtained from raw data (e.g., images). Recent results indicate that the generic descriptors extracted from CNNs are extremely effective in object recognition and localization in natural images. In the medical imaging domain, in many detection, classification and segmentation tasks, deep learning has proved to be the state-of-the-art foundation, leading to improved accuracy. It has also opened new frontiers in data analysis with rates of progress not before experienced. For an overview on deep learning for medical image quantification and analysis see (Greenspan et al. 2016; Zhou et al. 2017a).

10.4.6 Image Segmentation

Segmentation of images involves automatically circumscribing regions within an image to generate ROIs in the image. The ROIs usually correspond to anatomically meaningful structures, such as organs or parts of organs, or they may be lesions or other types of regions in the image pertinent to the application. The structures may be delineated by their borders, in which case edge-detection techniques (such as edge-following algorithms) are used, or by their composition in the image, in which case region-detection techniques (such as texture analysis) are used (Haralick and Shapiro 1992). Neither of these techniques has been completely successful as fully automated image segmentation methods; regions often have discontinuous borders or nondistinctive internal composition. Furthermore, contiguous regions often overlap. These and other complications make segmentation the most difficult subtask of the medical image processing problem. Because segmentation is difficult for a computer, it is usually performed either by hand or in a semi-automated manner with assistance by a human through operator-interactive approaches (Fig. 10.20). In both cases, segmentation is time intensive, and it therefore remains a major bottleneck that prevents more widespread application of image processing techniques.

A great deal of progress has been made in automated segmentation in the brain, partially because the anatomic structures tend to be reproducibly positioned across subjects and the contrast delineation among structures is often good. In addition, MRI images of brain tend to be high quality. Several software packages are currently available for automatic segmentation, particularly for normal macroscopic brain anatomy in cor-

tical and sub-cortical regions (Collins et al. 1995; Friston et al. 1995; Subramaniam et al. 1997; Dale et al. 1999; MacDonald et al. 2000; Brain Innovation B.V. 2001; FMRIDB Image Analysis Group 2001; Van Essen et al. 2001; Hinshaw et al. 2002). The Human Brain Project's Internet Brain Segmentation Repository (Kennedy 2001) has been developing a repository of segmented brain images to use in comparing these different methods.

Popular segmentation techniques can be (1) region-based Vs. edge-based methods, (2) knowledge-based Vs. data-driven methods, and combined methods.

Region-Based Vs. Edge-Based

In region-based segmentation, voxels are grouped into contiguous regions based on characteristics such as intensity ranges, spatial statistics and similarity to neighboring voxels (Shapiro and Stockman 2001; Li et al. 2011b). In brain MR images, a common class separation is into: gray matter, white matter, cerebrospinal fluid and background. One then uses these classifications as a basis for further segmentation (Choi et al. 1991; Zijdenbos et al. 1996). Another region-based approach is called region-growing, in which regions are grown from seed voxels manually or automatically placed within candidate regions (Davatzikos and Bryan 1996; Modayur et al. 1997). The regions found by any of these approaches are often further processed by mathematical morphology operators (Haralick 1988) to remove unwanted connections and holes (Sandor and Leahy 1997). Other well-known techniques include active contour and level set models (Li et al. 2011b; Hoogi et al. 2017) graph-based models (Shattuck and Leahy 2001) and clustering-based methods (Li et al. 2011a).

Contrary to region-based techniques, Edge-based segmentation relies on detecting the gradients in the image. These gradients are considered as the organ boundary. However, edge-based technique is very sensitive to image noise and to inconsistent broken boundaries.

Other techniques can be considered as hybrid frameworks, in which both region statistics and gradients information are included (Chakraborty et al. 1996; Shao et al. 2008). All the above techniques are essentially low-level techniques that only look at local or global regions in the image data.

Model-Based and Data-Driven Segmentation

Segmentation methods are mostly divided into model-based and data-driven approaches. The former considers prior knowledge about the organ/medical images to be analyzed, while the latter is based only on the specific analyzed image data, with no prior examples or knowledge. Deformable models that are part of the model-based curve evolution approach and called "Snakes" (Kass et al. 1987; Davatzikos and Bryan 1996; Dale et al. 1999; MacDonald et al. 2000; Van Essen et al. 2001). These models can include knowledge of the expected anatomy of the organ. For example, the cost function employed in the method developed by MacDonald (MacDonald et al. 2000) includes a term for the expected thickness of the brain cortex. Thus, these methods can become somewhat knowledge-based, where knowledge of anatomy is encoded in the cost function. Level set is another form of curve evolution technique but on contrary to Snake, level set is an implicit approach (Li et al. 2011b; Hoo-Chang et al. 2016; Hoogi et al. 2017). In both Snakes and level set, the contour is deformed according to a cost function that should be minimized and includes both intrinsic terms regarding the contour itself (e.g. contour smoothness), and extrinsic terms that depends on the image data.

Clustering-Based Segmentation

The core operation in a segmentation task is the division of the image into a finite set of clusters/regions with similar statistics, which are smooth and homogeneous in their content and their representation. When posed in this way, segmentation can be regarded as a problem of finding clusters in a selected feature space. The segmentation task can be seen as a combination of two main processes: (a) The generation of an image representation over a selected feature space. This can be termed the modeling stage. The model components are often viewed as groups, or *clusters* in the high-dimensional space. (b) The assignment

of pixels to one of the model components or segments. In order to be directly relevant for a segmentation task, the clusters in the model should represent homogeneous regions of the image. In general, the better the image modeling, the better the segmentation produced. Since the number of clusters in the feature space is often unknown, segmentation can be regarded as an *unsupervised clustering* task in the high-dimensional feature space.

There are many works on **clustering algorithms.** We can categorize them into three broad classes: (a) deterministic algorithms, (b) probabilistic model-based algorithms, and (c) graph-theoretic algorithms. The simplest of these are the deterministic algorithms such as k-means (Bishop 1995), mean-shift (Comaniciu and Meer 2002), and agglomerative methods (Duda et al. 2001). For certain data distributions, i.e., distributions of pixel feature vectors in a feature space, such algorithms perform well. For example, k-means provides good results when the data are convex or blob-like and the agglomerative approach succeeds when clusters are dense and there is no noise. These algorithms, however, have a difficult time handling more complex structures in the data. The probabilistic algorithms, on the other hand, model the distribution in the data using parametric models (McLachlan and Peel 2000). Such models include autoregressive (AR) models, Gaussian mixture models (GMM), Markov random fields (MRF), conditional random fields, and others. Efficient ways of estimating these models are available using maximum likelihood algorithms such as the Expectation-maximization (EM) algorithm (Dempster et al. 1977). While probabilistic models offer a principled way to explain the structures present in the data, they could be restrictive when more complex structures are present.

Another type of clustering algorithms is non-parametric in that this class imposes no prior shape or structure on the data. Examples of these are graph-theoretic algorithms based on spectral factorization (e.g., (Ng et al. 2001, Shi and Malik 2000)). Here, the image data are modeled as a graph. The entire image data along with a global cost function are used to partition the graph, with each partition now becoming an image segment. In this approach, global considerations determine localized decisions. Moreover, such optimization procedures are often compute-intensive.

Consider an example application in brain image segmentation using parametric modeling and clustering. The tissue and lesion segmentation problem in Brain MRI is a well-studied topic of research. In such images, there is interest in three main tissue types: white matter (WM), gray matter (GM) and cerebro-spinal fluid (CSF). The volumetric analysis of such tissue types in various part of the brain is useful in assessing the progress or remission of various diseases, such as Alzheimer's disease, epilepsy, sclerosis and schizophrenia. A segmentation example is shown in Fig. 10.21. In this example, images from 3 MRI imaging sequences are input to the system, and the output is a segmentation map, with different colors representing three different normal brain tissues, as well as a separate color to indicate regions of abnormality (multiple-sclerosis lesions).

Various approaches to the segmentation task are reviewed in (Pham et al. 2000). Among the approaches used are pixel-level intensity based clustering, such as K-means and Mixture of Gaussians modeling (e.g., (Kapur et al. 1996)). In this approach, the intensity feature is modeled by a mixture of Gaussians, where each Gaussian is assigned a semantic meaning, such as one of the tissue regions (or lesion). Using pattern recognition methods and learning, the Gaussians can be automatically extracted from the data, and once defined, the image can be segmented into the respective regions.

Algorithms for tissue segmentation using pixel-level intensity-based classification often exhibit high sensitivity to various noise artifacts, such as intra-tissue noise, inter-tissue intensity contrast reduction, partial-volume effects and others. Due to the artifacts present, classical voxel-wise intensity-based classification methods, including the K-means modeling and Mixture of Gaussians modeling, often give unrealistic results, with tissue class regions appearing granular, fragmented, or violating anatomical constraints. Specific works can be found addressing various aspects

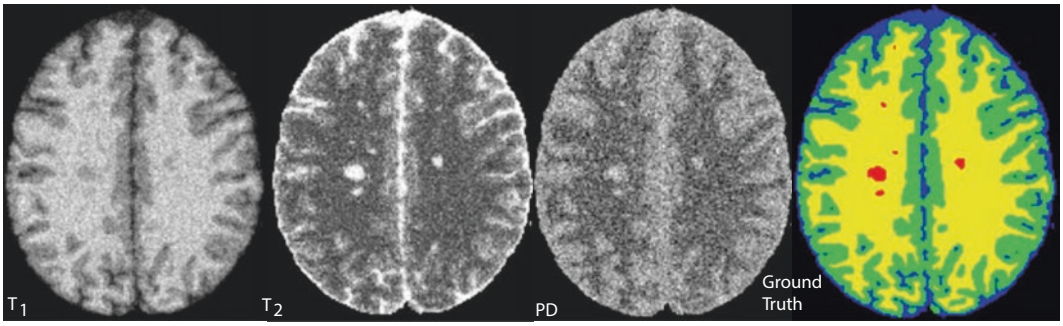


Fig. 10.21 Brain MRI segmentation example. Brain slice from multiple acquisition sequences (with 9% noise) was taken from BrainWEB (► http://www.bic.mni.mcgill.ca/brainweb/). From left to right: T1-, T2-, and proton density (PD)-weighted image. Segmentation of the images is shown on the right: Blue: CSF; Green: Gray matter (GM); Yellow: white matter (WM); Red: Multiple-sclerosis lesions (MSL). (Friefeld et al. 2009). (Reused with permission © Brainweb)

of these concerns (e.g., partial-volume effect quantification (Dugas-Phocion et al. 2004)). One way to address the smoothness issue is to add spatial constraints. This is often done during a pre-processing phase by using a statistical atlas, or as a post-processing step via Markov Random Field models. A statistical atlas provides the prior probability for each pixel to originate from a particular tissue class (e.g., (Van Leemput et al. 1999; Marroquin et al. 2002; Prastawa et al. 2004)).

Algorithms exist that use the maximum-a-posteriori (MAP) criterion to augment intensity information with the atlas. However, registration between a given image and the atlas is required, which can be computationally prohibitive (Rohlfing and Maurer Jr. 2003). Further, the quality of the registration result is strongly dependent on the physiological variability of the subject and may converge to an erroneous result in the case of a diseased or severely damaged brain. Finally, the registration process is applicable only to complete volumes. A single slice cannot be registered to the atlas. Therefore it cannot be segmented using these state-of-the-art algorithms. In (Greenspan et al. 2006) a robust, unsupervised, parametric method for segmenting 3D (or 2D) MR brain images with a high degree of noise and low contrast, is presented. A Constrained Gaussian Mixture Model (CGMM) framework is proposed, in which each tissue is modeled with multiple four-dimensional Gaussians, where each Gaussian represents a localized region (3 spatial features) and the intensity characteristic per region (T1 intensity feature). Incorporating the spatial information within the feature space is novel, as is using a large number of Gaussians per brain tissue to capture the complicated spatial layout of the individual tissues. Two key features of the proposed framework are: 1) combining global intensity modeling with localized spatial modeling, as an alternative scheme to MRF modeling, and 2) segmentation is entirely unsupervised; thus eliminating the need for atlas registration, or any intensity model standardization.

Segmentation can also be improved using a post-processing phase in which smoothness and immunity to noise can be achieved by modeling the interactions among neighboring voxels. Such interactions can be modeled using a Markov Random Field (MRF), and thus this technique has been used to improve segmentation (Held et al. 1997; Van Leemput et al. 1999; Zhang et al. 2001).

■ Segmentation Using Deep Learning

As noted earlier in this chapter, one of the fastest emerging research field over the last few years is deep learning. Deep learning can help outperform classical machine learning algorithms due to its ability to learn latent variables within the features space, features that the user can barely detect. On the other hand, deep learning requires a huge labeled training size that is not always available, which makes developing robust classification models challenging. However, there are methods to help

overcome these challenges, such as data augmentation (Greenspan et al. 2016), transfer learning (Hoo-Chang et al. 2016), and metric learning (Yang et al. 2010) that were specifically designed to handle these challenges. In transfer learning one uses a network that was pre-trained on another set of images, and we use the weights of this network as weights initialization for the current network that is analyzed. The initial weights are fine-tuned on the relevant current dataset, which should give better results than just using random weights. In metric learning, one actually learns the best metric that can represent and classify the data – instead of learning the classes themselves. In that way, we actually teach the network how to learn. In addition, metric learning is a kind of "image ontology," and as a result, it sketches the distances between different instances. Therefore, if a new test case will not be part of the training classes, it will not be misclassified to one of those classes.

The core idea of deep learning is to convolve the input image with different filters and within different scales (i.e. pooling), such as they will able to detect both low-level and high-level features. Many deep learning architectures are considered as patch-wise techniques.

Convolutional neural networks such as U-Net (Ronneberger et al. 2015; Trebeschi et al. 2017) and V-Net (Trebeschi et al. 2017) were designed specifically to deal with the typical challenges of the medical domain such as small amount of labeled data. Autoencoders, Variational Autoencoders and stacked-autoencoders can be used for image denoising and as for unsupervised feature extraction (Vincent et al. 2010; Bengio et al. 2013). Other methods were designed to handle with various of classifications tasks such as lesion detection (Wang et al. 2016), segmentation (Kayalibay et al. 2017; Trebeschi et al. 2017) and disease classification (Esteva et al. 2017).

10.4.7 Image Registration

The growing availability of 3-D and higher dimensionality structural and functional images leads to exciting opportunities for realistically observing the structure and function of the body. These opportunities are particularly widely exploited in brain imaging. Therefore, this section concentrates on 3-D brain imaging, with the recognition that many of the methods developed for the brain have been or will be applied to other areas as well.

The basic 2-D image processing operations of global processing, segmentation, feature detection, and classification generalize to higher dimensions, and are usually part of any image processing application. However, 3-D and higher dimensionality images give rise to additional informatics issues, which include image *registration* (which also occurs to a lesser extent in 2-D), *spatial* representation of anatomy, *symbolic* representation of anatomy, integration of spatial and symbolic anatomic representations in *atlases*, anatomical *variation*, and *characterization* of anatomy. All but the first of these issues deal primarily with anatomical structure, and therefore could be considered part of the field of structural informatics. They could also be thought of as being part of imaging informatics and **neuroinformatics**.

As noted previously, 3-D image volume data are represented in the computer by a 3-D volume array, in which each voxel represents the image intensity in a small volume of space. In order to depict anatomy accurately, the voxels must be accurately registered (or located) in the 3-D volume (*voxel registration*), and separately acquired image volumes from the same subject must be registered with each other (*volume registration*).

■ Voxel Registration

Imaging modalities such as CT, MRI, and confocal microscopy (▶ Sects. 10.2.3 and 10.2.5) are inherently 3-D: the scanner generally outputs a series of image slices that can easily be reformatted as a 3-D volume array, often following alignment algorithms that compensate for any patient motion during the scanning procedure. For this reason, almost all CT and MR manufacturers' consoles contain some form of three-dimensional reconstruction and visualization capabilities.

As noted in ▶ Sect. 10.4.4, two-dimensional images can be converted to 3-D volumes if they are closely spaced parallel

sections through a tissue or whole specimen and contain isotropic voxels. In this case, the problem is how to align the sections with each other. For whole sections (either frozen or fixed), the standard method is to embed a set of thin rods or strings in the tissue prior to sectioning to manually indicate the location of these **fiducials** on each section, then to linearly transform each slice so that the corresponding fiducials line up in 3-D (Prothero and Prothero 1986). An example of this technique is the Visible Human, in which a series of transverse slices were acquired, then reconstructed to give a full 3-D volume (Spitzer and Whitlock 1998) (► Chap. 22).

It is difficult to embed fiducial markers at the microscopic level, so intrinsic tissue landmarks are often used as fiducials, but the basic principle is similar. However, in this case tissue distortion may be a problem, so non-linear transformations may be required. For example Fiala and Harris (Fiala and Harris 2001) developed an interface that allows the user to indicate, on electron microscopy sections, corresponding centers of small organelles such as mitochondria. A non-linear transformation (warp) is then computed to bring the landmarks into registration.

An approach being pursued (among other approaches) by the National Center for Microscopy and Imaging Research (► http://ncmir.ucsd.edu/) combines reconstruction from thick serial sections with electron tomography (Soto et al. 1994). In this case the tomographic technique is applied to each thick section to generate a 3-D digital slab, after which the slabs are aligned with each other to generate a 3-D volume. The advantages of this approach over the standard serial section method are that the sections do not need to be as thin, and fewer of them need be acquired.

An alternative approach to 3-D voxel registration from 2-D images is stereo-matching, a technique developed in computer vision that acquires multiple 2-D images from known angles, finds corresponding points on the images, and uses the correspondences and known camera angles to compute 3-D coordinates of pixels in the matched images. The technique is being applied to the reconstruction of synapses from electron micrographs by a Human Brain Project collaboration between computer scientists and biologists at the University of Maryland (Agrawal et al. 2000).

■ Volume Registration

A related problem to that of aligning individual sections is the problem of aligning separate image volumes from the same subject, that is, *intra-subject* alignment. Because different image modalities provide complementary information, it is common to acquire more than one kind of image volume on the same individual. This approach has been particularly useful for brain imaging because each modality provides different information. For example, PET (► Sect. 10.2.3) provides useful information about function, but does not provide good localization with respect to the anatomy. Similarly, MRV and MRA (► Sect. 10.2.3) show blood flow but do not provide the detailed anatomy visible with standard MRI. By combining images from these modalities with MRI, it is possible to show functional images in terms of the underlying anatomy, thereby providing a common neuroanatomic framework.

The primary problem to solve in multimodality image fusion is volume registration—that is, the alignment of separately acquired image volumes. In the simplest case, separate image volumes are acquired during a single sitting. The patient's head may be immobilized, and the information in the image headers may be used to rotate and resample the image volumes until all the voxels correspond. However, if the patient moves, or if examinations are acquired at different times, other registration methods are needed. When intensity values are similar across modalities, registration can be performed automatically by intensity-based optimization methods (Woods et al. 1992; Collins et al. 1994). When intensity values are not similar (as is the case with MRA, MRV and MRI), images can be aligned to templates of the same modalities that are already aligned (Woods et al. 1993; Ashburner and Friston 1997). Alternatively, landmark-based methods can be used. The landmark-based methods are similar to those

used to align serial sections (see earlier discussion of voxel registration in this section), but in this case the landmarks are 3-D points. The Montreal Register Program (MacDonald 1993) is an example of such a program. Techniques and applications of volume registration in other domains have been described (Pelizzari 1998; Ferrante and Paragios 2017).

10.5 Image Interpretation and Computer Reasoning

The preceding sections of this chapter as well as ▶ Chap. 22 describe informatics aspects of image generation, storage, manipulation, and display of images. Rendering an interpretation is a crucial final stage in the chain of activities related to imaging. *Image interpretation* is this final stage in which the physician has direct impact on the clinical care process, by rendering a professional opinion as to whether abnormalities are present in the image and the likely significance of those abnormalities. The process of image interpretation requires "reasoning"— drawing inferences from facts; the facts are the image abnormalities detected and the known clinical history, and the inferred information is the diagnosis and management decision (what to do next, such as another test or surgery, etc.). Such reasoning usually entails uncertainty, and optimally would be carried out using probabilistic approaches (▶ Chap. 3), unless certain classic imaging patterns are recognized. In reality, radiology practice is usually carried out without formal probabilistic models that relate imaging observations to the likelihood of diseases. However, variation in practice is a known problem in image interpretation (Robinson 1997), and methods to improve this process are desirable.

Informatics methods can enhance radiological interpretation of images in two major ways: (1) image retrieval systems and (2) computer-based inference systems. The concept of *image retrieval* is similar to that of information retrieval (see ▶ Chap. 23), in which the user retrieves a set of documents pertinent to a question or information need. The information being sought when doing image retrieval is images with specific content—typically to find images that are similar in some ways to a query image (e.g., to find images in the PACS containing similar-appearing abnormalities to that in an image being interpreted). Finding images containing similar content is referred to as **content based image retrieval** (CBIR). By retrieving similar images and then looking at the diagnosis of those patients, the radiologist can gain greater confidence in interpreting the images from patients whose diagnosis is not yet known.

As with the task of medical diagnosis (▶ Chap. 24), radiological diagnosis can be enhanced using computer-based inference systems, the commonest type of which is *decision support systems*, which assist the physician in making clinical decisions. In computer inference (also referred to as "reasoning"), the machine takes in the available data (the images and possibly other clinical information), performs a variety of image processing methods (▶ Sect. 10.4), and uses one or more types of knowledge resources and/or mathematical models to render an output comprising either a decision or a ranked list of possible choices (e.g., diagnoses or locations on the image suspected of being abnormal).

In this section we describe informatics methods for image retrieval and computer inference with images.

10.5.1 Content-Based Image Retrieval

Since a key aspect of radiological interpretation is recognizing characteristic patterns in the imaging features which suggest the diagnosis, searching databases for similar images with known diagnoses could be an effective strategy to improving diagnostic accuracy. **CBIR** is the process of performing a match between images using their visual content. A query image can be presented as input to the system (or a combination of a query image and the patient's clinical record), and the system searches for similar cases in large archive settings (such as PACS) and returns a ranked list of such similar data (images).

This task requires an informative representation for the image data, along with similarity measures across image data. CBIR methods are already useful in non-medical applications such as consumer imaging and on the Web (Wang et al. 1997; Smeulders et al. 2000; Datta et al. 2008).

There has also been ongoing work to develop CBIR methods in radiology and several reviews on this subject have been published (Akgul et al. 2011; Endo et al. 2012; Kumar et al. 2013; Muramatsu 2018). The approach generally is based on deriving quantitative characteristics from the images (e.g., pixel statistics, spatial frequency content, etc.; ▶ Sect. 10.4.5), followed by application of similarity metrics to search databases for similar images (Lehmann et al. 2004; Muller et al. 2004; Greenspan and Pinhas 2007; Datta et al. 2008; Deserno et al. 2009; Napel et al. 2010; Faruque et al. 2013, 2015). The focus of the current work is on entire images, describing them with sets of numerical features, with the goal of retrieving similar images from medical collections (Hersh et al. 2009; Napel et al. 2010; Faruque et al. 2013, 2015) that provide benchmarks for image retrieval. However, in many cases only a particular region of the image is of interest when seeking similar images (e.g., finding images containing similar-appearing lesions to those in the query image). More recently, "localized" CBIR methods are being developed in which a part of the image containing a region of interest is analyzed (Deselaers et al. 2007; Rahmani et al. 2008; Napel et al. 2010).

There are several unsolved challenges in CBIR. First, CBIR has been largely focused on query based on single 2-D images; methods need to be developed for 3D retrieval in which a volume is the query "image." A second challenge is the need to integrate images with non-image clinical data to permit retrieval based on entire patient cases and not single images (e.g., the CBIR method should take into consideration the clinical history in addition to the image appearance in retrieving a similar "case").

Another limitation of current CBIR is that image semantics is not routinely included. The qualitative image features reported by the radiologist ("semantic features") are complementary to the quantitative data contained in image pixels. One approach to capturing image semantics is analyzing and processing "visual words" in images, captured as image patches or codebooks (▶ Sect. 10.4.5). These techniques have been shown to perform well in CBIR applications (Qiu 2002). Another approach to capture image semantics is to use the radiologist's imaging observations as image features. Several studies have found that combining the semantic information obtained from radiologists' imaging reports or annotations with the pixel-level features can enhance performance of CBIR systems (Ruiz 2006; Zhenyu et al. 2009; Napel et al. 2010). The knowledge representation methods described in ▶ Sects. 10.3.2 and 10.4.5 make it possible to combine these types of information.

10.5.2 Computer-Based Inference

Though image retrieval described above (and information retrieval in general) can be helpful to a practitioner interpreting images, it does not directly answer a specific question at hand, such as, "what is the diagnosis in this patient" or "what imaging test should I order next?" Answering such questions requires inference, either by the physician with all the available data, or by a computer, using physician inputs and the images. As the use of imaging proliferates and the number of images being produced by imaging modalities explodes, it is becoming a major challenge for practicing radiologists to integrate the multitude of imaging data, clinical data, and soon molecular data, to formulate an accurate diagnosis and management plan for the patient. Computer-based inference systems—specifically decision support systems—can help radiologists understand the biomedical import of this information and to provide guidance (Hudson and Cohen 2009).

There are two major approaches to computer-based inference using images: (1) using quantitative image features only (quan-

titative imaging computer inference systems), and (2) use knowledge associated with the images (knowledge-based computer inference systems).

Quantitative Imaging Computer Inference Systems

The process of deriving quantitative image features was described in ▶ Sect. 10.4.5. Quantitative imaging applications such as CAD and CADx use these quantifiable features extracted from medical images for a variety of decision support applications, such as the assessment of an abnormality to suggest a diagnosis, or to evaluate the severity, degree of change, or status of a disease, injury, or chronic condition. In general, the quantitative imaging computer reasoning systems apply a mathematical model (e.g., a *classifier*) or other machine learning methods to obtain a decision output based on the imaging inputs.

There are three types of systems that make inferences using quantitative imaging data, *computer-assisted detection* (CAD), *computer-assisted diagnosis* (CADx), and *computerized prediction* systems. In CAD, the computer locates ROIs in the image where abnormalities are suspected and the radiologist must evaluate their medical significance. This is generally accomplished using quantitative image analysis methods (▶ Sect. 10.4.5). In CADx, the computer is given an ROI corresponding to a suspected abnormality (possible with associated clinical information) and it outputs the likely diagnoses and possibly management recommendations (ideally with some sort of confidence rating as well as explanation facility). Ideally, the confidence of the algorithm in making this diagnosis is also provided as well as explanation or transparency to the user to understand how that diagnosis was determined from the facts. These systems generally use both quantitative imaging methods (▶ Sect. 10.4.5) as well as computer reasoning methods that leverage knowledge associated with the image (▶ Sect. 10.3.2). In computerized predication, a computational model based on analysis of the images (potentially integrated with other data) makes a clinical prediction about the patient.

CAD

In CAD applications, the goal is *detection* of abnormalities that are visible in the image, to scan the image and identify suspicious regions that may represent regions of disease in the patient. A common use for CAD is *screening*, the task of reviewing many images and identifying those that are suspicious and require closer scrutiny by a radiologist (e.g., mammography interpretation). Most CAD applications comprise an image processing pipeline (▶ Sect. 10.4) that uses global processing, segmentation, image quantitation with feature extraction, and classification to determine whether an image should be flagged for careful review by a radiologist or pathologist. In CAD and in screening in general, the goal is to detect disease; thus, the tradeoff favors having false positive instead of missing false negatives. Thus CAD systems tend to flag a reasonable number of normal images (false positives) and they miss very few abnormal images (false negatives). If the number of flagged images is small compared with the total number of images, then automated screening procedures can be economically viable. On the other hand, too many false positives are time-consuming to review and lessens user confidence in the CAD system; thus for CAD to be viable, they must minimize the number of false positives as well as false negatives.

CAD techniques for screening have been applied successfully to many different types of images (Doi 2007), including mammography images for identifying mass lesions and clusters of microcalcifications, chest X-rays and CT of the chest to detect small cancerous nodules, and volumetric CT images of the colon ("virtual colonscopy") to detect polyps. In addition, CAD methods have been applied to Papanicolaou (Pap) smears for cancerous or precancerous cells (Giger and MacMahon 1996), as well as to many other types of non-radiologic images.

As noted in ▶ Sect. 10.4.5, detection tasks in images can be accomplished using pre-defined image features or using learned features (deep learning), which is becoming a very popular approach to CAD with encour-

aging results (Firmino et al. 2014; Shin et al. 2016).

CADx

In CADx applications, a suspicious region in the image has already been identified (by the radiologist of a CAD application), and the goal is to evaluate it to render a diagnosis or differential diagnosis. CADx systems usually need to be provided an ROI, or they need to segment the image to locate specific organs and lesions in order to perform analysis of quantitative image features that are extracted from the ROI and use that to render a diagnosis. However, recently, CADx systems have begun to be developed using deep learning, which generally does not require any ROI, since the models are built using the raw image data (Al-Antari et al. 2018; Ishioka et al. 2018; Lee et al. 2018; Nishio et al. 2018).

In general, a mathematical model is created to relate the quantitative (or semantic) features to the likely diagnoses. These models are built either using pre-defined image features or using learned features (▶ Sect. 10.4.5). Most of the historical CADx systems have been built using pre-defined image features (Doi 2007), but if a sufficient number of labeled cases is available for training, deep learning appears to hold much promise for developing CADx systems (Chen et al. 2017; Hosny et al. 2018).

A particularly important emerging role for CADx systems is in the diagnosis of infection by the SARS-CoV-2 virus (COVID-19). The COVID-19 pandemic created extremely rapid and widespread person-to-person transmission of the disease (World Health Organization 2020). Definitive diagnosis is made using the reverse transcriptase polymerase chain reaction (RT-PCR) test. Since test can take up to 2 days to complete, and given the shortage of RT-PCR test kits, there was an urgent need for alternative and rapid methods to identify COVID-19 patients. Imaging (radiography or CT) are commonly used to identify pneumonia, but imaging has not, to date, been used to establish a diagnosis of COVID-19 since imaging findings are not specific. Thus, routine screening CT for the identification of COVID-19 pneumonia is currently not recommended by most radiology societies (Simpson et al. 2020). A number of recent works have been undertaken to detect COVID-19 pneumonia in patient using deep learning (Wang and Wong 2020, Greenspan et al. 2020). Systems that integrate clinical data with images may be particularly promising (Mei et al. 2020). Nonetheless, given the accuracy of current CADx systems for COVID-19 likely have insufficient accuracy, and new diagnostic tests are being developed that are processed in a shorter time (Billingsley 2020), imaging will not likely have a sustained major role in diagnosis. Other applications for CADx systems in COVID-19 were reviewed in (Kumar et al. 2020). Perhaps the most exciting role for computerized systems in this disease will be for making clinical predictions, such as survival, need for intensive care, ventilator support, and ultimate survival (Liang et al. 2020; Liu et al. 2020; Luo et al. 2020; Sperrin et al. 2020; Wynants et al. 2020; Yang et al. 2020; Yuan et al. 2020).

A limitation of using only pre-defined image features or unsupervised learned image features is that these models do not encode domain knowledge that may be critical to the accuracy of a CADx system; the presumption of using only image features is that all the knowledge needed for the diagnostic classification task is represented in image data itself. However in some cases, it is very useful to encode knowledge in a CADx system. Probabilistic models provide a strategy for incorporating domain knowledge and have been shown to be effective (Burnside et al. 2000, 2004a, b, 2006, 2007; Lee et al. 2009; Liu et al. 2009; Liu et al. 2011). Image features are generated based on the underlying disease, so there is probabilistic dependence on the disease and the quantitative and perceived imaging features. In fact, it can be argued that radiological interpretation is fundamentally a Bayesian task (Lusted 1960; Ledley and Lusted 1991; Donovan and Manning 2007) (see ▶ Chaps. 3 and 22), and thus decision-support strategies based on Bayesian models may be quite effective.

CADx can be very effective in practice, reducing variation and improving positive predictive value of radiologists (Burnside

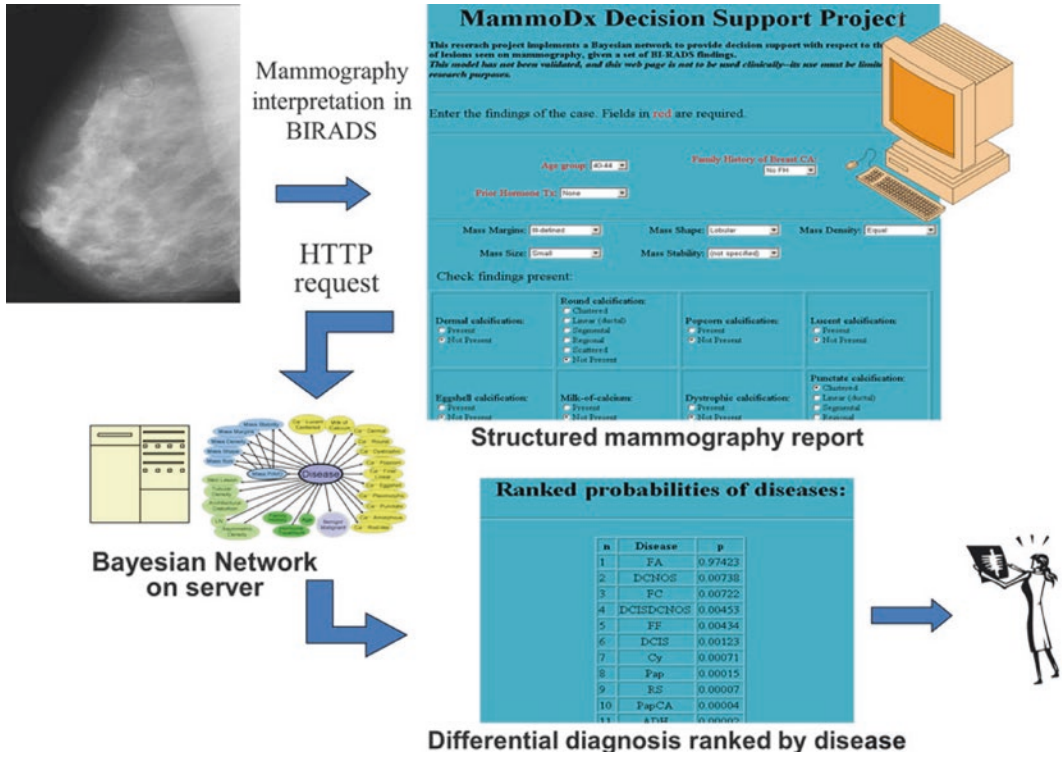


Fig. 10.22 Bayesian network-based system for decision support in mammography CADx. The radiologist interpreting the image enters the radiology observations and clinical information (patient history) in a structured reporting Web-based data capture form to render the report. This form is sent to a server which inputs the observations into the Bayesian network to calculate posterior probabilities of disease. A list of diseases, ranked by the probability of each disease, is return to the user who can make a decision based on a threshold of probability of malignancy, or based on shared decision making with the patient. (Figure reprinted with permission from (Rubin 2011). © Radiological Society of North America)

et al. 2006). Deploying CADx systems, however, can be challenging. Since the inputs to CADx generally need to be structured (semantic features from the radiologist and/or quantitative features from the image), a means of capturing the structured image information as part of the routine clinical workflow is required. A promising approach is to combine structured reporting with CADx (Fig. 10.22); the radiologist records the imaging observations with a data capture form, which provides the structured image content required to the CADx system. Ideally the output would be presented immediately to the radiologist as the report is generated so that the output of decision support can be incorporated into the radiology report. Such implementations will be greatly facilitated by informatics methods to extract and record the image information in structured and standard formats and with controlled terminologies (▶ Sect. 10.3.2).

Computerized Prediction

The goal of computerized prediction using images is to analyze characteristics of the disease manifest in the image and use that (without or with additional clinical data) to make predictions about the disease (e.g., life expectancy of the patient, whether or not the patient's disease will respond to a particular treatment, or whether the disease will recur or progress at some time in the future). Many methods have been developed to predict such future event, using both predefined image features and unsupervised feature learning (▶ Sect. 10.4.5) (Huang et al. 2016; Jun et al. 2016; Li et al. 2016; Nie et al. 2016; Bogowicz et al. 2017; Fave et al. 2017; van Timmeren et al. 2017; Wu et al. 2017;

Zhou et al. 2017b; Betancur et al. 2018; Cha et al. 2018; Gastounioti et al. 2018; Shi et al. 2018).

Knowledge-Based Image Inference Systems

The CAD and CADx systems do not require processing radiological knowledge (e.g., anatomic knowledge) in order to carry out their tasks; they are based on quantitative modeling of relationships of images features to diagnoses. However, not all image-based reasoning problems are amenable to this approach. In particular, knowledge-based tasks such as reasoning about anatomy, physiology, and pathology—tasks that entail symbolic manipulations of biomedical knowledge and application of logic—are best handled using different methods, such as ontologies and logical inference (see ▶ Chap. 24).

Knowledge-based computer reasoning applications use knowledge representations, generally ontologies, in conjunction with rules of logic to deduce information from asserted facts (e.g., from observations in the image). For example, an anatomy ontology may express the knowledge that "if a segment of a coronary artery is severed, then branches distal to the severed branch will not receive blood," and "the anterior and lateral portions of the right ventricle are supplied by branches of the right coronary artery, with little or no collateral supply from the left coronary artery." Using this knowledge, and recognition via image processing that the right coronary artery is severed in an injury, a computer reasoning application could deduce that the anterior and lateral portions of the right ventricle will become ischemic (among other regions; ◘ Fig. 10.23). In performing this reasoning task, the application uses the knowledge to draw correct conclusions by manipulating the anatomical concepts and relationships using the rules of logical inference during the reasoning process.

Computer reasoning with ontologies is performed by one of two methods: (1) *reasoning by ontology query* and (2) *reasoning by logical inference*. In reasoning by ontology query, the application traverse relationships that link particular entities in the ontology to directly answer particular questions about how those entities relate to each other. For example, by traversing the *part-of* relationship in an anatomy ontology, a reasoning application can infer that the left ventricle and right ventricle are part-of the chest (given that the ontology asserts they are each part of the heart, and that the heart is part of the chest), without our needing to specify this fact explicitly in the ontology.

In reasoning by logical inference, ontologies that encode sufficient information ("explicit semantics") to apply generic reasoning engines are used. The Web Ontology Language (OWL) (Bechhofer et al. 2004; Smith et al. 2004; Motik et al. 2008) is an ontology language recommended by the World Wide Web Consortium (W3C) as a standard language for the Semantic Web (WorldWideWeb Consortium W3C Recommendation 10 Feb 2004). OWL is similar to other ontology languages in that it can capture knowledge by representing the entities ("classes") and their attributes ("properties"). In addition, OWL provides the capability of defining "formal semantics" or meaning of the entities in the ontology. Entities are defined using logic statements that provide assertions about entities ("class axioms") using **description logics** (DL) (Grau et al. 2008). DLs provide a formalism enabling developers to define precise semantics of knowledge in ontologies and to perform automated deductive reasoning (Baader et al. 2003). For example, an anatomy ontology in OWL could provide precise semantics for "hemopericardium," by defining it as a pericardial cavity that contains blood.

Highly optimized computer reasoning engines ("reasoners") have been developed for OWL, helping developers to incorporate reasoning efficiently and effectively in their applications (Tsarkov and Horrocks 2006; Motik et al. 2009). These reasoners work with OWL ontologies by evaluating the asserted logical statements about classes and their properties in the original ontology (the "asserted ontology"), and they create a new ontology structure that is deduced from the asserted knowledge (the "inferred ontology"). This reasoning process is referred to as "automatic classification." The inferences obtained

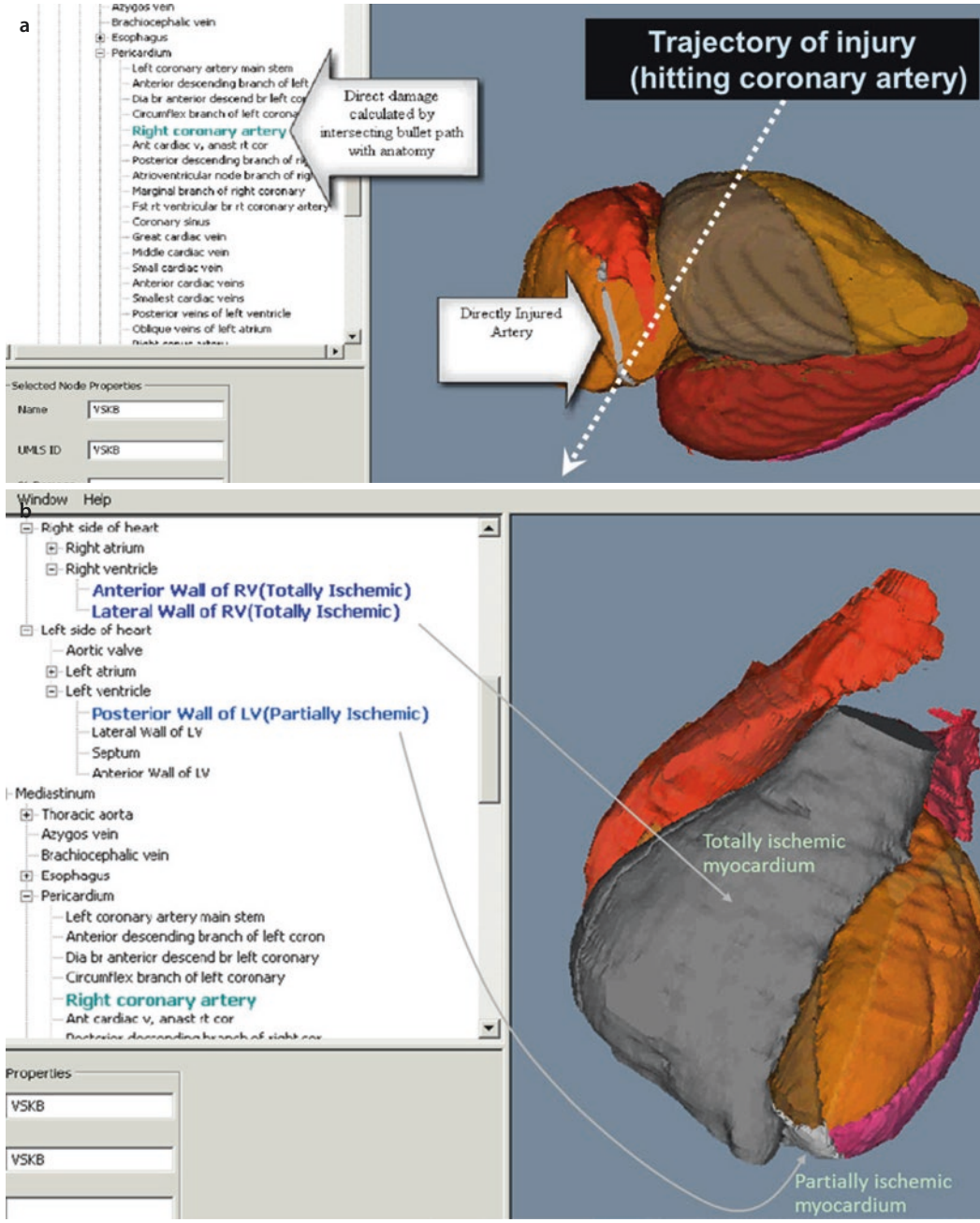


Fig. 10.23 Knowledge-based reasoning with images in a task to predict the portions of the heart that will become ischemic after a penetrating injury that injures particular anatomic structures. The application allows the user to draw a trajectory of penetrating injury on the image, a 3-D rendering of the heart obtained from segmented CT images. The reasoning application automatically carries out two tasks. **a** The application first deduces the anatomic structures that will be injured consequent to the trajectory (arrow, right) by interrogating semantic annotations on the image based on the trajectory of injury (injured anatomic structures shown in bold in the left panel). **b** The anatomic structures that are predicted to be initially injured are displayed in the volume rendering (dark gray = total ischemia; light gray = partial ischemia). In this example, the right coronary artery was injured, and the reasoning application correctly inferred there will be total ischemia of the anterior and lateral wall of the right ventricle and partial ischemia of the posterior wall of the left ventricle

from the reasoning process are obtained by querying the inferred ontology and looking for classes (or individuals) that have been assigned to classes of interest in the ontology. For example, an application was created to infer the consequences of cardiac injury in this manner (Fig. 10.23).

Several knowledge-based image reasoning systems have been developed that use ontologies as the knowledge source to process the image content and derive inferences from them. These include: (1) reasoning about the anatomic consequences of penetrating injury, (2) inferring and simulating the physiological changes that will occur given anatomic abnormalities seen in images, (3) automated disease grading/staging to infer the grade and/or stage of disease based on imaging features of disease in the body (4) surgical planning by deducing the functional significance of disruption of white matter tracts in the brain, (5) inferring the types of information users seek based on analyzing query logs of image searches, and (6) inferring the response of disease in patients to treatment based on analysis of serial imaging studies. We briefly describe these applications.

Reasoning about anatomic consequences of penetrating injury

In this system, images were segmented and semantic annotations applied to identify cardiac structures. An ontology of cardiac anatomy in OWL was used to encode knowledge about anatomic structures and the portions of them that are supplied by different arterial branches. Using knowledge about part-of relationships and connectivity, the application uses the anatomy ontology to infer the anatomic consequences of injury that are recognized on the input images (Fig. 10.23) (Rubin et al. 2004, 2005, 2006a).

Inferring and simulating the physiological changes

Morphological changes in anatomy have physiological consequences. For example, if a hole appears in the septum dividing the atria or ventricles of the heart (a septal defect), then blood will flow abnormally between the heart chambers and will produce abnormal physiological blood flow. The simulation community has created mathematical models to predict the physiological signals, such as time-varying pressure and flow, given particular parameters in the model such as capacitance, resistance, etc. The knowledge in these mathematical models can be represented ontologically, in which the entities correspond to nodes in the simulation model; the advantage is that a graphical representation of the ontology, corresponding to a graphical representation of the mathematical model, can be created. Morphological alterations seen in images can be directly translated into alterations in the ontological representation of the anatomic structures, and simultaneously can update the simulation model appropriately to simulate the physiological consequences of the morphological anatomic alteration (Rubin et al. 2006b). Such knowledge-based image reasoning methods could greatly enable functional evaluation of the static abnormalities seen in medical imaging.

Automated disease grading/staging

A great deal of image-based knowledge is encoded in the literature and not readily available to clinicians needing to apply it. A good example of this is the criteria used to grade and stage disease based on imaging criteria. For example, there are detailed criteria specified for staging tumors and grading the severity of disease. This knowledge has been encoded in OWL ontologies and used to automate grading of brain gliomas (Marquet et al. 2007) and staging of cancer (Dameron et al. 2006) based on the imaging features detected by radiologists. This ontology-based paradigm could provide a good model for delivering current biomedical knowledge to practitioners "just-in-time" to help them grade and stage disease as they view images and record their observations.

Surgical planning

Understanding complex anatomic relationships and their functional significance in the patient is crucial in surgical planning, par-

ticular for brain surgery, since there are many surgical approaches possible, and some will have less severe consequences to patients than others. It can be challenging to be aware of all these relationships and functional dependencies; thus, surgical planning is an opportune area to develop knowledge-based image reasoning systems. The anatomic and functional knowledge can be encoded in an ontology and used by an application to plan the optimal surgical approach. In recent work, such an ontological model was developed to assess the functional sequelae of disruptions of motor pathways in the brain, which could be used in the future to guide surgical interventions (Talos et al. 2008; Rubin et al. 2009b).

Inferring types of information users seek from images

Knowledge-based reasoning approaches have been used to evaluate image search logs on Web sites that host image databases to ascertain the types of queries users submit. RadLex (► Sect. 10.3.2) was used as the ontology, and by mapping the queries to leaf classes in RadLex and then traversing the subsumption relations, the types of queries could be deduced by interrogating the higher-level classes in RadLex (such as "visual observation" and "anatomic entity") (Rubin et al. 2011).

Inferring the response of disease treatment

As mentioned above, the complex knowledge required to grade and stage disease can be represented using an ontology. Similarly, the criteria used to assess the response of patients to treatment is also complex, evolving, and dependent on numerous aspects of image information. The knowledge needed to apply criteria of disease response assessment have been encoded ontologically, specifically in OWL, and used to determine automatically the degree of cancer response to treatment in patients (Levy et al. 2009, Levy and Rubin 2011). The inputs to the computerized reasoning method are the quantitative information about lesions seen in the images, recorded as semantic annotations using the AIM information model (► Sect. 10.3.2). This application demonstrates the potential for a streamlined workflow of radiology image interpretation and lesion measurement automatically feeding into decision support to guide patient care.

10.6 Conclusions

This chapter focuses on methods for computational representation and for processing images in biomedicine, with an emphasis on radiological imaging and the extraction and characterization of anatomical structure and abnormalities. It has been emphasized that the content of images is complex—comprising both quantitative and semantic information. Methods of making that content explicit and computationally-accessible have been described, and they are crucial to enable computer applications to access the "biomedical meaning" in images; presently, the vast archives of images are poorly utilized because the image content is not explicit and accessible. As the methods to extract quantitative and semantic image information become more widespread, image databases will be as useful to the discovery process as the biological databases (they will even likely become linked), and an era of "data-driven" and "high-throughput imaging" will be enabled, analogous to modern "high-throughput" biology. In addition, the computational imaging methods will lead to applications that leverage the image content, such as CAD/CADx and knowledge-based image reasoning that use image content to improve physicians' capability to care for patients.

Though this chapter has focused on radiology, we stress that the biomedical imaging informatics methods presented are generalizable and either have been or will be applied to other domains in which visualization and imaging are becoming increasingly important, such as microscopy, pathology, ophthalmology, and dermatology. As new imaging modalities increasingly become available for imaging other and more detailed body regions, the techniques presented in this chapter will increasingly be applied in all areas of biomedicine. For example, the development

of molecular imaging methods is analogous to functional brain imaging, since functional data, in this case from gene expression rather than cognitive activity, can be mapped to an anatomical substrate.

Thus, the general biomedical imaging informatics methods described here will increasingly be applied to diverse areas of biomedicine. As they are applied, and as imaging modalities continue to proliferate, a growing demand will be placed on leveraging the content in these images to characterize the clinical phenotype of disease and relate it to genotype and clinical data from patients to enhance research and clinical care.

Suggested Reading

Brinkley, J. F. (1991). Structural informatics and its applications in medicine and biology. *Academic Medicine, 66*(10), 589–591 Short introduction to the field.

Brinkley, J. F., & Rosse, C. (2002). Imaging and the human Brain project: A review. *Methods of Information in Medicine, 41*, 245–260 Review of image processing work related to the brain. Much of the brain-related material for this chapter was adapted from this article.

Deserno, T. M. (2011). *Biomedical image processing*. Berlin: Springer-Verlag. Edited book of current approaches to variety of biomedical image processing tasks.

Dreyer, K. J., & Geis, J. R. (2017 Dec). When machines think: Radiology's next frontier. *Radiology, 285*(3), 713–718.

Gonzalez, R. C., Woods, R. E., et al. (2009). *Digital Image processing using MATLAB*. S.I., Gatesmark Publishing. A comprehensive overview of image processing methods focusing on Matlab examples.

Pham, D. L., Xu, C., & Prince, J. L. (2000). Current methods in medical image segmentation. *Annual Review of Biomedical Engineering, 2*, 315–337 Overview of medical image segmentation.

Potchen, E. J. (2000). Prospects for progress in diagnostic imaging. *Journal of Internal Medicine, 247*(4), 411–424 Nontechnical description of newer imaging methods such as cardiac MRI, diffusion tensor imaging, fMRI, and molecular imaging. Current and potential use of these methods for diagnosis.

Robb, R. A. (2000). *Biomedical imaging, visualization, and analysis*. New York: Wiley-Liss. Overview of biomedical imaging modalities and processing techniques.

Sahiner, B., Pezeshk, A., Hadjiiski, L. M., Wang, X., Drukker, K., Cha, K. H., Summers, R. M., & Giger, M. L. (2019). Deep learning in medical imaging and radiation therapy. *Medical Physics, 46*(1), e1–e36.

Shapiro, L. G., & Stockman, G. C. (2001). *Computer vision*. Upper Saddle River: Prentice-Hall. Detailed description of many of the representations and methods used in image processing. Not specific to medicine, but most of the methods are applicable to medical imaging.

Soffer, S., Ben-Cohen, A., Shimon, O., Amitai, M. M., Greenspan, H., & Klang, E. (2019 Mar). Convolutional neural networks for radiologic images: A Radiologist's guide. *Radiology, 290*(3), 590–606.

Sonka, M., & Fitzpatrick, J. M. (2000). *Handbook of medical imaging, volume 2. Medical image processing and analysis*. Bellingham: SPIE Press. Overview of biomedical imaging modalities and processing techniques.

Wang, S., & Summers, R. M. (2012). Machine learning and radiology. *Medical Image Analysis, 16*(5), 933–951.

Yoo, T. S. (2004). *Insight into images: Principles and practice for segmentation, registration, and image analysis*. Wellesley: A K Peters Comprehensive overview of digital image processing methods with examples from the Insight Toolkit (ITK).

Questions for Discussion

1. How might you create an image processing pipeline to build an image-analysis program looking for abnormal cells in a PAP smear? How would you collect and incorporate semantic features into the program?
2. Why is segmentation so difficult to perform? Give two examples of ways by which current systems avoid the problem of automatic segmentation.
3. How might you build a decision-support system that is based on searching the hospital image archive for similar images and returning the diagnosis associated

with the most similar images? How might you make use of the semantic information in images in images to improve the accuracy of retrieval?

4. What are limitations of deep learning compared with other machine learning methods for creating image interpretation systems? When might a classical machine learning approach be better than deep learning?
5. Give an example of how knowledge about the problem to be solved (e.g., local anatomy in the image) could be used in future systems to aid in automatic segmentation.
6. Both images and free text share the characteristic that they are unstructured information; image processing methods to make the biomedical content in images explicit are very similar to related problems in natural language processing (NLP; ► Chap. 8). How are image processing methods and NLP similar in terms of (1) computer representation of the raw content? (2) representation of the semantic content? (3) processing of the content (e.g., what is the NLP equivalent of segmentation, or the image processing equivalent of named entity recognition)?

References

Abou-Elkacem, L., Bachawal, S. V., & Willmann, J. K. (2015). Ultrasound molecular imaging: Moving toward clinical translation. *European Journal of Radiology, 84*(9), 1685–1693.

Agrawal, M., Harwood, D., Duraiswami, R., Davis, L. S., & Luther, P. W. (2000). Three-dimensional ultrastructure from transmission electron microscope tilt series. *Proceedings, Second Indian Conference on Vision, Graphics and Image Processing*. Bangaore, India.

Aine, C. J. (1995). A conceptual overview and critique of functional neuroimaging techniques in humans I. MRI/fMRI and PET. *Critical Reviews in Neurobiology, 9*, 229–309.

Akgul, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., & Acar, B. (2011). Content-based image retrieval in radiology: Current status and future directions. *Journal of Digital Imaging, 24*(2), 208–222.

Al-Antari, M. A., Al-Masni, M. A., Choi, M. T., Han, S. M., & Kim, T. S. (2018). A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *International Journal of Medical Informatics, 117*, 44–54.

Alberini, J. L., Edeline, V., Giraudet, A. L., Champion, L., Paulmier, B., Madar, O., Poinsignon, A., Bellet, D., & Pecking, A. P. (2011). Single photon emission tomography/computed tomography (SPET/CT) and positron emission tomography/computed tomography (PET/CT) to image cancer. *Journal of Surgical Oncology, 103*(6), 602–606.

André, B., Vercauteren, T., Perchant, A., Buchner, A. M., Wallace, M. B., & Ayache, N. (2009). Introducing space and time in local feature-based Endomicroscopic image retrieval. Medical content-based retrieval for clinical decision support. B. Caputo, H. Mller, T. Syeda-Mahmood et al. Berlin, Heidelberg, Springer. *Lecture Notes in Computer Science, 5853*, 18–30.

Appel, B. (2001). Nomenclature and classification of lumbar disc pathology. *Neuroradiology, 43*(12), 1124–1125.

Armstrong, R. A. (2010). Review paper. Quantitative methods in neuropathology. *Folia Neuropathologica, 48*(4), 217–230.

Ashburner, J., & Friston, K. J. (1997). Multimodal image coregistration and partitioning - a unified framework. *NeuroImage, 6*(3), 209–217.

Avni. (2009). Addressing the ImageClef 2009 challenge using a patch-based visual words representation %U. http://www.clef-campaign.org/2009/working_notes/avni-paperCLEF2009.pdf. Working notes CLEF2009.

Baader, F. E., McGuinness, D. E., Nardi, D. E., Schneider, P. P. E., & Calvanese, D. E. (Eds.). (2003). *The description logic handbook: Theory, implementation and applications*. Cambridge University Press: New York.

Baker, J. A., Kornguth, P. J., Lo, J. Y., Williford, M. E., & Floyd, C. E., Jr. (1995). Breast cancer: Prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology, 196*(3), 817–822.

Baumann, B., Gotzinger, E., Pircher, M., Sattmann, H., Schuutze, C., Schlanitz, F., Ahlers, C., Schmidt-Erfurth, U., & Hitzenberger, C. K. (2010). Segmentation and quantification of retinal lesions in age-related macular degeneration using polarization-sensitive optical coherence tomography. *Journal of Biomedical Optics, 15*(6), 061704.

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., & Stein, L. A. (2004). OWL web ontology language reference. Technical report REC-owl-ref-20040210, the WorldwideWeb consortium, Available from http://www.w3.org/TR/2004/REC-owl-ref-20040210/

Becich, M. J. (2000). The role of the pathologist as tissue refiner and data miner: The impact of functional genomics on the modern pathology laboratory and the critical roles of pathology informatics and bioinformatics. *Molecular Diagnosis, 5*(4), 287–299.

Bengio, Y., Yao, L., Alain, G., & Vincent, P. (2013). Generalized Denoising auto-encoders as generative models. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 26* (pp. 899–907). Curran Associates, Inc.

Bennett, T. J., & Barry, C. J. (2009). Ophthalmic imaging today: An ophthalmic photographer's viewpoint – A review. *Clinical & Experimental Ophthalmology, 37*(1), 2–13.

Betancur, J., Commandeur, F., Motlagh, M., Sharir, T., Einstein, A. J., Bokhari, S., Fish, M. B., Ruddy, T. D., Kaufmann, P., Sinusas, A. J., Miller, E. J., Bateman, T. M., Dorbala, S., Di Carli, M., Germano, G., Otaki, Y., Tamarappoo, B. K., Dey, D., Berman, D. S., & Slomka, P. J. (2018). Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: A multicenter study. *JACC: Cardiovascular Imaging*.

Bidgood, W. D., Jr., & Horii, S. C. (1992). Introduction to the ACR-NEMA DICOM standard. *Radiographics, 12*(2), 345–355.

Billingsley, A. (2020). The Latest in Coronavirus (COVID-19) Testing Methods and Availability. Retrieved 5/17/2020, from https://www.goodrx.com/blog/coronavirus-covid-19-testing-updates-methods-cost-availability/

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford/New York: Clarendon Press; Oxford University Press.

Biswal, S., Resnick, D. L., Hoffman, J. M., & Gambhir, S. S. (2007). Molecular imaging: Integration of molecular imaging into the musculoskeletal imaging practice. *Radiology, 244*(3), 651–671.

Bittorf, A., Bauer, J., Simon, M., & Diepgen, T. L. (1997). Web-based training modules in dermatology. *MD Computing, 14*(5), 371–376. 381.

Bloom, F. E., & Young, W. G. (1993). *Brain browser*. New York: Academic.

Bodenreider, O. (2008). Biomedical ontologies in action: Role in knowledge management, data integration and decision support. *Yearbook of Medical Informatics*, 67–79.

Bogowicz, M., Riesterer, O., Stark, L. S., Studer, G., Unkelbach, J., Guckenberger, M., & Tanadini-Lang, S. (2017). Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma. *Acta Oncologica, 56*(11), 1531–1536.

Bosch, A., Munoz, X., Oliver, A., & Marti, J. (2006). Modeling and classifying breast tissue density in mammograms. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, 2*, 1552–1558.

Brain Innovation B.V. (2001). BrainVoyager. From http://www.BrainVoyager.de/

Brinkley, J. F. (1993). The potential for three-dimensional ultrasound. In F. A. Chervenak, G. C. Isaacson, & S. Campbell (Eds.), *Ultrasound in obstetrics and gynecology*. Boston: Little, Brown and Company.

Brinkley, J. F., Bradley, S. W., Sundsten, J. W., & Rosse, C. (1997). The digital anatomist information system and its use in the generation and delivery of web-based anatomy atlases. *Computers and Biomedical Research, 30*, 472–503.

Brinkley, J. F., Wong, B. A., Hinshaw, K. P., & Rosse, C. (1999). Design of an anatomy information system. *Computer Graphics and Applications, 19*(3), 38–48.

Brown, D. B., Gould, J. E., Gervais, D. A., Goldberg, S. N., Murthy, R., Millward, S. F., Rilling, W. S., Geschwind, J. F. S., Salem, R., Vedantham, S., Cardella, J. F., Soulen, M. C., Techn, S. I. R., & Tumor, I. W. G. I.-G. (2009). Transcatheter therapy for hepatic malignancy: Standardization of terminology and reporting criteria (reprinted from J Vasc Interv Radiol, vol 18, pg 1469-1478, 2007). *Journal of Vascular and Interventional Radiology, 20*(7), S425–S434.

Burnside, E., Rubin, D., & Shachter, R. (2000). A Bayesian network for mammography. *Proceedings of the AMIA Symposium*, 106–110.

Burnside, E. S., Rubin, D. L., & Shachter, R. D. (2004a). Using a Bayesian network to predict the probability and type of breast cancer represented by microcalcifications on mammography. *Studies in Health Technology and Informatics, 107*(Pt 1), 13–17.

Burnside, E. S., Rubin, D. L., Shachter, R. D., Sohlich, R. E., & Sickles, E. A. (2004b). A probabilistic expert system that provides automated mammographic-histologic correlation: Initial experience. *AJR. American Journal of Roentgenology, 182*(2), 481–488.

Burnside, E. S., Rubin, D. L., Fine, J. P., Shachter, R. D., Sisney, G. A., & Leung, W. K. (2006). Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: Initial experience. *Radiology, 240*(3), 666–673.

Burnside, E. S., Ochsner, J. E., Fowler, K. J., Fine, J. P., Salkowski, L. R., Rubin, D. L., & Sisney, G. A. (2007). Use of microcalcification descriptors in BI-RADS 4th edition to stratify risk of malignancy. *Radiology, 242*(2), 388–395.

Burnside, E. S., Davis, J., Chhatwal, J., Alagoz, O., Lindstrom, M. J., Geller, B. M., Littenberg, B., Shaffer, K. A., Kahn, C. E., Jr., & Page, C. D. (2009). Probabilistic computer model developed from clinical data in national mammography database format to classify mammographic findings. *Radiology, 251*(3), 663–672.

Buxton, R. B. (2009). Introduction to functional magnetic resonance imaging: Principles and techniques. Cambridge. In *UK*. New York: Cambridge University Press.

caBIG In-vivo Imaging Workspace. (2008). Annotation and Image Markup (AIM). Retrieved December 26, 2008, from https://cabig.nci.nih.gov/tools/AIM

Cabrera Fernandez, D., Salinas, H. M., & Puliafito, C. A. (2005). Automated detection of retinal layer structures on optical coherence tomography images. *Optics Express, 13*(25), 10200–10216.

Caputo, B., Tornmasi, T., & Orabona, F. (2008). Discriminative cue integration for medical image annotation. *Pattern Recognition Letters, 29*(15), 1996–2002.

Carneiro, G., Chan, A. B., Moreno, P. J., & Vasconcelos, N. (2007). Supervised learning of semantic classes for image annotation and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(3), 394–410.

Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., Moffat, J., Golland, P., & Sabatini, D. M. (2006). CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biology, 7*(10), R100.

Caviness, V. S., Meyer, J., Makris, N., & Kennedy, D. N. (1996). MRI-based topographic parcellation of human neocortex: An anatomically specified method with estimate of reliability. *Journal of Cognitive Neuroscience, 8*(6), 566–587.

Cha, Y. J., Jang, W. I., Kim, M. S., Yoo, H. J., Paik, E. K., Jeong, H. K., & Youn, S. M. (2018). Prediction of response to stereotactic radiosurgery for Brain metastases using convolutional neural networks. *Anticancer Research, 38*(9), 5437–5445.

Chakraborty, A., Staib, L. H., & Duncan, J. S. (1996). Deformable boundary finding in medical images by integrating gradient and region information. *IEEE Transactions on Medical Imaging, 15*(6), 859–870.

Channin, D. S., Mongkolwat, P., Kleper, V., & Rubin, D. L. (2009a). Computing human image annotation. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1*, 7065–7068.

Channin, D. S., Mongkolwat, P., Kleper, V., Sepukar, K., & Rubin, D. L. (2009b). The caBIG annotation and image markup project. *Journal of Digital Imaging*.

Chen, S., Liu, W., Qin, J., Chen, L., Bin, G., Zhou, Y., Wang, T., & Huang, B. (2017). Research progress of computer-aided diagnosis in cancer based on deep learning and medical imaging. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi, 34*(2), 314–319.

Choi, H. S., Haynor, D. R., & Kim, Y. (1991). Partial volume tissue classification of multichannel magnetic resonance images – a mixel model. *IEEE Transactions on Medical Imaging, 10*(3), 395–407.

Cimino, J. J. (1996). Review paper: Coding systems in health care. *Methods of Information in Medicine, 35*(4–5), 273–284.

Clarysse, P., Friboulet, D., & Magnin, I. E. (1997). Tracking geometrical descriptors on 3-D deformable surfaces: Application to the left-ventricular surface of the heart. *IEEE Transactions on Medical Imaging, 16*(4), 392–404.

Cohen, J. D. (2001). FisWidgets. 2001, from http://neurocog.lrdc.pitt.edu/fiswidgets/

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3-D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography, 18*(2), 192–205.

Collins, D. L., Holmes, D. J., Peters, T. M., & Evans, A. C. (1995). Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping, 3*, 190–208.

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(5), 603–619.

Corina, D. P., Poliakov, A. V., Steury, K., Martin, R. F., Brinkley, J. F., Mulligan, K. A., & Ojemann, G. A. (2000). Correspondences between language cortex identified by cortical stimulation mapping and fMRI. *Neuroimage (Human Brain Mapping Annual Meeting, June 12–16), 11*(5), S295.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29*, 162–173.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision And Pattern Recognition (CVPR'05)*.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage, 9*(2), 179–194.

Dameron, O., Roques, E., Rubin, D. L., Marquet, G., & Burgun, A. (2006). Grading lung tumors using OWL-DL based reasoning. *9th International Protégé Conference*. Stanford, CA.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys, 40*(2).

Davatzikos, C., & Bryan, R. N. (1996). Using a deformable surface model to obtain a shape representation of the cortex. *IEEE Transactions on Medical Imaging, 15*(6), 785–795.

de Figueiredo, E. H., Borgonovi, A. F., & Doring, T. M. (2011). Basic concepts of MR imaging, diffusion MR imaging, and diffusion tensor imaging. *Magnetic Resonance Imaging Clinics of North America, 19*(1), 1–22.

de Sisternes, L., Simon, N., Tibshirani, R., Leng, T., & Rubin, D. L. (2014). Quantitative SD-OCT imaging biomarkers as indicators of age-related macular degeneration progression. *Investigative Ophthalmology & Visual Science, 55*(11), 7093–7103.

de Sisternes, L., Jonna, G., Greven, M. A., Chen, Q., Leng, T., & Rubin, D. L. (2017a). Individual Drusen segmentation and repeatability and reproducibility of their automated quantification in optical coherence tomography images. *Translational Vision Science & Technology, 6*(1), 12.

de Sisternes, L., Jonna, G., Moss, J., Marmor, M. F., Leng, T., & Rubin, D. L. (2017b). Automated intraretinal segmentation of SD-OCT images in normal and age-related macular degeneration eyes. *Biomedical Optics Express, 8*(3), 1926–1949.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the

EM algorithm. *Journal of the Royal Statistical Society Series B, 39*, 1–38.

Deng, J., Dong, W., Socher, R., Li, L., Kai, L., & Li, F.-F. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*.

Depeursinge, A., Foncubierta-Rodriguez, A., Van De Ville, D., & Muller, H. (2014). Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities. *Medical Image Analysis, 18*(1), 176–196.

Deselaers, T., Hegerath, A., Keysers, D., & Ney, H. (2006). Sparse patch-histograms for object classification in cluttered images. In: *DAGM 2006, Pattern Recognition, 27th DAGM Symposium, Lecture Notes in Computer Science*, pp. 202–211.

Deselaers, T., Muller, H., Clough, P., Ney, H., & Lehmann, T. M. (2007). The CLEF 2005 automatic medical image annotation task. *International Journal of Computer Vision, 74*(1), 51–58.

Deserno, T. M., Antani, S., & Long, R. (2009). Ontology of gaps in content-based image retrieval. *Journal of Digital Imaging, 22*(2), 202–215.

Deshpande, N., Needles, A., & Willmann, J. K. (2010). Molecular ultrasound imaging: Current status and future directions. *Clinical Radiology, 65*(7), 567–581.

Dhenain, M., Ruffins, S. W., & Jacobs, R. E. (2001). Three-dimensional digital mouse atlas using high-resolution MRI. *Developmental Biology, 232*(2), 458–470.

DICOM Standards Committee. (2017). DICOM PS3.21 2017e – Transformations between DICOM and other representations; A.6 AIM v4 to DICOM TID 1500 mapping. From http://dicom.nema.org/medical/Dicom/2017e/output/chtml/part21/sect_A.6.html

DICOM Standards Committee – Working Group 8 – Structured Reporting. (2017). Digital Imaging and Communications in Medicine (DICOM); sup 200 – Transformation of NCI annotation and image markup (AIM) and DICOM SR measurement templates. From ftp://medical.nema.org/medical/dicom/Supps/LB/sup200_lb_AIM_DICOMSRTID1500.pdf

Diepgen, T. L., & Eysenbach, G. (1998). Digital images in dermatology and the dermatology online atlas on the world wide web. *The Journal of Dermatology, 25*(12), 782–787.

Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics, 31*(4–5), 198–211.

Donovan, T., & Manning, D. J. (2007). The radiology task: Bayesian theory and perception. *The British Journal of Radiology, 80*(954), 389–391.

D'Orsi, C. J., & Newell, M. S. (2007). BI-RADS decoded: Detailed guidance on potentially confusing issues. *Radiologic Clinics of North America, 45*(5), 751–763.

Drude, N., Tienken, L., & Mottaghy, F. M. (2017). Theranostic and nanotheranostic probes in nuclear medicine. *Methods, 130*, 14–22.

Drury, H. A., & Van Essen, D. C. (1997). Analysis of functional specialization in human cerebral cortex using the visible man surface based atlas. *Human Brain Mapping, 5*, 233–237.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: Wiley.

Dugas-Phocion, G., Ballester, M. A. G., Malandain, G., Lebrun, C., & Ayache, N. (2004). Improved EM-based tissue segmentation and partial volume effect quantification in multi-sequence brain MRI. *Medical Image Computing and Computer-Assisted Intervention – Miccai 2004, Pt 1, Proceedings, 3216*, 26–33.

Durot, I., Wilson, S. R., & Willmann, J. K. (2018). Contrast-enhanced ultrasound of malignant liver lesions. *Abdominal Radiology (NY), 43*(4), 819–847.

Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J., The, C. C., Hermsen, M., Manson, Q. F., Balkenhol, M., Geessink, O., Stathonikos, N., van Dijk, M. C., Bult, P., Beca, F., Beck, A. H., Wang, D., Khosla, A., Gargeya, R., Irshad, H., Zhong, A., Dou, Q., Li, Q., Chen, H., Lin, H. J., Heng, P. A., Hass, C., Bruni, E., Wong, Q., Halici, U., Oner, M. U., Cetin-Atalay, R., Berseth, M., Khvatkov, V., Vylegzhanin, A., Kraus, O., Shaban, M., Rajpoot, N., Awan, R., Sirinukunwattana, K., Qaiser, T., Tsang, Y. W., Tellez, D., Annuscheit, J., Hufnagl, P., Valkonen, M., Kartasalo, K., Latonen, L., Ruusuvuori, P., Liimatainen, K., Albarqouni, S., Mungal, B., George, A., Demirci, S., Navab, N., Watanabe, S., Seno, S., Takenaka, Y., Matsuda, H., Phoulady, H. A., Kovalev, V., Kalinovsky, A., Liauchuk, V., Bueno, G., Fernandez-Carrobles, M. M., Serrano, I., Deniz, O., Racoceanu, D., & Venancio, R. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA, 318*(22), 2199–2210.

Endo, M., Aramaki, T., Asakura, K., Moriguchi, M., Akimaru, M., Osawa, A., Hisanaga, R., Moriya, Y., Shimura, K., Furukawa, H., & Yamaguchi, K. (2012). Content-based image-retrieval system in chest computed tomography for a solitary pulmonary nodule: Method and preliminary experiments. *International Journal of Computer Assisted Radiology and Surgery, 7*(2), 331–338.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118.

Eysenbach, G., Bauer, J., Sager, A., Bittorf, A., Simon, M., & Diepgen, T. (1998). An international dermatological image atlas on the WWW: Practical use for undergraduate and continuing medical education, patient education and epidemiological research. *Studies in Health Technology and Informatics, 52*(Pt 2), 788–792.

Faruque, J., Rubin, D. L., Beaulieu, C. F., & Napel, S. (2013). Modeling perceptual similarity measures in CT images of focal liver lesions. *Journal of Digital Imaging, 26*(4), 714–720.

Faruque, J., Beaulieu, C. F., Rosenberg, J., Rubin, D. L., Yao, D., & Napel, S. (2015). Content-based image retrieval in radiology: Analysis of variability in human perception of similarity. *Journal of Medical Imaging (Bellingham), 2*(2), 025501.

Fave, X., Zhang, L., Yang, J., Mackin, D., Balter, P., Gomez, D., Followill, D., Jones, A. K., Stingo, F., Liao, Z., Mohan, R., & Court, L. (2017). Delta-radiomics features for the prediction of patient outcomes in non-small cell lung cancer. *Scientific Reports, 7*(1), 588.

Federative Committee on Anatomical Terminology (1998). *Terminologia Anatomica*. Stuttgart, Thieme.

Fedorov, A. (2012). 3D Slicer Annotation Image Markup. Accessed 11/11/2012, from http://www.na-mic.org/Wiki/index.php/Projects:QIN:3D_Slicer_Annotation_Image_Markup

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., & Kikinis, R. (2012). 3D Slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging, 30*(9), 1323–1341.

Fei-Fei, L., & Perona, P. (2003) A Bayesian hierarchical model for learning natural scene categories. *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 524–531.

Ferrante, E., & Paragios, N. (2017). Slice-to-volume medical image registration: A survey. *Medical Image Analysis, 39*, 101–123.

Fiala, J. C., & Harris, K. M. (2001). Extending unbiased stereology of brain ultrastructure to three-dimensional volumes. *Journal of the American Medical Association, 8*(1), 1–16.

Figurska, M., Robaszkiewicz, J., & Wierzbowska, J. (2010). Optical coherence tomography in imaging of macular diseases. *Klinika Oczna, 112*(4–6), 138–146.

Firmino, M., Morais, A. H., Mendoca, R. M., Dantas, M. R., Hekis, H. R., & Valentim, R. (2014). Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects. *Biomedical Engineering Online, 13*, 41.

FMRIDB Image Analysis Group. (2001). FSL – The FMRIB Software Libarary. From http://www.fmrib.ox.ac.uk/fsl/index.html

Fougerousse, F., Bullen, P., Herasse, M., Lindsay, S., Richard, I., Wilson, D., Suel, L., Durand, M., Robson, S., Abitbol, M., Beckmann, J. S., & Strachan, T. (2000). Human-mouse differences in the embryonic expression of developmental control genes and disease genes. *Human Molecular Genetics, 9*(2), 165–173.

Fox, P. T. (Ed.). (2001). *Human brain mapping*. New York: John Wiley & Sons.

Frackowiak, R. S. J., Friston, K. J., Frith, C. D., Dolan, R. J., & Mazziotta, J. C. (Eds.). (1997). *Human brain function*. Academic Press: New York.

Freton, A., & Finger, P. T. (2011). Spectral domain-optical coherence tomography analysis of choroidal osteoma. *The British Journal of Ophthalmology*.

Friefeld, O., Greenspan, H., & Jacob, G. (2009). Multiple sclerosis lesion detection using constrained GMM and curve evolution. *Journal of Biomedical Imaging, 2009*, 1–13.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping, 2*, 189–210.

Gabril, M. Y., & Yousef, G. M. (2010). Informatics for practicing anatomical pathologists: Marking a new era in pathology practice. *Modern Pathology, 23*(3), 349–358.

Gastounioti, A., Oustimov, A., Hsieh, M. K., Pantalone, L., Conant, E. F., & Kontos, D. (2018). Using convolutional neural networks for enhanced capture of breast parenchymal complexity patterns associated with breast cancer risk. *Academic Radiology, 25*(8), 977–984.

George, J. S., Aine, C. J., Mosher, J. C., Schmidt, D. M., Ranken, D. M., Schlitz, H. A., Wood, C. C., Lewine, J. D., Sanders, J. A., & Belliveau, J. W. (1995). Mapping function in human brain with magnetoencephalography, anatomical magnetic resonance imaging, and functional magnetic resonance imaging. *Journal of Clinical Neurophysiology, 12*(5), 406–431.

Gerstner, E. R., & Sorensen, A. G. (2011). Diffusion and diffusion tensor imaging in brain cancer. *Seminars in Radiation Oncology, 21*(2), 141–146.

Gevaert, O., Xu, J., Hoang, C., Leung, A., Quon, A., Rubin, D. L., Napel, S., & Plevritis, S. (2011). Integrating medical images and transcriptomic data in non-small cell lung cancer. AACR 102nd annual meeting. Orlando, FL.

Gevaert, O., Mitchell, L. A., Xu, J., Yu, C., Rubin, D., Zaharchuk, G., Napel, S., & Plevritis, S. (2012a). Radiogenomic analysis indicates MR images are potentially predictive of EGFR mutation status in glioblastoma multiforme. AACR 103nd annual meeting. Chicago, IL.

Gevaert, O., Hoang, C. D., Leung, A. N., Xu, J., Quon, A., Rubin, D. L., Napel, S., & Plevritis, S. (2012b). Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data–methods and preliminary results. *Radiology*, 264:387–396. PMID:22723499. PMCID:PMC3401348.

Giger, M. L. (2018). Machine learning in medical imaging. *Journal of the American College of Radiology, 15*(3 Pt B), 512–520.

Giger, M., & MacMahon, H. (1996). Image processing and computer-aided diagnosis. *The Radiologic Clinics of North America, 34*(3), 565–596.

Gimenez, F., Xu, J., Liu, T. T., Beaulieu, C., Rubin, D. L., Napel, S., & Liu, Y. (2011a). Prediction of radiologist observations using computational image features: Method and preliminary results. *Ninety-seventy annual scientific meeting of the RSNA*, Chicago, IL.

Gimenez, F., Xu, J., Liu, Y., Liu, T. T., Beaulieu, C., Rubin, D. L, & Napel, S. (2011b). On the Feasibility

of Predicting Radiological Observations from Computational Imaging Features of Liver Lesions in CT scans. *First IEEE Conference on Healthcare Informatics, Imaging, and Systems Biology (HISB), IEEE Computer Society*, San Jose, CA.

Goldberg, S. N., Grassi, C. J., Cardella, J. F., Charboneau, J. W., Dodd, G. D., 3rd, Dupuy, D. E., Gervais, D. A., Gillams, A. R., Kane, R. A., Lee, F. T., Jr., Livraghi, T., McGahan, J., Phillips, D. A., Rhim, H., Silverman, S. G., Solbiati, L., Vogl, T. J., Wood, B. J., Vedantham, S., & Sacks, D. (2009). Image-guided tumor ablation: Standardization of terminology and reporting criteria. *Journal of Vascular and Interventional Radiology, 20*(7 Suppl), S377–S390.

Gombas, P., Skepper, J. N., Krenacs, T., Molnar, B., & Hegyi, L. (2004). Past, present and future of digital pathology. *Orvosi Hetilap, 145*(8), 433–443.

Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2009). *Digital image processing using MATLAB*. S.I., Gatesmark Publishing.

Grau, B., Horrocks, I., Motik, B., Parsia, B., Patelschneider, P., & Sattler, U. (2008). Chapter 3: Description logics. In B. Porter, V. Lifschitz, & F. Van Harmelen (Eds.), *Handbook of knowledge representation* (p. 1005). Amsterdam/Boston: Elsevier: xxviii.

Greenspan, H., & Pinhas, A. T. (2007). Medical image categorization and retrieval for PACS using the GMM-KL framework. *IEEE Transactions on Information Technology in Biomedicine, 11*(2), 190–202.

Greenspan, H., Ruf, A., & Goldberger, J. (2006). Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. *IEEE Trans Med Imaging, 25*(9): 1233–1245.

Greenspan, H., Avni, U., Konen, E., Sharon, M., & Goldberger, J. (2011). X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. *IEEE Transactions on Medical Imaging, 30*(3), 733–746.

Greenspan, H., Ginneken, B. V., & Summers, R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging, 35*(5), 1153–1159.

Greenspan, H., Estepar, R. S., Niessen, W. J., Siegel, E., & Nielsen, M. (2020). Position paper on COVID-19 imaging and AI: From the clinical needs and technological challenges to initial AI solutions at the lab and national level towards a new era for AI in healthcare. *Medical Image Analysis, 66*.

Grossmann, P., Stringfield, O., El-Hachem, N., Bui, M. M., Rios Velazquez, E., Parmar, C., Leijenaar, R. T., Haibe-Kains, B., Lambin, P., Gillies, R. J., & Aerts, H. J. (2017). Defining the biological basis of radiomic phenotypes in lung cancer. *eLife, 6*.

Hansell, D. M., Bankier, A. A., MacMahon, H., McLoud, T. C., Muller, N. L., & Remy, J. (2008). Fleischner society: Glossary of terms for thoracic imaging. *Radiology, 246*(3), 697–722.

Hansen, L. K., Nielsen, F. A., Toft, P., Liptrot, M. G., Goutte, C., Strother, S. C., Lange, N., Gade, A., Rottenberg, D. A., & Paulson, O. B. (1999). Lyngby – Modeler's Matlab toolbox for spatio-temporal analysis of functional neuroimages. *NeuroImage, 9*(6), S241.

Haralick, R. M. (1988). *Mathematical morphology*. University of Washington.

Haralick, R. M., & Shapiro, L. G. (1992). *Computer and robot vision*. Reading: Addison-Wesley.

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-3*(6), 610–621.

Harney, A. S., & Meade, T. J. (2010). Molecular imaging of in vivo gene expression. *Future Medicinal Chemistry, 2*(3), 503–519.

Hasan, K. M., Walimuni, I. S., Abid, H., & Hahn, K. R. (2010). A review of diffusion tensor magnetic resonance imaging computational methods and software tools. *Computers in Biology and Medicine*.

Heiss, W. D., & Phelps, M. E. (Eds.). (1983). *Positron emission tomography of the brain*. Berlin/New York: Springer.

Held, K., Rota Kops, E., Krause, B. J., Wells, W. M., 3rd, Kikinis, R., & Muller-Gartner, H. W. (1997). Markov random field segmentation of brain MR images. *IEEE Transactions on Medical Imaging, 16*(6), 878–886.

Hersh, W., Muller, H., & Kalpathy-Cramer, J. (2009). The ImageCLEFmed medical image retrieval task test collection. *Journal of Digital Imaging, 22*(6), 648–655.

Hinshaw, K. P., Poliakov, A. V., Martin, R. F., Moore, E. B., Shapiro, L. G., & Brinkley, J. F. (2002). Shape-based cortical surface segmentation for visualization brain mapping. *NeuroImage, 16*(2), 295–316.

Hoang, C., Napel, S., Gevaert, O., Xu, J., Rubin, D. L., Leung, A., Merritt, R., Whyte, R., Shrager, J., & Plevritis, S. (2011). *NSCLC gene profiles correlate with specific CT characteristics: Image-omics*. Philadelphia: American Association for Thoracic Surgery (AATS).

Hoffman, J. M., & Gambhir, S. S. (2007). Molecular imaging: The vision and opportunity for radiology in the future. *Radiology, 244*(1), 39–47.

Hohne, K., Bomans, M., Pommert, A., Riemer, M., Schiers, C., Tiede, U., & Wiebecke, G. (1990). 3-D visualization of tomographic volume data using the generalized voxel model. *The Visual Computer, 6*(1), 28–36.

Hohne, K. H., Bomans, M., Riemer, M., Schubert, R., Tiede, U., & Lierse, W. (1992). A volume-based anatomical atlas. *IEEE Computer Graphics and Applications*, 72–78.

Hohne, K. H., Pflesser, B., Riemer, M., Schiemann, T., Schubert, R., & Tiede, U. (1995). A new representation of knowledge concerning human anatomy and function. *Nature Medicine, 1*(6), 506–510.

Hoo-Chang, S., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging, 35*(5), 1285–1298.

Hoogi, A., Subramaniam, A., Veerapaneni, R., & Rubin, D. L. (2017). Adaptive estimation of active contour parameters using convolutional neural networks and texture analysis. *IEEE Transactions on Medical Imaging, 36*(3), 781–791.

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. (2018). Artificial intelligence in radiology. *Nature Reviews. Cancer, 18*(8), 500–510.

Horii SC. Image acquisition. Sites, technologies, and approaches. *Radiol Clin North Am,* 1996;34:469–494. PMID:8657867.

Hu, Z., Abramoff, M. D., Kwon, Y. H., Lee, K., & Garvin, M. K. (2010a). Automated segmentation of neural canal opening and optic cup in 3D spectral optical coherence tomography volumes of the optic nerve head. *Investigative Ophthalmology & Visual Science, 51*(11), 5708–5717.

Hu, Z., Niemeijer, M., Abramoft, M. D., Lee, K., & Garvin, M. K. (2010b). Automated segmentation of 3-D spectral OCT retinal blood vessels by neural canal opening false positive suppression. *Medical Image Computing and Computer-Assisted Intervention, 13*(Pt 3), 33–40.

Huang, Y. Q., Liang, C. H., He, L., Tian, J., Liang, C. S., Chen, X., Ma, Z. L., & Liu, Z. Y. (2016). Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *Journal of Clinical Oncology, 34*(18), 2157–2164.

Hudson, D. L., & Cohen, M. E. (2009). Multidimensional medical decision making. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1*, 3405–3408.

International Anatomical Nomenclature Committee. (1989). *Nomina Anatomica*. Edinburgh: Churchill Livingstone.

Ishioka, J., Matsuoka, Y., Uehara, S., Yasuda, Y., Kijima, T., Yoshida, S., Yokoyama, M., Saito, K., Kihara, K., Numao, N., Kimura, T., Kudo, K., Kumazawa, I., & Fujii, Y. (2018). Computer-aided diagnosis of prostate cancer on magnetic resonance imaging using a convolutional neural network algorithm. *BJU International, 122*(3), 411–417.

Jiang, Y.-G., Ngo, C.-W., & Yang, J. (2007). Towards optimal bag-of-features for object categorization and semantic video retrieval. *Proceedings of the 6th ACM international conference on Image and video retrieval*. Amsterdam, The Netherlands, ACM, pp. 494–501.

Johnson, K. A., & Becker, J. A.. (2001). The Whole Brain Atlas. 2001, from http://www.med.harvard.edu/AANLIB/home.html

Jokerst, J. V., & Gambhir, S. S. (2011). Molecular imaging with theranostic nanoparticles. *Accounts of Chemical Research, 44*(10), 1050–1060.

Jun, W., Xia, L., Di, D., Jiangdian, S., Min, X., Yali, Z., & Jie, T. (2016). Prediction of malignant and benign of lung tumor using a quantitative radiomic method. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2016*, 1272–1275.

Jurie, F., & Triggs, B. (2005). Creating efficient codebooks for visual recognition. *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 – Volume 01, IEEE Computer Society*: 604–610%@ 600–7695-2334-X-7601.

Kahn, C. E., Jr., Langlotz, C. P., Burnside, E. S., Carrino, J. A., Channin, D. S., Hovsepian, D. M., & Rubin, D. L. (2009). Toward best practices in radiology reporting. *Radiology, 252*(3), 852–856.

Kahn, C. E., & Rubin, D. L. (2009). Automated semantic indexing of figure captions to improve radiology image retrieval. *Journal of the American Medical Informatics Association, 16*(3), 380–386.

Kang, J. H., & Chung, J. K. (2008). Molecular-genetic imaging based on reporter gene expression. *Journal of Nuclear Medicine, 49*(Suppl 2), 164S–179S.

Kapur, T., Grimson, W. E., Wells, W. M., 3rd, & Kikinis, R. (1996). Segmentation of brain tissue from magnetic resonance images. *Medical Image Analysis, 1*(2), 109–127.

Kass, M., Witkin, A., & Terzopoulos, D. (1987). Snakes: Active contour models. *International Journal of Computer Vision, 1*(4), 321–331.

Katragadda, C., Finnane, A., Soyer, H. P., Marghoob, A. A., Halpern, A., Malvehy, J., Kittler, H., Hofmann-Wellenhof, R., Da Silva, D., Abraham, I., Curiel-Lewandrowski, C., & G. International Society of Digital Imaging of the Skin -International Skin Imaging Collaboration. (2016). Technique standards for skin lesion imaging: A Delphi consensus statement. *JAMA Dermatology*.

Kayalibay, B., Jensen, G., & van der Smagt, P. (2017). CNN-based segmentation of medical imaging data. arXiv:1701.03056 [cs].

Kennedy, D. (2001). Internet brain segmentation repository. 2001, from http://neuro-www.mgh.harvard.edu/cma/ibsr

Kevles, B. (1997). *Naked to the bone: Medical imaging in the twentieth century*. New Brunswick: Rutgers University Press.

Kimborg, D. Y., & Aguirre, G. K.. (2002). A flexible architecture for neuroimaging data analysis and presentation. From http://www.nimh.nih.gov/neuroinformatics/kimberg.cfm

King, W., Proffitt, J., Morrison, L., Piper, J., Lane, D., & Seelig, S. (2000). The role of fluorescence in situ hybridization technologies in molecular diagnostics and disease management. *Molecular Diagnosis, 5*(4), 309–319.

Klinger, C. (2010). AIM on ClearCanvas Workstation Documentation. Accessed 11/11/2012, from https://

wiki.nci.nih.gov/display/AIM/AIM+on+ClearCanvas+Workstation+Documentation

Kontos, D., Summers, R. M., & Giger, M. (2017). Special section guest editorial: Radiomics and deep learning. *Journal of Medical Imaging (Bellingham), 4*(4), 041301.

Korner, M., Weber, C. H., Wirth, S., Pfeifer, K. J., Reiser, M. F., & Treitl, M. (2007). Advances in digital radiography: Physical principles and system overview. *Radiographics, 27*(3), 675–686.

Koslow, S. H., & Huerta, M. F. (Eds.). (1997). *Neuroinformatics: An overview of the human Brain project*. Mahwah: Lawrence Erlbaum.

Kremkau, F. W. (2006). *Diagnostic ultrasound principles and instruments*. St. Louis: Saunders Elsevier.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc., pp. 1097–1105.

Kulikowski, C. A. (1997). Medical imaging informatics: Challenges of definition and integration. *Journal of the American Medical Informatics Association, 4*(3), 252–253.

Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., Forster, K., Aerts, H. J., Dekker, A., Fenstermacher, D., Goldgof, D. B., Hall, L. O., Lambin, P., Balagurunathan, Y., Gatenby, R. A., & Gillies, R. J. (2012). Radiomics: The process and the challenges. *Magnetic Resonance Imaging, 30*(9), 1234–1248.

Kumar, A., Kim, J., Cai, W., Fulham, M., & Feng, D. (2013). Content-based medical image retrieval: A survey of applications to multidimensional and multimodality data. *Journal of Digital Imaging, 26*(6), 1025–1039.

Kumar, A., Gupta, P. K., & Srivastava, A. (2020). A review of modern technologies for tackling COVID-19 pandemic. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews, 14*(4), 569–573.

Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G., Granton, P., Zegers, C. M., Gillies, R., Boellard, R., Dekker, A., & Aerts, H. J. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer, 48*(4), 441–446.

Langlotz, C. P. (2006). RadLex: A new method for indexing online educational materials. *Radiographics, 26*(6), 1595–1597.

Langlotz, C. P. (2009). Structured radiology reporting: Are we there yet? *Radiology, 253*(1), 23–25.

Larabell, C. A., & Nugent, K. A. (2010). Imaging cellular architecture with X-rays. *Current Opinion in Structural Biology, 20*(5), 623–631.

Le Bihan, D., Mangin, J. F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., & Chabriat, H. (2001). Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging, 13*(4), 534–546.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Ledley, R. S., & Lusted, L. B. (1991). Reasoning foundations of medical diagnosis. *MD Computing, 8*(5), 300–315.

Lee, D. H. (2003). Magnetic resonance angiography. *Advances in Neurology, 92*, 43–52.

Lee, J. K. T. (2006). *Computed body tomography with MRI correlation*. Philadelphia: Lippincott Williams & Wilkins.

Lee, Y., Kim, N., Cho, K. S., Kang, S. H., Kim, D. Y., Jung, Y. Y., & Kim, J. K. (2009). Bayesian classifier for predicting malignant renal cysts on MDCT: Early clinical experience. *AJR. American Journal of Roentgenology, 193*(2), W106–W111.

Lee, J. H., Baek, J. H., Kim, J. H., Shim, W. H., Chung, S. R., Choi, Y. J., & Lee, J. H. (2018). Deep learning-based computer-aided diagnosis system for localization and diagnosis of metastatic lymph nodes on ultrasound: A pilot study. *Thyroid, 28*(10), 1332–1338.

Lehmann, T. M., Guld, M. O., Thies, C., Fischer, B., Spitzer, K., Keysers, D., Ney, H., Kohnen, M., Schubert, H., & Wein, B. B. (2004). Content-based image retrieval in medical applications. *Methods of Information in Medicine, 43*(4), 354–361.

Leong, F. J., & Leong, A. S. (2003). Digital imaging applications in anatomic pathology. *Advances in Anatomic Pathology, 10*(2), 88–95.

Levy, M. A., & Rubin, D. L. (2008). Tool support to enable evaluation of the clinical response to treatment. *American Medical Informatics Association Annual Symposium Proceedings*, 399–403.

Levy, M. A., & Rubin, D. L. (2011a). Computational approaches to assist in the evaluation of cancer treatment response. *Imaging in Medicine, 3*(2), 233–246.

Levy, M. A., & Rubin, D. L. (2011b). Current and future trends in imaging informatics for oncology. *Cancer Journal, 17*(4), 203–210.

Levy, M. A., O'Connor, M. J., & Rubin, D. L. (2009). Semantic reasoning with image annotations for tumor assessment. *American Medical Informatics Association Annual Symposium Proceedings, 2009*, 359–363.

Levy, M. A., Freymann, J. B., Kirby, J. S., Fedorov, A., Fennessy, F. M., Eschrich, S. A., Berglund, A. E., Fenstermacher, D. A., Tan, Y., Guo, X., Casavant, T. L., Brown, B. J., Braun, T. A., Dekker, A., Roelofs, E., Mountz, J. M., Boada, F., Laymon, C., Oborski, M., & Rubin, D. L. (2012). Informatics methods to enable sharing of quantitative imaging research data. *Magnetic Resonance Imaging, 30*(9), 1249–1256.

Lexe, G., Monaco, J., Doyle, S., Basavanhally, A., Reddy, A., Seiler, M., Ganesan, S., Bhanot, G., & Madabhushi, A. (2009). Towards improved cancer diagnosis and prognosis using analysis of gene expression data and computer aided imaging. *Experimental Biology and Medicine (Maywood, N.J.), 234*(8), 860–879.

Li, B. N., Chui, C. K., Chang, S., & Ong, S. H. (2011a). Integrating spatial fuzzy clustering with level set

methods for automated medical image segmentation. *Computers in Biology and Medicine, 41*(1), 1–10.

Li, C., Huang, R., Ding, Z., Gatenby, J. C., Metaxas, D. N., & Gore, J. C. (2011b). A level set method for image segmentation in the presence of intensity Inhomogeneities with application to MRI. *IEEE Transactions on Image Processing, 20*(7), 2007–2016.

Li, H., Zhu, Y., Burnside, E. S., Huang, E., Drukker, K., Hoadley, K. A., Fan, C., Conzen, S. D., Zuley, M., Net, J. M., Sutton, E., Whitman, G. J., Morris, E., Perou, C. M., Ji, Y., & Giger, M. L. (2016). Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. *NPJ Breast Cancer, 2*.

Liang, W., Liang, H., Ou, L., Chen, B., Chen, A., Li, C., Li, Y., Guan, W., Sang, L., Lu, J., Xu, Y., Chen, G., Guo, H., Guo, J., Chen, Z., Zhao, Y., Li, S., Zhang, N., Zhong, N., He, J., & C. China Medical Treatment Expert Group for. (2020). Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Internal Medicine*.

Lichtenbelt, B., Crane, R., & Naqvi, S. (1998). *Introduction to volume rendering*. Prentice Hall: Upper Saddle River.

Liu, Y. I., Kamaya, A., Desser, T. S., & Rubin, D. L. (2009). A controlled vocabulary to represent sonographic features of the thyroid and its application in a Bayesian network to predict thyroid nodule malignancy. *Summit on Translat Bioinforma, 2009*, 68–72.

Liu, Y. I., Kamaya, A., Desser, T. S., & Rubin, D. L. (2011). A bayesian network for differentiating benign from malignant thyroid nodules using sonographic and demographic features. *AJR. American Journal of Roentgenology, 196*(5), W598–W605.

Liu, Z., Jin, C., Wu, C. C., Liang, T., Zhao, H., Wang, Y., Wang, Z., Li, F., Zhou, J., Cai, S., Zeng, L., & Yang, J. (2020). Association between initial chest CT or clinical features and clinical course in patients with coronavirus disease 2019 pneumonia. *Korean Journal of Radiology, 21*(6), 736–745.

Lohmann, P., Kocher, M., Steger, J., & Galldiks, N. (2018). Radiomics derived from amino-acid PET and conventional MRI in patients with high-grade gliomas. *The Quarterly Journal of Nuclear Medicine and Molecular Imaging, 62*(3), 272–280.

Lowe, D. (1999). Object recognition from local scale invariant features. *Proceedings of the International Conference on Computer Vision*, Greece, pp. 1150–1157.

Lowe, H. J., Antipov, I., Hersh, W., & Smith, C. A. (1998). Towards knowledge-based retrieval of medical images. The role of semantic indexing, image content representation and knowledge-based retrieval. *Proc AMIA Symp*, pp. 882–886.

Luo, Z., Wang, N., Liu, P., Guo, Q., Ran, L., Wang, F., Tang, Y., & Li, Q. (2020). Association between chest CT features and clinical course of coronavirus disease 2019. *Respiratory Medicine, 168*, 105989.

Lusted, L. B. (1960). Logical analysis in roentgen diagnosis. *Radiology, 74*, 178–193.

MacDonald, D. (1993). Register, McConnel Brain Imaging Center, Montreal Neurological Institute.

MacDonald, D., Kabani, N., Avis, D., & Evans, A. C. (2000). Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage, 12*(3), 340–356.

Margolis, D. J., Hoffman, J. M., Herfkens, R. J., Jeffrey, R. B., Quon, A., & Gambhir, S. S. (2007). Molecular imaging techniques in body imaging. *Radiology, 245*(2), 333–356.

Marquet, G., Dameron, O., Saikali, S., Mosser, J., & Burgun, A. (2007). Grading glioma tumors using OWL-DL and NCI thesaurus. *American Medical Informatics Association Annual Symposium Proceedings*, 508–512.

Marroquin, J. L., Vemuri, B. C., Botello, S., Calderon, F., & Fernandez-Bouzas, A. (2002). An accurate and efficient bayesian method for automatic segmentation of brain MRI. *IEEE Transactions on Medical Imaging, 21*(8), 934–945.

Martin, R. F., & Bowden, D. M. (2001). *Primate Brain maps: Structure of the macaque Brain*. New York: Elsevier Science.

Martin, R. F., Mejino, J. L. V., Bowden, D. M., Brinkley, J. F., & Rosse, C. (2001). Foundational model of neuroanatomy: implications for the Human Brain Project. *Proc AMIA Annu Fall Symp*. Washington, DC, pp. 438–442.

Marwede, D., Schulz, T., & Kahn, T. (2008). Indexing thoracic CT reports using a preliminary version of a standardized radiological lexicon (RadLex). *Journal of Digital Imaging, 21*(4), 363–370.

Massoud, T. F., & Gambhir, S. S. (2003). Molecular imaging in living subjects: Seeing fundamental biological processes in a new light. *Genes and Development, 17*, 545–580.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Mechouche, A., Golbreich, C., Morandi, X., & Gibaud, B. (2008). Ontology-based annotation of brain MRI images. *AMIA Annu Symp Proc*, pp. 460–464.

Mehta, T. S., Raza, S., & Baum, J. K. (2000). Use of Doppler ultrasound in the evaluation of breast carcinoma. *Seminars in Ultrasound, CT, and MR, 21*(4), 297–307.

Mei, X., H.-C. Lee, K. Diao, M. Huang, B. Lin, C. Liu, Z. Xie, Y. Ma, P. M. Robson, M. Chung, A. Bernheim, V. Mani, C. Calcagno, K. Li, S. Li, H. Shan, J. Lv, T. Zhao, J. Xia, Q. Long, S. Steinberger, A. Jacobi, T. Deyer, M. Luksza, F. Liu, B. P. Little, Z. A. Fayad and Y. Yang (2020). Artificial intelligence-enabled rapid diagnosis of COVID-19 patients. medRxiv: 2020.2004.2012.20062661.

Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *2016 Fourth International Conference on 3D Vision (3DV)*.

Min, J. J., & Gambhir, S. S. (2008). Molecular imaging of PET reporter gene expression. *Handbook of Experimental Pharmacology, 185 Pt 2*, 277–303.

MIT Technology Review. (2013). Deep Learning. With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart. Retrieved 10/5/2018, from https://www.technologyreview.com/s/513696/deep-learning/

Modayur, B., Prothero, J., Ojemann, G., Maravilla, K., & Brinkley, J. F. (1997). Visualization-based mapping of language function in the brain. *NeuroImage, 6*, 245–258.

Motik, B., Grau, B. C., Horrocks, I., Parsia, B., Patel-Schneider, P., & Sattler, U. (2008). OWL 2: The next step for OWL. *Journal of Web Semantics, 6*(4), 309–322.

Motik, B., Shearer, R., & Horrocks, I. (2009). Hypertableau reasoning for description logics. *Journal of Artificial Intelligence Research, 36*, 165–228.

Muller, H., Michoux, N., Bandon, D., & Geissbuhler, A. (2004). A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International Journal of Medical Informatics, 73*(1), 1–23.

Muramatsu, C. (2018). Overview on subjective similarity of images for content-based medical image retrieval. *Radiological Physics and Technology*.

Nagarkar, D. B., Mercan, E., Weaver, D. L., Brunye, T. T., Carney, P. A., Rendi, M. H., Beck, A. H., Frederick, P. D., Shapiro, L. G., & Elmore, J. G. (2016). Region of interest identification and diagnostic agreement in breast pathology. *Modern Pathology, 29*(9), 1004–1011.

Napel, S. A., Beaulieu, C. F., Rodriguez, C., Cui, J., Xu, J., Gupta, A., Korenblum, D., Greenspan, H., Ma, Y., & Rubin, D. L. (2010). Automated retrieval of CT images of liver lesions on the basis of image similarity: Method and preliminary results. *Radiology, 256*(1), 243–252.

Napel, S., Hoang, C., Xu, J., Gevaert, O., Rubin, D. L., Plevritis, S., Xu, Y., Leung, A., & Quon, A. (2011). Computational and semantic annotation of CT and PET images and integration with genomic assays of tumors in non-small cell lung cancer (NSCLC) for decision support and discovery: method and preliminary results. *Ninety-seventy annual scientific meeting of the RSNA*, Chicago, IL.

National Cancer Institute. (2012). Annotation and image markup on clear canvas. Retrieved Accessed 11/7/2012, from https://wiki.nci.nih.gov/display/AIM/Annotation+and+Image+Markup+-+AIM

National Library of Medicine. (1999). *Medical subject headings – Annotated alphabetic list*. Bethesda: U.S. Department of Health and Human Services, Public Health Service.

Neumann, H., Kiesslich, R., Wallace, M. B., & Neurath, M. F. (2010). Confocal laser endomicroscopy: Technical advances and clinical applications. *Gastroenterology, 139*(2), 388–392. 392 e381-382.

Ng, A. Y., Jordan, M., & Weiss, Y. (2001). On spectral clustering: analysis and an algorithm. In: Advances in Neural Information Processing Systems (NIPS 13).

Nie, D., Zhang, H., Adeli, E., Liu, L., & Shen, D. (2016). 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. *Medical Image Computing and Computer-Assisted Intervention, 9901*, 212–220.

Nielsen, B., Albregtsen, F., & Danielsen, H. E. (2008). Statistical nuclear texture analysis in cancer research: A review of methods and applications. *Critical Reviews in Oncogenesis, 14*(2–3), 89–164.

Nishio, M., Sugiyama, O., Yakami, M., Ueno, S., Kubo, T., Kuroda, T., & Togashi, K. (2018). Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. *PLoS One, 13*(7), e0200721.

Niu, S., de Sisternes, L., Chen, Q., Leng, T., & Rubin, D. L. (2016). Automated geographic atrophy segmentation for SD-OCT images using region-based C-V model via local similarity factor. *Biomedical Optics Express, 7*(2), 581–600.

Nowak, E., Jurie, F., & Triggs, B. (2006). Sampling strategies for bag-of-features image classification. *Computer Vision – Eccv 2006, Pt 4, Proceedings, 3954*, 490–503.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(7), 971–987.

Organization for Human Brain Mapping (2001). *Annual Conference on Human Brain Mapping*. Brighton, United Kingdom.

Paddock, S. W. (1994). To boldly glow. Applications of laser scanning confocal microscopy in developmental biology. *BioEssays, 16*(5), 357–365.

Panwar, N., Huang, P., Lee, J., Keane, P. A., Chuan, T. S., Richhariya, A., Teoh, S., Lim, T. H., & Agrawal, R. (2016). Fundus photography in the 21st century–A review of recent technological advances and their implications for worldwide healthcare. *Telemedicine Journal and E-Health, 22*(3), 198–208.

Pawlus, A., Sokolowska-Dabek, D., Szymanska, K., Inglot, M. S., & Zaleska-Dorobisz, U. (2015). Ultrasound Elastography–Review of techniques and its clinical applications in pediatrics–Part 1. *Advances in Clinical and Experimental Medicine, 24*(3), 537–543.

Pelizzari, C. A. (1998). Image processing in stereotactic planning: Volume visualization and image registration. *Medical Dosimetry, 23*(3), 137–145.

Perkins, G., Renken, C., Martone, M. E., Young, S. J., Ellisman, M., & Frey, T. (1997). Electron tomography of neuronal mitochondria: Three-dimensional structure and organization of cristae and menbrane contacts. *Journal of Structural Biology, 119*(3), 260–272.

Permuth, J. B., Choi, J., Balarunathan, Y., Kim, J., Chen, D. T., Chen, L., Orcutt, S., Doepker, M. P., Gage, K., Zhang, G., Latifi, K., Hoffe, S., Jiang, K., Coppola, D., Centeno, B. A., Magliocco, A., Li, Q., Trevino, J., Merchant, N., Gillies, R., Malafa, M., & Florida Pancreas, C. (2016). Combining radiomic features with a miRNA classifier may improve prediction of malignant pathology for pancreatic intraductal papillary mucinous neoplasms. *Oncotarget, 7*(52), 85785–85797.

Pham, D. L., Xu, C. Y., & Prince, J. L. (2000). Current methods in medical image segmentation. *Annual Review of Biomedical Engineering, 2*, 315.

Pieper, S., Halle, M., & Kikinis, R. (2004). 3D SLICER. *IEEE International Symposium on Biomedical Imaging ISBI, 2004*, 632–635.

Plevritis, S., Gevaert, O., Xu, J., Hoang, C., Leung, A., Xu, Y., Quon, A., Rubin, D. L., & Napel, S. (2011). Rapid Identification of Prognostic Imaging Biomarkers for Non-small Cell Lung Carcinoma (NSCLC) by integrating image features and gene expression and leveraging public gene expression databases. *Ninety-seventy annual scientific meeting of the RSNA*, Chicago, IL.

Pouratian, N., Sheth, S. A., Martin, N. A., & Toga, A. W. (2003). Shedding light on brain mapping: Advances in human optical imaging. *Trends in Neurosciences, 26*(5), 277–282.

Prastawa, M., Gilmore, J., Lin, W. L., & Gerig, G. (2004). Automatic segmentation of neonatal brain MRI. *Medical Image Computing and Computer-Assisted Intervention – Miccai 2004, Pt 1, Proceedings, 3216*, 10–17.

Prothero, J. S., & Prothero, J. W. (1986). Three-dimensional reconstruction from serial sections IV. The reassembly problem. *Computers and Biomedical Research, 19*(4), 3610373.

Pujara, A. C., Kim, E., Axelrod, D., & Melsaether, A. N. (2018). PET/MRI in breast cancer. *Journal of Magnetic Resonance Imaging*.

Pysz, M. A., Gambhir, S. S., & Willmann, J. K. (2010). Molecular imaging: Current status and emerging strategies. *Clinical Radiology, 65*(7), 500–516.

Qiu, G. (2002). Indexing chromatic and achromatic patterns for content-based colour image retrieval. *Pattern Recognition, 35*(8), 1675–1686.

Rahmani, R., Goldman, S. A., Zhang, H., Cholleti, S. R., & Fritts, J. E. (2008). Localized content-based image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(11), 1902–1912.

Rajpurkar, P., J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren and A. Y. Ng (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. CoRR abs/1711.05225.

Ray, P. (2011). Multimodality molecular imaging of disease progression in living subjects. *Journal of Biosciences, 36*(3), 499–504.

Ray, P., & Gambhir, S. S. (2007). Noninvasive imaging of molecular events with bioluminescent reporter genes in living subjects. *Methods in Molecular Biology, 411*, 131–144.

Rector, A. L., Nowlan, W. A., & Glowinski, A. (1993). Goals for concept representation in the GALEN project. *Proceedings of the 17th Annual Symposium on Computer Applications in Medical Care (SCAMC 93). C. Safran*. New York: McGraw Hill, pp. 414–418.

Ribaric, S., Todorovski, L., Dimec, J., & Lunder, T. (2001). Presentation of dermatological images on the internet. *Computer Methods and Programs in Biomedicine, 65*(2), 111–121.

Ritchie, C. J., Edwards, W. S., Cyr, D. R., & Kim, Y. (1996). Three-dimensional ultrasonic angiography using power-mode Doppler. *Ultrasound in Medicine and Biology, 22*(3), 277–286.

Robinson, P. J. (1997). Radiology's Achilles' heel: Error and variation in the interpretation of the Rontgen image. *The British Journal of Radiology, 70*(839), 1085–1098.

Rohlfing, T., & Maurer, C. R., Jr. (2003). Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Transactions on Information Technology in Biomedicine, 7*(1), 16–25.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional networks for biomedical image segmentation*. Springer International Publishing.

Rosen, G. D., Williams, A. G., Capra, J. A., Connolly, M. T., Cruz, B., Lu, L., Airey, D. C., Kulkarni, K., & Williams, R. W. (2000). The Mouse Brain Library @ www.mbl.org. Int. Mouse Genome Conference 14, p. 166.

Ross, B., & Bluml, S. (2001). Magnetic resonance spectroscopy of the human brain. *Anatomical Record (New Anat.), 265*(2), 54–84.

Rosse, C. (2000). Terminologia Anatomica; considered from the perspective of next-generation knowledge sources. *Clinical Anatomy, 14*, 120–133.

Rosse, C., & Mejino, J. L. V. (2003). A reference ontology for bioinformatics: The foundational model of anatomy. *Journal of Bioinformatics, 36*(6), 478–500.

Rosse, C., Mejino, J. L., Modayur, B. R., Jakobovits, R. M., Hinshaw, K. P., & Brinkley, J. F. (1998a). Motivation and organizational principles for anatomical knowledge representation: The digital anatomist symbolic knowledge base. *Journal of the American Medical Informatics Association, 5*(1), 17–40.

Rosse, C., Shapiro, L. G., & Brinkley, J. F. (1998b). The Digital Anatomist foundational model: Principles for defining and structuring its concept domain. *Proceedings, American Medical Informatics Association Fall Symposium*. Orlando, Florida, pp. 820–824.

Rubin, D. L. (2008). Creating and curating a terminology for radiology: Ontology modeling and analysis. *Journal of Digital Imaging, 21*(4), 355–362.

Rubin, D. L. (2011). Measuring and improving quality in radiology: Meeting the challenge with informatics. *Radiographics, 31*(6), 1511–1527.

Rubin, D. L., & Napel, S. (2010). Imaging informatics: Toward capturing and processing semantic information in radiology images. *Yearbook of Medical Informatics*, 34–42.

Rubin, D. L., & Snyder, A. (2011). *ePAD: A cross-platform semantic image annotation tool ninety-seventh annual scientific meeting of the RSNA*. Chicago, IL.

Rubin, D. L., Bashir, Y., Grossman, D., Dev, P., & Musen, M. A. (2004). Linking ontologies with three-dimensional models of anatomy to predict the effects of penetrating injuries. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 5*, 3128–3131.

Rubin, D. L., Bashir, Y., Grossman, D., Dev, P., & Musen, M. A. (2005). Using an ontology of human anatomy to inform reasoning with geometric models. *Studies in Health Technology and Informatics, 111*, 429–435.

Rubin, D. L., Dameron, O., Bashir, Y., Grossman, D., Dev, P., & Musen, M. A. (2006a). Using ontologies linked with geometric models to reason about penetrating injuries. *Artificial Intelligence in Medicine, 37*(3), 167–176.

Rubin, D. L., Grossman, D., Neal, M., Cook, D. L., Bassingthwaighte, J. B., & Musen, M. A. (2006b). Ontology-based representation of simulation models of physiology. *AMIA Annu Symp Proc*, pp. 664–668.

Rubin, D. L., Mongkolwat, P., Kleper, V., Supekar, K., & Channin, D. S. (2008a). Medical imaging on the semantic web: Annotation and image markup. *2008 AAAI Spring Symposium Series, Semantic Scientific Knowledge Integration*. Stanford University.

Rubin, D. L., Rodriguez, C., Shah, P., & Beaulieu, C. (2008b). iPad: Semantic annotation and markup of radiological images. *American Medical Informatics Association Annual Symposium Proceedings*, 626–630.

Rubin, D. L., Mongkolwat, P., & Channin, D. S. (2009a). A semantic image annotation model to enable integrative translational research. *Summit on Translat Bioinforma, 2009*, 106–110.

Rubin, D. L., Talos, I. F., Halle, M., Musen, M. A., & Kikinis, R. (2009b). Computational neuroanatomy: Ontology-based representation of neural components and connectivity. *BMC Bioinformatics, 10*(Suppl 2), S3.

Rubin, D. L., Korenblum, D., Yeluri, V., Frederick, P., & Herfkens, R. J. (2010). Semantic annotation and image markup in a commercial PACS workstation. *Scientific Paper, Ninety-sixth annual scientific meeting of the RSNA*. Chicago, IL.

Rubin, D. L., Flanders, A., Kim, W., Siddiqui, K. M., & Kahn, C. E., Jr. (2011). Ontology-assisted analysis of web queries to determine the knowledge radiologists seek. *Journal of Digital Imaging, 24*(1), 160–164.

Ruiz, M. E. (2006). Combining image features, case descriptions and UMLS concepts to improve retrieval of medical images. *American Medical Informatics Association Annual Symposium Proceedings*, 674–678.

Sandor, S., & Leahy, R. (1997). Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Transactions on Medical Imaging, 16*(1), 41–54.

Schaltenbrand, G., & Warren, W. (1977). *Atlas for Stereotaxy of the human Brain*. Stuttgart: Thieme.

Scheckenbach, K., Colter, L., & Wagenmann, M. (2017). Radiomics in head and neck cancer: Extracting valuable information from data beyond recognition. *ORL: Journal for Otorhinolaryngology and Its Related Specialties, 79*(1–2), 65–71.

Schimel, A. M., Fisher, Y. L., & Flynn, H. W., Jr. (2011). Optical coherence tomography in the diagnosis and management of diabetic macular edema: Time-domain versus spectral-domain. *Ophthalmic Surgery, Lasers & Imaging, 42*(4), S41–S55.

Schultz, E. B., Price, C., & Brown, P. J. B. (1997). Symbolic anatomic knowledge representation in the read codes version 3: Structure and application. *Journal of the American Medical Informatics Association, 4*, 38–48.

Schwartz, L. H., Panicek, D. M., Berk, A. R., Li, Y., & Hricak, H. (2011). Improving communication of diagnostic radiology findings through structured reporting. *Radiology, 260*(1), 174–181.

Seidenari, S., Pellacani, G., & Grana, C. (2003). Computer description of colours in dermoscopic melanocytic lesion images reproducing clinical assessment. *The British Journal of Dermatology, 149*(3), 523–529.

Sensor Systems Inc. (2001). "MedEx." From http://medx.sensor.com/products/medx/index.html

Shao, L., Zhang, H., & de Haan, G. (2008). An overview and performance evaluation of classification-based least squares trained filters. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society, 17*(10), 1772–1782.

Shapiro, L. G., & Stockman, G. C. (2001). *Computer vision*. Prentice Hall: Upper Saddle River.

Shattuck, D. W., & Leahy, R. M. (2001). Automated graph-based analysis and correction of cortical volume topology. *IEEE Transactions on Medical Imaging, 20*(11), 1167–1177.

Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering, 19*, 221–248.

Shi, J. B., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(8), 888–905.

Shi, B., Grimm, L. J., Mazurowski, M. A., Baker, J. A., Marks, J. R., King, L. M., Maley, C. C., Hwang, E. S., & Lo, J. Y. (2018). Prediction of occult invasive disease in ductal carcinoma in situ using deep learning features. *Journal of the American College of Radiology, 15*(3 Pt B), 527–534.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures,

dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging, 35*(5), 1285–1298.
Simpson, S., Kay, F. U., Abbara, S., Bhalla, S., Chung, J. H., Chung, M., Henry, T. S., Kanne, J. P., Kligerman, S., Ko, J. P., & Litt, H. (2020). Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. *Journal of Thoracic Imaging*.
Singh, A., Massoud, T. F., Deroose, C., & Gambhir, S. S. (2008). Molecular imaging of reporter gene expression in prostate cancer: An overview. *Seminars in Nuclear Medicine, 38*(1), 9–19.
Sivic, J., & Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. *Proceedings of the International Conference on Computer Vision., 2*, 1470–1477.
Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349–1380.
Smith, M. K., Welty, C., & McGuinness, D. (2004). OWL Web Ontology Language Guide. http://www.w3.org/TR/owl-guide/
Smith, M. Q., Staley, C. A., Kooby, D. A., Styblo, T., Wood, W. C., & Yang, L. (2009). Multiplexed fluorescence imaging of tumor biomarkers in gene expression and protein levels for personalized and predictive medicine. *Current Molecular Medicine, 9*(8), 1017–1023.
Sohrab, M. A., Smith, R. T., & Fawzi, A. A. (2011). Imaging characteristics of dry age-related macular degeneration. *Seminars in Ophthalmology, 26*(3), 156–166.
Solomon, M., Liu, Y., Berezin, M. Y., & Achilefu, S. (2011). Optical imaging in cancer research: Basic principles, tumor detection, and therapeutic monitoring. *Medical Principles and Practice, 20*(5), 397–415.
Soto, G. E., Young, S. J., MArtone, M. E., Deerinick, T. J., Lamont, S. L., Carragher, B. O., Hamma, K., & Ellisman, M. H. (1994). Serial section electron tomography: A method for three-dimensional reconstruction of large structures. *NeuroImage, 1*, 230–243.
Spackman, K. A., Campbell, K. E., & Cote, R. A. (1997). SNOMED-RT: A reference terminology for health care. *Proceedings, AMIA Annual Fall Symposium. D. R. Masys. Philadelphia, Hanley and Belfus*, pp. 640–644.
Sperrin, M., Grant, S. W., & Peek, N. (2020). Prediction models for diagnosis and prognosis in Covid-19. *BMJ, 369*, m1464.
Spitzer, V. M., & Whitlock, D. G. (1998). The visible human dataset: The anatomical platform for human simulation. *The Anatomical Record, 253*(2), 49–57.
Stensaas, S. S., & Millhouse, O. E.. (2001). Atlases of the Brain. From http://medstat.med.utah.edu/kw/brain_atlas/credits.htm
Subramaniam, B., Hennessey, J. G., Rubin, M. A., Beach, L. S., & Reiss, A. L. (1997). Software and methods for quantitative imaging in neuroscience: the Kennedy Krieger Institute Human Brain Project. In S. H. Koslow & M. F. Huerta (Eds.), *Neuroinformatics: an overview of the Human Brain Project* (pp. 335–360). Mahwah: Lawrence Erlbaum.
Sundsten, J. W., Conley, D. M., Ratiu, P., Mulligan, K. A., & Rosse, C.. (2000). Digital Anatomist web-based interactive atlases. From http://www9.biostr.washington.edu/da.html
Swanson, L. W. (1999). *Brain maps: Structure of the rat brain*. New York: Elsevier Science.
Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical Publishers.
Talos, I. F., Rubin, D. L., Halle, M., Musen, M., & Kikinis, R. (2008). A prototype symbolic model of canonical functional neuroanatomy of the motor system. *Journal of Biomedical Informatics, 41*(2), 251–263.
Toga, A. W. (2001). UCLA Laboratory for Neuro Imaging (LONI). From http://www.loni.ucla.edu/
Toga, A. W., Ambach, K. L., & Schluender, S. (1994). High-resolution anatomy from in situ human brain. *NeuroImage, 1*(4), 334–344.
Toga, A. W., Santori, E. M., Hazani, R., & Ambach, K. (1995). A 3-D digital map of rat brain. *Brain Research Bulletin, 38*(1), 77–85.
Toga, A. W., Frackowiak, R. S. J., & Mazziotta, J. C. (Eds.). (2001). *Neuroimage: A journal of brain function*. Academic Press: New York.
Tommasi, T., Caputo, B., Welter, P., Güld, M. O., & Deserno, T. M. (2010). Overview of the CLEF 2009 medical image annotation track. *Proceedings of the 10th international conference on Cross-language evaluation forum: multimedia experiments*. Corfu, Greece, Springer-Verlag, pp. 85–93.
Toomre, D., & Bewersdorf, J. (2010). A new wave of cellular imaging. *Annual Review of Cell and Developmental Biology, 26*, 285–314.
Trebeschi, S., Griethuysen, J. J. M. V., Lambregts, D. M. J., Lahaye, M. J., Parmar, C., Bakers, F. C. H., Peters, N. H. G. M., Beets-Tan, R. G. H., & Aerts, H. J. W. L. (2017). Deep learning for fully-automated localization and segmentation of rectal Cancer on multiparametric MR. *Scientific Reports, 7*(1), 5301.
Tsarkov, D., & Horrocks, I. (2006). FaCT++ description logic reasoner: System description. *Automated Reasoning, Proceedings, 4130*, 292–297.
Valdora, F., Houssami, N., Rossi, F., Calabrese, M., & Tagliafico, A. S. (2018). Rapid review: Radiomics and breast cancer. *Breast Cancer Research and Treatment, 169*(2), 217–229.
Van Essen, D. C., & Drury, H. A. (1997). Structural and functional analysis of human cerebral cortex using a surface-basec atlas. *The Journal of Neuroscience, 17*(18), 7079–7102.
Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., & Anderson, C. H. (2001). An integrated software suite for surface-based analysis of

cerebral cortex. *Journal of the American Medical Association, 8*(5), 443–459.

Van Leemput, K., Maes, F., Vandermeulen, D., & Suetens, P. (1999). Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging, 18*(10), 897–908.

Van Noorden, S. (2002). Advances in immunocytochemistry. *Folia Histochemica et Cytobiologica, 40*(2), 121–124.

van Timmeren, J. E., Leijenaar, R. T. H., van Elmpt, W., Reymen, B., Oberije, C., Monshouwer, R., Bussink, J., Brink, C., Hansen, O., & Lambin, P. (2017). Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. *Radiotherapy and Oncology, 123*(3), 363–369.

Vapnik, V. N. (2000). *The nature of statistical learning theory*. New York: Springer.

Varma, M., & Zisserman, A. (2003). Texture classification: are filter banks necessary?, in. *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 91–698.

Venkatesh, S. K., & Ehman, R. L. (2015). Magnetic resonance elastography of abdomen. *Abdominal Imaging, 40*(4), 745–759.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked Denoising autoencoders: Learning useful representations in a deep network with a local Denoising criterion. *Journal of Machine Learning Research, 11*(Dec), 3371–3408.

Vreeman, D. J., & McDonald, C. J. (2005). Automated mapping of local radiology terms to LOINC. *AMIA Annu Symp Proc*, pp. 769–773.

Vreeman, D. J., Abhyankar, S., Wang, K. C., Carr, C., Collins, B., Rubin, D. L., & Langlotz, C. P. (2018). The LOINC RSNA radiology playbook – A unified terminology for radiology procedures. *Journal of the American Medical Informatics Association, 25*(7), 885–893.

Wang, L., & Wong, A.. (2020). COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images.

Wang, J. Z., Wiederhold, G., Firschein, O., & Wei, S. X. (1997). Content-based image indexing and searching using Daubechies' wavelets. *International Journal on Digital Libraries, 1*(4), 311–328.

Wang, D., A. Khosla, R. Gargeya, H. Irshad and A. H. Beck (2016). Deep learning for identifying metastatic breast cancer. arXiv:1606.05718 [cs, q-bio].

Wang, K. C., Patel, J. B., Vyas, B., Toland, M., Collins, B., Vreeman, D. J., Abhyankar, S., Siegel, E. L., Rubin, D. L., & Langlotz, C. P. (2017). Use of radiology procedure codes in health care: The need for standardization and structure. *Radiographics, 37*(4), 1099–1110.

Weaver, O., & Leung, J. W. T. (2018). Biomarkers and imaging of breast cancer. *AJR. American Journal of Roentgenology, 210*(2), 271–278.

Weiss, D. L., & Langlotz, C. P. (2008). Structured reporting: Patient care enhancement or productivity nightmare? *Radiology, 249*(3), 739–747.

Weissleder, R., & Mahmood, U. (2001). Molecular imaging. *Radiology, 219*, 316–333.

Weissleder, R., Schwaiger, M. C., Gambhir, S. S., & Hricak, H. (2016). Imaging approaches to optimize molecular therapies. *Science Translational Medicine, 8*(355), 355ps316.

Wessels, J. T., Yamauchi, K., Hoffman, R. M., & Wouters, F. S. (2010). Advances in cellular, subcellular, and nanoscale imaging in vitro and in vivo. *Cytometry. Part A, 77*(7), 667–676.

Willmann, J. K., van Bruggen, N., Dinkelborg, L. M., & Gambhir, S. S. (2008). Molecular imaging in drug development. *Nature Reviews. Drug Discovery, 7*(7), 591–607.

Wilson, T. (1990). *Confocal Microscopy*. San Diego: Academic Press Ltd.

Wong, B. A., Rosse, C., & Brinkley, J. F. (1999). Semiautomatic scene generation using the Digital Anatomist Foundational Model. *Proceedings, American Medical Informatics Association Fall Symposium*. Washington, D.C., pp. 637–641.

Woods, R. P., Cherry, S. R., & Mazziotta, J. C. (1992). Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography, 16*, 620–633.

Woods, R. P., Mazziotta, J. C., & Cherry, S. R. (1993). MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography, 17*, 536–546.

World Health Organization. (2020.) WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020. Retrieved 5/15/2020, from https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19%2D%2D-11-march-2020

WorldWideWeb Consortium. (W3C Recommendation 10 Feb 2004). OWLWeb Ontology Language Reference.

Wu, S., Zheng, J., Li, Y., Yu, H., Shi, S., Xie, W., Liu, H., Su, Y., Huang, J., & Lin, T. (2017). A Radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder Cancer. *Clinical Cancer Research, 23*(22), 6904–6911.

Wynants, L., Van Calster, B., Bonten, M. M. J., Collins, G. S., Debray, T. P. A., De Vos, M., Haller, M. C., Heinze, G., Moons, K. G. M., Riley, R. D., Schuit, E., Smits, L. J. M., Snell, K. I. E., Steyerberg, E. W., Wallisch, C., & van Smeden, M. (2020). Prediction models for diagnosis and prognosis of covid-19 infection: Systematic review and critical appraisal. *BMJ, 369*, m1328.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging, 9*(4), 611–629.

Yang, L., Jin, R., Mummert, L., Sukthankar, R., Goode, A., Zheng, B., Hoi, S. C. H., & Satyanarayanan, M. (2010). A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(1), 30–44.

Yang, Z., Shi, J., He, Z., Lu, Y., Xu, Q., Ye, C., Chen, S., Tang, B., Yin, K., Lu, Y., & Chen, X. (2020). Predictors for imaging progression on chest CT from coronavirus disease 2019 (COVID-19) patients. *Aging (Albany NY), 12*(7), 6037–6048.

Yasaka, K., Akai, H., Kunimatsu, A., Kiryu, S., & Abe, O. (2018). Deep learning with convolutional neural network in radiology. *Japanese Journal of Radiology, 36*(4), 257–272.

Yoo, T. S. (2004). *Insight into images : Principles and practice for segmentation, registration, and image analysis*. Wellesley: A K Peters.

Yu, F., & Ip, H. H. (2008). Semantic content analysis and annotation of histological images. *Computers in Biology and Medicine, 38*(6), 635–649.

Yuan, M., Yin, W., Tao, Z., Tan, W., & Hu, Y. (2020). Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. *PLoS One, 15*(3), e0230548.

Zaleska-Dorobisz, U., Pawlus, A., Szymanska, K., Lasecki, M., & Ziajkiewicz, M. (2015). Ultrasound Elastography–Review of techniques and its clinical applications in pediatrics–Part 2. *Advances in Clinical and Experimental Medicine, 24*(4), 725–730.

Zalis, M. E., Barish, M. A., Choi, J. R., Dachman, A. H., Fenlon, H. M., Ferrucci, J. T., Glick, S. N., Laghi, A., Macari, M., McFarland, E. G., Morrin, M. M., Pickhardt, P. J., Soto, J., & Yee, J. (2005). CT colonography reporting and data system: A consensus proposal. *Radiology, 236*(1), 3–9.

Zhang, Y. Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging, 20*(1), 45–57.

Zhang, H., Ingham, E. S., Gagnon, M. K., Mahakian, L. M., Liu, J., Foiret, J. L., Willmann, J. K., & Ferrara, K. W. (2017). In vitro characterization and in vivo ultrasound molecular imaging of nucleolin-targeted microbubbles. *Biomaterials, 118*, 63–73.

Zhenyu, H., Yanjie, Z., Tonghui, L., & Jianguo, Z. (2009). Combining text retrieval and content-based image retrieval for searching a large-scale medical image database in an integrated RIS/PACS environment, SPIE.

Zhou, S. K., Greenspan, H., & Shen, D. (2017a). *Deep learning for medical image analysis*. London/San Diego: Elsevier/Academic Press.

Zhou, Y., He, L., Huang, Y., Chen, S., Wu, P., Ye, W., Liu, Z., & Liang, C. (2017b). CT-based radiomics signature: A potential biomarker for preoperative prediction of early recurrence in hepatocellular carcinoma. *Abdominal Radiology (NY), 42*(6), 1695–1704.

Zhou, M., Scott, J., Chaudhury, B., Hall, L., Goldgof, D., Yeom, K. W., Iv, M., Ou, Y., Kalpathy-Cramer, J., Napel, S., Gillies, R., Gevaert, O., & Gatenby, R. (2018). Radiomics in brain tumor: Image assessment, quantitative feature descriptors, and machine-learning approaches. *AJNR. American Journal of Neuroradiology, 39*(2), 208–216.

Zijdenbos, A. P., Evans, A. C., Riahi, F., Sled, J., Chui, J., & Kollokian, V. (1996). Automatic quantification of multiple sclerosis lesion volume using stereotactic space. *Proc. 4th Int. Conf. on Visualization in Biomedical Computing*. Hamburg. pp. 439–448.

Zimmerman, S. L., Kim, W., & Boonn, W. W. (2011). Informatics in radiology: Automated structured reporting of imaging findings using the AIM standard and XML. *Radiographics, 31*(3), 881–887.

Personal Health Informatics

Robert M. Cronin, Holly Jimison, and Kevin B. Johnson

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_11

Learning Objectives

After reading this chapter, you should know the answers to these questions:

1. What is the role of the patient or consumer in healthcare decisions?
2. How does patient empowerment come into play in the various care delivery settings and phases of health care?
3. What are some examples of sensors that can be used to assist in personal health management?
4. How can you ensure patient privacy and the security of patient generated data in the home and environment?
5. What are the various features of personal health technologies (e.g., personal health records, mobile applications, etc.)?
6. How do individuals obtain various types of health information?

11.1 Introduction

Complexity and collaboration characterize health care in the early twenty-first century. Complexity arises from our deeper and more sophisticated understanding of health and disease, including the addition of molecular/genomic processes and social/behavioral determinants. Complexity also arises from the myriad of new treatments available for many diseases, and emerging data about the role of nutrition, exercise, sleep, and stress in preserving health. Collaboration begins with the realization that successful attainment of optimal wellbeing and effective management of disease processes necessitate active engagement of clinicians, laypersons, support systems, and society as a whole. Collaboration extends beyond the societally-focused opportunities into the healthcare system itself, where care is more fragmented, leading to greater needs for collaboration and communication.

Now more than ever before, the healthcare system recognizes the role of the person who interacts with this system, who is increasingly interested in engaging or called upon to engage through various states of health and disease. This recognition has given rise to a movement that underpins the birth of personal health informatics, as described below.

11.2 Patient-Centered Care and Personal Health Informatics

Patient-centered care has become a core component of medical care. From early work in the 1960s and 70s (Balint 1969; Waitzkin and Stoeckle 1972), to the concept of the chronic disease model (Bodenheimer et al., 2002c; Coleman et al. 2009), the National Academy of Medicine landmark report *Crossing the Quality Chasm* (Institute of Medicine (US) Committee on Quality of Health Care in America 2001) making "patient-centered" one of the six aims of health care, and the development and incorporation of the medical home (Kellerman and Kirk 2007) and patient engagement (Dentzer 2013),[1] patient-centered care has taken center stage in medicine.

As a result of this visibility, healthcare institutions, health planners, congressional representatives, and hospital public relations departments are among many promoters of patient-centered care, a concept rooted in "deep respect for patients as unique living beings, and the obligation to care for them on their terms"(Epstein and Street Jr 2011). To be patient-centered, one must accept people seeking care as persons with a unique social world, who should be listened to, informed, respected, and involved in their own care—and whose wishes are heard, if not acted upon, during their healthcare journey. Patient-centered care complements evidence-based medicine by including patient preferences into the decision making about treatment options. Berwick, described three maxims of patient-

1 Health CfA. A New Definition of Patient Engagement. What is Engagement and Why is it Important? 2010. Available from: ► http://www.cfah.org/file/CFAH_Engagement_Behavior_Framework_current.pdf

centered care: (1) "The needs of the patient come first."(2) "Nothing about me without me." (3) "Every patient is the only patient" (Berwick 2009).

Patient-centered care represents a shift in the physician's role from paternalistic and authoritative to collaborative--leveraging the perspectives of people and their support system, whether the support system consists of family, caregivers, or even technology, as partners in making decisions and delivering care. Patient-centered care requires the entire healthcare team to be more mindful, informative, and empathic, and for patients to actively participate in their care. Patient-centered care encourages inclusiveness and engagement for shared decision making among the stakeholders to develop a comprehensive care plan aligned with the whole person.

The maxims of patient-centered care form the basis of the field personal health informatics, as originally proposed by Warner Slack and Tom Ferguson in 1993 (Demiris 2016). In particular, these maxims translate into the following desiderata:

- People are able to access care that is coordinated and collaborative
- Care is focused on the whole person, not just the physical comfort
- Care considers people's values, culture, and socioeconomic status
- People and their support system are active partners in care, not passive listeners
- People's goals within the healthcare system align with the system's mission, values, and quality metrics
- People and their caregivers participate in shared decision making with their providers and play a role in the decisions at the personal, population, and system level
- Sharing health information with people and caregivers enables informed decision making
- Support systems' presence in the care setting are encouraged and facilitated

As described in ▶ Sect. 13.5, these desiderata provide a new lens through which the whole of biomedical informatics should be viewed.

11.2.1 Using Biomedical Informatics to Impact Patient-Centered Medicine

To conclude this section, we provide a few illustrative examples of how patient-centered care and personal health informatics can help shape the care delivered.

1. Patient-centered ambulatory care.

   Routine ambulatory care, by its very nature, focuses on the whole person, and not just their diagnosis. Caring for the whole person requires the ability to utilize resources such as social workers, financial counselors, mental health providers, transportation, peer support programs, daily living assistance, and language and literacy education and resources. Making the provider aware of the needs of the person being cared for can be enabled through electronic health records (EHRs) and clinical decision support that work within the provider workflow, potentially utilizing data provided by the patient or the patient's social and family network. In addition, using tools like a patient portal, the healthcare system can provide alerts and reminders to patients for care such as the Influenza vaccine, nutrition counseling, and medication refill reminders.

   Access to care can be facilitated through telemedicine and telehealth (see ▶ Chap. 20), as well as through apps that could enable daily living assistance and peer support programs (see ▶ Chap. 19). All of these applications need to consider language and literacy (health and technology), as well as challenges created by asynchronous healthcare-related discussions.

   Example: Jane uses her mobile app to remind her when to fill her asthma medications and communicate with her social worker and financial counselor to help her purchase those medications. She uses the same app to talk with her support system and other people with asthma who can understand and support her journey through the stages of her disease.

2. Patient-centered acute care and care transitions.

   Acute care settings are characterized by sudden and rapid changes in patient status, *ad hoc* appointments, and frequent handoffs of care. To help patients understand what is happening to them and to facilitate decision-making, patients often request unrestricted and continuous access to their social support network in this setting. Patients and their support system should be present during rounds, which are performed at the bedside, and at shift changes. Family interaction should take place in an environment that is as comfortable as possible, equipped with access to information and to experts as necessary. Information technology can help make this scenario possible through inpatient personal health records, where patients and their families can view information in real time to help participate in rounds and make informed decisions (Huerta et al. 2017; O'Leary et al. 2016; Prey et al. 2014). Mobile devices can improve communication, care plan management, and knowledge transfer, even when family members are not present. Other personal informatics tools, such as wearable sensors or smart scales, could be introduced during a hospital stay to develop behavior changes that could carry over into the home environment (Steinberg et al. 2013).

   Example: Through her mobile phone, Beverly's family was present remotely in the hospital during rounds to ask the health care team questions. Her family also encouraged Beverly to learn to use the hospital's smart scale, which can monitor her weight and keep her heart failure under control. Beverly now uses her smart scale daily at home to send updates to her family and healthcare providers who can encourage her and help her stay out of the hospital.

3. Patient-centered care at home.

   A significant amount of health care can and should occur in the home setting. This setting is where social, financial, supporting, and behavioral factors can affect a person's health care and management. Applications like patient portals or personal health records can enable a person to review lab results and clinical notes, communicate with their healthcare providers, schedule appointments, pay bills, and obtain educational information. Wearable technologies enable self-management and monitoring of critical information, such as weight, blood pressure, glucose levels, and medication adherence (Marcolino et al. 2018). Finally, informatics applications could deliver information that can help augment knowledge about disease status that can enable people to make informed decisions about the need to escalate home care or schedule a return visit (Asnani et al. 2016).

   Example: Richard knows a lot about his heart failure, but when given the right personal health informatics tools, he learned about signs and symptoms when his heart failure gets worse. He avoids costly readmissions to the hospital by remembering to take his medications daily thanks to reminders from his wearable technologies, and now knows when to communicate with his providers through his patient portal when his heart failure worsens.

4. Personalized medicine.

   At its core, the maxims of patient-centered care require that any management plan provided by healthcare providers to patients will need to be personalized (see ► Chap. 28). Medications, procedures, supportive or curative plans all should be tailored to the person/family receiving them. However, the augmentation of our knowledge about the impact of a person's omics (genomics, proteomics, metabolomics, etc.), and environment can now be used to personalize therapy (Collins 2004).

   Example: Using sensors on James' phone along with his genetic information, his providers can use targeted medications to treat his cancer and can help him meet his treatment goal of golfing again.

11.2.2 Limitations of Patient-Centered Care

As with any change in the locus of control, the transition from authoritative to collaborative decision-making that is the hallmark of patient-centered care raises some concerns. For example, there are concerns that patient-centered medicine conflicts with evidence-based medicine (Weaver 2015). One of the current challenges in medicine is to bring these two worlds together, which could be accomplished using more sophisticated searching of the literature or mining EHR data to uncover evidence supporting deviations in care (Gallego et al. 2015). Another concern is that physicians are stewards of social resources, but some would argue that physicians do not know the social responsibility of patient-centered medicine (Berwick 2009). A third concern is the juxtaposition between a patient needing and wanting improved access to the healthcare system, and a healthcare system that is already both expensive to run, in part because of fragmentation and attempts to improve access (see ▶ Chap. 29) (Enthoven 2009).

Aside from clinician concerns, there are design constraints on the healthcare system that limit patient-centered medicine. First, the system must support a shift in the locus of control of care decisions towards patients and their caregivers. Developing health informatics tools such as delivering information through mHealth and providing decision aids for people can facilitate informed decision making about care. Second, transparency of care options and their associated costs and outcomes needs to extend to all components of care, including research, and education. While informatics researchers are working on these issues, further work is needed in this area, especially in providing transparent care coordination choices and data liquidity. Third, individualization and customization need to be design targets, creating health care and health informatics systems and tools that can adapt to the individual needs and circumstances of people. Research in personal health informatics demonstrates the importance of personalization and individualization, but significant work remains. Finally, it is critical to train young healthcare professionals in the expectations of their profession related to patient-centered care. Education informatics can help bridge gaps between what young health professionals currently know and practice and what a patient-centered care model could look like with an appropriate curriculum of educational modules, knowledge testing, and practice modeling.

11.3 Historical Perspective of Personal Health Informatics

In essence, what we think of as consumer/patient engagement reflects a shift from the person as the silent recipient of ministrations from a wise, beneficent clinician, to an active collaborator whose values, preferences, and lifestyle not only alter predisposition to certain illnesses but also shape the characteristics of desirable treatments. In this section, we will describe the ancient concept of paternalism, the rise of patient-centered medicine, and how personal health informatics has supported and enabled the ability of the person who used to have health care enacted upon them, to be an engaged and active participant in their health care.

11.3.1 Paternalism and Professionalism of Medicine and Informatics

Paternalism is thought to go back as far as the history of medicine. Hippocrates was the father of medical paternalism as he wrote that physicians should conceal most things from the patient including the patient's present and future condition. He believed that medical knowledge be kept secret from patients. The Hippocratic oath, which is recited today by medical students, is silent about the communication between the doctor and patient relevant to the patient's treatment. Paternalistic medicine continued through medieval times where patients were told to honor doctors since doc-

tors received their authority from God and patients must promise obedience.

Paternalistic medicine is about keeping information primarily in the hands of the physician and medical system and in certain cases giving misinformation to patients to keep accurate information from them. Healthcare decisions are made by the physician and medical system, and patients are expected to abide by these decisions with no or minimal input. In the 16th and 17th centuries, some physicians started to acknowledge that patients might have a voice in their care. However, doctors of eminence, like Dr. Benjamin Rush, wrote that doctors could yield to patients in matters of little consequence, but maintain an inflexible authority over them in matters essential to life.

Paternalism is still present today. Most biomedical publications are inaccessible and costly without a subscription. Getting complete medical records can be difficult, even when through conventional information technology applications. Informed consent, in many cases, is not sufficiently explained to the person having the treatment for the person to understand all the risks, benefits, other options, and details of what is being done. The discharge process is perhaps the best example of present-day paternalism, where patients have little say about their readiness for discharge and often are made responsible complying with complicated discharge instructions. Patients who believe they are ready before their care team agrees, or who are dissatisfied with their care, are required to sign a form attesting to leaving against medical advice. The final source of paternalism comes from guidelines that healthcare providers get to guide their decisions. These guidelines, discharge forms, and other processes typically do not involve patients in their creation.

Although early information systems were almost exclusively provider-centric, recent advances in computer system availability has prompted the development of less paternalistic tools that may be used by people and families, as described below.

11.3.2 The Rise of Patient-Centered Medicine and Personal Health Informatics

Enid Balint coined the term patient-centered care in 1969 (Balint 1969). A few pioneers of patient-centered medicine include Barbara Korsch who explored the listening skills of physicians in training (Korsch 1989), John Ware who discovered the components of patient satisfaction (Ware Jr et al. 1983), Debra Roter and Judith Hall who described the properties and dysfunction of doctor-patient communication and how to improve this communication (Roter and Hall 2006), Howard Waitzkin and John Stoeckle who demonstrated how to tap into patient's views and knowledge of their symptoms to what causes them could lead to improved doctor-patient relationships (Waitzkin and Stoeckle 1972), Michael Barry, Jack Fowler, Al Mulley, Joseph Henderson, and Jack Wennberg, who developed shared decision making theory and technology and associations with improved outcomes (Barry et al. 1995), and Judith Hibbard who helped us understand patient's desires for knowledge (Hibbard et al. 2007) and advanced our knowledge and tools about patient engagement and activation (Hibbard and Greene 2013; Hibbard et al. 2004). Other landmarks in this paradigm shift included Engel's proposal to "take into account the patient" (Engel 1977), Cassell's transcriptions of clinical encounters, which provided a basis to understand the doctor–patient relationship (Cassell 1985), and Kleinman's definitions of "disease" and "illness" as the patient's subjective experience of feeling ill (Kleinman 1988).

Personal Health Informatics can be traced back to the early twentieth century, where the U.S. Federal Children's Bureau served as a major source of health information for the public. Mothers could write to this federal agency, asking questions about normal child development, nutrition, and disease management. Written materials, such as letters and pamphlets served as the primary mechanism for delivering information that supported lay

people in their handling of health challenges. Patient education companies such as Krames would partner with organizations like the American Heart Association to provide general printed material on heart disease or with the American Cancer Society to provide information on cancer.

Personal Health Informatics applications followed a similar trend as patient-centered medicine with early applications in the 1950's and 1960's. Collen and colleagues at Kaiser Permanente created one of the earliest patient data collection applications--a health appraisal system that prompted for patient data and returned a systematic risk appraisal (Collen et al. 1964). Warner Slack and colleagues at the University of Wisconsin used a mainframe computer system as a health assessment tool. Patients sat at a cathode ray tube (CRT) terminal and responded to text questions, receiving a printed summary of their health appraisal at the end of the session (Slack et al. 1966). Figure 11.1 shows an early example of the mainframe-based tool developed by Slack. At Massachusetts General Hospital in the late 1950's, computer-driven telephone systems were used to conduct home-based follow-up with post-surgery cardiac patients, calling them daily to obtain pulse readings. This was followed by an era of interactive video systems that augmented delivery of information and helped patients understand the risks and benefits associated with treatment options, but also to help define their values for possible future health outcomes. The prime examples of this type of system originated with the Foundation for Informed Medical Decision Making. As early as 1973, Wennberg and others discovered that the rates of many expensive surgeries and other treatments would vary from location to location throughout the U.S. (Wennberg and Gittelsohn 1973). This variation seemed to occur for medical conditions where there were multiple viable treatment options and choices depended more on physician and resource availability than need or patient characteristics. Barry et al. focused on developing interactive video consumer decision aids that focused on these conditions (e.g., prostate cancer, breast cancer, back pain, etc.), discovering that patient preferences and priorities for possible health outcomes could vary dramatically from person to person, and could be critical to defining an optimal decision (Barry et al. 1995).

Fig. 11.1 Early use of computing in consumer health informatics, here taking a medical history directly from a patient (Slack WV et al. (1966). *NEJM* with permission)

In the 1980s, clinicians and health educators capitalized on the increasingly common personal computers as vehicles for health education. As shown in Fig. 11.2 The Body Awareness Resource Network (BARN), developed in the 1980s by Gustafson and colleagues at the University of Wisconsin, engaged adolescents in game-like interactions to help them learn about growth and development, develop healthy attitudes towards avoidance of risky behaviors, and rehearse strategies for negotiating the complex interpersonal world of adolescence (Bosworth et al. 1983).

Another influential early personal health informatics application was the Comprehensive Health Enhancement Support System (CHESS) developed by Gustafson and colleagues in 1989 at the University of Wisconsin (Gustafson et al. 2001). CHESS

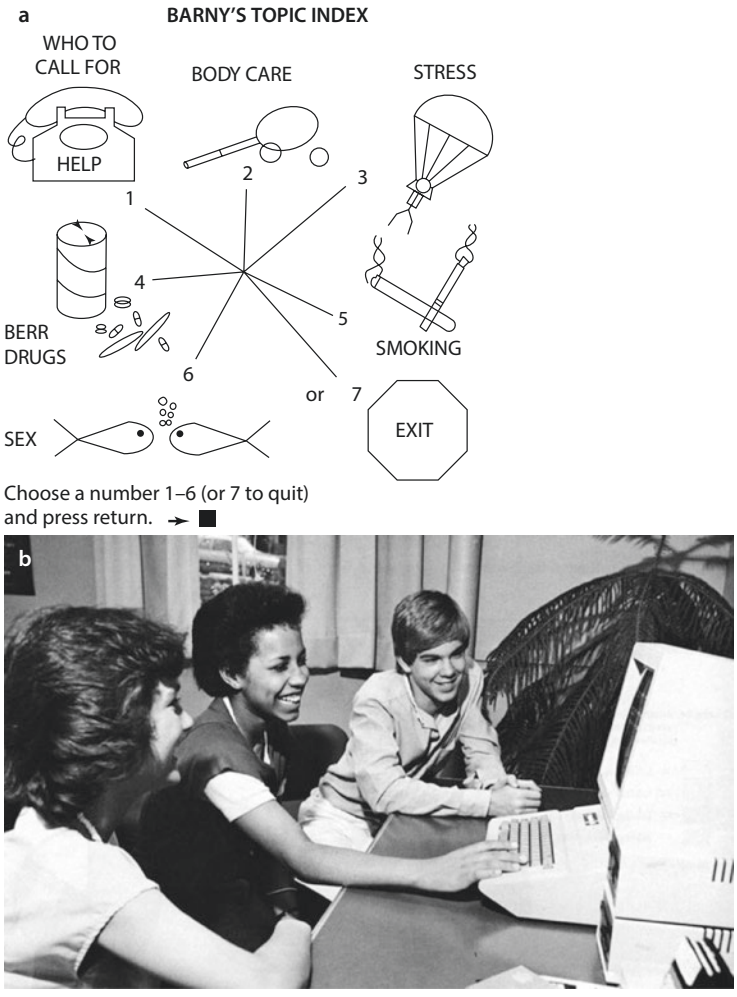


Fig. 11.2 BARN topic index and use by teens (Bosworth et al. 1983). This picture shows teens interacting with a game on an early graphical computer. The figure on the left is the topic index as displayed on the screen. (With permission from Gustafson, D, personal communication)

provided women with breast cancer information through curated articles and directories of cancer services, decision-making through charts, decision aids, and action plans, and emotional support through online support groups. Ferguson was also heavily influential in the 1980's and 90's in creating and analyzing online social support groups for patients (Ferguson 1996).

As the Internet became more available in homes, Internet support groups gathered momentum. One such example, Hopkins Teen Central developed by johnson et al. (2001), allowed otherwise isolated children with cystic fibrosis to meet virtually and to discuss health and developmental issues that impacted healthy decision making. The idea of the Internet support group became an active area of development, continuing through today (Eysenbach et al. 2004). During this time, computer games also increased in uptake, and while time-consuming and expensive to build, they were relatively easy to disseminate, and were associated with measurable changes in knowledge (Lieberman 1988) and, in some cases, symptom management (Patel et al. 2006; Redd et al. 1987).

Some major areas of growth in personal health informatics over the past 20 years include home telehealth, mobile health, personal health records, and personal genomics. The growth of home telehealth technologies has grown rapidly over the past 20 years (see ▶ Chap. 20). Some notable randomized controlled trials include Informatics for Diabetes Education and Telemedicine

(IDEATel), the Telemonitoring Study for Chronic Obstructive Pulmonary Disease (COPD), and the Tele-ERA study. Large randomized controlled trials demonstrated improved effects of these interventions. Personal health records started in the late 1990s with the Patient-Centered Access to Secure Systems Online (PCASSO) portal (Masys and Baker 1997). Tethered personal health records, commonly referred to as patient portals, have been implemented by hundreds of institutions, with increasing adoption being driven by governmental policy such as the Affordable Care Act and Meaningful Use in the US, and the Power of Information strategy in the UK. There has been increasing literature demonstrating increased uptake and use of patient portals, and also improvement in patient satisfaction, communication, and outcomes (Ammenwerth et al. 2012; Goldzweig et al. 2012). With the advent of mobile technologies, such as smart phones and tablets, mobile health (mHealth) has exploded over the past 5 years (see ▶ Chap. 19). While significant work has been done in the U.S. utilizing this technology, a significant push of mHealth has occurred in low- and middle- income countries because of the ubiquitous nature and low cost of mobile phones as compared to other forms of technology. With the increased ability to record and review daily activities through mobile technologies, a movement called the Quantified Self has evolved (Appelboom et al. 2014). The Quantified Self movement, driven by the theory of patient engagement, is a fast growing practice of self-monitoring driven by technological advances and breakthroughs in miniaturization of wearable and environmental sensors. Personal genomics first became available in the early 2000s because of direct to consumer genetic testing. Companies like 23andMe, ▶ Ancestry.com, and Pathway genomics provide the ability to test one's own genetic composition at home to discover genetic risk for diseases like breast cancer, ancestry, and pharmacogenomic information. Issues with regulatory bodies such as the Food and Drug Administration has prevented all information about genetic risk to be provided to customers of these companies for concern about false information (see ▶ Chap. 28).

Patient-centered medicine and Personal Health Informatics have become and will continue to be a bigger and more important driving force in medicine. In 1998 the Institute of Medicine (IOM) established a major program on Quality of Health Care in America and placed patient-centered as one of the six aims in their landmark paper on improving the quality of health care, *Crossing the Quality Chasm* (Institute of Medicine (US) Committee on Quality of Health Care in America 2001). Many studies have demonstrated improvement in care using patient-centered medicine including classic medical outcomes (Epstein and Street Jr 2011), improved shared decision making (Golomb et al. 2007), and reducing unnecessary surgical operations.[2] Personal Health Informatics has also demonstrated important improvements in medical care (Gibbons et al. 2009). In a Medline search of "patient-centered care", tens of thousands of articles have been published about patient centered medicine, with only 59 from 1950 to 1992. A similar search for "consumer health informatics" OR "personal health informatics" demonstrates hundreds of articles, with only 4 articles from 1950 to 1992. In recognition of the growth of scientific studies in this domain, in 2008 the MeSH term "Consumer Health Information" was introduced, defined as "information intended for potential users of medical and health care services" (Demiris 2016). As people increase their engagement in their health, and technologies improve their ability to do so, personal health informatics will become a bigger and more important part of biomedical informatics and the subdisciplines within it.

2 O'Connor A, Stacey, D, Rovner, D, Holmes-Rovner, M, Tetroe, J, Llewellyn-Thomas, H, Entwistle, V., Tait, V, Rostom, A, Fiset, V, Barry, M. Institute for Healthcare Improvement: Patient decision aids for balancing the benefits and harms of health care options: A systematic review and meta-analysis 2018 [cited 2018 June 22]. Available from: ▶ http://www.ihi.org/resources/Pages/Publications/PatientDecisionAidsforBalancingBenefitsHarmsofHealthCareOptions.aspx

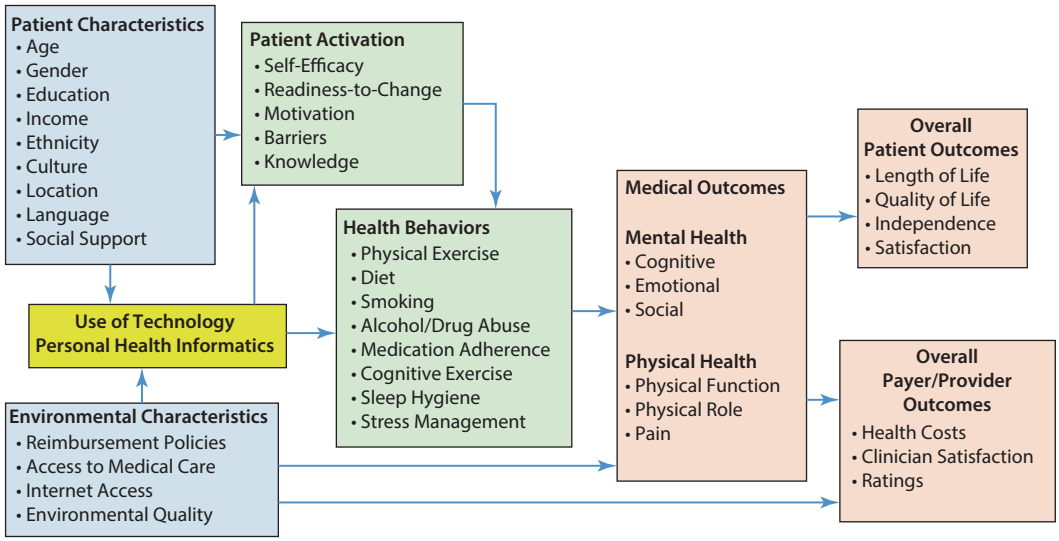


Fig. 11.3 Analytic framework showing the interplay between social, cultural, and behavioral features and the opportunities for personal health informatics on outcomes

11.4 Important Concepts in Personal Health Informatics

In this section we review how social and cultural, economic and financial, education, language and literacy, environmental and behavior factors influence and mediate health outcomes for patients and consumers of health care. Figure 11.3 shows a framework for thinking about how Personal Health Technology can influence outcomes from a variety of stakeholder's points of view within a context of social and cultural factors (Jimison et al. 2008; Keselman et al. 2008).

Interactive consumer health technology applications have had an increasingly important role in health care. Work based on the IOM's Crossing the Quality Chasm report (Institute of Medicine (US) Committee on Quality of Health Care in America 2001) focused on supporting self-management by encouraging providers to use education and other interventions to systematically increase patients' skills, confidence, and empowerment in managing their health problems (Holman and Lorig 2004). Two specific initiatives include patient-centered care and informatics. As described in ▶ Sect. 13.2, patient-centered care aims to inform and involve patients and their families in decision making and self-management, coordinate and integrate care, provide physical comfort and emotional support, understand patients' concepts of illness and their cultural beliefs, and understand and apply principles of disease prevention and behavioral change appropriate for diverse populations. Informatics aims to communicate, manage knowledge, and support the use of information technology for decision making (Jimison et al. 2008). Examples of these informatics tools include home monitoring systems with interactive disease-management or self-management technology, educational or decision-aid software that is interactively customized to the patient's needs, online patient support groups, tailored interactive health reminder systems where interactions are linked with electronic health records, and patient-physician electronic messaging. These types of tools may be implemented on a variety of platforms using Web/Internet technology, touch screen kiosks, mobiles phones, or combinations of these. The location where individuals access the information also widely varies – ranging from clinics, hospitals, home, workplace, or any mobile location. However, many factors relating to how the technology is deployed to the consumer or health system can influence access, usability, and effectiveness. Many studies have demonstrated health

outcome disparities related to race and ethnicity, income, and education. With the increasing availability and use of personal health technologies, we have an opportunity to reduce these disparities with appropriate targeted design choices. The following sections identify some of the design challenges.

11.4.1 Health Literacy and Numeracy

Literacy skills play an important role in navigating the healthcare system, in learning about health and medical concerns, and in using personal health technologies. Health literacy is defined as "the degree to which individuals can obtain, process, and understand the basic health information and services they need to make appropriate health decisions"(Berkman et al. 2011). Several skills are required for an individual to appropriately integrate healthcare information and function effectively in the healthcare environment. One must be able to understand written material (print literacy), understand graphs and numerical quantitative information (numeracy), and be able to both speak and listen effectively (oral literacy). Low health literacy is a significant problem in the United States. In 2003, approximately 80 million adults in the United States (36 percent) had limited health literacy. Certain population subgroups have higher rates of limited health literacy. For instance, rates are higher among older adults, minorities, individuals who have not completed high school, adults who spoke a language other than English before starting school, and people living in poverty. Highlighting the health impact of low health literacy, a 2004 systematic evidence review found a relationship between low health literacy and poor health outcomes (Berkman et al. 2011). Specifically, lower health literacy (measured by reading skills) was associated with lower health-related knowledge and comprehension, higher hospitalization rates, poor global health measures, and certain chronic diseases. There are important considerations that personal health informatics must address to overcome these literacy issues such as adaptable interfaces and intuitive icons and graphics; however, designers need to give these considerations priority.

11.4.2 Digital Divide

Concerns about a "digital divide" between the "haves" and "have-nots" have long existed, mainly focused on economic access to the technology. And certainly, access to information technology is now seen as an important component to quality health care. Thus, disparity in access leads to disparity in health outcomes. However, there are many potential causes of a digital divide in addition to income. Studies have shown links to education, ethnicity, gender, urban/rural geography, age, and culture (Carr 2007; Ernest III et al. 2004; Kontos et al. 2014; Mossberger et al. 2006; Neter and Brainin 2012; Wensheng 2002). The terms "digital native" and "digital immigrant" are often used to characterize those generations of people who were born into a digital world, versus those who have experienced the migration from non-digital to digital information. Mobile phone and smartphone access has changed the digital divide trend somewhat recently with over 95% of the global population having access to a mobile phone (Fehske et al. 2011), and a rapidly growing number in all sectors having access to smartphones. Interestingly, in the U.S. the rate of smartphone adoption among Blacks and Hispanics outpaces that among Whites. With more personal health informatics systems using mobile phone interfaces, we may better address the digital divide issue in the future. The most challenging issues at this point relate to available mobile and smartphone bandwidth and education.

11.4.3 Chronic Conditions

Approximately 120 million Americans have one or more chronic illnesses, accounting for 70 to 80 percent of healthcare costs. Twenty-five percent of Medicare recipients have four or more chronic conditions, accounting for

two thirds of Medicare expenditures (Hoffman et al. 1996; Wagner 2001). Most patients with chronic conditions such as hypertension, diabetes, hyperlipidemia, congestive heart failure, asthma, and depression are not treated adequately, and the burden of chronic illness is magnified by the fact that chronic conditions often occur as comorbidities (Bodenheimer et al. 2002c; Wagner et al. 2001). One key element of systems-oriented chronic care models is support of patient self-management in the home environment (Bodenheimer et al. 2002b). Such self-management support can reduce hospitalizations, emergency department use, and overall managed care costs (Bodenheimer et al. 2002a; Coleman and Newton 2005; Lorig et al. 2001; Renders et al. 2001; Whitlock et al. 2000).

11.4.4 Conditions Associated with Aging

A great many elderly persons receiving care have functional limitations, such as reduced sensory, cognitive, or motor capabilities and may require disease management for multiple chronic conditions. Although personal health informatics has the potential to empower patients to become more active in the care process, the elderly may be disadvantaged unless the designers of both software and hardware technology consider their needs explicitly (Demiris 2016). Usability and accessibility issues are important quality criteria for Web-based and mobile interventions (see ▶ Chap. 5), but often neglected by designers and evaluators (Eysenbach et al. 2002).

11.4.5 Behavior Management

Many patients have believed in the concept of health prevention through wellness activities (active lifestyle development, stress reduction, weight control) long before healthcare professionals endorsed this mode of self-care. With the advent of preventive medicine and data supporting the role that wellness activities can play in maintaining health, many of these tenets have become a part of the armamentarium for disease management and are an area of discovery supported by the Agency for Healthcare Research and Quality (AHRQ) and other Federal institutes.

A complete discussion of foundational models of behavior change is beyond the scope of this chapter, but key works are listed in ◘ Table 11.1. Researchers and educators capitalize on these theories to reduce risky behaviors (e.g., cigarette smoking, unprotected sexual intercourse, and unhealthy eating) and to promote desirable health behaviors (i.e., referred to as behavior change).

Beginning in 2007, the Robert Wood Johnson Foundation, in the Project HealthDesign Initiative (Brennan et al. 2007), catalyzed the development of personal health applications, with the belief that a properly developed common platform would be essential to the spread of intelligent, interoperable and theoretically-based behavior change tools. This initiative demonstrated many tools that could help consumers with behavior change. These demonstrations leveraged the widespread adoption of "smart" phone technology across geographic and socioeconomic divides. This widespread adoption, coupled with easy to use software development environments, enables the development of personal health applications that operate as stand-alone or integrated tools available to most consumers.

11.5 The Impact of Personal Health Informatics on Biomedical Informatics

It is almost axiomatic in biomedical informatics that the introduction of a new discipline or information user category into the healthcare system induces change. Such is the case with the inclusion of person-centered care into biomedical informatics. The unique aspects of what distinguishes health system-generated, provider-generated, and person-generated data from each other permeate all aspects of biomedical informatics. These unique charac-

Table 11.1 Models of health behavior change

Name and source	Summary
Self-efficacy [Bandura 1977]	An individual's impression of one's own knowledge and skill to perform any task, based on prior success, physical ability, and outside sources of persuasion. Predicts the amount of effort a person will expend to change behavior. It is a key component of other theories, such as the Theory of Planned Behavior.
Social cognitive theory [Bandura 1989]	Behavior change is determined by personal, environmental and behavioral elements, which are interdependent.
Theory of planned behavior [Ajzen 1985]	A link between attitudes and behavior. It asserts that behaviors viewed positively and supported by others (subjective norm) are more likely to have higher levels of motivation and more likely to be performed.
Transtheoretical/stages of change model [Prochaska 2005]	This model asserts that behavioral change is a 5-step process, between which a person may oscillate before achieving complete change.
Patient engagement/patient activation [Dentzer 2013; Hibbard 2004]	Patient engagement describes the actions one can take to achieve maximum benefits from the healthcare services available. Patient activation is a person's knowledge, confidence, and skills used to manage their health. Improved engagement and activation have been associated with improved healthcare outcomes
Early games targeted at behavior change (described above) attempted to remove the stigma (attitude) associated with engaging in healthy behaviors (such as taking medications to combat a chronic illness)	

teristics impact both the development of methods to acquire and store data, as well as applications of these methods to impact care and discovery. To provide a frame of reference about the impact of person-centered care on the field, we will describe the magnitude of change using selected examples covered elsewhere in the book.

11.5.1 Data Science

The field of data science has been impacted greatly by personal health informatics methods and advances. This discipline, which is focused on the use of scientific processes, algorithms and systems to extract knowledge and insights from both structured and unstructured data, addresses a number of concerns inherent in so-called "big data" (Provost and Fawcett 2013). These data come in large amounts (volume), with often fast speeds of arrival and (velocity), varying formats (variety), varying timing (variability), unclear accuracy/validity (veracity), and significant risks to patient privacy (vulnerability). Data collected by patients who may be less familiar with healthcare terminology add a layer of complexity to the variety, veracity, and vulnerability challenges inherent in these data. Data collected using commonly available sensor technologies add to the velocity and volume challenges, but also present new opportunities. These, and other new challenges imposed by the addition of personal data into the healthcare/discovery system are new opportunities for personal health informatics research.[3]

11.5.2 Precision Medicine

Perhaps no area has received greater attention in the last 10 years than precision medicine. The goal of precision medicine as noted

3 Technology OotNCfHI. Conceptualizing a Data Infrastructure for the Capture, Use, and Sharing of Patient-Generated Health Data in Care Delivery and Research through 2024, 2018. Available from: ► https://www.healthit.gov/sites/default/files/onc_pghd_final_white_paper.pdf

in the Mission statement of the NIH's *All of Us* initiative, is "to enable a new era of medicine through research, technology and polices that empower patients, researchers and providers to work together toward development of individualized treatments" (All of Us Research Program Investigators 2019) (► https://obamawhitehouse.archives.gov/precision-medicine) (see ► Chap. 28). Patient empowerment in this new era of medicine will take many forms, and will question how we help patients with concepts such as contributing personal data to empower discovery, understanding a radically new way to subcategorize diseases according to patient-specific attributes, learning a new "trust" model to understand why two patients with the same disease may have very different treatment plans, and even recognizing an expanding role for consumers in establishing and critiquing health policy (Adams and Petersen 2016; Juengst et al. 2012).

11.5.3 Ethical, Legal and Social Issues

As our understanding about the implications of adopting a person-centered care philosophy mature, so do the ways in which this philosophy should govern the treatment of the person and her data and information. This philosophy introduces new ethical, legal and social issues that are deserving of thought to maximize the effective and safe use of these data and applications (see ► Chap. 12). For example, concerns raised by the access to decision-making applications became the impetus for thoughtful discussion about the timing of FDA regulation of mHealth apps (Lye et al. 2018).

One of the most exciting, though potentially alarming consequences of our extensive use of the Web for shopping, communicating, and learning is that each of us leaves behind a profile of who we are, what we like or dislike, what we know or don't know, and what we want or already have. When combined with data mining and natural language processing techniques, it is possible to create highly targeted and predictive personal knowledge. Search engines regularly exploit this opportunity to create a profile of each searcher and to improve the relevance of retrieved results (Frisse 1996). We can expect the use of these data to impact how medical care is personalized. Data created by consumers, coupled with ubiquitous computing, might provide just–in-time nutritional consults, over the counter medication advice, or advice that might prevent illnesses, such as convenient locations to receive a flu vaccine or when to begin medications for seasonal allergies (Guo et al. 2016; Pellegrini et al. 2018; Swendeman et al. 2015).

This technology, which is likely to improve the user experience and functionality available with consumer-facing technologies, also has significant downside risks to patient privacy. In particular, data from email, internet searches, support group chats, genome risk prediction sites, and mobile or cloud-based apps all may be directly identifiable, or may be combined with data from other sources to be identifiable. Together, they may be used to create a profile of an individual, his or her health status, the health status of related individuals, or other profiles, all of which may be useful for targeted advertising, insurance risk, employability, or other purposes. Biomedical informatics research in the person-centered care era focuses on such topics as the boundaries of HIPAA protection, genomic privacy, and re-identification risk.

A digital divide can lead to serious ethical issues. If technological interventions are only available or usable to a segment of the population, this imbalance can threaten the impact of these technologies on improving healthcare for all and increase healthcare disparities. Certain interventions, such as mHealth and the Internet of Things technologies, require a certain digital literacy. Installing and maintaining these devices with the tremendous amount of data they generate can be daunting to populations with lower literacy and numeracy as well as those with low proficiency levels for problem solving in technology-rich environments. Also, mobile phones made by different companies can have different capabilities. At the present time, Android phones are widely used by people with lower

socioeconomic status. If Apple iOS apps are shown to be more effective than Android, limited access may increase the disparities and effect of these interventions for many people. Finally, many mHealth apps are only available in English, which further increases the digital divide.

While the direction that personal health informatics will take in the future is at best, educated speculation, it is clear that as long as patient-provider partnerships are endorsed, technology will be a third partner in ensuring that activated people manage their health and disease effectively.

11.5.4 Communication

Inherent in all aspects of information management is the recognition about the audience to whom information is being communicated. Perhaps no field has a larger gap between what is understood by its professionals and its consumers than health care. Numerous studies have demonstrated issues that persist with the age-old challenge of how to educate people about diseases. These challenges are magnified as the scope of consumer engagement broadens to include many of the topics listed above, and as we use data and evidence for care enters everyday discourse with patients (McCormack et al. 2013). Furthermore, the separation of information communication to and from medical professionals creates opportunities for misunderstanding at best, and inappropriate actions being taken by patients at worse (Isaacs and Creinin 2003; Morgan 2013). This is an area ripe for research and evaluation by biomedical informatics professionals, much of which is already underway in areas such as methods to circumvent literacy and numeracy challenges. It is clear that there is a correlation between health literacy and quality of life (Zheng et al. 2018), but also clear that more research needs to be done to understand how to communicate in the face of this reality (Newnham et al. 2017; Fisher et al. 2016).

11.5.5 Mobile Health Care (mHealth)

Perhaps the most significant change in the landscape of personal health informatics has been the adoption of "smartphone" technology into society. Smartphones are mobile phones that perform many functions found on present-day computers. They typically contain a touch screen interface, camera, Internet access, short-range wireless interconnection technology, and an operating system capable of executing downloaded applications. Smartphone ownership has grown worldwide, with an estimated 81% ownership by adults in the United States.[4] When combined with a new generation of wireless or connected peripheral technologies (imaging tools, wearable sensors, monitoring systems, etc.) downloadable applications (apps) have revolutionized information collection and use by people, and have defined a new field called mobile health care (mHealth) (see ▶ Chap. 19) (Cameron et al. 2017).

One of the main byproducts of the mHealth era has been a radical improvement in consumer empowerment, coupled with information sharing that enables individual groups to make "informed" decisions about their care with or without the assistance of a healthcare professional. With these new capabilities come enormous opportunities for biomedical informatics to influence the entire healthcare system. Terms such as "quantified self" (Dudley et al. 2015) and "Internet of Things" (Dimitrov 2016) begin to characterize the potential of mHealth.

It is through the use of mHealth applications that the notion of personal health informatics has grown beyond the individual's clinical needs to population-level care needs.

4 Pew Research Center. Smartphone ownership is growing rapidly around the world, but not always equally. Available from: ▶ https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/. Accessed November 9, 2019.

Indeed, innovations such as Apple's ResearchKit © and the explosion of wearable fitness trackers connected to social networks are designed for smartphones and mHealth technologies. These technologies also are being positioned improve the structure of healthcare delivery, through innovations such as appointment self-scheduling, direct-to-consumer e-consults, and peripheral devices that make home diagnoses commonplace (Topol 2015; Kawano et al. 2012).

Like any foundational change in biomedical informatics, the advent of mHealth creates new paradigms for concepts such as usability and usefulness. Unique characteristics of people (literacy, numeracy, language differences) must be kept in mind and considered, along with the capabilities of people and their living, working, and social environments. Research in user-centered design, usability assessment, failure modes and effects analysis, and other techniques to assure safe and effective use are increasingly critical to advances in mHealth (see ► Chap. 5) (Overdijkink et al. 2018; Matthew-Maich et al. 2016).

Another significant challenge for mHealth is integration into clinical workflows. If healthcare providers are not prescribing mHealth apps or using their data, patients may be less likely to use them if they cannot engage their provider in shared decision making. In other personal health informatics tools, such as patient portals, adoption and promotion of patient portal usage by providers leads to increased usage by patients (Cronin et al. 2015). It will be critical to improve usefulness of the vast amount of data generated by mHealth, aid provider and patients in choosing and using mHealth apps, and determine the appropriate touch points of these interventions between providers and patients, which will enable the potential of mHealth in the future.

11.5.6 Social Network Systems

Social network systems, epitomized by Facebook (► www.facebook.com), are online virtual communities where participants describe themselves with member-entered attributes, establish or break connections to other members, communicate, and share information. This simple strategy creates a virtually unlimited method to connect similar people to one another, and has been shown to be an effective tool to connect people with specific health needs (Moorhead et al. 2013). The for-profit online health-related social networking community Patients Like Me has demonstrated that individuals with a severe chronic disease—amyotrophic lateral sclerosis—are highly willing, even without compensation, to contribute data and observations to a patient community (Frost and Massagli 2008) to accelerate learning about their disease. The site has no ties to the conventional healthcare system and short-circuits the traditional research enterprise, rewarding participants, not just researchers, with knowledge. The patient outcomes of diverse therapies are collected using crowd sourcing, where patients contribute their information to a common database that can be queried to obtain summaries of an aggregated experience of their peers.

Social networking Web sites share most or all of the features of electronic support groups, and even some data commonly provided through a portal (through creating an affiliation with a group who externalizes public or private information). Social networking platforms combined with personal health records provide a means for social network members to share and aggregate data obtained from the traditional health system, and to do so in a private manner (Eysenbach 2008; Weitzman et al. 2011a).

One of the features of health-related online social networks is the rapid dissemination of information across a network; however, there is great variability in the quality of discourse on health-related social networking sites. Conversations may be moderated, in certain cases by a health coach (Jimison et al. 2007). Conversations also may be unmoderated and commercial influences may enter the discourse without transparency. There are also concerns around privacy. Compared with the restrictive institutional consents and compacts with patients that limit use of data and specimens under federal regulations applica-

ble to much federally sponsored research, online social networks are generally governed by no more than a terms of use statement, often subject to change without notice in 30 days. These privacy policies may be difficult to find and not written in language accessible by a population with a broad range of health literacy (Weitzman et al. 2011b). Industry standards governing safety and privacy of online health-related social networking are yet to emerge.

11.5.7 Application Example: EHR Portals

As the electronic health record gains acceptance, its relevance to individual people also grows. Many hospitals and clinics have begun providing direct patient access to the clinical record. These portals are defined as person-facing systems tethered to electronic health records, allowing them views of clinical or claims data in institutional electronic health record systems or payer systems (Tang et al. 2006; Kim and Johnson 2002). Portals provide motivated people with a way to electronically access sections of their records to recall salient instructions or obtain results of tests. Some of the first such personal health portals were Columbia's PatCIS system (Cimino et al. 2002) and Beth Israel Deaconess's PatientSite, developed in 1999 (Weingart et al. 2006). Two of the most widely deployed portals are Epic's MyChart (Serrato et al. 2007) and MyHealtheVet (Nazi and Woods 2008). Many of these portals provide capabilities besides simply viewing EHR information, such as secure physician-patient messaging, appointment scheduling, providing educational information, and viewing and managing medical bills. One of the more recent additions to this set of capabilities is the OpenNotes effort started by MyHealtheVet, which exposes every progress note to patients, instead of exposing only discharge summaries. As new data types are provided to the healthcare system by people in support of their health, we can expect the data types exposed through EHR portals to change, along with new EHR portal application capabilities.

11.5.8 Application Example: Personal Health Records

According to the Markle Foundation, a personal health record (PHR) is "an electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment" Connecting for Health Personal Health Working Group (2003). Like the EHR portal, the PHR has become the foundation for developing tools to store data and to facilitate its reuse in ways people find engaging.

The idea of a personal repository for medical information is far from new. Families with infants have used an immunization record book for decades. The immunization blue book is a quintessential, efficient system with portable information that supports entry by multiple providers and storage by the patient. Clayton Christensen, who invented the concept of "disruptive innovation," summarizes the widely held promise of this technology in his book, the *Innovator's Prescription: A Disruptive Solution for Health Care* (Christensen et al. 2009). "We cannot overstate how important PHRs are to the efficient functioning of a low-cost, high quality healthcare system." PHRs enable users to acquire copies of their data from every site of care. In some ways, this model advances information flow far more than models requiring inter-institutional data sharing agreements. Data from two competing healthcare networks may reside in the same PHR without cumbersome agreements between those two networks. The patient asserts her claim to the data for each network independently. This model of data aggregation may promote data liquidity far more than competing approaches, such as health information exchanges, which require centralized management of data sharing agreements between networks and institutions (Adler-Milstein et al. 2008).

In part fueled by the knowledge gained during the Robert Wood Johnson Foundation's Project HealthDesign, various companies in the early 2000's developed commercially available personal health records. While some of these remain viable, many from that era were discontinued, largely due to the complex nature of establishing interoperability with external data sources, as well as unsustainable financial models.[5] Recently, however, as data liquidity and standards promoting interoperability have been mandated through Federal legislation, there has been a resurgence of activity to create PHRs from both small start-up companies and large EHR vendors. The future is still uncertain, but this suite of applications continues to be a likely foundation for more sophisticated services and applications used by people in support of their health or illness management (Staccini et al. 2018).

11.5.9 Application Example: Sensors for Home Monitoring and Tailored Health Interventions

With rapidly advancing technologies, sensors that measure and monitor are everywhere. We see an increasing population interested in monitoring their activity levels with wrist devices that now measure movement (converted into steps or calories burned), heart rate, and electrodermal activity (for stress level). There are wireless weight scales that are useful for weight management of everyone, and for fluid management of heart failure patients. Even smartphones have a myriad of embedded sensors useful for managing one's health (e.g., GPS for location context, motion for activity level, light and noise level for sleep management, and voice statistics for mood management).

In health care there is an increasing need to manage chronic conditions more effectively by empowering patients and family caregivers with more active roles in self management. Sensors in the home and environment (including wearables) provide important input to algorithms that infer patient state and deliver tailored feedback and motivational messaging. ◘ Figure 11.4 shows patient-generated sensor data being aggregated in a local device (typically a smartphone) and transferred with strict security protocols to a secure server. The inference algorithms in real time then generate messaging and summary content for the patient, as well as a health coach, clinician, and remote or local family caregiver (Pavel et al. 2015).

For disease management interventions typical sensors include wireless weight scales for fluid management, wireless blood pressure cuffs for cardiac disease, blood glucose meters for patients with diabetes, and peak flow meters for those with asthma. Additionally, many disease management protocols include weight management, physical exercise and medication management. ◘ Figure 11.5a–c show examples of sensor technology used in home health settings. Motion sensors, as shown in ◘ Fig. 11.5a, can be used to determine real-time patient location for inferring context as well as for measuring walking speed statistics (a useful cognitive indicator) (Hagler et al. 2010). Sleep quality can now be measured with varying accuracy with techniques ranging from accerometers in wrist fitness trackers to pressure sensitive bed strips placed under the mattress that detect heart rate, heart rate variability (for stress recovery at night), respiration, as well as total sleep time and sleep efficiency. Another more complex approach to sensing in the home involves imaging, as shown with the interactive video exercise in ◘ Fig. 11.5c. In this case, data from the Kinect camera is used to detect movement compared to goal state and provide just-in-time feedback to the user (Jimison et al. 2015; Obdrzalek et al. 2012; Ofli et al.

5 Google Official Blog. An update on Google Health and Google PowerMeter. Available at: ▶ https://googleblog.blogspot.com/2011/06/update-on-google-health-and-google.html. Accessed November 9, 2019.

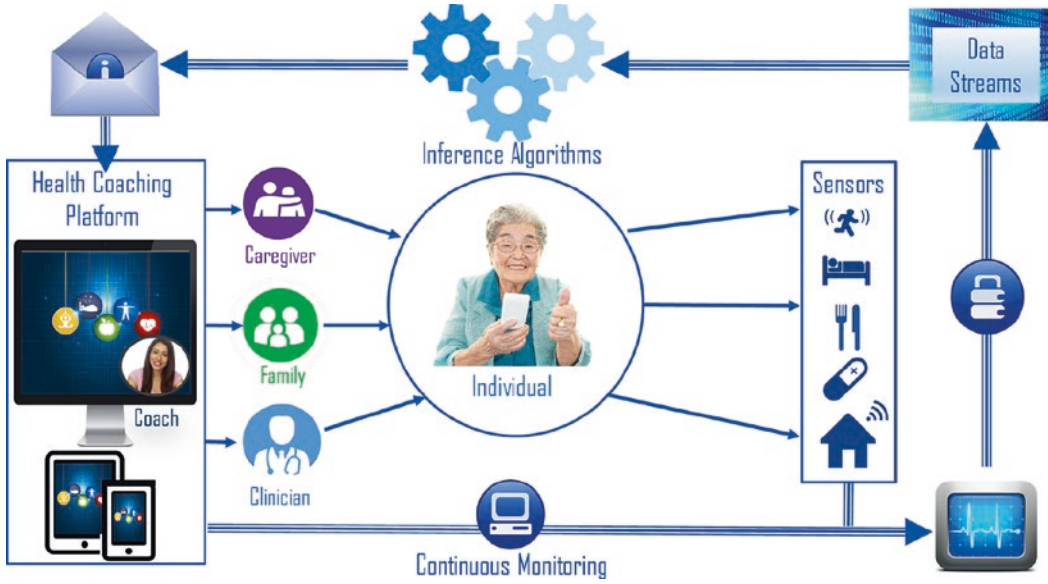


Fig. 11.4 This diagram shows how sensor data from the home or environment, generated by a patient who may have a chronic condition or an individual with an interest in improving health, can be used as input to a coaching platform to provide tailored motivation and feedback

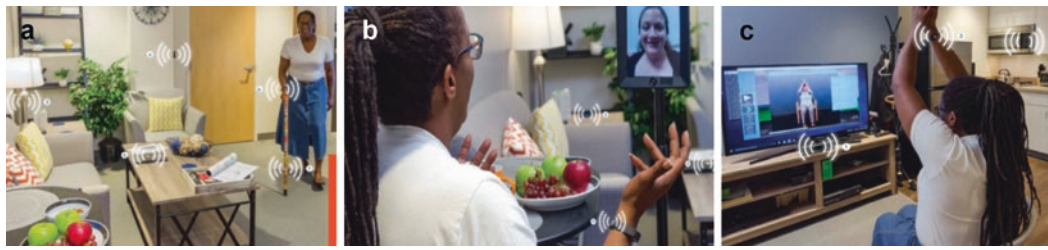


Fig. 11.5 This series of images shows examples of sensors and technology used to gain information about an individual's state and provide tailored just-in-time feedback. **a** shows a motion sensor near the door, sensors on a smart cane, a presence lamp, wireless blood pressure cuff, and an Amazon Echo. **b** shows an individual interacting with a nurse care manager using a remote controlled Double Robot. **c** shows in-home chair exercises with real-time feedback using data from the Kinect camera. (With permission from the Consortium on Technology for Proactive Care at Northeastern University. Photo courtesy of Dr. Holly Jimison)

2016). Inferences on strength, flexibility, balance and endurance can be monitored over time and provided to both the user and clinician. Interactive voice messaging systems (e.g., Amazon Echo or Google Home) have an important role in home health interventions, both as communication devices, but also as sensors of voice affect. Finally, interactions with computers, tablets and smartphones provide valuable information on cognitive performance, both specifically with adaptive cognitive computer games (Hagler et al. 2014; Jimison et al. 2010), but also with indirect measurements over time of motor speed, search time and cognitive load (Hagler et al. 2011, 2014).

The streaming data from a variety of sensors in the home and environment can be overwhelming from a clinical perspective. Figure 11.6 shows a sample phone interface for coordinating information from a variety of sensors (Williamson 2015). This example shows a summary main screen for a patient or caregiver with feedback on calendar to-do's, adherence to goals, medication taking (an issue in taking medications noted

Fig. 11.6 This example shows a main screen for a patient or caregiver summary information. (Reprinted with permission of author, S. Williamson)

by a red "X"), level of socialization, cognitive function, and sleep quality (soft warning noted by an orange "!"). In this case, clicking on an icon opens a screen with further detail. The main goal of sensor-based systems for health is to facilitate health management and adherence to an individual's health goals using known principles of health behavior change. This type of technology enables a scalable and potentially cost-effective approach to providing continuity of care. It addresses use of an often untapped resource of both patient and family caregiver participation as part of the care team.

11.6 Future Opportunities and Challenges

As is exemplified by the previous section, the opportunities for personal health informatics to improve health outcomes are plentiful. Even worldwide, the access to mobile phones is becoming nearly ubiquitous, and the affordability of health sensors and devices for continuous monitoring and just-in-time intervention is also improving rapidly. However, we also see upcoming challenges in the areas of payment models and equity.

11.6.1 Reimbursement and Business Models

Many countries have global budgets for health care, usually managed at a regional level, where it is possible to allocate funds for cost-effective health interventions that personal health technologies may enable. However, medical care reimbursement in the United States is moving slowly towards value-based care. Healthcare systems in the U.S. require incentives and fairly short-term business model demonstrations to modify their workflow and hiring practices for a new model of value-based care. This model of care would bring healthcare consumers and family members as integral members of the care team, facilitated by personal health technologies.

11.6.2 Opportunities for Innovation

As health care moves from being clinic-centric and hospital-centric to person-centric and more proactive, there are many opportunities for new advances in personal health informatics to facilitate this change and improve health outcomes. As mentioned earlier in this chapter, advances in the assessment of person state through new always-on sensors and improved computational modeling will allow more tailored and timely messaging and interventions.

Virtual Reality and Augmented Reality are important innovations that can transform the way that individuals, especially older adults, are cared for. Artificial Intelligence innovations that could lead to more tailored messages for a person's health and wellness could overcome barriers such as remembering to take their medications by targeting cues to improve care. Finally, fusing the information

from sensors could allow for improved assessment of people and their health.

Many of the innovations, however, will need to be social and protocol-based. For example, new workflow and hiring practices will be needed to compensate for the data and information these innovations will create. An increased emphasis on proactive person-centered care to improve outcomes and reduce costs will necessitate better use of community health workers and health behavior change coaches that interface with both the patient and the clinical team. One of the most exciting, though potentially alarming consequences of our extensive use of the Web for shopping, communicating, and learning is that each of us leaves behind a profile of who we are, what we like or dislike, what we know or don't know, and what we want or already have. When combined with data mining and natural language processing techniques, it is possible to create highly targeted and predictive personal knowledge. Data created by consumers, coupled with ubiquitous computing, might provide just–in-time nutritional consults, over the counter medication advice, or advice that might prevent illnesses, such as convenient locations to receive a flu vaccine or when to begin medications for seasonal allergies. We can expect the use of these massive data sets (also called "big data") to impact how medical care is personalized. While the direction that consumer health informatics will take in the future is at best, educated speculation, it is clear that as long as patient-provider partnerships are endorsed, technology will be a third partner in ensuring that activated consumers manage their health and disease effectively.

Suggested Readings

Berwick, D. M. (2009). What 'patient-centered' should mean: confessions of an extremist. Health Affairs, 28(4), w555–ww65. This paper is written by Dr. Donald Berwisk, an influential proponent of patient-centered health care and former administrator of the Centers for Medicare and Medicaid Services (CMS). The paper describes a number of maxims of patient-centered care.

Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health & Human Services. Conceptualizing a data infrastructure for the capture and use of patient-generated health data. Available at: https://aspe.hhs.gov/conceptualizing-data-infrastructure-capture-and-use-patient-generated-health-data. Accessed 1 Nov 2019. This paper describes a project sponsored by the Office of the National Coordinator for Health Information Technology (ONC) regarding a data infrastructure for patients to share their data with caregivers, providers, researchers, and others according to their preferences.

Prey, J. E., Woollen, J., Wilcox, L., Sackeim, A. D., Hripcsak, G., Bakken, S., et al. (2014). Patient engagement in the inpatient setting: A systematic review. Journal of the American Medical Informatics Association, 21(4), 742–750. This paper reviews literature involving patient engagement in the hospital setting. The authors identify challenges such as inconsistent use of termionology and gaps ini knowledge regarding impact on health outcomes and cost-effectiveness.

Topol, E. J. (2015). The patient will see you now: The future of medicine is in your hands. New York: Basic Books. This popular book describes the author's vision for medicine based on patient-centered health care, mobile health, and consumer health informatics.

Questions for Discussion

1. What is the role of the health system in monitoring the quality of discourse on online social networks?
2. What is the optimal model for personal health records? Should personal health records display advertisements?
3. Which populations of consumers would be most likely to use personal health records?
4. Which consumer technologies do you think will be most influential in consumer-focused health informatics?
5. What is the right balance between privacy of personal health information and ready access to it? For example, for an unconscious patient in the emergency department?

References

Adams, S. A., & Petersen, C. (2016). Precision medicine: Opportunities, possibilities, and challenges for patients and providers. *Journal of the American Medical Informatics Association, 23*(4), 787–790. https://doi.org/10.1093/jamia/ocv215.

Adler-Milstein, J., McAfee, A. P., Bates, D. W., & Jha, A. K. (2008). The state of regional health information organizations: Current activities and financing. *Health Affairs (Millwood), 27*(1), w60–w69. https://doi.org/10.1377/hlthaff.27.1.w60.

All of Us Research Program Investigators. (2019). The "All of Us" research program. New England Journal of Medicine, 381(7), 668–676. Chicago.

Ammenwerth, E., Schnell-Inderst, P., & Hoerbst, A. (2012). The impact of electronic patient portals on patient care: A systematic review of controlled trials. *Journal of Medical Internet Research, 14*(6). https://doi.org/10.2196/jmir.2238.

Appelboom, G., LoPresti, M., Reginster, J.-Y., Sander Connolly, E., & Dumont, E. P. (2014). The quantified patient: A patient participatory culture. *Current Medical Research & Opinion. Taylor & Francis, 30*(12), 2585–2587.

Asnani, M. R., Quimby, K. R., Bennett, N. R., & Francis, D. K. (2016). Interventions for patients and caregivers to improve knowledge of sickle cell disease and recognition of its related complications. *Cochrane Database of Systematic Reviews, 10*, Cd011175. https://doi.org/10.1002/14651858.CD011175.pub2.

Balint, E. (1969). The possibilities of patient-centered medicine. *The Journal of the Royal College of General Practitioners, 17*(82), 269.

Barry, M. J., Fowler, F. J., Mulley, A. G., Henderson, J. V., & Wennberg, J. E. (1995). Patient reactions to a program designed to facilitate patient participation in treatment decisions for benign prostatic hyperplagia. *Medical Care, 33*(8), 771–782.

Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., Viera, A., Crotty, K., et al. (2011). Health literacy interventions and outcomes: An updated systematic review. *Evidence Report Technology Assessment (Full Report), 199*(1), 941. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23126607

Berwick, D. M. (2009). What 'patient-centered' should mean: Confessions of an extremist. *Health Affairs, 28*(4), w555–w565.

Bodenheimer, T., Lorig, K., Holman, H., & Grumbach, K. (2002a). Patient self-management of chronic disease in primary care. *JAMA, 288*(19), 2469–2475.

Bodenheimer, T., Wagner, E. H., & Grumbach, K. (2002b). Improving primary care for patients with chronic illness. *JAMA, 288*(14), 1775–1779.

Bodenheimer, T., Wagner, E. H., & Grumbach, K. (2002c). Improving primary care for patients with chronic illness: the chronic care model, Part 2. *JAMA, 288*(15), 1909–1914. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12377092

Bosworth, K., Gustafson, D. H., Hawkins, R. P., Chewning, B., & Day, T. (1983). Adolescents, health education, and computers: The body awareness resource network (BARN). *Health Education, 14*(6), 58–60.

Brennan, P. F., Downs, S., Casper, G., & Kenron, D. (2007). *Project HealthDesign: Stimulating the next generation of personal health records.* Paper presented at the AMIA annual symposium proceedings.

Cameron, J. D., Ramaprasad, A., & Syn, T. (2017). An ontology of and roadmap for mHealth research. *International Journal of Medical Informatics, 100*, 16–25. https://doi.org/10.1016/j.ijmedinf.2017.01.007.

Carr, D. (2007). The global digital divide. *Contexts, 6*(3), 58.

Cassell, E. J. (1985). *Talking with patients: Clinical technique* (Vol. 2). Cambridge, MA: Mit Press.

Christensen, C. M., Grossman, J. H., & Hwang, J. (2009). *The innovator's prescription : A disruptive solution for health care*. New York: McGraw-Hill.

Cimino, J., Patel, V., & Kushniruk, A. (2002). The patient clinical information system (PatCIS): Technical solutions for and experience with giving patients access to their electronic medical records. *International Journal of Medical Informatics, 68*(1–3), 113–127.

Coleman, M. T., & Newton, K. S. (2005). Supporting self-management in patients with chronic illness. *American Family Physician, 72*(8), 1503–1510.

Coleman, K., Austin, B. T., Brach, C., & Wagner, E. H. (2009). Evidence on the Chronic Care Model in the new millennium. *Health Affiars (Millwood), 28*(1), 75–85. https://doi.org/10.1377/hlthaff.28.1.75.

Collen, M. F., Rubin, L., Neyman, J., Dantzig, G. B., Baer, R. M., & Siegelaub, A. (1964). Automated multiphasic screening and diagnosis. *American Journal of Public Health and the Nations Health, 54*(5), 741–750.

Collins, F. S. (2004). The case for a US prospective cohort study of genes and environment. *Nature, 429*(6990), 475–477. https://doi.org/10.1038/nature02628.

Connecting for Health Personal Health Working Group. (2003). The Personal Health Working Group: final report. Markle Foundation.

Cronin, R. M., Davis, S. E., Shenson, J. A., Chen, Q., Rosenbloom, S. T., & Jackson, G. P. (2015). Growth of secure messaging through a patient portal as a form of outpatient interaction across clinical specialties. *Applied Clinical Informatics, 6*(2), 288–304. https://doi.org/10.4338/ACI-2014-12-RA-0117.

Demiris, G. (2016). Consumer health informatics: Past, present, and future of a rapidly evolving domain. *Yearbook of Medical Informatics, 25*(Suppl 1), S42.

Dentzer, S. (2013). Rx for the 'blockbuster drug'of patient engagement. *Health Affairs (Project Hope), 32*(2), 202.

Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. *Healthcare Informatics Research, 22*(3), 156–163. https://doi.org/10.4258/hir.2016.22.3.156.

Dudley, J. T., Listgarten, J., Stegle, O., Brenner, S. E., & Parts, L. (2015). Personalized medicine: From genotypes, molecular phenotypes and the quantified self, towards improved medicine. In *Pacific symposium on biocomputing co-chairs* (pp. 342–346). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25592594

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*(4286), 129–136.

Enthoven, A. C. (2009). Integrated delivery systems: The cure for fragmentation. *American Journal of Managed Care, 15*(10 Suppl), S284–S290. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20088632

Epstein, R. M., & Street, R. L., Jr. (2011). The values and value of patient-centered care. *Annals of Family Medicine, 9*(2), 100–103. https://doi.org/10.1370/afm.1239.

Ernest, J., III, Wilson, E. J., & Wilson III, E. J., III. (2004). *The information revolution and developing countries*. Cambridge, MA: MIT Press.

Eysenbach, G. J. (2008). Medicine 2.0: Social networking, collaboration, participation, apomediation, and openness. *Journal of Medical Internet Research, 10*(3), e22.

Eysenbach, G., Powell, J., Kuss, O., & Sa, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the world wide web: A systematic review. *JAMA, 287*(20), 2691–2700.

Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., & Stern, A. (2004). Health related virtual communities and electronic support groups: Systematic review of the effects of online peer to peer interactions. *BMJ, 328*(7449), 1166.

Fehske, A., Fettweis, G., Malmodin, J., & Biczok, G. (2011). The global footprint of mobile communications: The ecological and economic perspective. *IEEE Communications Magazine, 49*(8).

Ferguson, T. (1996). *Health online: How to find health informaiton, support groups, and self help communities in cyberspace. The Millenium Whole Earth Catalogue*. Reading, MA: Da Capo Press.

Fisher, A., Manicavasagar, V., Kiln, F., & Juraskova, I. (2016). Communication and decision-making in mental health: A systematic review focusing on bipolar disorder. *Patient Education and Counseling, 99*(7), 1106–1120. https://doi.org/10.1016/j.pec.2016.02.011.

Frisse, M. E. (1996). What is the Internet learning about you while you are learning about the Internet? *Academic Medicine, 71*(10), 1064–1067. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9177639

Frost, J. H., & Massagli, M. P. (2008). Social uses of personal health information within PatientsLikeMe, an online patient community: What can happen when patients have access to one another's data. *Journal of Medical Internet Research, 10*(3), e15. https://doi.org/10.2196/jmir.1053.

Gallego, B., Walter, S. R., Day, R. O., Dunn, A. G., Sivaraman, V., Shah, N., et al. (2015). Bringing cohort studies to the bedside: Framework for a 'green button' to support clinical decision-making. *Journal of Comparative Effectiveness Research, 4*(3), 191–197. https://doi.org/10.2217/cer.15.12.

Gibbons, M. C., Wilson, R. F., Samal, L., Lehman, C. U., Dickersin, K., Lehmann, H. P., et al. (2009). Impact of consumer health informatics applications. *Evidence Report/Technology Assessment, 188*, 1.

Goldzweig, C. L., Towfigh, A. A., Paige, N. M., Orshansky, G., Haggstrom, D. A., Beroes, J. M., et al. (2012). *Systematic review: Secure messaging between providers and patients, and patients' access to their own medical record: Evidence on health outcomes, satisfaction, efficiency and attitudes*. Washington, DC: Department of Veterans Affairs (US).

Golomb, B. A., McGraw, J. J., Evans, M. A., & Dimsdale, J. E. (2007). Physician response to patient reports of adverse drug effects. *Drug Safety, 30*(8), 669–675.

Guo, J., Zhou, X., Sun, Y., Ping, G., Zhao, G., & Li, Z. (2016). Smartphone-based patients' activity recognition by using a self-learning scheme for medical monitoring. *Journal of Medical Systems, 40*(6), 140. https://doi.org/10.1007/s10916-016-0497-2.

Gustafson, D. H., Hawkins, R., Pingree, S., McTavish, F., Arora, N. K., Mendenhall, J., et al. (2001). Effect of computer support on younger women with breast cancer. *Journal of General Internal Medicine, 16*(7), 435–445.

Hagler, S., Austin, D., Hayes, T. L., Kaye, J., & Pavel, M. (2010). Unobtrusive and ubiquitous in-home monitoring: A methodology for continuous assessment of gait velocity in elders. *IEEE Transactions on Biomedical Engineering, 57*(4), 813–820.

Hagler, S., Jimison, H., & Pavel, M. (2011). *Unobtrusive in-home assessment by means of everyday computer mouse usage.* Paper presented at the Engineering in Medicine and Biology Society (EMBC), 2011 annual international conference of the IEEE.

Hagler, S., Jimison, H. B., & Pavel, M. (2014). Assessing executive function using a computer game: Computational modeling of cognitive processes. *IEEE Journal of Biomedical and Health Informatics, 18*(4), 1442–1452.

Hibbard, J. H., & Greene, J. (2013). What the evidence shows about patient activation: Better health outcomes and care experiences; fewer data on costs. *Health Affiars (Millwood), 32*(2), 207–214. https://doi.org/10.1377/hlthaff.2012.1061.

Hibbard, J. H., Stockard, J., Mahoney, E. R., & Tusler, M. (2004). Development of the Patient Activation Measure (PAM): Conceptualizing and measuring activation in patients and consumers. *Health Services Research, 39*(4 Pt 1), 1005–1026. https://doi.org/10.1111/j.1475-6773.2004.00269.x.

Hibbard, J. H., Peters, E., Dixon, A., & Tusler, M. (2007). Consumer competencies and the use of comparative quality information: It isn't just about literacy. *Medical Care Research and Review, 64*(4), 379–394.

Hoffman, C., Rice, D., & Sung, H.-Y. (1996). Persons with chronic conditions: Their prevalence and costs. *JAMA, 276*(18), 1473–1479.

Holman, H., & Lorig, K. (2004). Patient self-management: A key to effectiveness and efficiency in care of chronic disease. *Public Health Reports, 119*(3), 239–243.

Huerta, T. R., McAlearney, A. S., & Rizer, M. K. (2017). Introducing a patient portal and electronic tablets to inpatient care. *Annals of Internal Medicine, 167*(11), 816–817. https://doi.org/10.7326/m17-1766.

Institute of Medicine (US) Committee on Quality of Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press, ©2001.

Isaacs, J. N., & Creinin, M. D. (2003). Miscommunication between healthcare providers and patients may result in unplanned pregnancies. *Contraception, 68*(5), 373–376. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14636942

Jimison, H., Pavel, M., Larimer, N., & Mullen, P. (2007). *A general architecture for computer based health coaching*. Paper presented at the international conference on technology & aging.

Jimison, H., Gorman, P., Woods, S., Nygren, P., Walker, M., Norris, S., & Hersh, W. (2008). Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. *Evidence Report Technology Assessment (Full Report), 175*, 1–1422.

Jimison, H. B., McKanna, J., Ambert, K., Hagler, S., Hatt, W. J., & Pavel, M. (2010). *Models of cognitive performance based on home monitoring data.* Paper presented at the Engineering in Medicine and Biology Society (EMBC), 2010 annual international conference of the IEEE.

Jimison, H. B., Hagler, S., Kurillo, G., Bajcsy, R., & Pavel, M. (2015). *Remote health coaching for interactive exercise with older adults in a home environment.* Paper presented at the Engineering in Medicine and Biology Society (EMBC), 2015 37th annual international conference of the IEEE.

Johnson, K. B., Ravert, R. D., & Everton, A. (2001). Hopkins Teen Central: Assessment of an internet-based support system for children with cystic fibrosis. Pediatrics, 107(2), e24-e24.

Juengst, E. T., Flatt, M. A., & Settersten, R. A., Jr. (2012). Personalized genomic medicine and the rhetoric of empowerment. *The Hastings Center Report, 42*(5), 34–40. https://doi.org/10.1002/hast.65.

Kawano, K., Suzuki, M., & Araki, K. (2012). Use of digital patient photographs and electronic medical record data as diagnostic tools in Japan. *Journal of Medical Systems, 36*(5), 3321–3326. https://doi.org/10.1007/s10916-012-9824-4.

Kellerman, R., & Kirk, L. (2007). Principles of the patient-centered medical home. *American Family Physician, 76*(6), 774.

Keselman, A., Logan, R., Smith, C. A., Leroy, G., & Zeng-Treitler, Q. (2008). Developing informatics tools and strategies for consumer-centered health communication. *Journal of the American Medical Informatics Association, 15*(4), 473–483.

Kim, M. I., & Johnson, K. B. (2002). Personal health records: Evaluation of functionality and utility. *Journal of the American Medical Informatics Association, 9*(2), 171–180. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11861632.

Kleinman, A. (1988). *The illness narratives: Suffering, healing, and the human condition*. New York: Basic Books.

Kontos, E., Blake, K. D., Chou, W.-Y. S., & Prestin, A. (2014). Predictors of eHealth usage: Insights on the digital divide from the Health Information National Trends Survey 2012. *Journal of Medical Internet Research, 16*(7), e172.

Korsch, B. M. (1989). Current issues in communication research. *Health Communication, 1*(1), 5–9.

Lieberman, M. A. (1988). The role of self-help groups in helping patients and families cope with cancer. *CA: a Cancer Journal for Clinicians, 38*(3), 162–168.

Lorig, K. R., Ritter, P., Stewart, A. L., Sobel, D. S., Brown, B. W., Jr., Bandura, A., et al. (2001). Chronic disease self-management program: 2-year health status and health care utilization outcomes. *Medical Care*, 1217–1223.

Lye, C. T., Forman, H. P., Daniel, J. G., & Krumholz, H. M. (2018). The 21st Century Cures Act and electronic health records one year later: Will patients see the benefits? *Journal of the American Medical Informatics Association, 25*(9), 1218–1220. https://doi.org/10.1093/jamia/ocy065.

Marcolino, M. S., Oliveira, J. A. Q., D'Agostino, M., Ribeiro, A. L., Alkmim, M. B. M., & Novillo-Ortiz, D. (2018). The impact of mHealth interventions: Systematic review of systematic reviews. *JMIR mHealth and uHealth, 6*(1), e23. https://doi.org/10.2196/mhealth.8873.

Masys, D. R., & Baker, D. B. (1997). Patient-Centered Access to Secure Systems Online (PCASSO): A secure approach to clinical data access via the World Wide Web. In *Proceedings: a conference of the American Medical Informatics Association/ ... AMIA annual fall symposium. AMIA fall symposium* (pp. 340–343). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9357644

Matthew-Maich, N., Harris, L., Ploeg, J., Markle-Reid, M., Valaitis, R., Ibrahim, S., et al. (2016). Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: A scoping review. *JMIR mHealth and uHealth, 4*(2), e29. https://doi.org/10.2196/mhealth.5127.

McCormack, L., Haun, J., Sorensen, K., & Valerio, M. (2013). Recommendations for advancing health literacy measurement. *Journal of Health Communication, 18*(Suppl 1), 9–14. https://doi.org/10.1080/10810730.2013.829892.

Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research, 15*(4), e85. https://doi.org/10.2196/jmir.1933.

Morgan, S. (2013). Miscommunication between patients and general practitioners: Implications for clinical practice. *Journal of Primary Health Care, 5*(2), 123–128. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23748393

Mossberger, K., Tolbert, C. J., & Gilbert, M. (2006). Race, place, and information technology. *Urban Affairs Review, 41*(5), 583–620.

Nazi, K. M., & Woods, S. S. (2008). MyHealtheVet PHR: A description of users and patient portal use. In *AMIA annual symposium proceedings* (p. 1182). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18999142

Neter, E., & Brainin, E. (2012). eHealth literacy: Extending the digital divide to the realm of health information. *Journal of Medical Internet Research, 14*(1), e19.

Newnham, H., Barker, A., Ritchie, E., Hitchcock, K., Gibbs, H., & Holton, S. (2017). Discharge communication practices and healthcare provider and patient preferences, satisfaction and comprehension: A systematic review. *International Journal for Quality in Health Care, 29*(6), 752–768. https://doi.org/10.1093/intqhc/mzx121.

O'Leary, K. J., Lohman, M. E., Culver, E., Killarney, A., Randy Smith, G., Jr., & Liebovitz, D. M. (2016). The effect of tablet computers with a mobile patient portal application on hospitalized patients' knowledge and activation. *Journal of the American Medical Informatics Association, 23*(1), 159–165. https://doi.org/10.1093/jamia/ocv058.

Obdrzalek, S., Kurillo, G., Ofli, F., Bajcsy, R., Seto, E., Jimison, H., & Pavel, M. (2012). *Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population.* Paper presented at the Engineering in Medicine and Biology Society (EMBC), 2012 annual international conference of the IEEE.

Ofli, F., Kurillo, G., Obdržálek, Š., Bajcsy, R., Jimison, H., & Pavel, M. (2016). Design and evaluation of an interactive exercise coaching system for older adults: Lessons learned. *IEEE Journal of Biomedical and Health Informatics, 20*(1), 201.

Overdijkink, S. B., Velu, A. V., Rosman, A. N., van Beukering, M. D., Kok, M., & Steegers-Theunissen, R. P. (2018). The usability and effectiveness of mobile health technology-based lifestyle and medical intervention apps supporting health care during pregnancy: Systematic review. *JMIR mHealth and uHealth, 6*(4), e109. https://doi.org/10.2196/mhealth.8834.

Patel, A., Schieble, T., Davidson, M., Tran, M. C., Schoenberg, C., Delphin, E., & Bennett, H. (2006). Distraction with a hand-held video game reduces pediatric preoperative anxiety. *Pediatric Anesthesia, 16*(10), 1019–1027.

Pavel, M., Jimison, H. B., Korhonen, I., Gordon, C. M., & Saranummi, N. (2015). Behavioral informatics and computational modeling in support of proactive health management and care. *IEEE Transactions on Biomedical Engineering, 62*(12), 2763–2775.

Pellegrini, C. A., Conroy, D. E., Phillips, S. M., Pfammatter, A. F., McFadden, H. G., & Spring, B. (2018). Daily and seasonal influences on dietary self-monitoring using a smartphone application. *Journal of Nutrition Education and Behavior, 50*(1), 56–61 e51. https://doi.org/10.1016/j.jneb.2016.12.004.

Prey, J. E., Woollen, J., Wilcox, L., Sackeim, A. D., Hripcsak, G., Bakken, S., et al. (2014). Patient engagement in the inpatient setting: A systematic review. *Journal of the American Medical Informatics Association, 21*(4), 742–750. https://doi.org/10.1136/amiajnl-2013-002141.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data, 1*(1), 51–59. https://doi.org/10.1089/big.2013.1508.

Redd, W. H., Jacobsen, P. B., Die-Trill, M., Dermatis, H., McEvoy, M., & Holland, J. C. (1987). Cognitive/attentional distraction in the control of conditioned nausea in pediatric cancer patients receiving chemotherapy. *Journal of Consulting and Clinical Psychology, 55*(3), 391.

Renders, C. M., Valk, G. D., Griffin, S. J., Wagner, E. H., & Assendelft, W. J. (2001). Interventions to improve the management of diabetes in primary care, outpatient, and community settings: A systematic review. *Diabetes Care, 24*(10), 1821–1833.

Roter, D., & Hall, J. A. (2006). *Doctors talking with patients/patients talking with doctors: Improving communication in medical visits* (Vol. 2). Westport: Praeger.

Serrato, C. A., Retecki, S., & Schmidt, D. E. (2007). MyChart-A new mode of care delivery: 2005 personal health link research report. *The Permanente Journal, 11*(2), 14–20. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21461088

Slack, W. V., Hicks, P., Reed, C. E., & Van Cura, L. J. (1966). A computer-based medical-history system. *New England Journal of Medicine, 274*(4), 194–198.

Staccini, P., Lau, A. Y. S., & Section Editors for the IMIA Yearbook Section on Education and Consumer Health Informatics. (2018). Findings from 2017 on consumer health informatics and education: Health data access and sharing. *Yearbook of Medical Informatics, 27*(1), 163–169. https://doi.org/10.1055/s-0038-1641218.

Steinberg, D. M., Tate, D. F., Bennett, G. G., Ennett, S., Samuel-Hodge, C., & Ward, D. S. (2013). The efficacy of a daily self-weighing weight loss intervention using smart scales and e-mail. *Obesity (Silver Spring), 21*(9), 1789–1797. https://doi.org/10.1002/oby.20396.

Swendeman, D., Comulada, W. S., Ramanathan, N., Lazar, M., & Estrin, D. (2015). Reliability and validity of daily self-monitoring by smartphone application for health-related quality-of-life, antiretroviral adherence, substance use, and sexual behaviors among people living with HIV. *AIDS and Behavior, 19*(2), 330–340. https://doi.org/10.1007/s10461-014-0923-8.

Tang, P. C., Ash, J. S., Bates, D. W., Overhage, J. M., & Sands, D. Z. (2006). Personal health records: Definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics Association, 13*(2), 121–126. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16357345

Topol, E. J. (2015). *The patient will see you now : The future of medicine is in your hands*. New York: Basic Books.

Wagner, E. H. (2001). Meeting the needs of chronically ill people: Socioeconomic factors, disabilities, and comorbid conditions are obstacles. *British Medical Journal. British Medical Journal Publishing Group, 323*(7319), 945–946.

Wagner, E. H., Austin, B. T., Davis, C., Hindmarsh, M., Schaefer, J., & Bonomi, A. (2001). Improving chronic illness care: Translating evidence into action. *Health Affairs, 20*(6), 64–78.

Waitzkin, H., & Stoeckle, J. D. (1972). The communication of information about illness. In *Psychosocial aspects of physical illness* (Vol. 8, pp. 180–215). Basel: Karger Publishers.

Ware, J. E., Jr., Snyder, M. K., Wright, W. R., & Davies, A. R. (1983). Defining and measuring patient satisfaction with medical care. *Evaluation and Program Planning, 6*(3–4), 247–263.

Weaver, R. R. (2015). Reconciling evidence-based medicine and patient-centred care: Defining evidence-based inputs to patient-centred decisions. *Journal of Evaluation in Clinical Practice, 21*(6), 1076–1080. https://doi.org/10.1111/jep.12465.

Weingart, S., Rind, D., Tofias, Z., & Sands, D. (2006). Who uses the patient internet portal? *Journal of the American Medical Informatics Association, 13*(1), 91–95.

Weitzman, E. R., Adida, B., Kelemen, S., & Mandl, K. D. (2011a). Sharing data for public health research by members of an international online diabetes social network. *PloS One, 6*(4), e19256.

Weitzman, E. R., Kelemen, S., & Mandl, K. D. (2011b). Surveillance of an online social network to assess population-level diabetes health status and healthcare quality. *Online Journal of Public Health Informatics, 3*(3). https://doi.org/10.5210/ojphi.v3i3.3797.

Wennberg, J., & Gittelsohn, A. (1973). Small area variations in health care delivery: A population-based health information system can guide planning and regulatory decision-making. *Science, 182*(4117), 1102–1108.

Wensheng, W. (2002). *Bridging the digital divide inside China.* Paper presented at the 3rd annual conference of the Association of Internet Researchers, October 13–16, Maastricht, NL.

Whitlock, W. L., Brown, A., Moore, K., Pavliscsak, H., Dingbaum, A., Lacefield, D., et al. (2000). Telemedicine improved diabetic management. *Military Medicine, 165*(8), 579–584.

Williamson, S. S. (2015). *The user-centered design of a mobile app for distance caregivers of older adults that live in smart home environments*.

Zheng, M., Jin, H., Shi, N., Duan, C., Wang, D., Yu, X., & Li, X. (2018). The relationship between health literacy and quality of life: A systematic review and meta-analysis. *Health and Quality of Life Outcomes, 16*(1), 201. https://doi.org/10.1186/s12955-018-1031-7.

Ethics in Biomedical and Health Informatics: Users, Standards, and Outcomes

Kenneth W. Goodman and Randolph A. Miller

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_12

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- Why is ethics important to informatics?
- What are the leading ethical issues that arise in health care informatics?
- What are examples of appropriate and inappropriate uses and users for health-related software?
- Why does the establishment of standards touch on ethical issues?
- Why does system evaluation involve ethical issues?
- What challenges does informatics pose for patient and provider confidentiality?
- How can the tension between the obligation to protect confidentiality and that to share data be minimized?
- How might computational health care alter the traditional provider–patient relationship?
- What ethical issues arise at the intersection of informatics and managed care?
- What are the leading ethical and legal issues in the debate over governmental regulation of health care computing tools?

12.1 Ethical Issues in Biomedical and Health Informatics

» More and more the tendency is towards the use of mechanical aids to diagnosis; nevertheless, the five senses of the doctor do still, and must always, play the preponderating part in the examination of the sick patient. Careful observation can never be replaced by the tests of the laboratory. The good physician now or in the future will never be a diagnostic robot. – Scottish surgeon Sir William Arbuthnot-Lane (Lane 1936)

Human values should govern research and practice in the health professions. Health informatics, like other health professions, encompasses issues of appropriate and inappropriate behavior, of honorable and disreputable actions and intentions, of right and wrong. Students and practitioners of the health sciences, including informatics, share an important obligation to explore the moral underpinnings, ethical challenges and social issues related to their research and practice.

Ethical questions in medicine, nursing, human subjects research, psychology, social work, and affiliated fields continue to evolve and increase in number; nevertheless, the key issues are generally well known. Major questions in general bioethics have long been addressed in numerous professional, scholarly, and educational contexts. Ethical issues in health informatics are, for the most part, less familiar, even though certain of them have received attention for decades (Szolovits and Pauker 1979; Miller et al. 1985a; de Dombal 1987). Indeed, informatics now constitutes a source of some of the most important and interesting ethical debates in all the health professions. It has even been suggested that biomedical informatics raises so many such issues it could itself be used as the basis for a bioethics curriculum (Goodman 2017).

People often assume that the confidentiality of electronically stored patient information is the most important ethical issue in informatics. Although confidentiality and privacy are indeed of vital interest and significant concern, the field is rich with other ethical issues, including the appropriate selection and use of informatics tools in clinical settings; the determination of who should use such tools; the role of system evaluation; the obligations of system developers, maintainers, and vendors; the appropriate standards for interacting with industry; and the use of computers to track clinical outcomes to guide future practice. In addition, informatics engenders many important legal and regulatory questions.

To consider ethical issues in health care informatics is to explore a significant intersection among several professions—health care informatics per se, health care delivery and administration, applied computing and systems engineering, and ethics—each of which constitutes a vast field of inquiry. Fortunately, growing interest in bioethics and computation-related ethics has produced a starting point for such exploration. An initial ensemble of guiding principles, or ethical criteria, has emerged to orient decision making in health care informatics. These criteria are of practi-

cal utility to health informatics, and often have broader implications for all of biomedical informatics.

12.2 Health-Informatics Applications: Appropriate Use, Users, and Contexts

Application of computer-based technologies in the health professions can build on previous experience in adopting other devices, tools, and methods. Before clinicians perform most health-related interventions (e.g., diagnostic testing, prescription of medications, surgical and other therapeutic procedures), they generally evaluate appropriate evidence, standards, available technologies, presuppositions, and values. Indeed, the very evolution of the health professions entails the evolution of evidence, of standards, of available technologies, of presuppositions, and of values.

To answer the clinical question, "What should be done in this case?" one must pay attention to a number of subsidiary questions, such as:

1. What is the problem?
2. What resources are available and what am I competent to do?
3. What will maintain or improve this patient's care?
4. What will otherwise produce the most desirable results (e.g., in public health)?
5. How strong are my beliefs in the accuracy of my answers to questions 1 through 4?

Similar considerations determine the appropriate use of informatics tools.

12.2.1 The Standard View of Appropriate Use

Excitement and enthusiasm often accompany initial use of new tools in clinical settings. Negative emotions are also common (Sittig et al. 2005). Based on the uncertainties that surround any new technology, scientific evidence counsels caution and prudence. As in other clinical areas, evidence and reason determine the appropriate level of caution. For instance, there is considerable evidence that electronic laboratory information systems improve access to clinical data when compared with manual, paper-based test-result distribution methods. To the extent that such systems improve care at an acceptable cost in time and money, there is an obligation to use computers to store and retrieve clinical laboratory results. There is a small but growing body of evidence that existing **clinical expert systems** can improve patient care in a small number of practice environments at an acceptable cost in time and money (Kuperman and Gibson 2003). Nevertheless, such systems cannot yet uniformly improve care in typical, more general practice settings, at least not without careful attention to the full range of managerial as well as technical issues affecting the particular care delivery setting in which they are used (Kaplan and Harris-Salamone 2009; Holroyd-Leduc et al. 2011; Shih et al. 2011).

Clinical expert systems (see ▶ Chap. 24) attempt to provide decision support for diagnosis, therapy, and/or prognosis in a more detailed and sophisticated manner than do simple reminder systems (Duda and Shortliffe 1983). A necessary adjunction of expert systems – creation and maintenance of their related knowledge bases – still involves leading-edge research and development. Humans for the most part remain superior to electronic systems in understanding patients and their problems, in efficiently interacting with patients to ascertain pertinent past history and current symptoms across the spectrum of clinical practice, in the interpretation and representation of data, and in clinical synthesis. Humans might in the future however not hold the upper hand in these tasks, and claims of their superiority must continually be tested empirically (Blois 1980).

What has been called the "standard view" of computer-assisted clinical diagnosis (Miller 1990a; cf. Friedman 2009) holds in part that human cognitive processes, being more suited to the complex task of diagnosis than machine intelligence, should not be overridden or trumped by computers. The standard view states that when adequate (and even exem-

plary) decision-support tools are developed, they should be viewed and used as supplementary and subservient to human clinical judgment: They *support* decisions by human beings; they do not *make* decisions. Progress should be measured in terms of whether clinicians using a CDS tool perform better on specific tasks than the same clinicians without the tool (Miller 1990a; cf. Friedman 2009). These tools should assume subservient roles because the clinician caring for the patient knows and understands the patient's situation and can make compassionate judgments better than computer programs. Furthermore, clinicians, and not machine algorithms, are the entities which the state licenses, and specialty boards accredit, to practice medicine, surgery, nursing, pharmacy, and other health-related activities.

Corollaries of the standard view are that (1) practitioners have an obligation to use any computer-based tool responsibly, through adequate user training and by developing an understanding of the system's abilities and limitations; and (2) practitioners must not abrogate their clinical judgment reflexively when using computer-based decision aids.

The skills required for diagnosis are in many respects different from those required for the acquisition, storage, and retrieval of laboratory data. There is no contradiction in urging extensive use of efficient, non-burdensome laboratory information systems, and, for the time being, cautious deployment of expert diagnostic decision-support tools (i.e., not permitting their use in settings in which knowledgeable clinicians cannot immediately override faulty advice). Nevertheless, U.S. policy under the HITECH act of 2009 (as discussed in ▶ Chap. 29), led to widespread adoption of less-than ideal electronic health record systems. In many settings, those systems engendered less efficient, burdensome, error-prone care delivery and physician burnout (cf. Halamka and Tripathi 2017).

More over, the standard view addresses a key aspect of the question, "How and when should computers be used in clinical practice?" by capturing important moral intuitions about error avoidance and evolving standards. Error avoidance and the benefits that follow from it shape the obligations of practitioners. In computer-software use, as in all other areas of clinical practice, good intentions alone are insufficient to insulate recklessness from culpability. Thus, the standard view may be seen as a tool for both error avoidance and ethically optimized action.

Ethical software use, then, should be evaluated against a broad background of evidence for actions that produce favorable outcomes. Because informatics is a science in ongoing ferment, system improvements and evidence of such improvements are constantly emerging. Clinicians have an obligation to be familiar with this evidence after attaining minimal acceptable levels of familiarity with informatics in general and with the clinical systems they use in particular (◘ Fig. 12.1).

12.2.2 Appropriate Users and Educational Standards

Efficient and effective use of health care informatics systems requires prior system evaluations demonstrating utility, education and training of new users, monitoring of experience, and appropriate, timely updating. Indeed, such requirements resemble those for other tools used in health care and in other domains. Inadequate preparation in the use of tools is an invitation to catastrophe. When the stakes are high and the domain large and complex—as is the case in the health professions—education and training take on moral significance.

Who should use a health care-related computer application? Consider expert decision-support systems as an example. An early paper on ethical issues in informatics noted that potential users of such systems include physicians, nurses, physicians' assistants, paramedical personnel, students of the health sciences, patients, and insurance and government evaluators (Miller et al. 1985a). Are members of all these groups appropriate users? One cannot answer the question until one precisely specifies the intended use for the system (i.e., the particular clinical questions the system will address). The appropriate level

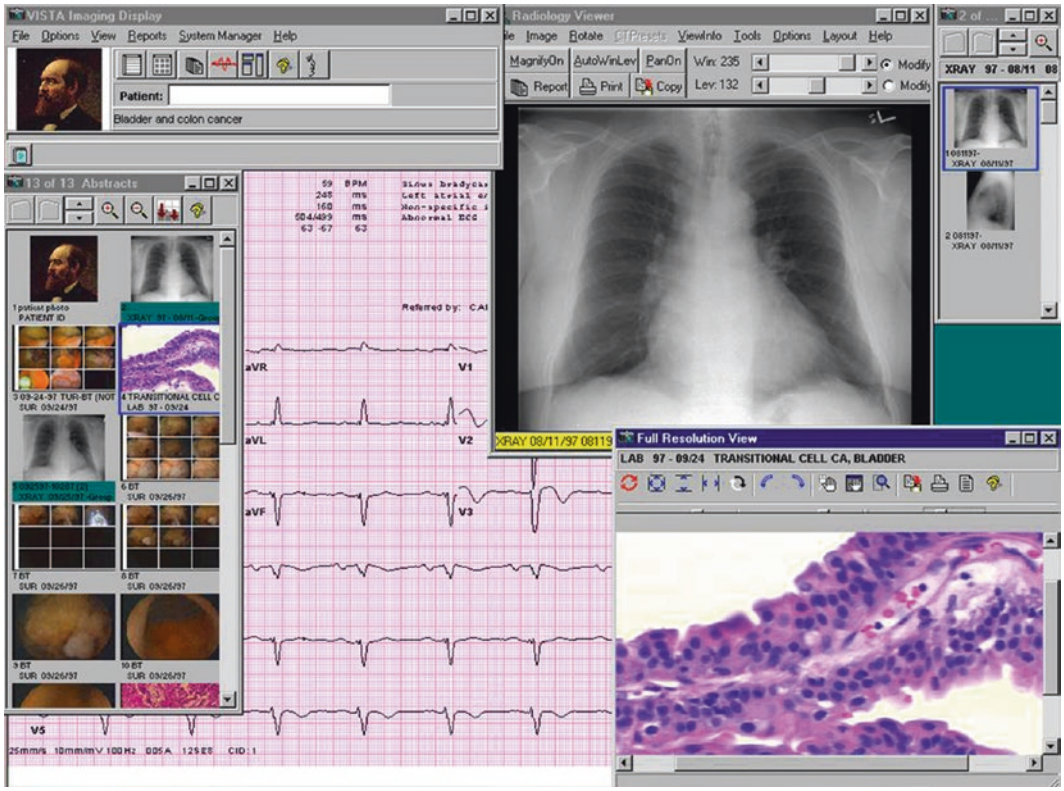


Fig. 12.1 The U.S. Department of Veterans Affairs in the 1970s developed the highly regarded "Veterans Health Information Systems and Technology Architecture" (VistA), once the largest electronic health record system in the United States. This fictitious screen shot demonstrates some of the system's functions and utilities. (Credit: Courtesy of U.S. Department of Veterans Affairs, Veterans Health Administration Office of Informatics and Analytics)

12

of training must be correlated with the question at hand. At one end of an appropriate-use spectrum, we can posit that medical and nursing students should employ decision-support systems for educational purposes; this assertion is relatively free of controversy once it has been verified that such tools convey accurately a sufficient quantity and quality of educational content. But it is less clear that patients, administrators, or insurance company gatekeepers, for example, should use expert decision-support systems for assistance in making diagnoses, in selecting therapies, or in evaluating the appropriateness of health professionals' actions or determining their reimbursement. To the extent that some systems present general medical advice in generally understandable but sufficiently nuanced formats, such as once was the case with Dr. Benjamin Spock's 1950s era print-based child-care primer, one might condone system use by laypersons. There are additional legal concerns related to negligence and product liability, however, when health-related products are sold directly to patients rather than to licensed practitioners, and when such products give patient-specific counsel rather than general clinical advice (Miller et al. 1985a).

Suitable use of a software program that helps a user to suggest diagnoses, to select therapies, or to render prognoses must be plotted against an array of goals and best practices for achieving those goals, including consideration of the characteristics and requirements of individual patients. For example, the multiple, interconnected inferential strategies required for arriving at an accurate diagnosis depend on knowledge of facts;

experience with procedures; and familiarity with human behavior, motivation, and values. **Diagnosis** is a process rather than an event (Miller 1990a), so even well-validated diagnostic systems must be used appropriately in the overall context of patient care.

To use a **diagnostic decision-support system** (► Chap. 24), a clinician must be able to recognize when the computer program has erred, and, when it is accurate, what the output means and how it should be interpreted. This ability requires knowledge of both the diagnostic sciences and the software applications, and the strengths and limitations of each. After assigning a diagnostic label, the clinician must communicate the diagnosis, prognosis, and implications to a patient, and must do so in ways both appropriate to the patient's educational background and conducive to future treatment goals. It is not enough to be able to tell patients that they have cancer, human immunodeficiency virus (HIV), diabetes, or heart disease and then simply hand over a prescription. The care provider must also offer context when available, comfort when needed, and hope as appropriate. For instance, the reason many organizations have required counseling both before and after HIV and genetic testing is not to vex busy health professionals but to ensure that comprehensive, high-quality care, rather than mere diagnostic labeling, is delivered.

This discussion points to the following set of ethical principles for appropriate use of decision-support systems:

1. A computer program should be used in clinical practice only after appropriate evaluation of its efficacy and the documentation that it performs its intended task at an acceptable cost in time and money.
2. Users of most clinical systems should be health professionals who are qualified to address the question at hand on the basis of their licensure, clinical training, and experience. Software systems should be used to augment or supplement, rather than to replace or supplant, such individuals' decision making.
3. All uses of informatics tools, especially inpatient care, should be preceded by adequate training and instruction, which should include review of applicable product evaluations.

Such principles and claims should be viewed as analogous to other standards or rules in clinical medicine and nursing.

12.2.3 Obligations and Standards for System Developers and Maintainers

Users of clinical programs must rely on the work of other people who are often far removed from the context of use. As with all complex technologies, users depend on the developers and maintainers of a system and must trust evaluators who have validated a system for clinical use. Health care software applications are among the most complex tools in the technological armamentarium. Although this complexity imposes certain obligations on end users, it also commits a system's developers, designers, and maintainers to adhere to reasonable standards and, indeed, to acknowledge their moral responsibility for doing so.

12.2.3.1 Ethics, Standards, and Scientific Progress

The very idea of a **standard of care** embodies a number of complex assumptions linking ethics, evidence, outcomes, and professional training. To say that nurses or physicians must adhere to a standard is to say, in part, that they ought not stray from procedures previously shown or generally believed to work better than alternatives. The difficulty lies in how to determine if a procedure or device "works better" than another. Such determinations in the health sciences constitute progress, and provide evidence that we now know more. Criteria for weighing such evidence, albeit short of proof in most cases, are applied. For example, evidence from well-designed randomized controlled trials merits greater trust than evidence derived from uncontrolled ret-

rospective studies (see ▶ Chap. 13). Typically, verification by independent investigators must occur before placing the most recent study results into common practice.

People who develop, maintain, and sell health care computing systems and their components have obligations that parallel those of system users. These obligations include holding patient care as the foremost value. The duty to limit or prevent harm to patients applies to system developers as well as to practitioners. Although this principle is easy to suggest and, generally, to defend, it invites subtle, and sometimes overt, resistance from people for whom profit or fame are primary motivators. (This is of course also true for other medical devices, processes and industries.) To be sure, quests for fame and fortune often produce good outcomes and improved care, at least eventually. Even so, some approaches fail to take into account the role of intention as a moral criterion (cf. Goodman et al. 2010).

In medicine, nursing, and psychology, a number of models of the **professional–patient relationship** place trust and advocacy at the apex of a hierarchy of values. Such a stance cannot be maintained if goals and intentions other than patient well-being are (generally) assigned primacy. The same principles apply to those who produce and attend to health care information systems. Because these systems are health care systems—and are not devices for accounting, entertainment, real estate, and so on—and because system underperformance can cause pain, disability, illness, and death, it is essential that the threads of trust run throughout the fabric of clinical system design and maintenance.

System purchasers, users, and patients must rely upon developers and maintainers to recognize the potentially grave consequences of errors or carelessness, trust them to care about the uses to which the systems will be put, and rely upon them to value the reduced suffering of other people at least as much as they value their own personal gain. This reliance emphatically does not entail that system designers and maintainers are blameworthy or unethical if they hope and strive to profit from their diligence, creativity, and effort. Rather, it implies that no amount of financial benefit for a designer or builder can counterbalance bad outcomes or ill consequences that result from recklessness, avarice, or inattention to the needs of clinicians and their patients. Purchasers and users should require demonstrations that systems are worthy of such trust and reliance before placing patients at risk, and that safeguards (human and mechanical) are in place to detect, alert, and rectify situations in which systems underperform.

Quality standards should stimulate scientific progress and innovation while safeguarding against system error and abuse. These goals might seem incompatible, but they are not. Let us postulate a standard that requires timely updating and testing of knowledge bases that are used by decision-support systems. To the extent that database accuracy is needed to maximize the accuracy of inferential engines, it is trivially clear how such a standard will help to prevent or reduce decision-support mistakes. Furthermore, the standard should be seen to foster progress and innovation in the same way that any insistence on best possible accuracy helps to protect scientists and clinicians from pursuing false leads, or wasting time in testing poorly wrought hypotheses. It will not do for database maintainers to insist that they are busy doing the more productive or scientifically stimulating work of improving knowledge representation, say, or database design. Although such tasks are important, they do not supplant the tasks of updating and testing tools in their current configuration or structure. Put differently, scientific and technical standards are perfectly able to stimulate progress while taking a cautious or even conservative stance toward permissible risk in patient care.

This approach has been described as **progressive caution**: "Medical informatics is, happily, here to stay, but users and society have extensive responsibilities to ensure that we use our tools appropriately. This might cause us to move more deliberately or slowly than some would like" (Goodman 1998).

A more recent concern, with both ethical and legal implications, is the responsibility of software developers to design and implement

software programs that cannot easily be hacked by malicious code writers. This concern goes beyond privacy and confidentiality issues (discussed below), and includes the possibility that medical devices with embedded software might be nefariously "reprogrammed" in a manner that might cause harm to patients (see, for example, Pugh et al. 2018 and Sackner-Bernstein 2017). A more detailed discussion of this topic appears under the ▶ Sect. 12.5 below.

12.2.3.2 System Evaluation as an Ethical Imperative

Any move toward "best practices" in biomedical informatics is shallow and feckless if it does not include a way to measure whether a system performs as intended. This and related measurements provide the ground for quality control and, as such, are the obligations of system developers, maintainers, users, administrators, and perhaps other players (see ▶ Chap. 13).

» Medical computing is not merely about medicine or computing. It is about the introduction of new tools into environments with established social norms and practices. The effects of computing systems in health care are subject to analysis not only of accuracy and performance but of acceptance by users, of consequences for social and professional interaction, and of the context of use. We suggest that system evaluation can illuminate social and ethical issues in medical computing, and in so doing improve patient care. That being the case, there is an ethical imperative for such evaluation (Anderson and Aydin 1998).

To give a flavor of how a comprehensive evaluation program can ethically optimize implementation and use of an informatics system, consider these ten criteria for system scrutiny (Anderson and Aydin 1994):

1. Does the system work as designed?
2. Is it used as anticipated?
3. Does it produce the desired results?
4. Does it work better than the procedures it replaced?
5. Is it cost effective?
6. How well have individuals been trained to use it?
7. What are the anticipated long-term effects on how organizational units interact?
8. What are the long-term effects on the delivery of medical care?
9. Will the system have an impact on control in the organization?
10. To what extent do effects depend on practice setting?

Another way to make this important point is by emphasizing that *people* use computer systems. Even the finest system might be misused, misunderstood, or mistakenly allowed to alter or erode previously productive human relationships. Evaluation of health information systems in their contexts of use should be taken as a moral imperative. Such evaluations require consideration of a broader conceptualization of "what works best" and must look toward improving the overall health care delivery system rather than only that system's technologically based components. These higher goals entail the creation of a corresponding mechanism for ensuring institutional oversight and responsibility (Miller and Gardner 1997a, b).

12.3 Privacy, Confidentiality, and Data Sharing

Some of the greatest challenges of the Information Age arise from placing computer applications in health care settings while upholding traditional principles and values. One challenge involves balancing two competing values: (1) free access to information, and (2) protection of patients' privacy and confidentiality.

Only computers can efficiently manage the now-vast amount of information generated during clinical encounters and other health care transactions (see ▶ Chap. 2); at least in principle, such information should be easily available to health professionals and others involved in the administration of the care-delivery system, so that they can provide effective, efficient care for patients. Yet, making this information readily available creates

greater opportunities for inappropriate access. Such access may be available to curious health care workers who do not need the information to fulfill job-related responsibilities, and, even more worrisome, to other people who might use the information to harm patients physically, emotionally, or financially. Clinical system administrators must balance the goals of protecting confidentiality by restricting use of computer systems and improving care by assuring the integrity and availability of data. These objectives are not incompatible, but there are trade-offs that cannot be avoided.

12.3.1 Foundations of Health Privacy and Confidentiality

Privacy and confidentiality are necessary for people to evolve and mature as individuals, to form relationships, and to serve as functioning members of society. Imagine what would happen if the local newspaper or gossip blog produced a daily report detailing everyone's actions, meetings, and conversations. It is not that most people have terrible secrets to hide but rather that the concepts of solitude, intimacy, and the desire to be left alone make no sense without the expectation that at least some of our actions and utterances will be kept private or held in confidence among a limited set of persons.

The "average" sentiment about the appropriate sphere of private vs. public may vary considerably from culture to culture, and even from generation to generation within any particular culture; and it may differ widely among persons within a culture or generation, and evolve for any particular person over a lifetime. Even the "born digital" generation, for which social media are a fixture of everyday life, has – and ought to have –its boundaries (Palfrey and Gasser 2010).

The terms *privacy* and *confidentiality* are not synonymous. As commonly used, "privacy" generally applies to people, including their desire not to suffer eavesdropping, whereas "confidentiality" is best applied to information. One way to think of the difference is as follows. If someone follows you and spies on you entering an AIDS clinic, your privacy is violated; if someone sneaks into the clinic without observing you in person and looks at your health care record, your record's confidentiality is breached. In discussions of the electronic health record, the term privacy may also refer to individuals' desire to restrict the disclosure of personal data (National Research Council 1997).

There are several important reasons to protect privacy and confidentiality. One is that privacy and confidentiality are widely regarded as rights of all people, and such protections help to accord them respect. On this account, people do not need to provide a justification for limiting access to their identifiable health data; privacy and confidentiality are entitlements that a person does not need to earn, to argue for, or to defend. Another reason is more practical: protecting privacy and confidentiality benefits both individuals and society. Patients who know that their identifiable health care information will not be shared inappropriately are more comfortable disclosing that information to clinicians. This trust is vital for the successful physician–patient, nurse–patient, or psychologist-patient relationship, and it helps practitioners to do their jobs. This insight is as old as the Hippocratic corpus.

Privacy and confidentiality protections also benefit public health. People who fear disclosure of personal information are less likely to seek out professional assistance, increasing the risks that contagion will be spread and maladies will go untreated. In addition, people still suffer discrimination, bias, and stigma when certain health data do fall into the wrong hands. Financial harm may occur if insurers are given unlimited access to family members' records, or access to patient data, because some insurers might be tempted to increase the price of insurance for individuals at higher risk of illness or discriminate in other ways if such price differentiation were forbidden by law. This is, in the United States, among the reasons the Patient Protection and Affordable Care Act of 2010 (U.S. Public Law 111-148), in prohibiting insurers from discrimination based on "pre-existing conditions," was so important – and why subsequent efforts on ideological grounds to overturn the act are dangerously erosive.

The ancient idea that physicians should hold health care information in confidence is therefore applicable whether the data are written on paper or processed in silicon. The obligations to protect privacy and to keep confidences fall to system designers and maintainers, to administrators, and, ultimately, to the physicians, nurses, and others who elicit the information in the first place. The upshot for all of them is this: protection of privacy and confidentiality is not an option, a favor, or a helping hand offered to patients with embarrassing health problems; it is a duty, regardless of the malady or the medium in which information about it is stored.

Some sound clinical practice and public health traditions run counter to the idea of absolute confidentiality. When a patient is hospitalized, it is expected that all appropriate (and no inappropriate) employees or affiliates of the institution—primary physicians, consultants, nurses, therapists, and technicians—will have access to the patient's medical records, when it is in the interest of the patient's care to do so. In most communities of the United States, the contacts of patients who have active tuberculosis or certain sexually transmitted diseases are routinely identified and contacted by public health officials so that the contacts may receive proper medical attention. Such disclosures serve the public interest and are and should be legal because they decrease the likelihood that more widespread harm to other individuals might occur through transmission of an infection unknowingly.

A separate but important public health consideration (discussed in more detail below) involves the ability of health care researchers to anonymously pool data (i.e., pool by removing individual persons' identifying information) from patient cases that meet specified conditions to determine the natural history of the disease and the effects of various treatments. Examples of benefits from such pooled data analyses range from the ongoing results generated by regional collaborative chemotherapy trials to the discovery, more than four decades ago, of the appropriateness of shorter lengths of stay for patients with myocardial infarction (McNeer et al. 1975). More recently, the need for robust **syndromic surveillance** has been asserted as necessary for adequate bioterrorism preparedness, for earlier detection of naturally occurring disease outbreaks, and, most dramatically, in the Coronavirus pandemic of 2020 (Ienca and Vayena 2020; see also ▶ Chap. 18).

12.3.2 Electronic Clinical and Research Data

Access to electronic patient records holds extraordinary promise for clinicians and for other people who need timely, accurate patient data (see ▶ Chap. 14). Institutions that do not yet deploy electronic health record systems have fallen behind; this may become blameworthy. Failure to use such systems may also disqualify institutions for reimbursements from public and private insurance, making it effectively an organizational death sentence. Conversely, systems that make it easy for clinicians to access data also make it easier for people in general to access the data, and electronic systems generally magnify number of persons whose information becomes available when a system security breach occurs. Some would consider failure to prevent inappropriate access as at least as blameworthy as failure to provide adequate and appropriate access.

Nonetheless, there is no contradiction between the obligation to maintain a certain standard of care (in this case, regarding minimal levels of computer use) and ensuring that such a technical standard does not imperil the rights of patients. Threats to confidentiality and privacy are fairly well known. They include economic abuses, or discrimination by third-party payers, employers, and others who take advantage of the ever-burgeoning market in health data; insider abuse, or record snooping by hospital or clinic workers who are not directly involved in a patient's care but examine a record out of curiosity, for instance; identity theft for insurance or other forms of financial fraud; and malevolent hackers, or people who, via networks or other means, copy, delete, or alter confidential information – or threaten to do so, a component of "ransomware" (see, e.g., Sittig and Singh

2016; and Slayton 2018). Moreover, widespread dissemination of information throughout the health care system often occurs without explicit patient consent. Health care providers, third-party payers, managers of pharmaceutical benefits programs, equipment suppliers, and oversight organizations collect large amounts of patient-identifiable health information for use in managing care, conducting quality and utilization reviews, processing claims, combating fraud, and analyzing markets for health products and services (National Research Council 1997).

The proper approach to such challenges is one that will ensure both that appropriate clinicians and other people have rapid, easy access to patient records and that others do not have access. Is that a contradictory burden? No. Is it easy to achieve both? No. There are many ways to restrict inappropriate access to electronic records, but all come with a cost. Sometimes the cost is explicit, as when it comes in the form of additional security software and hardware; sometimes it is implicit, as when procedures are required that increase the time commitment by system users.

A well-established standard way to view the landscape of protective measures is to divide it into technological methods and institutional or policy approaches (Alpert 1998):

12.3.2.1 Technological Methods

Computer systems per se can optimize some aspects of security. Typical systems verify that users are who they claim to be ("authenticating") with passwords, tokens or biometrics. Other controls limit access to people with a professional "need to know." Creating audit trails, or logs, to record who viewed confidential records enables authorized facility administrators, automated security auditing programs, and patients to later review who accessed what. Encryption can protect data in transit and at rest (in storage). These technical means are complemented by protecting the elements of the electronic infrastructure with physical barriers when operations allow it. Auditing works best when appropriately severe punishments are widely known to be policy, and when policy breaches are uniformly punished in a semi-public manner.

Technological efforts to improve health system security have emerged as a kind of sub-specialty in health informatics, with system developers, computer scientists and others working to improve confidentiality protections. This often entails both better firewalls against intrusion and software to prevent re-identification of stored data with the individuals to whom the data apply (see, for example, Malin and Goodman 2018).

12.3.2.2 Policy Approaches

In its landmark report, the National Research Council (1997) recommended that hospitals and other health care organizations create security and confidentiality committees and establish education and training programs. These recommendations parallel an approach that had worked well elsewhere in hospitals for matters ranging from infection control to bioethics. The U.S. Health Insurance Portability and Accountability Act (HIPAA) requires the appointment of privacy and security officials, special policies, and the training of health care workforce members who have access to health information systems. The European Union's General Data Protection Regulation (GDPR) requires new accountability and governance measures, standards for access by people to data and information about them, and rules for use of that data and information.

Such measures are all the more important when health data are accessible through networks. The rapid growth of **integrated delivery networks (IDNs)** (see ► Chap. 16) and Health Information Exchanges, for example, illustrate the need not to view health data as a well into which one drops a bucket but rather as an irrigation system that makes its contents available over a broad—sometimes an extremely broad—area. It is not yet clear whether privacy and confidentiality protections that are appropriate in hospitals will be fully effective in a ubiquitously networked environment, but it is a start. System developers, users, and administrators are obliged to identify appropriate measures in light of the particular risks associated with a given imple-

mentation. There is no excuse for failing to make this a top priority throughout the data storage and sharing environment.

12.3.2.3 Electronic Data and Human Subjects Research

The use of patient information for **clinical research** and for quality assessment raises interesting ethical challenges. The presumption of a right to confidentiality seems to include the idea that patient records are inextricably linked to patient names or to other identifying data. In an optimal environment, then, patients can monitor who is looking at their records. But if all unique identifiers have been stripped from the records, is there any sense in talking about confidentiality?

The benefits to **public health** loom large in considering record-based research (► Chap. 18). A valuable benefit of the electronic health record is the ability to access vast numbers of patient records to estimate the incidence and prevalence of various maladies, to track the efficacy of clinical interventions, and to plan efficient resource allocation (see ► Chap. 18). Such research and planning would, however, impose onerous or intractable burdens if informed, or valid consent had to be obtained from every patient whose record was represented in the sample. Using confidentiality to impede or forbid such research fails to benefit patients at the same time it sacrifices beneficial scientific investigations.

A more practical course is to establish safeguards that better balance the ethical obligations to privacy and confidentiality against the social goals of public health and systemic efficiency. This balancing can be pursued via a number of paths. The first is to establish mechanisms to **anonymize** the information in individual records or to decouple the data contained in the records from any unique patient identifier. This task is not always straightforward; it can be remarkably difficult to anonymize data such that, when coupled with other data sets, the individuals are not at risk of re-identification. A relatively rare disease diagnosis coupled with demographic data such as age and gender, or geographic data such as a postal code, may act as a surrogate unique identifier; that is, detailed information can in combination serve as a data fingerprint that picks out an individual patient even though the patient's name, Social Security number, or other (official) unique identifiers have been removed from the record. Challenges and opportunities related to de-identifying and re-identifying data are among the most interesting, difficult and important in all health computing (Atreya et al. 2013; Benitez and Malin 2010; Malin and Sweeney 2004; Malin et al. 2011; Sweeney 1997; Tamersoy et al. 2012).

Such challenges point to a second means of balancing ethical goals in the context of database research: the use of institutional panels, such as **medical record committees** or **institutional review boards**. Submission of database research to appropriate institutional scrutiny is one way to make the best use of more or less anonymous electronic patient data. Competent panel members should be educated in the research potential of electronic health records, as well as in ethical issues in epidemiology and public health. Scrutiny by such committees can also give appropriate weight to competing ethical concerns in the context of internal research for quality control, outcomes monitoring, and so on (Goodman 1998; Miller and Gardner 1997a, b).

12.3.2.4 Challenges in Bioinformatics

Safeguards are increasingly likely to be challenged as genetic information makes its way into the health care record (see ► Chaps. 11 and 28). The risks of bias, discrimination, and social stigma increase dramatically as **genetic data** become available to clinicians and investigators. Indeed, genetic information "goes beyond the ordinary varieties of medical information in its predictive value" (Macklin 1992). Genetic data also may be valuable to people predicting outcomes, allocating resources, and the like. In addition, genetic data are rarely associated with only a single person; they may provide information about relatives, including relatives who do not want to know about their genetic risk factors or potential maladies, as well as relatives who would love dearly to know more about their kin's genome. There is still much work to be

done in sorting out and addressing the ethical issues related to electronic storage, sharing, and retrieval of genetic data (Goodman 1996, 2016a).

Bioinformatics or **computational biology** provides an exciting ensemble of new tools to increase our knowledge of genetics, genetic diseases, and public health. Use of these tools is accompanied by responsibilities to attend to the ethical issues raised by new methods, applications, and consequences (Goodman and Cava 2008). Identifying and analyzing these issues are among the key tasks of those who work at the intersection of ethics and health information technology. The future of genetics and genomics is utterly computational, with data storage and analysis posing some of greatest financial and scientific challenges. For instance:

- How, to what extent, and by whom should genomic databases be used for clinical or public health decision support?
- Are special rules needed to govern the study of information in digital genetic repositories (or are current human subjects research protection rules adequate)?
- Does data mining software present new challenges when applied to human genetic information?
- What policies are required to guide and inform the communication of patient-specific and incidental findings?
- Are special protections and precautions needed to address and transmit findings about population subgroups?

It might be that the tools and uses of computational biology will eventually offer ethical challenges—and opportunities—as important, interesting and compelling as any technology in the history of the health sciences. Significantly, this underscores the importance of arguments to the effect that attention to ethics must accompany attention to science. Victories of health science research and development will be undermined by any failures to address corresponding ethical challenges. We must strive to identify, analyze, and resolve or mitigate important ethical issues.

12.4 Social Challenges and Ethical Obligations

The expansion of **evidence-based medicine** and, in the United States, of managed care (now sometimes called **accountable care** since the passage of health reform legislation in 2010; see ▶ Chap. 29) places a high premium on the tools of health informatics. The need for data on clinical outcomes is driven by a number of important social and scientific factors. Perhaps the most important among these factors is the increasing unwillingness of governments and insurers to pay for interventions and therapies that do not work or that do not work well enough to justify their cost.

Health informatics helps clinicians, administrators, third-party payers, governments, researchers, and other parties to collect, store, retrieve, analyze, and scrutinize vast amounts of data—though the task of documenting this is itself a matter of research on what has come to be called "meaningful use." The functions of health informatics might be undertaken not for the sake of any individual patient but rather for cost analysis and review, quality assessment, scientific research, and so forth. These functions are critical, and if computers can improve their quality or accuracy, then so much the better.

Challenges arise when intelligent applications are mistaken for decision-making surrogates or when institutional or public policy recommends or favors computer output over human cognition. This may be seen as a question or issue arising under the rubric of "appropriate uses and users." That is, by whom, when, and under what constraints may we elicit and invoke computational analysis in shaping or applying public policy? The question whether an individual physician or multispecialty group, say, should be hired or retained or reimbursed or rewarded is information-intensive. The question that follows, however, is the key one: How should the decision-making skills of human and machine be used, and balanced (cf. Glaser 2010)?

12.4.1 Vendor Interactions

Motivated if not inspired by both technological necessity and financial opportunity, humble private practices and sprawling medical centers have—or should have—begun the transition from a paper patient record to an electronic one. The need to make such a transition is not in dispute: paper (and handwriting) are hard to store, find, read and analyze. **Electronic Health Records** (EHR) are not, or should not be. While there are important debates about the speed of the transition and regarding software quality, usability and ability to protect patient safety, it is widely agreed that the recording and storage of health information must be electronic.

Public policy has attempted to overcome some of the reluctance to make the change because of financial concerns. Notably, the U.S. Health Information Technology for Economic and Clinical Health (HITECH) Act, a part of the American Recovery and Reinvestment Act of 2009 (Blumenthal 2010), authorized some $27 billion in incentives for EHR adoption. These incentives helped address but did not eliminate financial concerns in that they offset only some of the cost of converting to an e-system. Still, while a number of companies had previously found opportunity in developing hospital and other clinical information systems, HITECH accelerated the pace (see ► Chap. 29).

The firms that make and sell EHRs are not regulated in the same way as those that manufacture pharmaceutical products or medical devices (see ► Sect. 12.5.3). In an increasingly competitive environment, this has led to controversy about the nature of vendor interactions with the institutions that buy their products. An EHR system for a mid-sized hospital can cost upwards of $100 million over time, including consulting services, hardware and training. It follows that it is reasonable to ask what values should guide such vendor interactions with clients, and whether they should be similar to or different from values that govern other free-market dealings.

While many or most contracts between vendors and hospitals are confidential, it has been reported that some HIT vendors require contract language that indemnifies system developers for personal injury claims or malpractice, even if the vendor is at fault; some vendors require system purchasers to agree not to disclose system errors except to the vendor (Koppel and Kreda 2009). Such provisions elicit concern to the extent they place or appear to place corporate interests ahead of patient safety and welfare. In this case, a working group chartered by AMIA, the society for informatics professionals (see ► Chap. 1), issued a report that provided guidance on a number of vendor interaction issues (Goodman et al. 2010). Importantly, the working group comprised industry representatives as well as scientists and other academics. The group's recommendations included these:

- Contracts should not contain language that prevents system users, including clinicians and others, from using their best judgment about what actions are necessary to protect patient safety. This includes freedom to disclose system errors or flaws, whether introduced or caused by the vendor, the client, or a third party. Disclosures made in good faith should not constitute violations of HIT contracts. This recommendation neither entails nor requires the disclosure of trade secrets or of intellectual property.
- Because vendors and their customers share responsibility for patient safety, contract provisions should not attempt to circumvent fault and should recognize that both vendors and purchasers share responsibility for successful implementation. For example, vendors should not be absolved from harm resulting from system defects, poor design or usability, or hard-to-detect errors. Similarly, purchasers should not be absolved from harm resulting from inadequate training and education, inadequate resourcing, customization, or inappropriate use.

While some of the debates that led to those conclusions were about political economy (regulation vs. free enterprise) as much as ethics (right vs. wrong), the opportunity for rapprochement in the service of a patient-centered approach may be seen as an affirmation of the utility of an applied ethics process in the evolution of health information technology.

12.4.2 Computational Prognosis

Consider the utility of **prognostic scoring systems** that use physiologic and mortality data to compare new critical-care patients with thousands of previous patients (Knaus et al. 1991). Such systems allow hospitals to track the performance of their critical-care units by, say, comparing the previous year's outcomes to this year's or by comparing one hospital to another. If, for instance, patients with a particular profile tend to survive longer than their predecessors, then it might be inferred that **critical care** has improved. Such scoring systems can be useful for internal research and for quality management (Fig. 12.2).

Now suppose that most previous patients with a particular physiologic profile have died in critical-care units; this information might be used to identify ways to improve care of such patients—or it might be used in support of arguments to contain costs by denying care to subsequent patients fitting the profile (since they are likely to die anyway).

An argument in support of such an application might be that decisions to withdraw or withhold care are often and customarily made on the basis of subjective and fragmented evidence; so it is preferable to make such decisions on the basis of objective data of the sort that otherwise underlie sound clinical practice. Such **outcomes data** are precisely what fuels the engines of managed care, wherein health professionals and institutions compete on the basis of cost and outcomes. Why should society, or a managed-care organization, or an insurance company pay for critical care when seemingly objective evidence exists that such care will not be efficacious? Contrarily, consider the effect on future scientific insights of denying care to such patients. Scientific progress is often made by noticing that cer-

Fig. 12.2 "Severity adjusted daily data" in fictitious APACHE® Outcomes screen shot. Using prognostic scoring systems, clinicians in critical-care units can monitor events and interventions and administrators can manage staffing based on patient acuity. Clinicians can also use such systems to predict mortality, raising a number of ethical issues. This image shows 10 CCU patients. For the second one in the leftmost column, for instance, the "acute physiology score" is 128; the risk of hospital mortality is 96% and the risk of ICU mortality is 92%. (Credit: Courtesy of Cerner Corporation, with permission)

tain patients do better under certain circumstances, and investigation of such phenomena leads to better treatments. If all patients meeting certain criteria were denied therapy on the basis of a predictive tool, it would become a self-fulfilling prophecy for a much longer time that all such patients would not do well (Miller 1997).

Now consider use of a decision-support system to evaluate, review, or challenge decisions by human clinicians; indeed, imagine an insurance company using a diagnostic expert system to determine whether a physician should be reimbursed for a particular procedure. If the expert system has a track record for accuracy and reliability, and if the system "disagrees" with the human's diagnosis or treatment plan, then the insurance company can contend that reimbursement for the procedure would be a mistake. Why pay for a procedure that is not indicated, at least according to a computational analysis?

In the two examples just offered (a prognostic scoring system is used to justify termination of treatment to conserve resources, and a diagnostic expert system is used to deny a physician reimbursement for procedures deemed inappropriate), there seems to be justification for adhering to the computer output. There are, however, three reasons why it is problematic to rely exclusively on clinical computer programs to guide policy or practice in these ways:

1. As we argued earlier with the standard view of computational diagnosis (and, by easy extension, prognosis), human cognition is, at least for a while longer, still superior to machine intelligence. Moreover, the act of rendering a diagnosis or prognosis is not merely a statistical or computational operation performed on uninterpreted data. Rather, identifying a malady and predicting its course requires understanding a complex ensemble of causal relations, interactions among a large number of variables, and having a store of salient background knowledge—considerations that have thus far failed to be grasped, assessed, and effectively blended into decisions made by computer programs.
2. Decisions about whether to treat a given patient are often value laden and must be made relative to treatment goals. In other words, it might be that a treatment will improve the quality of life but not extend life, or vice versa (Youngner 1988). Whether such treatment is appropriate cannot be determined scientifically or statistically (Brody 1989). The decisions ultimately depend on human preferences—those of the provider or, even more importantly, the patient.
3. Applying computational operations on aggregate data to individual patients runs the risk of including individuals in groups they resemble but to which they do not actually belong. Of course, human clinicians run this risk all the time—the challenge of inferring correctly that an individual is a member of a set, group, or class is one of the oldest problems in logic and in the philosophy of science. The point is that computers have not solved this problem, yet, and allowing policy to be guided by simple or unanalyzed correlations constitutes a conceptual error.

The idea is not that diagnostic or prognostic computers are always wrong—we know that they are not—but rather there are numerous instances in which we do not know whether they are right. It is one thing to allow aggregate data to guide policy; doing so is just using scientific evidence to maximize good outcomes. But it is altogether different to require that a policy disallow individual **clinical judgment** and expertise.

Informatics can contribute in many ways to health care reform. Indeed, computer-based tools can help to illuminate ways to reduce costs, to optimize clinical outcomes, and to improve care. Scientific research, quality assessment, and the like are, for the most part, no longer possible without computers. But it does not follow that the insights from such research apply in all instances to the myriad variety of actual clinical cases at which competent human clinicians excel.

The Coronavirus crisis of 2020 provided an opportunity to assess and review the use of

prognostic scoring systems to inform or guide triage and rationing. Controversy inherent in resource allocation under conditions of scarcity was magnified when decisions about ventilator allocation, for instance, were made based on a prognostic score rendered by utilities resident in electronic health records (Truog et al. 2020). Although a strong case can be made that such use was permissible during a crisis and in the absence of anything better, an equally strong case must be made that this opportunity to assess and review the use of prognostic scoring systems not be squandered. The *in situ* use of this informatics tool should be scrutinized and studied.

12.4.3 Effects of Informatics on Traditional Relationships

Patients are often frightened and vulnerable. Treating illness, easing fear, and respecting vulnerability are among the core obligations of physicians, nurses, and other clinicians. Health informatics has the potential to complement these traditional duties and the relationships that they entail. We have pointed out that medical decisions are shaped by nonscientific considerations. This point is important when we assess the effects of informatics on human relationships. Thus:

> The practice of medicine or nursing is not exclusively and clearly scientific, statistical, or procedural, and hence is not, so far, computationally tractable. This is not to make a hoary appeal to the "art and science" of medicine; it is to say that the science is in many contexts inadequate or inapplicable: Many clinical decisions are not exclusively medical—they have social, personal, ethical, psychological, financial, familial, legal, and other components; even art might play a role. (Miller and Goodman 1998)

12.4.3.1 Professional–Patient Relationships

If computers, databases, and networks can improve physician–patient or nurse–patient relationships, perhaps by improving communication, then we shall have achieved a happy result. If reliance on computers impedes the abilities of health professionals to establish trust and to communicate compassionately, however, or further contributes to the dehumanization of patients (Shortliffe 1993, 1994), then we may have paid too dearly for our use of these machines.

Suppose that a physician uses a decision-support system to test a diagnostic hypothesis or to generate differential diagnoses, and suppose further that a decision to order a particular test or treatment is based on that system's output. A physician who is not able to articulate the proper role of computational support in his decision to treat or test will risk alienating those patients who, for one reason or another, will be disappointed, angered, or confused by the use of computers in their care. To be sure, the physician might just withhold this information from patients, but such deception carries its own threats to trust in the relationship.

Patients are not completely ignorant about the processes that constitute human decision making. What they do understand, however, may be subverted when their doctors and nurses use machines to assist delicate cognitive functions. We must ask whether patients should be told the accuracy rate of decision support automata—when they have yet to be given comparable data for humans. Would such knowledge improve the informed-consent process, or would it "constitute another befuddling ratio that inspires doubt more than it informs rationality?" (Miller and Goodman 1998).

To raise such questions is consistent with promoting the responsible use of computers in clinical practice. The question whether computer use will alienate patients is an empirical one; it is a question for which, despite many initial studies, we lack conclusive data to answer. (For example, we cannot yet state definitively whether all categories of patients will respond well to all specific types of e-mail messages from their doctors. Nevertheless, as a moral principle discussed above, one should not convey a new diagnosis of a malignancy via email.) To address the question now anticipates potential future problems. We must ensure that the exciting

potential of health informatics is not subverted by our forgetting that the practice of medicine, nursing, and allied professions is deeply human and fundamentally intimate and personal.

12.4.3.2 Consumer Health Informatics

The growth of the World Wide Web and the commensurate evolution of clinical and health resources on the Internet also raise issues for professional–patient relationships. **Consumer health informatics**—technologies focused on patients as the primary users—makes vast amounts of information available to patients (see ▶ Chap. 11). There is also, however, misinformation—even outright falsehoods and quackery—posted on some sites. If physicians and nurses have not established relationships based on trust, the erosive potential of apparently authoritative Internet resources can be great. Physicians once accustomed to newspaper-inspired patient requests for drugs and treatments now face ever-increasing demands that are informed by Web browsing. Consequently, the following issues have gained in ethical importance for more than a decade:

- Peer review: How and by whom is the quality of a Web site to be evaluated? Who is responsible for the accuracy of information communicated to patients?
- Online consultations: There is no standard of care yet for online medical consultations. What risks do physicians and nurses run by giving advice to patients whom they have not met or examined in person? This question is especially important in the context of **telemedicine** or **remote-presence health care**, the use of video teleconferencing, image transmission, and other technologies that allow clinicians to evaluate and treat patients in other than face-to-face situations (see ▶ Chap. 20). Use of telehealth tools became ubiquitous during the Coronavirus pandemic of 2020. This too presents an opportunity to evaluate widespread adoption in context.
- Support groups: Internet support groups can provide succor and advice to the sick, but there is a chance that someone who might benefit from seeing a physician will not do so because of anecdotes and information otherwise attained. How should this problem be addressed?

That a resource is touted as worthwhile does not make it so. We lack evidence to illuminate the utility of consumer health informatics and its effects on professional–patient relationships. Such resources cannot be ignored given their ubiquity, and they often are useful for improving health. But we insist that here—as with decision support, appropriate use and users, evaluation, and privacy and confidentiality—there is an ethical imperative to proceed with caution. Informatics, like other health technologies, will thrive if our enthusiasm is open to greater evidence and is wed to deep reflection on human values.

12.4.3.3 Personal Health Records

At the same time as institutions have moved to computer-based health records systems, the tools available to individuals to keep their own health records have been making a similar transition. Electronic **personal health record** (PHR) systems, whether designed for use on a decoupled storage device or accessible over the Web, are now available from a rapidly expanding set of organizations (see ▶ Chap. 11) (◘ Figs. 12.3 and 12.4). Indeed, increasingly many patients access aspects of their health and medical records through "portals" established by EHR vendors.

PHRs provide a storage base for data once kept on paper (or in the patient's head) and repeatedly extracted with each institutional encounter for inclusion in that entity's records system, typically:

- Allergies, current medications
- Current health status and major health issues (if any)
- Major past health episodes and the condition of oneself and (sometimes) relatives
- Vaccinations, surgeries and other treatments

All these data can be kept on something simple (and un-networked) like a flash drive. It is

becoming more common to store the data on a Web site, where PHR data can also be linked to other health information relevant to the person. The PHR data can also be linked via a portal to a health care provider institution's records, to allow updating in both directions, or be free of any such tie. A flash drive can be forgotten or lost, whereas a Web site can be centrally updated and uniformly available via any properly authenticated device on the Internet.

Fig. 12.3 Project HealthDesign (Brennan et al. 2010) was a landmark program sponsored by the Robert Wood Johnson Foundation's Pioneer Portfolio and intended to foster development of personal health records. Here is a barcode scanner that recognizes medication labels. Designed by researchers at the University of Colorado at Denver, the "Colorado Care Tablet" allows elderly users to track prescriptions with such scanners and portable touch-screen tablets. (Credit: Courtesy of Project HealthDesign; Creative Commons Attribution 3.0 Unported License)

12

Traditional insurers and health care providers are duty-bound by privacy laws and regulations to protect the information under their control. PHRs have a somewhat shakier set of protections given their relatively short history. The legal obligations of institutions that provide PHRs, but do not fully manage the content of those records nor their use, as well as the obligations (if any) of the individuals who "manage" their own health records, remain to be resolved (Cushman et al. 2010).

PHRs are now commonly linked to so-called "personal health applications" (PHAs) which provide ways of moving beyond simple static storage of one's medical history. Most provide some sort of primitive decision support, if only in linking to additional information about a particular disease or condition. Others include more ambitious decision-support functionality. All the concerns about the accuracy of Web-based information recur in this context, with concerns about the reliability of decision support added to that. Compounding concerns about accuracy are the inherent limitations of the "owner-operator": If it can be difficult for trained health care providers to evaluate the quality of advice rendered by a decision support system, the challenges for patients will be commensurately greater.

Traditional health care institutions may see the PHR as a device for patient empowerment because it adds a way for persons to keep track of their own data; but they can also be

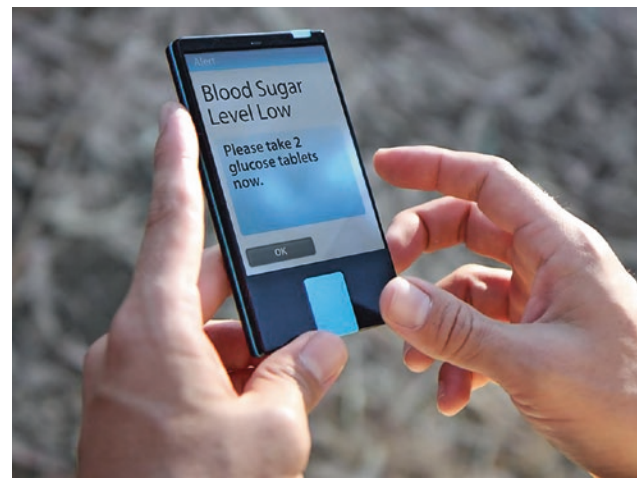


Fig. 12.4 A portable blood glucose communicator is part of the personal health record system developed by the T.R.U.E. Research Foundation of Washington, DC. The diabetes management application analyzes, summarizes, displays and makes individualized recommendations on nutritional data, physical activity data, prescribed medications, continuous blood glucose data, and self-reported emotional state. (Credit: Courtesy of Project HealthDesign; Creative Commons Attribution 3.0 Unported License)

used as a way of preserving "loyalty" to a particular institution in the health care system. It has been proposed that PHRs be subject to standards allowing "interoperability"—in this case, easy movement from one type of PHR to another—to prevent leveraging it as an impediment to patients' movements when they wish to change providers or other preferences change. Whether such standards will evolve enough to make it easy to move from one PHR to another remains to be seen, given economic incentives to impede patient movement (just in case a patient is financially desirable because of insurance status or personal wealth).

Whether PHRs will reach the majority of patients is uncertain. For persons who must chronically manage complex treatment regimens for themselves or for dependents, PHRs and their associated applications may be compelling. Persons who deal with less complex or transient conditions may prefer to leave records management to their providers. In the context of health care, PHRs have the potential to replicate the "digital divide," exacerbating rather than reducing health disparities. Persons with higher levels of income and education may differentially benefit from PHRs by more readily making fuller use of them. In the absence of robust policy protections, some minors may be reluctant to use PHRs as long as parents or guardians retain access.

12.5 Legal and Regulatory Matters

The use of clinical computing systems in health care raises a number of interesting and important legal and regulatory questions.

12.5.1 Difference Between Law and Ethics

Ethical and legal issues often overlap. Ethical considerations apply in attempts to determine what is good or meritorious and which behaviors are desirable or correct in accordance with higher principles. Legal principles are generally derived from ethical ones but deal with the practical regulation of morality or behaviors and activities. Many legal principles deal with the inadequacies and imperfections in human nature and the less-than-ideal behaviors of individuals or groups. Ethics offers conceptual tools to evaluate and guide moral decision making. Laws directly tell us how to behave (or not to behave) under various specific circumstances and prescribe remedies or punishments for individuals who do not comply with the law. Historical precedent, matters of definition, issues related to detectability and enforceability, and evolution of new circumstances affect legal practices more than they influence ethical requirements.

12.5.2 Legal Issues in Biomedical Informatics

Prominent legal issues related to the use of software applications in clinical practice and in biomedical research include liability under tort law; potential use of computer applications as expert witnesses in the courtroom; legislation governing privacy and confidentiality; and copyrights, patents, and intellectual property issues.

12.5.2.1 Liability Under Tort Law

In the United States and in many other nations, principles of tort law govern situations in which harm or injuries result from the manufacture and sale of goods and services (Miller et al. 1985a). Because there are few, if any, U.S. legal precedents directly involving harm or injury to patients resulting from use of clinical software applications (as opposed to a small number of well-documented instances where software associated with medical devices has caused harm), the following discussion is hypothetical. The principles involved are, however, well established with voluminous legal precedents outside the realm of clinical software.

A key legal distinction is the difference between products and services. **Products** are physical objects, such as stethoscopes, that go through the processes of design, manufacture, distribution, sale, and subsequent use by pur-

chasers. **Services** are intangible activities provided to consumers at a price by (presumably) qualified individuals.

The practice of clinical medicine has been deemed a service through well-established legal precedents. On the other hand, clinical software applications can be viewed as either goods ("products") (software programs designed, tested, debugged, placed on DVDs or other media, and distributed physically to purchasers) or services (applications that present data or provide advice to practitioners engaged in a service such as delivering health care). There are few legal precedents to determine unequivocally how software will be viewed by the courts, and it is possible that clinical software programs will be treated as goods under some circumstances and as services under others. It might be the case that that software purchased and running in a private office to handle patient records or billing would be deemed a product, but the same software mounted on shared, centralized computers and accessed over the Internet (and billed on a monthly basis) would be offering a service.

Three ideas from tort law potentially apply to the clinical use of software systems:

(1) **Harm by intention**—when a person injures another using a product or service to cause the damage, (2) the **negligence theory**, and (3) **strict product liability** (Miller et al. 1985a). Providers of goods and services are expected to uphold the standards of the community in producing goods and delivering services. When individuals suffer harm due to substandard goods or services, they may sue the service providers or goods manufacturers to recover damages. **Malpractice** litigation in health care is based on negligence theory.

Because the law views delivery of health care as a service (provided by clinicians), it is clear that negligence theory will provide the minimum legal standard for clinicians who use software during the delivery of care. Patients who are harmed by clinical practices based on imperfect software applications may sue the health care providers for negligence or malpractice, just as patients may sue attending physicians who rely on the imperfect advice of a human consultant (Miller et al. 1985a). Similarly, a patient might sue a practitioner who has not used a decision-support system when it can be shown that use of the decision-support system is part of the current standard of care, and that use of the program might have prevented the clinical error that occurred (Miller 1989). It is not clear whether the patients in such circumstances could also successfully sue the software manufacturers, as it is the responsibility of the licensed practitioner, and not of the software vendor, to uphold the standard of care in the community through exercising sound clinical judgment. Based on a successful malpractice suit against a clinician who used a clinical software system, it might be possible for the practitioner to sue the manufacturer or vendor for negligence in manufacturing a defective clinical software product, but cases of this sort have not yet been filed. If there were such suits, it might be difficult for a court to discriminate between instances of improper use of a blameless system and proper use of a less-than-perfect system.

In contrast to negligence, strict product liability applies only to harm caused by defective products and is not applicable to services. The primary purpose of strict product liability is to compensate the injured parties rather than to deter or punish negligent individuals (Miller et al. 1985a). For strict product liability to apply, three conditions must be met:

1. The product must be purchased and used by an individual.
2. The purchaser must suffer physical harm as a result of a design or manufacturing defect in the product.
3. The product must be shown in court to be "unreasonably dangerous" in a manner that is the demonstrable cause of the purchaser's injury.

Note that negligence theory allows for adverse outcomes. Even when care is delivered in a competent, caring, and compassionate manner, some patients with some illnesses will not do well. Negligence theory protects providers from being held responsible for all individuals who suffer bad outcomes. As long as the quality of care has met the prevailing standards, a practitioner should not be found liable in a

malpractice case (Miller et al. 1985a). Strict product liability, on the other hand, is not as forgiving or understanding.

No matter how good or exemplary a manufacturer's designs and manufacturing processes might be, if even one in ten million products is defective, and that one product defect is the cause of a purchaser's injury, then the purchaser may collect damages (Miller et al. 1985a). The plaintiff needs to show only that the product was unreasonably dangerous and that its defect led to harm. In that sense, the standard of care for strict product liability is 100-percent perfection. To some extent, appropriate product labeling (e.g., "Do not use this metal ladder near electrical wiring") may protect manufacturers in certain strict product liability suits in that clear, visible labeling may educate the purchaser to avoid "unreasonably dangerous" circumstances. Appropriate labeling standards may similarly benefit users and manufacturers of clinical expert systems (Geissbuhler and Miller 1997).

Health care software programs sold to clinicians who use them as decision-support tools in their practices are likely to be treated under negligence theory as services. When advice-giving clinical programs are sold directly to patients, however, and there is less opportunity for intervention by a licensed practitioner, it is more likely that the courts will treat them as products, using strict product liability, because the purchaser of the program is more likely to be the individual who is injured if the product is defective. (As personal health records become more common, this legal theory may well be tested.)

A growing number of software "bugs" in medical devices have been reported to cause injury to patients (Majchrowski 2010; Levis 2014). The U.S. Food and Drug Administration (FDA) has traditionally viewed software embedded within medical devices, such as cardiac pacemakers and implantable insulin pumps, as part of the physical device, and so regulates such software as part of the device (FDA 2011). The courts are likely to view such software using principles of strict product liability (Miller and Miller 2007). Most recently, the FDA has contemplated wider regulatory scope (see ► Sect. 12.5.3).

Corresponding to potential strict product liability for faulty software embedded in medical devices is potential negligence liability if such software can easily be "hacked" (Robertson 2011). Malicious code writers might mimic external software-based "radio" controllers for pacemakers and insulin pumps and reprogram them to cause harm to patients. While such "hackers" should face criminal prosecution if they cause harm by intention, the device manufacturers have a responsibility to make it difficult to change the software code embedded in devices without proper authorization.

12.5.2.2 Privacy and Confidentiality

The ethical basis for privacy and confidentiality in health care is discussed in ► Sect. 12.3.1. For a long time, the legal state of affairs for privacy and confidentiality of electronic health records was chaotic (as it remains for written records, to some extent). This state of affairs in the U.S. had not significantly changed in the three decades since it was described in a classic New England Journal of Medicine article (Curran et al. 1969).

However, a key U.S. law, the Health Insurance Portability and Accountability Act (HIPAA), has prompted significant change. HIPAA's privacy standards became effective in 2003 for most health care entities, and its security standards followed 2 years later; a breach-notification rule was expanded in 2010 and HITECH provisions were incorporated in 2013. A major impetus for the law was that the process of "administrative simplification" via electronic recordkeeping, prized for its potential to increase efficiency and reduce costs, would also pose threats to patient privacy and confidentiality. Coming against a backdrop of a variety of noteworthy cases in which patient data were improperly—and often embarrassingly—disclosed, the law was also seen as a badly needed tool to restore confidence in the ability of health professionals to protect confidentiality. While the law has been accompanied by debate both on the adequacy of its measures and the question whether compliance was unnecessarily burdensome, it nevertheless established the first nationwide health privacy protections. At its

core, HIPAA embodies the idea that individuals should have access to their own health data, and more control over uses and disclosures of that health data by others. Among its provisions, the law requires that patients be informed about their privacy rights, including a right of access; that uses and disclosures of "protected health information" generally be limited to exchanges of the "minimum necessary"; that uses and disclosures for other than treatment, payment and health care operations be subject to patient authorization; and that all employees in "covered entities" (institutions that HIPAA legally affects) be educated about privacy and information security.

As noted above, the HITECH Act provided substantial encouragement for Electronic Health Record (EHR) development, particularly the encouragement of billions of dollars in federal subsidies for "meaningful use" of EHRs. However, HITECH also contained many changes to HIPAA privacy and security requirements, strengthening the regulations that affect the collection, use and disclosure of health information not only by covered entities, but also the "business associates" (contractors) of those covered entities, and other types of organizations engaged in health information exchange.

The Office of Civil Rights in the U.S. Department of Health and Human Services remains the entity primarily charged with HIPAA enforcement, but there is now a role for states' attorneys general as well as other agencies such as the Federal Trade Commission. HITECH increases penalty levels under HIPAA and includes a mandate for investigations and periodic audits, shifting the enforcement balance away from voluntary compliance and remediation plans.

HITECH's changes to HIPAA, those from other federal laws such as the Genetic Information Nondiscrimination Act of 2008 (GINA) and the Patient Safety and Quality Improvement Act of 2005, and the new attention to information privacy and security in most states' laws, comprise significant changes to the legal-regulatory landscape for health information.

12.5.2.3 Copyright, Patents, and Intellectual Property

Intellectual property protection afforded to developers of software programs, biomedical knowledge bases, and World Wide Web pages remains an underdeveloped area of law. Although there are long traditions of copyright and patent protections for non-electronic media, their applicability to computer-based resources is not clear. **Copyright law** protects intellectual property from being copied verbatim, and **patents** protect specific methods of implementing or instantiating ideas. The number of lawsuits in which one company claimed that another copied the functionality of its copyrighted program (i.e., its "look and feel") has grown, however, and it is clear that copyright law does not protect the "look and feel" of a program beyond certain limits. Consider, for example, the unsuccessful suit in the 1980s by Apple Computer, Inc., against Microsoft, Inc., over the "look and feel" of Microsoft Windows as compared with the Apple Macintosh interface (which itself resembled the earlier Xerox Alto interface).

It is not straightforward to obtain copyright protection for a list that is a compilation of existing names, data, facts, or objects (e.g., the telephone directory of a city), unless you can argue that the result of compiling the compendium creates a unique object (e.g., a new organizational scheme for the information) (Tysyer 1997). Even when the compilation is unique and copyrightable, the individual components, such as facts in a database, might not be copyrightable. That they are not copyrightable has implications for the ability of creators of biomedical databases to protect database content as intellectual property. How many individual, unprotected facts can someone copy from a copyright-protected database before legal protections prevent additional copying?

A related concern is the intellectual-property rights of the developers of materials made available through the World Wide Web. Usually, information made accessible to the public that does not contain copyright annotations is considered to be in the public domain. It is tempting to build from the work

of other people in placing material on the Web, but copyright protections must be respected. Similarly, if you develop potentially copyrightable material, the act of placing it on the Web, in the public domain, would allow other people to treat your material as not protected by copyright. Resolution of this and related questions may await workable commercial models for electronic publication on the World Wide Web, whereby authors could be compensated fairly when other people use or access their materials. Electronic commerce might eventually provide copyright protection (and perhaps revenue) similar to age-old models that now apply to paper-based print media; for instance, to use printed books and journals, you must generally borrow them from a library, purchase them or access them under Creative Commons or similar open-access platforms.

12.5.3 Regulation and Monitoring of Computer Applications in Health Care

In the mid-1990s, the U.S. Food and Drug Administration (FDA) held public meetings to discuss new methods and approaches to regulating clinical software systems as medical devices. In response, a consortium of professional organizations related to health care information (AMIA, the Center for Health Care Information Management, the Computer-Based Patient Record Institute, the American Health Information Management Association, the Medical Library Association, the Association of Academic Health Science Libraries, and the American Nurses Association) drafted a position paper published in both summary format and as a longer discussion with detailed background and explanation (Miller and Gardner 1997a, b). The position paper was subsequently endorsed by the boards of directors of all the organizations (except the Center for Health Care Information Management) and by the American College of Physicians Board of Regents.

The consortium recommended the following:

- Recognition of four categories of clinical system risks and four classes of monitoring and regulatory actions that can be applied based on the level of risk in a given setting.
- Local oversight of clinical software systems, whenever possible, through the creation of autonomous **software oversight committees**, in a manner partially analogous to the institutional review boards that are federally mandated to oversee protection of human subjects in biomedical research. Experience with prototypical software-oversight committees at pilot sites should be gained before any national dissemination.
- Adoption by health care-information system developers of a code of good business practices.
- Recognition that budgetary, logistic, and other constraints limit the type and number of systems that the FDA can regulate effectively.
- Concentration of FDA regulation on those systems posing highest clinical risk, with limited opportunities for competent human intervention, and FDA exemption of most other clinical software systems.

The recommendations for combined local and FDA monitoring are summarized in Table 12.1. We do not yet know whether improved outcomes would occur if vendors were to give qualified (i.e., informatics-capable) institutional purchasers greater local control over system functionality.

Section 618 of the 2012 Food and Drug Administration Safety and Innovation Act (FDASIA), Public Law 112- 144, mandated a new generation of oversight guidelines for clinical software. Pursuant to the legislation, the FDA, the Office of the National Coordinator for Health Information Technology, and the Federal Communications Commission held public hearings and conducted workshops. The ensuing April 2014 FDASIA Health IT report (FDASIA 2014)

Table 12.1 Consortium recommendations for monitoring and regulating clinical software systems

	Regulatory class			
Variable	**A**	**B**	**C**	**D**
Supervision by FDA	Exempt from regulation	Excluded from regulation	Simple registration and postmarket surveillance required	Premarket approval and postmarket surveillance required
Local software oversight committee	Optional	Mandatory	Mandatory	Mandatory
Role of software oversight committee	Monitor locally	Monitor locally instead of monitoring by FDA	Monitor locally and report problems to FDA as appropriate	Assure adequate local monitoring without replicating FDA activity
Software risk category				
0: Informational or generic systems[a]	All software in category	–	–	–
1: Patient-specific systems that provide low-risk assistance with clinical problems[b]	–	All software in category	–	–
2: Patient-specific systems that provide intermediate-risk support on clinical problems[c]	–	Locally developed or locally modified systems	Commercially developed systems that are not modified locally	–
3: High-risk, patient-specific systems[d]	–	Locally developed, non commercial systems	–	Commercial systems

Reproduced with permission from Author(s). Miller and Gardner (1997a). ©American College of Physicians
[a]Includes systems that provide factual content or simple, generic advice (such as "give flu vaccine to eligible patients in mid-autumn") and generic programs, such as spreadsheets and databases
[b]Systems that give simple advice (such as suggesting alternative diagnoses or therapies without stating preferences) and give ample opportunity for users to ignore or override suggestions
[c]Systems that have higher clinical risk (such as those that generate diagnoses or therapies ranked by score) but allow users to ignore or override suggestions easily; net risk is therefore intermediate
[d]Systems that have great clinical risk and give users little or no opportunity to intervene (such as a closed-loop system that automatically regulates ventilator settings)

met Congress' requirements to propose "strategy and recommendations on an appropriate, risk-based regulatory framework pertaining to health-information technology, including mobile medical applications, that promotes innovation, protects patient safety, and avoids regulatory duplication." The Report implemented many recommendations from the National Academy of Medicine's 2012 report, "Health IT and Patient Safety: Building Safer Systems for Better Care" (IOM 2012a).

The FDASIA Health IT Report specified three categories of risk based on system functionality, rather than software product category or on implementation platform (FDASIA 2014). The functionality categories are:

1. Administrative health IT functions. Non-exhaustive examples given in the Report include billing and claims processing, inventory management, and scheduling. The FDASIA Report categorizes those

functions posing little or no risk to the patient, and exempt from additional oversight.
2. Health management IT functions. Non-exhaustive examples cited in the Report include encounter documentation, electronic access to clinical results, non-device-related clinical decision support, medication management, provider order entry, knowledge management, and electronic communication, including health information exchange. The FDASIA Report asserts that the actual safety risks posed by this category are for the most part outweighed by their potential benefits, and require limited national-level oversight. Whereas the FDA previously played a major regulatory and monitoring role for applications with this category of functionality, the FDASIA Report transfers responsibility for such oversight to a collaboration between ONC and commercial vendors.
3. Medical device health IT functions. Non-exhaustive examples listed in the Report include computer-assisted detection software, notification of real-time alarms from bedside monitors, and robotic surgery systems. The FDA will maintain oversight responsibility for device-related clinical software.

A given application or product may involve more than one of the functionality categories.

In concordance with the National Academy of Medicine 2012 recommendations (IOM 2012a), the FDASIA Report also created a public-private Health IT Safety Center, to be coordinated by ONC. It will promote innovations regarding patient safety and identify interventions to improve safety, including education about best practices.

12.5.4 Software Certification and Accreditation

If, as above, (1) there is an ethical obligation to evaluate health information systems in the contexts in which they are being used, and if, as we just saw, (2) there are good reasons to consider the adoption of software oversight committees or something similar, then it is worthwhile to consider the ethical utility of efforts to review and endorse medical software and systems.

Established in 2004, the Certification Commission for Health Information Technology, in collaboration with the Office of the National Coordinator for health information technology, assesses electronic health records according to an array of criteria, in part to determine their success in contributing to "meaningful use." These criteria address matters ranging from electronic provider order entry and electronic problem lists to decision support and access control (cf. Classen et al. 2007; Wright et al. 2009). The criteria, tests and test methods are developed in concert with the National Institute of Standards and Technology. Practices and institutions that want to receive government incentive payments must adopt certified electronic health record technologies.

Conceived under the American Recovery and Reinvestment Act, these processes aim to improve outcomes, safety and privacy. Whether they can accomplish this—as opposed to celebrate technology for its own sake—is an excellent source of debate (Hartzband and Groopman 2008). What should be uncontroversial is that any system of regulation, review or certification must be based on and, as a matter of process emphasize, certain values. These might include, among others, patient-centeredness, ethically optimized data management practices, and what we have here commended as the "standard view," that is, human beings and not machines practice medicine, nursing and psychology.

The move to certification has unfortunately engendered precious little in the way of ethical analysis, however. To make any system of regulation, review or certification ethically credible, government and industry leaders must eventually make explicit that attention to ethics is a core component of their efforts.

An ethical approach to certification of clinical applications should entail "in vivo"

(on the front lines of clinical care) as well as "in vitro" (laboratory-based) testing (IOM 2012a). As stated in ▶ Sect. 12.2.2, "A computer program should be used in clinical practice only after appropriate evaluation of its efficacy and the documentation that it performs its intended task at an acceptable cost in time and money." Federal and other certification programs currently address, in vitro, whether certain pre-specified technical capabilities exist within a software application. They do not, however, determine whether a given software package will be usable at an affordable cost in time and money in vivo, post-installation.

Almost all vendors' comprehensive, institutional-level clinical information systems now pass the minimal federal certification standards. Those certification standards test algorithmic functionality while neglecting to assess real-world clinical impacts post-installation. The 2012 National Academy of Medicine report on Health IT and Patient Safety noted, "poor usability … is one of the single greatest threats to patient safety. … Evaluation of the impact of health IT on usability and on cognitive workload is important to determine unintended consequences and the potential for distraction, delays in care, and increased workload in general. Usability guidelines and principles focused on improving safety need to be put into practice" (IOM 2012a, p. 81). The certification standards also do not fully evaluate the accuracy or completeness of systems' underlying information/knowledge bases. Nor do they evaluate information/knowledge base accuracy and maintainability over time.

Many institutional-level systems are so expensive that they financially cripple the medical centers and clinicians' offices that adopt them. Many sites experience substantial decreases in post-installation revenue (usually transient, lasting several months, but sometimes persistent). System-induced disruptions of workflows diminish the number of patients who can be seen, and impair (at least temporarily) charge capture for billable services. The FDASIA Report minimizing decision support oversight fits better with past implementations, when academic system developers could directly and efficiently address local problems of system functionality. Present-day commercial systems are inflexible, opaque, and maintained by vendors from a distance. The high sticker price of such systems guarantees that, once purchased, institutions cannot afford to de-install and replace problematic systems. Government mandates to install such expensive, disruptive "certified" software systems appear to some as unethical. The certification process should be expanded to evaluate the pragmatic, local, post-installation aspects of system function "at an acceptable cost in time and money."

The diminished patient-care workflows engendered by cumbersome clinical software systems potentially increase the time and money costs of delivering quality healthcare – costs borne by patients, third-party payors, and the government. Clinicians often pay a higher price – beyond lost revenues. In their 2017 commentary, "The HITECH Era in Retrospect," Halamka and Tripathi stated, "we lost the hearts and minds of clinicians. … We expected interoperability without first building the enabling tools. In a sense, we gave clinicians suboptimal cars, didn't build roads, and then blamed them for not driving" (Halamka and Tripathi 2017). Verghese, Shah, and Harrington, in a 2018 *JAMA* Viewpoint, added: "The nationwide implementation of electronic medical records (EMRs) resulted in many unanticipated consequences … the redundancy of the notes, the burden of alerts, and the overflowing inbox has contributed to … physician reports of symptoms of burnout. … Most EMRs serve their front-line users quite poorly" (Vergese et al. 2018).

Installation of vendors' massive, complex, institutional clinical software products creates additional ethical dilemmas. Post-installation, clinicians lack a clear or deep understanding of how system functions affect their patients' care and safety. Institutions no longer possess the level of control/autonomy to change their systems locally, as they did decades ago when many academic medical centers had home-grown systems that they could manage and "evolve" at will. Prior to the advent of such large, complex clinical systems, clinicians directly responsible for a patient's care

personally knew and supervised all important, relevant decision-making. By contrast, after system installation, clinicians must obtain substantial system-related technical training and expertise to be able to access or alter the clinical knowledge underlying a vendor system's patient-specific recommendations.

Even more opaque are the hidden mechanisms for how specific knowledge is brought to bear during clinical decision support. For example, a system might have a "patient is pregnant" indicator. That indicator may trigger warnings when someone orders medications contra-indicated in pregnancy, or when physicians order radiological studies deemed unsafe for the condition. The ethical issue is that most vendor systems obscure the basis for how the "patient is pregnant" flag is set locally. For example, Hospital A might use the nurse's admission intake interview form to set the pregnancy flag. The underlying information gathered by the nurse in such a setting might be the patient's response to the question "Are you pregnant now?" The latter is inadequate for patient safety. Hospital B might use the result of a patient's beta-HCG test to set the pregnancy flag. While a more reliable indicator of pregnancy than hearsay from the patient, the beta-HCG test may not be ordered by every clinician in all relevant circumstances; hence, that mechanism may also be imperfect for decision support. A physician practicing at both Hospital A and Hospital B would be unlikely to know that the flags have different meanings at each site. Yet all the clinician can see is the status of the flag.

This ethical problem extends far beyond setting pregnancy flags. System vendors may claim to provide a wide range of patient-safety related decision support tools, but responsible care providers cannot trust system-generated advice when the triggers for decision support rules are potentially unreliable and inaccessible. Tort law requires clinician-providers to uphold the standard of care for their patients. Ignorance of the basis for important system-initiated clinical advice is inconsistent with upholding the standard of care. Certifying agencies should require that clinical system vendors make the basis for each decision-support rule transparent. Such transparency should require that, without special training, a clinical user could easily (and on request during system use) determine the underlying logic and data supporting each instance of patient-specific advice. Systems should also enable display of the evidential basis for other, more general advice – information that might become outdated over time. Vendors must expose, in non-technical terms, how decision-support triggers are locally determined at each site.

12.6 Summary and Conclusions

Ethical issues are important in biomedical informatics, and especially so in the clinical arena. An initial ensemble of guiding principles, or ethical criteria, has emerged to orient decision making:

1. Specially trained human beings (e.g., licensed practitioners) remain, so far, best able to provide health care for other human beings. Computer software developers should strive to warn caregivers whenever it appears that a mistake is imminent. However, because clinical practice involves as many exceptions as rules, software systems should not be allowed to overrule a clinician's decision once a warning has been issued.
2. Practitioners who use informatics tools should be clinically qualified and adequately trained in using the software products.
3. The tools themselves should be carefully evaluated and validated, in vitro and in vivo.
4. Health informatics tools and applications should be evaluated not only in terms of performance, including efficacy, but also in terms of their influences on institutions, institutional cultures, and workplace social forces.
5. Ethical obligations should extend to system developers, maintainers, and supervisors as well as to clinician users.
6. Education programs and security measures should be considered essential for

protecting confidentiality and privacy while improving appropriate access to personal patient information.
7. Adequate oversight should be maintained to optimize ethical use of electronic patient information for scientific and institutional research.

New sciences and technologies always raise interesting and important ethical issues. Much the same is true for legal issues, although in the absence of precedent or legislation any legal analysis will remain vague. Similarly, important challenges confront people who are trying to determine the appropriate role for government in regulating health care software. The lack of clear public policy for such software underscores the importance of ethical insight and education as the exciting new tools of biomedical and health informatics become more common.

Suggested Readings

Goodman, K. W. (2016b). Ethics, medicine, and information technology: Intelligent machines and the transformation of health care. Cambridge: Cambridge University Press. This volume contains material on informatics and human values, electronic health records, confidentiality and privacy, decision support, prognostic scoring systems, research, including the role of learning health systems, and governance of medical computer systems.

Goodman, K. W. (2020). Ethics and health informatics. International Yearbook of Medical Informatics. https://doi.org/10.1055/s-0040-1701966. This overview emphasizes issues raised by Big Data and Machine Learning/Artificial Intelligence.

Goodman, K. W., Berner, E. S., Dente, M. A., Kaplan, B., Koppel, R., Rucker, D., Sands, D. Z., & Winkelstein, P. (2011). Challenges in ethics, safety, best practices, and oversight regarding HIT vendors, their customers, and patients: a report of an AMIA special task force. Journal of the American Medical Informatics Association, 18(1), 77–81. This document is one of the first to examine issues related to the manufacturers and vendors of electronic health records, including their relationships with hospitals.

IOM (Institute of Medicine). (2012b). Health IT and patient safety: Building safer systems for better care. Washington, DC: The National Academies Press. A key National Academy of Medicine report linking appropriate use of EHRs and patient safety.

Miller, R. A. (1990b). Why the standard view is standard: People, not machines, understand patients' problems. Journal of Medicine and Philosophy. 15, 581–591. This contribution lays out the standard view of health informatics. This view holds, in part, that because only humans have the diverse skills necessary to practice medicine or nursing, machine intelligence should never override human clinicians.

Miller, R. A., Schaffner, K. F., Meisel, A. (1985b). Ethical and legal issues related to the use of computer programs in clinical medicine. Annals of Internal Medicine. 102, 529–536. This article constitutes a major early effort to identify and address ethical issues in informatics. By emphasizing the questions of appropriate use, confidentiality, and validation, among others, it sets the stage for all subsequent work.

Questions for Discussion

1. What is meant by the "standard view" of appropriate use of medical information systems? Identify three key criteria for determining whether a particular use or user is appropriate.
2. Can quality standards for system developers and maintainers simultaneously safeguard against error and abuse and stimulate scientific progress? Explain your answers. Why is there an ethical obligation to adhere to a standard of care?
3. Identify (a) two major threats to patient data confidentiality, and (b) policies or strategies that you propose for protecting confidentiality against these threats.
4. Many prognoses by human beings are subjective and are based on faulty memory or incomplete knowledge of previous cases. What are the two drawbacks

to using objective prognostic scoring systems to determine whether to allocate care to individual patients?

5. People who are educated about their illnesses tend to understand and to follow instructions, to ask insightful questions, and so on. How can the World Wide Web improve patient education? How, on the other hand, might Web access hurt traditional physician–patient and nurse–patient relationships?

Acknowledgments Reid Cushman, PhD, contributed to this chapter in an earlier edition. His comments are gratefully acknowledged.

References

Alpert, S. A. (1998). Health care information: Access, confidentiality, and good practice. In K. W. Goodman (Ed.), *Ethics, computing, and medicine: Informatics and the transformation of health care* (pp. 75–101). Cambridge: Cambridge University Press.

Anderson, J. G., & Aydin, C. E. (1994). Overview: Theoretical perspectives and methodologies for the evaluation of health care information systems. In J. G. Anderson, C. E. Aydin, & S. J. Jay (Eds.), *Evaluating health care information systems: Methods and applications* (pp. 346–354). Thousand Oaks: Sage.

Anderson, J. G., & Aydin, C. E. (1998). Evaluating medical information systems: Social contexts and ethical challenges. In K. W. Goodman (Ed.), *Ethics, computing, and medicine: Informatics and the transformation of health care* (pp. 57–74). Cambridge: Cambridge University Press.

Atreya, R. V., Smith, J. C., McCoy, A. B., Malin, B., & Miller, R. A. (2013). Reducing patient re-identification risk for laboratory results within research datasets. *Journal of the American Medical Informatics Association, 20*, 95–101.

Benitez, K., & Malin, B. (2010). Evaluating re-identification risks with respect to the HIPAA privacy rule. *Journal of the American Medical Informatics Association, 17*, 169–177.

Blois, M. S. (1980). Judgement and computers. *New England Journal of Medicine, 303*, 192–197.

Blumenthal, D. (2010). Launching HITECH. *New England Journal of Medicine, 362*(5), 382–385.

Brennan, P. F., Downs, S., & Casper, G. (2010). Project HealthDesign: Rethinking the power and potential of personal health records. *Journal of Biomedical Informatics, 43*(5 Suppl), S3–S5.

Brody, B. A. (1989). The ethics of using ICU scoring systems in individual patient management. *Problems in Critical Care, 3*, 662–670.

Classen, D. C., Avery, A. J., & Bates, D. W. (2007). Evaluation and certification of computerized provider order entry systems. *Journal of the American Medical Informatics Association: JAMIA, 14*(1), 48–55.

Curran, W. J., Stearns, B., & Kaplan, H. (1969). Privacy, confidentiality, and other legal considerations in the establishment of a centralized health-data system. *New England Journal of Medicine, 281*, 241–248.

Cushman, R., Froomkin, M. A., Cava, A., Abril, P., & Goodman, K. W. (2010). Ethical, legal and social issues for personal health records and applications. *Journal of Biomedical Informatics, 43*, S51–S55.

de Dombal, F. T. (1987). Ethical considerations concerning computers in medicine in the 1980s. *Journal of Medical Ethics, 13*, 179–184.

Duda, R. O., & Shortliffe, E. H. (1983). Expert systems research. *Science, 220*, 261–268.

FDA (U.S. Food and Drug Administration). (2011). *Infusion pump software safety research at FDA*. Retrieval 18 November 2020: http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/GeneralHospitalDevicesandSupplies/InfusionPumps/ucm202511.htm.

FDASIA. (2014). *FDASIA health IT report: Proposed strategy and recommendations for a risk-based framework*. Retrieval 18 November 2020: https://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHReports/UCM391521.pdf

Friedman, C. P. (2009). A "fundamental theorem" of biomedical informatics. *Journal of the American Medical Informatics Association, 16*(2), 169–170.

Geissbuhler, A. J., & Miller, R. A. (1997). Desiderata for product labeling of medical expert systems. *International Journal of Medical Informatics, 47*(3), 153–163.

Glaser, J. (2010). HITECH lays the foundation for more ambitious outcomes-based reimbursement. *The American Journal of Managed Care, 16*(12), SP19–SP23.

Goodman, K. W. (1996). Ethics, genomics and information retrieval. *Computers in Biology and Medicine, 26*, 223–229.

Goodman, K. W. (1998). Outcomes, futility, and health policy research. In K. W. Goodman (Ed.), *Ethics, computing, and medicine: Informatics and the transformation of health care* (pp. 116–138). Cambridge: Cambridge University Press.

Goodman, K. W. (2016a). *Ethics, medicine, and information technology: Intelligent machines and the transformation of health care*. Cambridge: Cambridge University Press.

Goodman, K. W. (2017). Health information technology as a universal donor to bioethics education.

Cambridge Quarterly of Healthcare Ethics, 26(2), 342–347.

Goodman, K. W., & Cava, A. (2008). Bioethics, business ethics, and science: Bioinformatics and the future of healthcare. *Cambridge Quarterly of Healthcare Ethics, 17*(4), 361–372.

Goodman, K. W., Berner, E. S., Dente, M. A., Kaplan, B., Koppel, R., Rucker, D., Sands, D. Z., & Winkelstein, P. (2010). Challenges in ethics, safety, best practices, and oversight regarding HIT vendors, their customers, and patients. *Journal of the American Medical Informatics Association, 18*(1), 77–81.

Halamka, J. D., & Tripathi, M. (2017). The HITECH era in retrospect. *New England Journal of Medicine, 310*, 907–909.

Hartzband, P., & Groopman, J. (2008). Off the record—Avoiding the pitfalls of going electronic. *New England Journal of Medicine, 358*(16), 1656–1658.

Holroyd-Leduc, J. M., Lorenzetti, D., Straus, S. E., Sykes, L., & Quan, H. (2011). The impact of the electronic medical record on structure, process and outcomes within primary care: A systematic review of the evidence. *Journal of the American Medical Informatics Association, 18*, 732–737.

Ienca, M., & Vayena, E. (2020). On the responsible use of digital data to tackle the COVID-19 pandemic. *Nature Medicine, 26*, 463–464.

IOM (Institute of Medicine). (2012a). *Health IT and patient safety: Building safer systems for better care*. Washington, DC: The National Academies Press.

Kaplan, B., & Harris-Salamone, K. D. (2009). Health IT success and failure: Recommendations from literature and an AMIA workshop. *Journal of the American Medical Informatics Association, 16*, 291–299.

Knaus, W. A., Wagner, D. P., & Lynn, J. (1991). Short-term mortality predictions for critically ill hospitalized adults: Science and ethics. *Science, 254*, 389–394.

Koppel, R., & Kreda, D. (2009). Health care information technology vendors' "hold harmless" clause: Implications for patients and clinicians. *JAMA, 301*, 1276–1278.

Kuperman, G. J., & Gibson, R. F. (2003). Computer physician order entry: Benefits, costs, and issues. *Annals of Internal Medicine, 139*(1), 31–39.

Lane, W. A. (1936). What the mouth reveals. *New Health, 11*, 34–35.

Levis, J. (Ed.). (2014). *H.I.T. or miss: Lessons learned from health information technology implementations*. Chicago: American Health Information Management Association.

Macklin, R. (1992). Privacy and control of genetic information. In G. J. Annas & S. Elias (Eds.), *Gene mapping: Using law and ethics as guides* (pp. 157–172). New York: Oxford University Press.

Majchrowski, B. (2010). Medical software's increasing impact on healthcare and technology management. *Biomedical Instrumentation and Technology, 44*(1), 70–74.

Malin, B., & Goodman, K. W. (2018). Between access and privacy: Challenges in sharing health data. *IMIA Yearbook of Medical Informatics, 27*, 55–59.

Malin, B., & Sweeney, L. (2004). How (not) to protect genomic data privacy in a distributed network: Using trail re-identification to evaluate and design anonymity protection systems. *Journal of Biomedical Informatics, 37*, 179–192.

Malin, B., Loukides, G., Benitez, K., & Clayton, E. W. (2011). Identifiability in biobanks: Models, measures, and mitigation strategies. *Human Genetics, 130*, 383–392.

McNeer, J. F., Wallace, A. G., Wagner, G. S., Starmer, C. F., & Rosati, R. A. (1975). The course of acute myocardial infarction: Feasibility of early discharge of the uncomplicated patient. *Circulation, 51*, 410–413.

Miller, R. A. (1989). Legal issues related to medical decision support systems. *International Journal of Clinical Monitoring and Computing, 6*, 75–80.

Miller, R. A. (1990a). Why the standard view is standard: People, not machines, understand patients' problems. *The Journal of Medicine and Philosophy, 15*, 581–591.

Miller, R. A. (1997). Predictive models for primary caregivers: Risky business? *Annals of Internal Medicine, 127*(7), 565–567.

Miller, R. A., & Gardner, R. M. (1997a). Summary recommendations for the responsible monitoring and regulation of clinical software systems. *Annals of Internal Medicine, 127*(9), 842–845.

Miller, R. A., & Gardner, R. M. (1997b). Recommendations for responsible monitoring and regulation of clinical software systems. *Journal of the American Medical Informatics Association, 4*, 442–457.

Miller, R. A., & Goodman, K. W. (1998). Ethical challenges in the use of decision-support software in clinical practice. In K. W. Goodman (Ed.), *Ethics, computing, and medicine: Informatics and the transformation of health care* (pp. 102–115). Cambridge: Cambridge University Press.

Miller, R. A., & Miller, S. M. (2007). Legal and regulatory issues related to the use of clinical software in health care delivery. In R. A. Greenes (Ed.), *Clinical decision support: The road ahead* (pp. 423–444). Boston: Elsevier.

Miller, R. A., Schaffner, K. F., & Meisel, A. (1985a). Ethical and legal issues related to the use of computer programs in clinical medicine. *Annals of Internal Medicine, 102*, 529–536.

National Research Council. (1997). *For the record: Protecting electronic health information*. Washington, D.C.: National Academy Press.

Palfrey, J., & Gasser, U. (2010). *Born digital: Understanding the first generation of digital natives*. New York: Basic Books.

Pugh, J., Pycroft, L., Sandberg, A., Aziz, T., & Savulescu, J. (2018). Brainjacking in deep brain stimulation and autonomy. *Ethics and Information Technology, 20*(3), 219–232.

Robertson, J. (2011). Insulin pumps, monitors vulnerable to hacking. Associated Press via various media, e.g., The Washington Times, August 4. Retrieval November 17, 2020: http://www.washingtontimes.com/news/2011/aug/4/insulin-pumps-monitors-vulnerable-to-hacking/?page=all.

Sackner-Bernstein, J. (2017). Design of hack-resistant diabetes devices and disclosure of their cyber safety. *Journal of Diabetes Science and Technology, 11*(2), 198–202.

Shih, S. C., McCullough, C. M., Wang, J. J., Singer, J., & Parsons, A. S. (2011). Health information systems in small practices. Improving the delivery of clinical preventive services. *American Journal of Preventive Medicine, 6*, 603–609.

Shortliffe, E. H. (1993). Doctors, patients, and computers: Will information technology dehumanize health-care delivery? *Proceedings of the American Philosophical Society, 137*(3), 390–398.

Shortliffe, E. H. (1994). Dehumanization of patient care. Are computers the problem or the solution? *Journal of the American. Medical Informatics Association, 1*, 76–78.

Sittig, D. F., Krall, M., Kaalaas-Sittig, J., & Ash, J. S. (2005). Emotional aspects of computer-based provider order entry: A qualitative study. *Journal of the American Medical Informatics Association, 12*(5), 561–567.

Sittig, D. F., & Singh, H. (2016). A socio-technical approach to preventing, mitigating, and recovering from ransomware attacks. *Applied Clinical Informatics, 7*(2), 624–632.

Slayton, T. B. (2018). Ransomware: The virus attacking the healthcare industry. *Journal of Legal Medicine, 38*(2), 287–311.

Sweeney, L. (1997). Weaving technology and policy together to maintain confidentiality. *The Journal of Law, Medicine Ethics, 25*, 98–110.

Szolovits, P., & Pauker, S. G. (1979). Computers and clinical decision making: Whether, how much, and for whom? *Proceedings of the IEEE, 67*, 1224–1226.

Tamersoy, A., Loukides, G., Nergiz, M. E., Saygin, Y., & Malin, B. (2012). Anonymization of longitudinal electronic medical records. *IEEE Transactions on Information Technology in Biomedicine, 16*, 413–423.

Truog, R. D., Mitchell, C., & Daley, G. Q. (2020). The toughest triage — Allocating ventilators in a pandemic. *New England Journal of Medicine, 382*, 1973–1975.

Tysyer, D. A. (1997). Database legal protection. *Bitlaw*. Retrieval 18 November 2020: http://www.bitlaw.com/copyright/database.html

Vergese, A., Shah, N. H., & Harrington, R. A. (2018). What this computer needs is a physician. *Journal of the American Medical Association, 319*(1), 19–20.

Wright, A., Sittig, D. F., Ash, J. S., Sharma, S., Pang, J. E., & Middleton, B. (2009). Clinical decision support capabilities of commercially-available clinical information systems. *Journal of the American Medical Informatics Association, 16*(5), 637–644.

Youngner, S. J. (1988). Who defines futility? *JAMA, 260*, 2094–2095.

Evaluation of Biomedical and Health Information Resources

Charles P. Friedman and Jeremy C. Wyatt

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_13

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- Why are empirical studies based on the methods of evaluation and technology assessment important to the successful implementation of information resources to improve health?
- What challenges make studies in informatics difficult to carry out? How are these challenges addressed in practice?
- Why can all evaluations be classified as empirical studies?
- What features do all evaluations have in common?
- What are the key factors to take into account as part of a process of deciding what are the most important questions to use to frame a study?
- What are the major assumptions underlying objectivist and subjectivist approaches to evaluation? What are the strengths and weaknesses of each approach?
- How does one distinguish measurement and demonstration aspects of objectivist studies, and why are both aspects necessary? In the demonstration aspect of objectivist studies, how are control strategies used to draw inferences?
- What steps are followed in objectivist and subjectivist studies? What techniques are employed by investigators to ensure rigor and credibility of their findings?
- Why is communication between investigators and stakeholders central to the success of any evaluation?

13.1 Introduction

Most people understand the term evaluation to mean an assessment of an organized, purposeful activity. Evaluations are usually conducted to answer questions or in anticipation of the need to make decisions (Wyatt and Spiegelhalter 1990; Ammenwerth 2015). Evaluations may be informal or formal, depending on the characteristics of the decision to be made and, particularly, how much is at stake. But all activities labeled as evaluation involve the empirical process of collecting information that is relevant to the decision at hand. For example, when choosing a holiday destination, members of a family may informally ask friends which Hawaiian island they prefer and browse various websites including those that provide ratings of specific destinations. After factoring in costs and convenience, the family reaches a decision. More formally, when a health care organization faces the choice of a new electronic health record system, the leadership will develop a plan to collect comparable data about competing systems, analyze the data according to the plan, and ultimately, through a predetermined process, make a decision.

The field of biomedical and health informatics focuses on the collection, processing, and communication of health-related information and the implementation of **information resources**—usually consisting of digital technology designed to interact with people—to facilitate these activities.[1] These information resources can collect, store, and process data related to the health of individual persons (institutional or personal electronic health records), manage and reason about biomedical knowledge (knowledge acquisition tools, knowledge bases, decision-support systems, and intelligent tutoring systems), and support activities related to public health (disease registries and vital statistics, disease outbreak detection and tracking). Thus, there is a vast range of biomedical and health information resources that can be foci of evaluation.

Information resources have many different aspects that can be studied (Friedman and Wyatt 2005; Chap. 3). Where safety is an issue, as it often is, (Fox 1993; Black et al. 2011; Russ et al. 2014), we might focus on

1 In this chapter, we will use the terms "information resource" and "information system" generally as synonyms. However, "information system" applies more specifically to applications of digital technology whereas a "resource" is a broad term that could, for example, include informal collegial consultations.

inherent characteristics of the resource, asking such questions as, "Are the code and architecture compliant with current software engineering standards and practices?" or "Is the data structure the optimal choice for this type of application?" Clinicians and patients, however, might ask more pragmatic questions such as, "Is the knowledge in this system completely up-to-date?" or "Who can access this information besides me?" Executives and public officials might wish to understand the effects of these resources on individuals and populations, asking questions such as, "Has this resource improved the quality of care?" or "What effects will a patient portal have on working relationships between practitioners and patients?" Thus, evaluation methods in biomedical informatics must address a wide range of issues, from technical characteristics of specific systems to systems' effects on people and organizations. The outcomes or effects attributable to the use of health information resources will almost always be a function of how individuals choose to use them, and the social, cultural, organizational, and economic context in which these uses take place (Lundsgaarde 1987).

For these reasons, there is no formula for designing and executing evaluations; every evaluation, to some significant degree, must be custom-designed. A major factor shaping the design of evaluations is the decisions the evaluation is expected to inform. In the end, choices about what evaluation questions to pursue and how to collect and analyze data to pursue them, are exquisitely sensitive to each study's special circumstances and constrained by the resources that are available for it. Evaluation is very much the art of the possible. But neither is evaluation an exercise in alchemy, pure intuition, or black magic. There exist many methods for evaluation that have stood the test of time and proved useful in practice. There is a literature on what methods work and under specific circumstances, and there are numerous published examples of successful evaluation studies. In this chapter, we will introduce many of these methods, and present frameworks that guide the application of methods to specific decision problems and study settings.

13.2 Why Are Formal Evaluation Studies Needed?

13.2.1 Computing Artifacts Have Special Characteristics

Why are empirical studies of information resources needed at all? Why is it not possible, for example, to model (and thus predict) the performance of information resources and their impact on users, and thus save a lot of time and effort? The answer lies, to a great extent, in the complexity of computational artifacts and their use. For some disciplines, specification of the structure of an artifact allows one to predict how it will function, and engineers can even design new objects with known performance characteristics directly from functional requirements. Examples of such artifacts are elevators and conventional road bridges. The principles governing the behavior of materials and structures made of specific materials are sufficiently well understood that a new elevator can be designed to a set of performance characteristics with the expectation that it will perform exactly as predicted. Laboratory testing of models of these devices is rarely needed. Field testing of the artifact, once built, is conducted to reveal relatively minor anomalies, which can be rapidly remedied, or to tune or optimize performance. However, when the object concerned is a computer-based resource, not an elevator, the story is different (Littlejohns et al. 2003). Software designers and engineers have theories linking the structure to the function of only the most trivial computer-based resources (Somerville 2002). Because of the complexity of computer-based systems themselves, their position as part of a complex socio-technical system including the users and the organization in which they work, and the lack of a comprehensive theory connecting structure and function, there is no way to know exactly how an information resource will perform until it is built and tested (Murray et al. 2004); and similarly there is no way to know that any revisions will bring about the desired effect until the next version of the resource is tested. It is also impossible to predict how even a perfectly

functioning information resource will impact user decisions or actions.

In sum, the only practical way to determine if a reasonably complex body of computer code does what it is intended to do is to test it in the laboratory and in the field. This testing can take many shapes and forms. The informal design, test, and revise activity that characterizes the development of all computer software is one such form of testing and results in software that usually functions as expected *by the developers*. More formal and exhaustive approaches to software design, verification and testing using synthetic test cases (e.g., Scott et al. 2011) and other approaches help to guarantee that the software will do what it was designed to do. Even these approaches, however, do not guarantee the success of the software when put into the hands of the intended end-users. This requires more formal studies of the types that will be described in this chapter, which can be undertaken before, during, and after the initial development of an information resource. Such evaluation studies can guide further development; indicate if the resource is likely to be safe for use in real health care, public health, research, or educational settings; or elucidate if it has the potential to improve the professional performance of the resource users and the health of individuals and populations. Many stakeholders wish to know if the resource, as actually used in practice, has had the intended beneficial effects.

Many other writings elaborate on the points offered here. Some of the earliest include Spiegelhalter (1983) and Gaschnig et al. (1983) who discussed these phases of evaluation by drawing analogies from the evaluation of new drugs or the conventional software life cycle, respectively. Wasson et al. (1985) discussed the evaluation of clinical prediction rules together with some useful methodological standards that apply equally to information resources. Many other authors since then have described, with differing emphases, the evaluation of health care information resources, often focusing on decision-support tools, which pose some of the most extreme challenges. One relevant book (Friedman and Wyatt 2005) discusses the challenges posed by evaluation in biomedical informatics and offers a wide range of methods described in considerable detail to help investigators explore and resolve these challenges. Other books have explored more technical, health technology assessment or organizational approaches to evaluation methods (Szczepura and Kankaanpaa 1996; van Gennip and Talmon 1995; Anderson et al. 1994; Brender 2005; Harasevich and Pickering 2017).

13.2.2 The Special Issue of Safety

Before disseminating any biomedical information resource that stores and communicates health data or knowledge and is designed to influence real-world practice or personal health decisions, it is important to verify that the resource is safe when used as intended. In the case of new drugs, European and US regulators have imposed a statutory duty on developers to perform extensive in vitro testing, and in vivo testing in animals, before any human receives a dose of the drug. Since 2000, the safety of biomedical information resources has come increasingly into the spotlight (Rigby et al. 2001; Koppel et al. 2005). Accordingly, testing of information resources is now being considered, with governmental agencies imposing risk-based regulatory frameworks and clearer classifications of medical devices (Slight and Bates 2014; FDASIA 2014; EU Regulatory Framework 2018). For biomedical information resources, safety tests analogous to those required for drugs would include assessment of the accuracy of the data stored and retrieved, measuring the accuracy of any risk estimate or advice from a decision support system, determining whether and how easily end-users can employ the resource for its intended purposes, and estimating how often the resource furnishes misleading or incorrect information (Eminovic et al. 2004). It may be necessary to repeat these assessments following any substantial modifications to the information resource, as the correction of safety-related problems may itself generate new problems or uncover previously unrecognized ones.

Determining if an information resource is safe and effective goes fundamentally to the

process of evaluation we address in this chapter. Almost all of the methodological issues we raise apply to safety assessments. Casual assessments that fail to address these issues will not resolve the safety question, and will not reveal safety defects that can be remedied. Many of these issues are issues of sampling that we introduce in ▶ Sect. 15.4.2. For example, the advice or other "output" generated by most information resources depends critically on the quality and quantity of data available to it and on the manner in which the resource is used by patients or practitioners. People or practitioners who are untrained, in a hurry, or exhausted at 3 A.M., are more likely to fail to enter key data that might lead to the resource generating misleading advice, or to fail to heed an alarm that is not adequately emphasized by the user interface. Coded data automatically entered into resources may be inaccurate, incomplete, or not coded in the manner anticipated by the resource. Thus, to generate valid results, functional tests must put the resources in actual users' hands under the most realistic conditions possible, or in the hands of people with similar knowledge, skills and experience if samples of intended users are not available. For example, a Facebook advertisement for an app to help women detect when conception was likely from body temperature readings was withdrawn in the UK because the accuracy quoted in publicity materials related to ideal use rather than use in everyday practice.[2] [BBC news story 29-8-18].

Other safety issues are, from a methodological perspective, issues of measurement that we address in ▶ Sect. 13.4.2. For example, should "usability" of an information resource be determined by documenting that the resource development process followed best practices to inculcate usability, asking end-users if they believed the resource was usable, or by documenting and studying their "click streams" to determine if end-users actually navigated the resource as the designers intended? There is no single clear answer to this question (see Jakob Nielsen's invaluable resource on user testing[3]), but we will see that all measurement processes have features that make their results more or less dependable and useful. We will also see that the measurement processes built into evaluation studies can themselves be designed to make the results of the studies more helpful to all stakeholders, including those focused on safety.

13.3 Two Universals of Evaluation

13.3.1 The Full Range of What Can Be Formally Studied

Deciding what to study is fundamentally a process of winnowing down from a universe of potential questions to a parsimonious set of questions that can be realistically addressed given the priorities, time, and resources available. This winnowing process can begin with the full range of what can potentially be studied. To both ensure that the most important questions do get "on the table" and to help eliminate the less important ones, it can be useful to start with a comprehensive list. While experienced evaluators do not typically begin study planning from this broadest perspective, it is always helpful to have a broad range of options in mind.

There are five major aspects of an information resource that can be studied:

1. Need for the resource: In advance of any development, investigators can study the status quo *absent* the resource, including the nature of problems the resource is intended to address and how frequently these problems arise. (When an information resource is already deployed, the "status quo" might be the currently deployed resource, and the resource under study is a proposed replacement for it or enhancement to it.)
2. Design and development process: Investigators study the skills of the devel-

2 ▶ https://www.bbc.co.uk/news/technology-45328965 (Accessed 11.20.19).

3 ▶ https://www.nngroup.com/articles/ (Accessed 11.20.19).

opment team, and the development methodologies employed by the team, to understand if the resulting resource is likely to function as intended.
3. Resource static structure: Here the focus of the evaluation includes specifications, flow charts, program code, and other representations of the resource that can be inspected without actually running it.
4. Resource usability and dynamic functions: The focus is on whether the resource has the potential to be beneficial: the degree to which intended end-users can navigate the resource and how it performs when it is used in pilots prior to full deployment.
5. Resource use, effect and impact: Finally, after deployment, the focus switches from the resource itself to the extent of its use and its effects on professional, patient or public users, and on health care organizations.

In a theoretically "complete" evaluation, sequential studies of a particular resource might address all of these aspects, over the life cycle of the resource. In the real world, however, it is difficult, and rarely necessary, to be so comprehensive. Over the course of its development and deployment, a resource may be studied many times, with the studies in their totality touching on many or most of these aspects. Some aspects of an information resource will be studied informally using anecdotal data collected via casual methods. Other aspects will be studied more formally in ways that are purposefully designed to inform specific development decisions and that involve systematic collection and analysis of data. Distinguishing those aspects that will be studied formally from those left for informal exploration is a challenging task facing all evaluators.

13.3.2 The Structure of All Evaluation Studies, Beginning with a Negotiation Phase

If the list offered in the previous section can be seen as the universe of what can be studied, Fig. 13.1 can be used as a framework for planning all evaluation studies. The first stage in any study is negotiation between the "investigators" (or "evaluators") who will be carrying out the study and the "stakeholders" who have interests in or otherwise will be concerned about the study results. Before a study can proceed, the key stakeholders who are supporting the study financially and providing other essential resources for it—such as the institution where the information resource is deployed—must be satisfied with the general plan. The negotiation phase identifies the broad aim and objectives of the study, what kinds of reports and other deliverables will result and by when, where the study personnel will be based, the resources available to conduct the study, and any constraints on what can be studied. When a study of an information resource is being conducted internally—that is, when all of the key stakeholders represent one organization that also employs the investigators—it is still very useful to have an internal negotiation to lay out details of the study.

The results of the negotiation phase are expressed in a document, generally known as a contract between the evaluators and the key stakeholders. The contact guides the planning and execution of the study and, in a very significant way, protects all parties from misunderstandings about intent and execution. Like any contract, an **evaluation contract** can be changed later with consent of all parties.

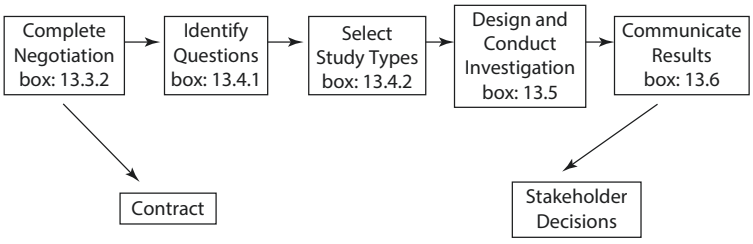


Fig. 13.1 Generic structure of all evaluation studies

Following the negotiation process and its reflection in a contract, the planning of the evaluation proceeds in a sequence of logical steps, starting with the formulation of specific questions to be addressed, then the selection of the type(s) of study that will be used, the investigation that entails the collection and analysis of data, and ultimately the communication back to the stakeholders of the findings, which typically inform a range of decisions. Although ◘ Fig. 13.1 portrays a one-way progression through this sequence of stages, in the real world of evaluation there are often detours and backtracks.

13.4 Deciding What to Study and What Type of Study to Do: Questions and Study Types

13.4.1 The Importance of Identifying Questions

Once the study's objective, scope and other applicable "ground rules" have been established, the real work of study planning can begin. The next step, as suggested by ◘ Fig. 13.1, is to convert the perspectives of the concerned parties, and what these individuals or groups want to know, into a finite, specific set of questions. It is important to recognize that, for any evaluation setting that is interesting enough to merit formal evaluation, the number of potential questions is infinite. This essential step of identifying a tractable number of questions has a number of benefits:

- It helps to crystallize thinking of both investigators and key members of the audience who are the stakeholders in the evaluation.
- It guides the investigators and stakeholders through the critical process of assigning priority to certain issues and thus productively narrowing the focus of a study.
- It converts broad statements of aim (e.g., "*to evaluate a new order communications system*") into specific questions that can potentially be answered (e.g., "*What is the impact of the order communications system on how clinical staff spend their time, the rate and severity of adverse drug events and the length of patient stay*?").
- It allows different stakeholders in the evaluation process—patients, professional groups, managers – to see the extent to which their own concerns are being addressed, and to ensure that these feed into the evaluation process.
- Most important, perhaps, it is hard if not impossible to develop investigative methods without first identifying questions, or at least focused issues, for exploration. The choice of methods follows from the evaluation questions: not from the novel technology powering the information resource or the type of resource being studied. Unfortunately, some investigators choose to apply the same set of the methods to any study, irrespective of the questions to be addressed, or even to limit the evaluation questions addressed to those compatible with the methods they prefer. We do not endorse this limiting approach.

Consider the distinction made earlier between informal evaluations that people undertake continuously as they make choices as part of their everyday personal or professional lives, and more formal evaluations that are planned and then executed according to that plan. In short, formal evaluations are those that conform to the architecture of ◘ Fig. 13.1. In these formal evaluations, the questions that actually get addressed survive a narrowing process that begins with a broad set of candidate questions. When starting a formal evaluation, therefore, a major decision is whom to consult to establish the questions that will get "on the table", how to log and analyze their views, and what weight to place on each of these views. There is always a wide range of potential players in any evaluation (see ▶ Box 13.1) and there is no formula that defines whom to consult or in what order. Through this process, the investigators apply their common sense and, with experience, learn to follow their instincts. The only universal mistake is to fail to consult one or more of the key stakeholders, especially those paying for the study or those ultimately

making the key decisions to be informed by the evaluation. It is often useful to establish a group to advise and guide the evaluators, a group with broad representation that will help ensure that study remains true to the interests and preferences of the stakeholders.

Through discussions with various stakeholder groups, the hard decisions regarding the questions to be addressed in the study are made. A significant challenge for investigators is the risk of getting swamped by detail, resulting from the multiplicity of questions that can be asked in any study. To manage through the process, it is important to reflect on the major issues identified after each round of discussions with stakeholders, and then identify the questions that map to these issues. Where possible, keep questions at the same level of granularity.

It is critical that the specific questions serving as the beacon guiding the study be determined and endorsed by all key stakeholders, before any significant decisions about the detailed design of the study are made. We will see later that evaluation questions can, in many circumstances, change over the course of a study; but that fact does not obviate the need to specify a set of questions at the outset. ▶ Appendix A describes two evaluation scenarios, and suggests some evaluation questions that may be appropriate for each.

Box 13.1 Some of the Potential Players in an Evaluation Study

- Those commissioning the evaluation study, who will typically have questions or decisions that rely on the data collected
- Those paying for the evaluation study
- Those paying for the development and/or deployment of the information resource
- End-users of the resource, who are often providers of data for the study
- Developers of the resource and their managers
- Care providers and their managers
- Staff responsible for resource implementation and user training
- Information technology staff and leaders in the organization where the resource is deployed
- Senior managers in the organization where the resource is deployed
- The patients whose care the resource may directly or indirectly influence
- Staff in ancillary services whose workload may be affected by resource deployment, for example laboratory or imaging departments following deployment of a diagnostic decision support system
- Quality improvement and safety professionals in the organization in which the resource is implemented

13.4.2 Selecting a Study Type

After developing the list of evaluation questions, the next step is to understand which study type(s) the evaluation questions naturally invoke. The study types we will introduce in this chapter are specific to the evaluation of information resources, and are particularly informative to the design of evaluation studies in biomedical informatics. These study types are described below and also summarized in ◘ Table 13.1. The second column of ◘ Table 13.1 links the study types to the aspect of the resource that is studied—as previously introduced in ▶ Sect. 13.3.1. Each study type is likely to appeal to certain interests of particular stakeholders, as suggested in the rightmost column of the table. A wide range of data collection and analysis methods, as discussed later in ▶ Sect. 13.5, can be used to answer the questions embraced by all nine study types. Choice of a study type typically does not constrain the methods that can be used to collect and analyze data. And we will see later in this chapter, and specifically in ▶ Sect. 13.5.2.2, that all of these study types are what can be called demonstration studies, in contrast to so-called measurement studies. Finally, a set of studies, exemplifying many of these study types, is introduced and described in ▶ Appendix B.

Table 13.1 Classification of demonstration study types by broad study question and the stakeholders most concerned

Study type	Aspect studied	Broad study question	Audience/stakeholders primarily interested in results
1. Needs assessment	Need for the resource	What is the problem?	Resource developers, funders of the resource
2. Design validation	Design and development process	Is the development method in accord with accepted practices?	Funders of the resource; professional and governmental certification agencies e.g., Food and Drug Administration, Office of the National Coordinator for HIT
3. Structure validation	Resource static structure	Is the resource appropriately designed to function as intended?	Professional indemnity insurers, resource developers; professional and governmental certification agencies
4. Usability test	Resource dynamic usability and function	Can intended users navigate the resource so it carries out intended functions?	Resource developers, users, funders
5. Laboratory function study	Resource dynamic usability and function	Does the resource have the potential to be beneficial?	Resource developers, funders, users, academic community
6. Field function study	Resource dynamic usability and function	Does the resource have the potential to be beneficial in the real world?	Resource developers, funders, users
7. Lab user effect study	Resource effect and impact	Is the resource likely to change user behavior?	Resource developers and funders, users
8. Field user effect study	Resource effect and impact	Does the resource change actual user behavior in ways that are positive?	Resource users and stakeholders, resource purchasers and funders
9. Problem impact study	Resource effect and impact	Does the resource have a positive impact on the original problem?	The universe of stakeholders

1. **Needs assessment** studies seek to clarify the information problem the resource is intended to solve. These studies take place before the resource is designed–usually in the setting where the resource is to be deployed, although simulated settings may sometimes be used. Ideally, the potential users of the resource will be studied while they work with real problems or cases, to understand better how information is used and managed, and to identify the causes and consequences of inadequate information flows. The investigator seeks to understand users' skills, knowledge and attitudes, as well as how they make decisions or take actions. To ensure that developers have a clear model of how a proposed information resource will fit with working practices and structures, they may also need to study health care or research processes, team functioning, or relevant aspects of the larger organization in which work is done (Wyatt et al. 2010).
2. **Design validation** studies focus on the quality of the processes of information resource design and development, for example by asking experts to review these processes. The experts may review docu-

ments, interview the development team, compare the suitability of the software engineering methodology and programming tools used with others that are available, and generally apply their expertise to identify potential shortcomings in the approach used to develop the software, as well as constructively to suggest how these shortcomings might be corrected.

3. **Structure validation** studies address the static form of the software, usually after a first prototype has been developed. This type of study is most usefully performed by an expert or a team of experts with experience in developing software for the problem domain and concerned users. For these purposes, the investigators need access to both summary and detailed documentation about the system architecture, the structure and function of each module, and the interfaces among them. The expert might focus on the appropriateness of the algorithms that have been employed and check that they have been correctly implemented by examining the code and its documentation. Experts might also examine the data structures (e.g., whether they are appropriately normalized) and knowledge bases (e.g., whether they are evidence-based, up to date, and modelled in a format that will support the intended analyses or reasoning). Most of this will be done by inspection and discussion with the development team. Sometimes specialized software may be used to test the structure of the resource (Somerville 10th edition 2015).

   Note that the study types listed up to this point do not require a functioning information resource. However, beginning with usability testing below, the study types require the existence of at least a functioning prototype.

4. **Usability testing** studies address whether intended users can actually operate or navigate the software, to determine whether the resource has the potential to be helpful to them (see also ▶ Chap. 5). In this type of study, testing of a prototype by typical users informs further development and should improve its usability. Although usability testing can be performed by obtaining opinions of usability experts who "test drive" the resource, usability can also be tested by deploying the resource in a laboratory or classroom setting, introducing users to it, and then allowing them either to navigate at will and provide unstructured comments or to attempt to complete some scripted tasks (see extensive material by Nielsen, ▶ www.useit.com). Data can be collected by the computer itself, from the user, by a live observer, via audio or video capture of users' actions and statements, or by specialized instrumentation such as eye-tracking tools. Many software developers have usability testing labs equipped with sophisticated measurement systems, staffed by experts in human computer interaction to carry out these studies—an indication of the importance increasingly attached to this type of study (Zhang et al. 2003; Saitwal et al. 2010).
5. **Laboratory function studies** go beyond usability to explore more specific aspects of the information resource, such as the quality of data captured, the speed of communication, the validity of the calculations carried out, or the appropriateness of the results or advice given. These functions relate less to the basic usability of the resource and more to how the resource performs in relation to what it is trying to achieve for the user or the organization. When carrying out any kind of function testing, real or proxy users are employed. The study results depend crucially on what problems the users are asked to solve, so the "tasks" employed in these studies (eg. case scenarios) should correspond as closely as possible to those to which the resource will be applied in real working life. *Such tasks very across a set of dimensions—for example, difficulty, problem domain, and urgency—so it is very important to employ in these studies a set of tasks that spans the range of these dimensions.*
6. **Field function studies** are a variant of laboratory function testing in which the

resource is "pseudo-deployed" in a real work place and employed by real users with real problems—but only up to a point. In field function tests, although the resource is used by real users with real tasks, the users have no immediate access to the output or results of their interaction with the resource that might influence their real decisions or actions, so no effects on these can occur. The output is recorded for later review by the investigators, and perhaps by the users themselves.

Studies of the effect or impact of information resources on users and problems are in many ways the most demanding. As the focus of study moves from function testing, which is always hypothetical, to possible effects on health decisions or care processes, the conduct of research, or educational practice, there is often the need to establish cause and effect and to submit studies to external review.

7. In **laboratory user effect studies**, simulated user decisions or actions are studied. Practitioners employ the resource in a laboratory setting and are asked what they "would do" with the results or advice the resource generates, but no action is taken. Laboratory user effect studies can be conducted with prototype or released versions of the resource, outside the practice environment. Although such studies involve individuals who are representative of the "end-user" population, the primary results of the study derive from simulated actions, so the care of patients or conduct of research is not affected by a study of this type. An example is a study in which junior physicians viewed realistic prescribing scenarios and interacted with a simulated prescribing tool while they were exposed to simulated prescribing alerts of various kinds and the rate of prescribing errors was measured (Scott et al. 2011).
8. In a **field user effect study**, the actual actions or decisions of the users of the resource are studied after the resource is formally deployed. This type of study provides an opportunity to test whether the resource is actually used by the intended users, whether they obtain accurate and useful information from it, and whether this use affects their decisions and actions in significant ways. In field user effect studies, the emphasis is on the behaviors and actions of users, and not the health outcomes or consequences of these behaviors. For example, one study examined the impact of SMS reminders on anti-retroviral medication adherence in Africans with HIV and showed a dramatic improvement (Lester et al. 2010).
9. **Problem impact studies** are similar to field user effect studies in many respects, but differ profoundly in the questions that are the focus of exploration. Problem impact studies examine the extent to which the original health problem that motivated creation or deployment of the information resource has been addressed. Often this requires investigation that looks beyond the actions of care providers, researchers, or patients to examine the consequences of these actions. In the Lester study of SMS alerts (Lester et al. 2010), increased adherence to antiretroviral therapy (a user action) was also accompanied by improved viral load suppression. However, user effects cannot be assumed to engender problem impacts. For example, an information resource designed to reduce medication errors may affect the behavior of clinicians who employ the resource relative to those who do not, but for a variety of reasons, the actual incidence of harmful medication episodes remains unchanged. In such an instance, clinical pharmacists who review orders may be catching and correcting these errors before patients are affected. In other examples, individuals may be motivated to exercise through interaction with a wearable information resource but fail to meet weight loss objectives because they cannot afford concomitant changes in their diets. In still other domains, an information resource may be widely used by researchers to access biomedical information, as determined by a user effect study, but a subsequent problem impact study may or may not reveal effects on scientific productivity.

New educational technology may change the ways students learn, but may or may not increase their performance on standardized examinations. Problem impact studies, as well as user effect studies, will be sensitive to unintended consequences. Sometimes, the solution to the target problem creates other, unintended and unanticipated problems that can affect perceptions of success. As electronic mail became an almost universal mode of communication, almost no one anticipated the problems of "spam" or "phishing".

13.4.3 Factors Distinguishing the Nine Study Types

Table 13.2 further distinguishes the nine study types, as described above, using a set of key differentiating factors discussed in detail in the paragraphs that follow.

The setting in which the study takes place Studies of the design process, the resource structure, and many resource functions are typically conducted outside the active

Table 13.2 Factors distinguishing the nine demonstration study types

Study type	Study setting	Version of the resource	Sampled users	Sampled tasks	What is observed
1. Needs assessment	Field	None, or pre-existing resource to be replaced	Anticipated resource users	Actual tasks	User skills, knowledge, decisions or actions; care processes, costs, team function or organization; patient outcomes
2. Design validation	Development lab	None	None	None	Quality of design method or team
3. Structure validation	Lab	Prototype or released version	None	None	Quality of resource structure, components, architecture
4. Usability test	Lab	Prototype or released version	Proxy, real users	Simulated, abstracted	Speed of use, user comments, completion of sample tasks
5. Laboratory function study	Lab	Prototype or released version	Proxy, real users	Simulated, abstracted	Speed and quality of data collected or displayed; accuracy of advice given…
6. Field function study	Field	Prototype or released version	Proxy, real users	Real	Speed and quality of data collected or displayed; accuracy of advice given…
7. Lab user effect study	Lab	Prototype or released version	Real users	Abstracted, real	Impact on user knowledge, simulated/ pretend decisions or actions
8. Field user effect study	Field	Released version	Real users	Real	Extent and nature of resource use. Impact on user knowledge, real decisions, real actions
9. Problem impact study	Field	Released version	Real users	Real	Impact on targeted health status

health care or decision environment, in a "laboratory" setting. Studies to elucidate the need for a resource and studies of its impact on users would usually take place in settings—known generically as the "field"—where health care practitioners, researchers, students, patients or administrators are making real choices in the real world. These studies can take place only in settings where the resource is available for use and where health care or health behavior activities occur and/or where other important decisions are made. To an investigator planning such studies, an important consideration that determines the kind of study possible is the degree of access to resource users in the field setting. If, as a practical matter, access to the field setting is very limited, then several study types listed in ◘ Tables 13.1 and 13.2 are either not possible, or the validity of the field studies that are possible will be reduced.

The version of the resource used For some kinds of studies, a simulated or prototype version of the resource may be sufficient (Scott et al. 2011; Russ et al. 2014), whereas for studies in which the resource is employed by intended users to support real decisions and actions, a fully robust and reliable version is needed (e.g., Lester et al. 2010).

The sampled resource users Information resources nearly always function through interaction with one or more such "users" who bring to the interaction their own domain knowledge and knowledge of how to operate the resource. Exceptions might include closed loop control systems such as smart insulin pumps and pharmacy robots, but even in these cases, humans who set the parameters for the otherwise autonomous operations of these devices can be seen as a form of "users". In some types of evaluation studies, the users of the resource are not the end users for whom the resource is ultimately designed, but are members of the development or evaluation teams, or other individuals who can be called "proxy users", chosen because they are conveniently available or because they are affordable. (For example, senior medical students can sometimes be used as proxies for more experienced physicians.) In other types of studies, the users are sampled from the end-users for whom the resource is ultimately designed. The type of users employed gives shape to a study and can affect its results profoundly. The usability of a resource is easily overestimated if the "users" in a study are those who designed or are otherwise familiar with the resource. As another example, volunteer users of a consumer-oriented resource such as a dieting app may be more motivated than the general population the resource is designed to benefit.

The sampled tasks For function, effect, and impact studies, the users included in the study actually interact with the resource. This requires tasks, often clinical or scientific decision problems, for the users to undertake. These tasks can be invented or simulated; they can be abstracted versions of real cases or problems, shortened to suit the specific purposes of the study; or they can be live cases or research problems as they present to resource users in their everyday work. Clearly, the kinds of tasks employed, and how they are sampled, in a study have serious implications for the study results and the conclusions that can be drawn from them.

The observations that are made All evaluation studies entail observations that generate data that are subsequently analyzed to generate the study results. As seen in ◘ Table 13.2, many different kinds of observations can be made.

In the paragraphs above we have introduced the term "sampled" for both tasks and users. It is important to establish that in real evaluation studies, tasks and users are always sampled from some real or hypothetical population. Choosing appropriate methods to sample users and tasks is a major challenge in evaluation study design since it is never possible, practical, or desirable to study everyone doing everything possible with an information resource. Sampling issues are addressed later in this chapter.

13.5 Conducting Investigations: Collecting and Drawing Conclusions from Data

13.5.1 Two Grand Approaches to Study Design, Data Collection, and Analysis

Several authors have developed classifications, or **typologies**, of evaluation methods or approaches. Among the best is that developed in 1980 by Ernest House (1980). Even though it is somewhat old, a major advantage of House's typology is that each approach is linked elegantly to an underlying philosophical model, as detailed in his book. This classification divides current practice into eight discrete approaches, four of which may be viewed as **objectivist** and four of which may be viewed as **subjectivist**. While the distinctions between the eight approaches House describes are beyond the scope of this chapter, the grand distinction between objectivist and subjectivist approaches is very important. Note that these approaches are not entitled "objective" and "subjective", because those labels carry strong and fundamentally misleading connotations of scientific precision in the former case and of idiosyncratic imprecision in the latter. We will see in this section how both objectivist (often called quantitative) and subjectivist (often called qualitative) approaches find rigorous application across the range of study types described earlier.

To appreciate the fundamental difference between the approaches, it is necessary to address their very different philosophical roots. The objectivist approaches derive from a **logical–positivist** philosophical orientation—the same orientation that underlies the classic experimental sciences. The major premises underlying the objectivist approaches are as follows:

- In general, the attributes of interest are properties of the resource under study, or the people interacting with it. More specifically, this premise suggests that the merit and worth of an information resource—the attributes of most interest in evaluation—can in principle be measured, with all observations yielding the same result. It also assumes that an investigator can measure these attributes without affecting how the resource under study functions or is used.
- Rational persons can and should agree on what attributes of a resource are important to measure and what results of these measurements would be identified as the most desirable, correct, or positive outcome. In informatics, making this assertion is tantamount to stating that perfection in resource or user performance can always be identified and that all rational individuals can be brought to consensus on what "perfection" is.
- Because numerical measurement allows precise statistical analysis of performance over time or performance in comparison with some alternative, numerical measurement is prima facie superior to a verbal description. Verbal, descriptive data (generally known as qualitative data) are thus useful in only preliminary studies to identify hypotheses for subsequent, more precise analysis using quantitative methods.
- Through these kinds of comparisons, it is possible to demonstrate to a reasonable degree that a resource is or is not superior to what it replaced, or to a competing resource.

Contrast these assumptions with the set of assumptions that derives from an **intuitionist–pluralist** or de-constructivist philosophical position that spawns a set of subjectivist approaches to evaluation:

- What is observed about a resource invariably depends in fundamental ways on the observer. Different observers of the same resource might legitimately come to different conclusions. Both can be objective in their appraisals even if they do not agree; it is not necessary that one is right and the other wrong. Important insight can derive from both, and from their juxtaposition.
- Merit and worth must be explored in context. The value of a resource emerges through study of the resource as it

functions in a particular decision-making environment.
- Individuals and groups can legitimately hold different perspectives on what constitutes the most desirable outcome of introducing a resource into an environment. There is no reason to expect them to agree, and it may be counterproductive to even try to lead them to consensus. An important aspect of an evaluation would be to document the ways in which they disagree.
- Verbal description can be highly illuminating. Qualitative data are valuable, in and of themselves, and can lead to conclusions as convincing as those drawn from quantitative data. The value of qualitative data, therefore, goes far beyond that of identifying issues for later more "precise" exploration using quantitative methods.
- Evaluation should be viewed as an exercise in argument or rhetoric, rather than as a demonstration, because every study can "appear equivocal when subjected to serious scrutiny" (House 1980).

The approaches to evaluation that derive from this subjectivist philosophical perspective may seem strange, imprecise, and unscientific when considered for the first time. This perception stems in large part from the widespread acceptance of the objectivist worldview in biomedicine. Over the last two decades, however, thanks to some early high quality studies (e.g., Forsythe et al. 1992; Sheikh et al. 2011; Wright et al. 2015) the importance and utility of these subjectivist approaches in evaluation have been established within biomedical informatics. It is important for people trained in classic experimental methods at least to understand, and possibly even to embrace, the subjectivist worldview if they are to conduct fully informative evaluation studies.

13.5.2 Conduct of Objectivist Studies

Figure 15.2 expands the generic process for conducting evaluation studies to illustrate the steps involved in conducting an objectivist study. Figure 13.2 illustrates the linear sequence in which the investigation portion of an objectivist evaluation study is typically carried out. We will focus in this chapter on issues of study design that are most challenging in objectivist studies and we will further focus on that subset of objectivist study designs which are comparative in nature. More details on the other aspects of objectivist studies are available in standard references on experimental design (Campbell and Stanley 1963; Rothman et al. 2008).

13.5.2.1 Structure and Terminology of Comparative Studies

Most objectivist evaluations performed in the world make a comparison of some type. For informatics, aspects of performance of

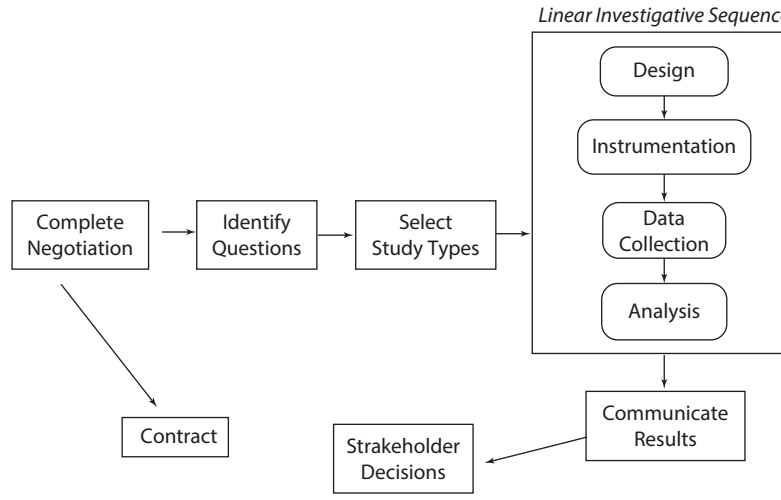


Fig. 13.2 Generic structure depicting an objectivist investigation

individuals, groups, or organizations *with* the information resource are often compared to those same aspects *without* the resource, with some alternative resource, or an alternate design of the same resource. After identifying a sample of participants for the study, the investigator assigns each participant, often randomly, to one or a set of conditions and some outcomes of interest are measured for each participant. The averaged values of these outcomes are then compared across the conditions. If all other factors are controlled, either directly through the design of the study or statistically through randomization, then any measured difference in the averaged outcomes can be attributed to the resource.

This relatively simple description of a comparative study belies the many issues that affect their design, execution, and ultimate usefulness. To understand these issues, we must first develop a precise terminology.

The **participants** in a study are the entities about which/whom data are collected. It is key to emphasize that participants are often people—for example, care providers or recipients—but also may be information resources, groups of people, or organizations. Because many of the activities in informatics are conducted in hierarchical settings with naturally occurring groups (a "physician's patients"; the "researchers in a laboratory"), investigators must, for a particular study, define the participants carefully and consistently.

Variables are specific characteristics of the participants or the setting that either are measured purposefully by the investigator or are self-evident properties that do not require measurement. Some variables may take a continuous range of values while others have a discrete set of levels, corresponding to each of the possible measured values. For example, in a hospital setting, physician members of a ward team can be classified as residents, fellows, or attending physicians. In this case, the variable "physician's level of qualification" is said to have three discrete "levels".

The **dependent variables** are those variables in the study that capture the outcomes of interest to the investigator. (For this reason, dependent variables are also called **outcome variables**.) A study may have one or more dependent variables. In a typical study, the dependent variables will be computed, for each participant, as an average over a number of tasks. For example, clinicians' diagnostic performance may be measured over a set of cases, or "tasks", that provide a range of diagnostic challenges.

The **independent variables** are included in a study to explain the measured values of the dependent variables. For example, whether an information resource is available, or not, to support certain clinical tasks could be the major independent variable in a study designed to evaluate the effects of that resource.

Measurement challenges almost always arise in the assessment of the outcome or dependent variable for a study. Often, for example, the dependent variable is some type of performance measure that invokes concerns about reliability (precision) and validity (accuracy) of measurement. The independent variables may also raise measurement challenges. When the independent variable is marital status, for example, the measurement problems are relatively straightforward. If the independent variable is an attitude or other "state of mind", such as computer or health literacy, profound measurement challenges can arise.

13.5.2.2 Issues of Measurement

Measurement is the process of assigning a value corresponding to the presence, absence, or degree of a specific attribute in a specific object. When we speak specifically of measurement, it is customary to use the term "object" to refer to the entity on which measurements are made. Measurement usually results in either (1) the assignment of a numerical score representing the extent to which the attribute of interest is present in the object, or (2) the assignment of an object to a specific category. Taking and recording the temperature (attribute) of a person (object) is an example of the process of measurement.

From the premises underlying objectivist studies (see ► Sect. 13.5.1), it follows that proper execution of such studies requires careful and specific attention to methods of measurement. It can never be assumed that

attributes of interest are measured without error. Accurate and precise measurement must not be an afterthought and indeed, most scientific progress occurred once challenging measurement problems, such as measuring the speed of light or mass of an electron, were solved. Measurement is of particular importance in biomedical informatics because, as a relatively young field, informatics does not have a well-established tradition of "variables worth measuring" or proven instruments for measuring them. By and large, people planning studies in informatics are faced first with the task of deciding what to measure and then with that of developing their own measurement methods. For most researchers, these tasks prove to be harder and more time-consuming than initially anticipated.

We can underscore the importance of measurement by establishing a formal distinction between studies undertaken to develop methods for making measurements, which we call measurement studies, and the subsequent use of these methods to address questions of direct importance in informatics, which we call demonstration studies. **Measurement studies** seek to determine how accurately and precisely an attribute of interest can be measured in a population of objects. In an ideal objectivist measurement, which never actually occurs, all observers will agree on the result of the measurement. Any disagreement is therefore due to error, which should be minimized. The more agreement among observers or across observations, the better the measurement. Measurement procedures developed and validated through measurement studies provide researchers with the measurement instruments they need to conduct **demonstration studies** that directly address questions of substantive and practical concern to the stakeholders for an evaluation study. Once we know how accurately we can measure an attribute using a particular procedure and instrument, we can employ the measured values of this attribute as a variable in a demonstration study to draw inferences about the performance, perceptions, or effects of an information resource. For example, once measurement studies have determined how accurately and precisely the usability of a class of information resources can be measured, using a specific measurement method, a subsequent demonstration study could explore which of two resources that are members of this class has greater usability.

A detailed discussion of measurement methods and issues is beyond the scope of this chapter but these topics are discussed in the Friedman and Wyatt textbook previously referenced. The bottom line is that investigators should know that their measurement methods will be adequate before they collect data for their studies. If the measures to be used do not have an established track record it is necessary to perform a measurement study, involving data collection on a small scale, to establish the adequacy of all measurement procedures (e.g., Ramnarayan et al. 2003; Demiris et al. 2000). Even if the measurement procedures of interest do have a track record in a particular setting, they may not perform equally well in a different environment, so a further measurement study may still be necessary. Researchers should always ask themselves, "How good are my measures in this particular setting?" whenever they are planning a study, before they proceed to the demonstration phase. The importance of measurement studies for informatics was first explained by Michaelis and co-workers (1990) and later expanded by Friedman and Abbas (2003). A study by Scott et al. (2019) documented that informatics studies continue to underappreciate the importance of measurement issues.

Whenever possible, investigators planning demonstration studies should employ established measurement methods with a "track record", by re-using them, rather than by developing their own. Increasingly, compendia of measurement instruments specifically for health informatics are available on the Internet.[4]

4 Examples include: ► https://www.gem-beta.org/public/home.aspx and ► https://healthit.ahrq.gov/health-it-tools-and-resources/evaluation-resources/health-it-survey-compendium-search (Both accessed 11.20.18).

13.5.2.3 Sampling Strategies

Selection of Participants

The participants selected for objectivist studies must resemble those to whom the evaluator and others responsible for the study wish to apply the results. For example, when attempting to quantify the likely impact of a clinical information resource on clinicians at large, there is no point in studying its effects on the clinicians who helped develop it, especially if they built it, as they are likely to be much more familiar with the resource than average practitioners. Characteristics of clinical participants that typically need to be taken into account include age, experience, role, type of work environment, attitude toward digital information resources, and extent of their involvement in the development of the resource. Analogous factors would apply to patients or health care consumers as participants.

Volunteer Effect

A common bias in the selection of participants is the use of volunteers. It has been established in many areas that people who volunteer as participants, whether to complete questionnaires, participate in psychology experiments, or test-drive new cars or other technologies, are atypical of the population at large (e.g., Pinsky et al. 2007). Although evaluations are often the "art of the possible", and all participants in studies are ultimately volunteers in the sense that no one can or should be coerced to participate, it is important to take steps to make the study participants as representative as possible of the resource's ultimate user community. A systematic approach to participant selection would first identify the full population of users and then sample from that population either randomly or sometimes purposively to be sure the sample includes participants with characteristics seen as essential to a thorough test of the resource. Once a sample is selected, follow-up to invitation letters and other mechanisms can achieve as close to 100% recruitment of the selected sample as possible. Relatively modest financial incentives can significantly boost participation rates.

Number of Participants Needed

The financial investment required for an evaluation study depends critically on the number of participants needed. The required number in turn depends on the purpose and design of the study. In usability studies, discussed below, a great deal can be learned from a relatively small sample.[5] In subjectivist studies, participant selection can be a dynamic process where study participants identify other participants. In objectivist user effect or problem impact studies, sample sizes are directed by the precision of the answer required from the study and the risk investigators are willing to take of failing to detect a significant effect. (All other things being equal, the larger the sample size, the greater the likelihood of detecting an effect of a specified size using a predetermined criterion for statistical significance.) Statisticians can advise on this point and carry out power analyses that estimate the sample-size required. Sometimes, in order to recruit the required number of participants, an element of volunteer effect must be tolerated; often there is a trade-off between obtaining a sufficiently large sample and ensuring that the sample is representative. Also, the impact of sample size on effect detection is non-linear. The value of adding, say, 10 more representative participants to a sample of 100 is far less than that of adding 10 more participants to a sample of 30.

Selection of Tasks

In the same way that participants must be carefully selected to resemble the people likely to use the information resource, any tasks the participants complete in the study must also resemble those that will generally be encountered where the information resource is deployed. Thus when evaluating a clinical order-entry system intended for general use, it would be unwise to use only complex cases from, for example, a pediatric intensive care setting. Although the order-entry

5 See ► https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ (Accessed 11.20.18).

system might well be of considerable benefit in intensive care cases, it is inappropriate to generalize results from such a limited sample to the full range of cases seen in ambulatory pediatrics. An instructive example is provided by the study of Van Way et al. (1982) who developed a scoring system for diagnosing appendicitis and studied the resource's accuracy using exclusively patients who had undergone surgery for suspected appendicitis. Studying this group of patients had the benefit of allowing the true cause of the abdominal pain to be obtained with near certainty as a by-product of the surgery itself. However, in these patients who had all undergone surgery for suspected appendicitis the symptoms were more severe and the incidence of appendicitis was five to ten times higher than for the typical patient for whom such a scoring system would be used. Thus, the accuracy obtained with postsurgical patients would be a poor estimate of the system's accuracy in routine clinical use.

If the performance of an information resource is measured on a number of hand-picked tasks, the functions it performs may appear spuriously complete and its usability overestimated. This is especially likely if these cases are similar to, or even identical with, a "training" set of tasks used to develop or tune the information resource before the evaluation is carried out. When a statistical model that powers an information resource is carefully adjusted to achieve maximal performance on training data, this adjustment may worsen its accuracy on a fresh set of data due to a phenomenon called overfitting (Wasson 1985; Srivastava et al. 2014; Ravi et al. 2017). Thus, it is important to obtain a new set of tasks and evaluate performance on this new test set, a process called cross-validation. Sometimes developers omit tasks from a sample if they do not fall within the scope of the information resource, for example if the final diagnosis for a case is not represented in a diagnostic system's knowledge base. This practice violates the principle that a test set should be representative of all tasks in which the information resource will be used, and will overestimate its accuracy in the real world.

13.5.2.4 Control Strategies in Comparative Studies

One of the most challenging questions in quantitative comparative study design is how to obtain control (Liu et al. 2011). In the context of informatics, control mechanisms seek to account for all factors in a study environment that are not attributable to the information resource. In the following sections, we review a series of control strategies. We employ, as a running example of an information resource under study, a reminder system that prompts physicians to order prophylactic antibiotics for orthopedic patients to prevent postoperative infections. In this example, the intervention is the deployment of the reminder system; the participants are the physicians; and the tasks are the surgical cases. The dependent variables are physicians' ordering of antibiotics (a user effect measure) and the rate of postoperative infections (a problem impact measure). As such, this is an example of a study which straddles two of the types in ◘ Table 13.1.

▪ Descriptive (Uncontrolled) Studies

In the simplest possible design, an uncontrolled or **descriptive study**, we deploy the reminder system and then make our measurements. There is no independent variable as such. Suppose that we discover that the overall postoperative infection rate is 5% and that physicians order prophylactic antibiotics in 60% of orthopedic cases. Although we have two measured dependent variables, it is hard to draw meaningful conclusions from these results. Although the results might be informative from a patient safety perspective, is not possible to draw any conclusions about the effect of the resource.

▪ Historically Controlled Experiments

As a first improvement to a descriptive study, from the perspective of control, consider a **historically controlled experiment**, sometimes called a **before–after study** (Wyatt & Wyatt 2003). The investigator makes baseline measurements of antibiotic ordering and postoperative infection rates before the information resource is installed, and then makes the same measurements after the information resource is in routine use. The independent variable is time and

has two levels: before and after resource installation. Let us say that, at baseline, the postoperative infection rates were 10% and physicians ordered prophylactic antibiotics in only 40% of cases; but the post-intervention figures are 5% and 60%, respectively (see Table 13.3).

The investigators may claim that the halving of the infection rate can be safely ascribed to the information resource, especially because it was accompanied by a substantial improvement in physicians' antibiotic prescribing. Many other factors might, however, have changed in the interim to cause these results, especially if there was a long interval between the baseline and post-intervention measurements. New staff could have been employed; the mix of patients could have changed; new prophylactic antibiotics may have been introduced; quality improvement meetings may have highlighted the infection problem and thus caused greater clinical awareness; and/or incentive programs may have been introduced to reward prescribing. Simply assuming that the reminder system alone caused the reduction in infection rates is naive.

Simultaneous Nonrandomized Controls

To address some of the problems with historical controls, we might use **simultaneous controls**, which require additional measurements to be made with a new group of physicians and their patients who are not influenced by the prophylactic antibiotic reminder system--but who *are* subject to any other changes taking place in the environment.

This study design would be a parallel group comparative study with simultaneous controls; if the physicians are given the option to choose whether to use the reminder system or not, it is a case control study. Table 13.4 gives hypothetical results of such a study, focusing on postoperative infection rates as a single outcome measure or dependent variable. The independent variables are time and group, both of which have two levels.

Table 13.4 illustrates improvement in the group where reminders were available, but no improvement—indeed a slight deterioration—where no reminders were available. This design provides suggestive evidence of an improvement that is most likely to be due to the reminder system.

However, even though the controls in this example are simultaneous, attribution of the effect to the information resources remains refutable because there may be some systematic, unknown difference between the clinicians and/or patients in the two groups. For example, if the two groups comprised the patients and clinicians in two adjacent wards, the difference in the infection rates could be attributable to differences between the wards. Perhaps hospital-staffing levels improved in some wards but not in others, or there was cross infection by a multiple-resistant organism only among the patients in the control ward. To overcome such criticisms, we could try to measure everything that happens to every patient in both wards and to build complete profiles of all staff to rule out systematic differences. Even then, attribution of the effect to the information resources would be vulnerable to the accusation that some variable that we did not measure—and did not even know about—explains the difference. An alternative strategy is to make the intervention and control groups statistically comparable by randomizing them.

Table 13.3 Results from a hypothetical before-after study of the impact of reminders on post operative infection rates

	Prescribing rate (%)	Infection rate (%)
Baseline	40	10
Post-intervention	60	5

Table 13.4 Results of a hypothetical non-randomized parallel group study of reminders and post op infection rates

	Reminder group (%)	Control group (%)
Baseline rate	10	10
Post-intervention rate	5	11

Simultaneous Randomized Controls

The crucial problem in the previous example is that, although the controls were simultaneous, there may have been systematic, unmeasured differences between them and the participants receiving the intervention (Liu and Wyatt 2011). A simple and effective way of removing systematic differences, whether due to known or unknown factors, is to randomly assign participants to control or intervention groups. Thus, we could randomly allocate one-half of the physicians to receive the antibiotic reminders and the remaining physicians to work as they did before. We would then measure and compare postoperative infection rates in patients managed by physicians in the reminder and control groups. Provided that the physicians care only for their assigned patients, any difference that is statistically "significant" (conventionally, a result that is statistically determined to have a probability of 0.05 or less of occurring by chance) can be attributed reliably to the reminders.

Table 13.5 shows the hypothetical results of such a study. The baseline infection rates in the patients managed by the two groups of physicians are similar, as we would expect, because the patients were allocated to the groups by chance. There is a greater reduction in infection rates in patients of reminder physicians compared with those of control physicians. Because strict random assignment means that there was no systematic difference in physician or patient characteristics between groups, the only systematic difference between the two groups of patients is receipt of reminders by their physicians.

Table 13.5 Results of a hypothetical randomized controlled trial of the impact of reminders on post op infection rates

	Reminder physicians (%)	Control physicians (%)
Baseline infection rate	11	10
Post intervention infection rate	6	9
Difference in infection rate	−5	−1

Provided that the number of patients is large enough to provide a sufficient number of events (post op infections) for these results to be statistically significant (about 250 infections observed overall), we would conclude with some confidence that providing physicians with reminders **caused** the reduction in infection rates. The small reduction, from baseline to installation, in infection rates in control cases is not unexpected, even in a perfectly randomized study. It could reflect changes in practice policy that affected both groups, some cross-talk among physicians who work in the same setting, or the Hawthorne Effect (whereby people's performance often improves when it is studied). These phenomena occur in the real world of evaluation and should be expected. However, because the pre-post difference in the reminder group was larger, an effect due to the information resource is likely to have occurred.

13.5.2.5 Drawing Conclusions from Observational Data: Real World Evidence

Demonstration studies using a planned, prospective data collection process share a number of challenges, including their cost, inevitable delays setting up the study and recruiting participants--and the concern that, because of the volunteer effect or other biases, the results will not confidently generalize to typical resource users, patients, or care settings. The use of so-called observational data from routinely generated patient care records--often linked to administrative or other data sources--could overcome many of these limitations and offer a more economical and faster method to address evaluation questions related to information resources. Ideally, such studies can be performed retrospectively with all necessary data drawn from existing data repositories and no further data collection required. These methods give rise to what is coming to be called Real World Evidence (RWE). As ever-increasing amounts of routinely collected data become available in coded digital forms, there is growing interest in RWE (Sherman et al. 2016).

RWE methods are increasingly important but also have important limitations. In the

absence of the kinds of controls described in the previous section of this chapter, it difficult to attribute any observed user effects or problem impacts to a specific cause. The analysis of observational data typically results in a pattern of correlations among the variables included in an analysis. This pattern of correlations must be interpreted with great care. For example, if Factor A (for example, extent of use of a decision support system) is correlated with outcome O (for example, fewer medication errors), A is not necessarily a direct cause of O. It may be the case that some Factor B (for example, clinical workload), which was not included in the study but which is correlated with A, is the true cause of O. There is, however, no way to know this, because Factor B was not included in the data set used for the study. This phenomenon is known as unmeasured confounding, and is the primary source of concern when putative causal conclusions are drawn from observational data.

Other concerns arise with the quality of data drawn from documentation of routine care. Data entered by care providers themselves under pressure of time will not be expected to be as accurate and precise as data entered by an undistracted research assistant paid to be careful observer. Also, incompleteness in observational data may not be randomly distributed, so simply increasing the sample size or the range of included variables may make matters worse, not better. A further consideration is confounding by indication, in which the (usually unconscious) prejudice of clinicians leads to biased prescribing or recommendations to patients to use new, expensive or risky apps, online consultations or other digital services for those patients with the best – or the worst – prognosis. In these cases, a straightforward analysis will overestimate effectiveness, so a technique called propensity scoring is needed (McMurry et al. 2015).

Methods from the field of econometrics can be very helpful for establishing causal relationships between variables in observational studies. One approach is instrumental variable (IV) methods that attempt to identify an "instrument" that usually determines whether a treatment is given (Davey Smith et al. 2007). A clinical example might arise when trying to estimate the benefit of bone marrow transplant (BMT) in acute myeloid leukemia in children without performing a randomized trial. It can be argued that one can compare mortality in this condition between children with and without a living sibling, since nearly all children with this leukemia and a live sibling will get a BMT and those without a sibling are much less likely to have a successful transplant. Since the presence of a living sibling is unrelated to whether a child has the disease, this kind of observational study might be almost as informative as a randomized trial for determining whether BMT is effective [3]. An informatics example would be if some diabetic patients are covered for online consultations by their health insurance while others are not. As long as we are satisfied that there are no systematic differences in disease severity or treatment adherence between the two patient groups, we could use the IV method to estimate the impact of online consultation on diabetes control and progression, assuming that those who are covered to use online consultations will usually take that option. However, the major challenge in designing an IV study is to identify an instrument that fulfils the following essential criteria: it usually dictates whether the intervention is given, does not affect the outcome except via the intervention, and is not correlated in any way with the outcome, or with other factors that cause it (Streeter et al. 2017; Gray et al. 2019).

In the age of "big data", observational studies generating Real World Evidence will become increasingly important, but from the perspective of evaluation in informatics, they will apply only to field studies, and principally to user effect and problem impact studies. The seven other study types will remain largely reliant on prospective methods, although prospective studies too can benefit from the existence of data marts by incorporating already-available data whenever possible. As data quality improves, data marts become more comprehensive, and methods to establish causation gain increased sophistication, RWE methods will continue to mature. For studies addressing health outcomes--what this chapter refers to as problem impact studies--it is almost inevitable that Real World Evidence approaches will

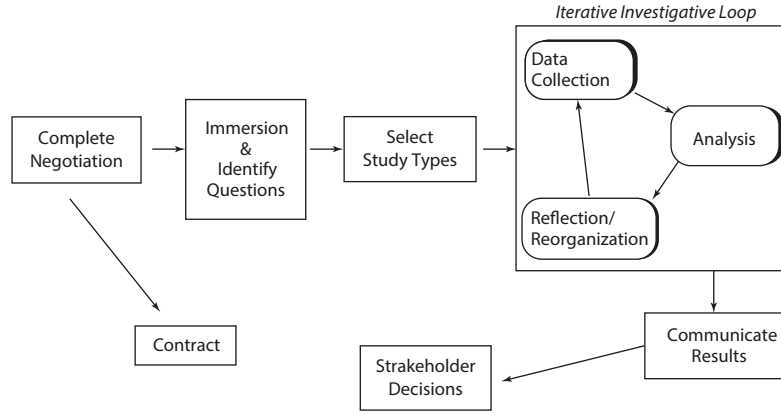


Fig. 13.3 Generic structure depicting a subjectivist investigation

assume an important place alongside randomized designs and in the best possible scenario, the two will complement each other.

13.5.3 Conduct of Subjectivist Studies

The objectivist comparative approaches to evaluation, described in the previous section, are useful for addressing some, but not all, of the interesting and important questions that challenge investigators in medical informatics. The subjectivist approaches described in this section address the problem of evaluation from a very different set of premises. They use different but equally rigorous methods. Figure 13.3 expands the generic process for conducting evaluation studies to illustrate the stages involved in conducting a subjectivist study and emphasizes the "iterative loop" of data collection, analysis and reflection as the major distinguishing characteristic of a subjectivist investigation. Another distinctive feature of subjectivist studies is an immersion in the environment where the resource has been or will be deployed. Because subjectivist approaches may be less familiar to readers, we describe subjectivist studies in more detail than we did their objectivist counterparts.

13.5.3.1 The Rationale for Subjectivist Studies

Subjectivist methods enable us to address the deeper questions that arise in informatics: the detailed, individualistic "whys" and "according to whoms" in addition to the aggregate "whethers" and "whats." Subjectivist approaches seek to represent the viewpoints of people who are users of the resource or are otherwise significant participants in the environment where the resource operates. The goal is illumination rather than judgment. The investigators seek to build an argument that promotes deeper understanding of the information resource or environment of which it is a part. The methods used derive largely from ethnography (Forsythe 1992; Ventres et al. 2006; Pope et al. 2013). The investigators immerse themselves physically in the environment (the "field") where the information resource is or will be operational, and collect data primarily through observations, interviews, or reviews of documents. The designs—or data-collection plans—of these studies are not rigidly predetermined and do not unfold in a fixed sequence. They develop dynamically and nonlinearly, as the investigators' experience in the field accumulates.

13.5.3.2 A Rigorous, but Different, Methodology

These subjectivist approaches to evaluation, like their objectivist counterparts, are empirical methods. Although it is easy to focus only on their differences, these two broad classes of evaluation approaches share many features. In all empirical studies, for example, data are collected with great care; the investigators are always aware of what they are doing and why. The data are then compiled, interpreted, and ultimately reported.

Investigators keep records of their procedures, and these records are open to audit by the investigators themselves or by individuals outside the study team. The principal investigator or evaluation-team leader is under an almost sacred scientific obligation to report the study methods. Failure to do so will invalidate a study. Both classes of approaches also share a dependence on theories that guide investigators to explanations of the observed phenomena, as well as to a dependence on the pertinent literature such as published studies that address similar phenomena or similar settings. In both approaches, there are rules of good practice that are generally accepted; it is therefore possible to distinguish a "good" study from a bad one.

There are, however, fundamental differences between objectivist and subjectivist approaches. First, subjectivist studies are **emergent** in design. Objectivist studies typically begin with a set of hypotheses or specific questions, and with a plan for addressing each member of this set. The investigator assumes that, barring major unforeseen developments, the plan will be followed exactly. Deviation would be seen as a potential source of bias. For example, an objectivist investigator who sees negative results emerging from the exploration of a particular question or use of a particular measurement instrument might be inclined to change strategies in hope of obtaining more positive findings. In contrast, subjectivist studies typically begin with general orienting issues that stimulate the early stages of investigation. Through these initial investigations, the important questions for further study emerge. The subjectivist investigator is willing, at virtually any point, to adjust future aspects of the study in light of the most recent information obtained, while carefully recording that this had happened and why. Subjectivist investigators tend to be incrementalists; they thoughtfully change their plans as necessary from day-to-day and have a high tolerance for ambiguity and uncertainty. In this respect, they are much like good software developers. Also like software developers, subjectivist investigators must develop the ability to recognize when a project is finished, when further benefit can be obtained only at too great a cost in time, money, or work.

A second feature of subjectivist studies is a **naturalistic** orientation: a reluctance to manipulate the setting of the study, which in informatics is typically the environment into which the information resource is introduced. Subjectivist studies do not alter the environment to study it. Control groups, placebos, purposeful altering of information resources to create contrasting interventions, and other techniques that are central to the construction of objectivist studies typically are not used. Subjectivist studies will, however, employ quantitative data for descriptive purposes and may offer quantitative comparisons when the research setting offers a "natural experiment" where such comparisons can be made without deliberate manipulation. For example, when physicians and nurses both use a clinical system to enter orders, the differing experiences of the two professional groups offer a natural basis for comparison. Subjectivist researchers are opportunists where pertinent information is concerned; they will use what they see as the best information available to illuminate a question under investigation.

A third important distinguishing feature of subjectivist studies is that their end product is a report written in narrative prose. While these reports may be lengthier than the statistical reports from objectivist studies, no technical understanding of quantitative research methodology is required to comprehend them. Results of subjectivist studies are therefore accessible—and may even be entertaining—to a broad community in a way that results of objectivist studies are not. Reports of subjectivist studies seek to engage their audience.

13.5.3.3 Natural History of a Subjectivist Study

Figure 13.3 illustrates the stages that characterize a subjectivist study (see also Chap. 9 in Friedman and Wyatt 2005). These stages constitute a general sequence, but, as we mentioned, subjectivist investigators must always be prepared to revise their thinking and possibly to return to earlier stages in light of new

data or insights resulting from its analysis. Backtracking is a legitimate step in this model.

1. *Negotiation of the ground rules of the study*: The understanding between the study team and the persons commissioning a study should embrace the general aims of the study; the kinds of methods to be used; access to various sources of information, including health care providers, patients, and various documents; and the format for interim and final reports. The aims of the study may be formulated in a set of initial **orienting questions**. Ideally, this understanding will be expressed in a memorandum of understanding, analogous to a contract.
2. *Immersion into the environment*: At this stage, the investigators begin spending time in the work environment. Their activities range from formal introductions to informal conversations, or to silent presence at meetings and other events. Investigators use the generic term **field** to refer to the setting, which may be multiple physical locations, where the work under study is carried out. Trust and openness between the investigators and the people in the field are essential elements of subjectivist studies to ensure full and candid exchange of information.

   Even as immersion is taking place, the investigator is already collecting data to sharpen the initial questions or issues guiding the study. Early discussions with people in the field, and other activities primarily targeted toward immersion, inevitably begin to shape the investigators' views. Almost from the outset, the investigator is typically addressing several aspects of the study simultaneously.
3. *Iterative loop*: At this point, the procedural structure of the study becomes akin to an iterative loop, as the investigator engages in cycles of data collection, analysis and reflection, "member checking", and reorganization. Data collection involves interview, observation, document analysis, and other methods. Data are collected on planned occasions, as well as serendipitously and spontaneously. The data are recorded carefully and are interpreted in the context of what is already known. Analysis and reflection entail the contemplation of the new findings during each cycle of the loop. Member checking is the sharing of the investigator's emerging thoughts and beliefs with the participants themselves. Reorganization results in a revised agenda for data collection in the next cycle of the loop.

   Although each cycle within the iterative loop is depicted as unidirectional, this representation is misleading. Net progress through the loop is clockwise, as shown in ◘ Fig. 13.3, but backward steps are natural and inevitable. They are not reflective of mistakes or errors. An investigator may, after conducting a series of interviews and studying what participants have said, decide to speak again with multiple participants to clarify their positions on a particular issue.
4. *Communicate results*: Subjectivist studies tend to have a multi-staged reporting and communication process. The first draft of the study report should itself be viewed as a research instrument. By sharing this report with a variety of individuals, the investigator obtains a major check on the validity of the findings. Typically, reactions to the preliminary report will generate useful clarifications and a general sharpening of the study findings. Because the report usually includes a prose narrative, it is vitally important that it be well written in language understandable by all intended audiences. Circulation of the report in draft, for comments by the intended recipients, can ensure that the final document communicates as intended. Use of anonymous quotations from interviews and documents makes a report highly vivid and meaningful to readers.

   The final report, once completed, should be distributed as negotiated in the original memorandum of understanding. Distribution is often accompanied by "meet the investigator" sessions that allow interested persons to ask the author of the report to expand or explain what has been written.

13.5.3.4 Subjectivist Data-Collection and Data-Analysis Methods

What data-collection strategies are in the subjectivist researcher's tool kit? There are several, and they are typically used in combination. We shall discuss each one, assuming a typical setting for a subjectivist study in biomedical informatics: the introduction of an information resource into patient care activities in a hospital.

■ Observation

The investigators typically immerse themselves into the setting under study in one of two ways. The investigator may act purely as a detached observer, becoming a trusted and unobtrusive feature of the environment but not a participant in the day-to-day work and thus reliant on multiple "informants" as sources of information. True to the naturalistic feature of this kind of study, great care is taken to diminish the possibility that the presence of the observer will skew any work activities or that the observer will be rejected outright by the team. An alternative approach is participant observation, where the investigator becomes a member of the work team. Participant observation is more difficult to engineer; it may require the investigator to have specialized training in the study domain. It is time consuming but can give the investigator a more vivid impression of life in the work environment. During both kinds of observation, data accrue continuously. These data are qualitative and may be of several varieties: statements by health care providers, patients, family members, administrative staff, and others; gestures and other nonverbal expressions of these same individuals; and characteristics of the physical setting that seem to affect the delivery of health care.

■ Interviews

Subjectivist studies rely heavily on interviews. Formal interviews are occasions where both the investigator and interviewee are aware that the answers to questions are being recorded (on paper or digitally) for direct contribution to the evaluation study. Formal interviews vary in their degree of structure. At one extreme is the unstructured interview, where there are no predetermined questions. Between the extremes is the semi structured interview, where the investigator specifies in advance a set of topics that he/she would like to address--but is flexible as to the order in which these topics are addressed, and is open to discussion of topics not on the pre-specified list. At the other extreme is the structured interview, with a schedule of questions that are always presented in the same words and in the same order. In general, the unstructured and semi structured interviews are preferred in subjectivist research. Informal interviews—spontaneous discussions between the investigators and members of a team that occur during routine observation—are also part of the data collection process. Informal interviews are invariably considered a source of important data. Group interviews, akin to focus groups, may also be employed (e.g., Haddow et al. 2011). Group interviews are very efficient ways to reach large numbers of participants, but investigators should not assume that individual participants will express in a group setting the same sentiments they will express if interviewed one-on-one.

Sampling also enters into the interview process. There are usually more participants to interview than resources to conduct them. Unlike in objectivist studies, where random sampling is a form of gold standard to inform statistical attributions of effects, subjectivist studies employ more purposeful sampling strategies. Investigators might actively seek interviewees they suspect to have unique or particularly insightful or influential opinions. They might remain in more frequent contact with key informants who, for various reasons, have the most insight into what is happening.

■ Document and Artifact Analysis

Every project produces a trail of papers and other artifacts. These include patient charts, the various versions of an information resource and its documentation, memoranda prepared by the project team, perhaps a cartoon hung on an office door. Unlike the day-to-day events of health care, these artifacts do not change once created or introduced. With appropriate permissions negotiated in advance, they can be examined retrospectively

and referred to repeatedly, as necessary, over the course of a study. Also included under this heading are **unobtrusive measures**, which are the records accrued as part of the routine use of the information resource. They include, for example, user log files of an information resource. Data from these measures are often quantifiable and analyzed quantitatively even though the overall study design is qualitative in nature.

Anything Else That Seems Useful

Subjectivist investigators are supreme opportunists. As questions of importance to a study emerge, the investigators will collect any information that they perceive as bearing on these questions. This data collection could include clinical chart reviews, questionnaires, tests, simulated patients, and other methods more commonly associated with the objectivist approaches.

When to end data collection is another challenge in otherwise open-ended subjectivist studies. "Saturation" is important principle to help investigators know when to stop. Stated simply, a data collection process is saturated when it becomes evident that, as more data are collected, no new findings or insights are emerging.

Analysis of Subjectivist Data

There are many alternative procedures for analysis of qualitative data. In general terms, the investigator looks for insights, themes or trends emerging from several different sources. He/she collates individual statements and observations by theme, as well as by source. Investigators typically use software especially designed to facilitate analysis of qualitative data.[6] Because they allow electronic recording of the data while the investigator is "in the field", tablets, smartphone Apps and other hand-held devices are changing the way subjectivist research is carried out.

The subjectivist analysis process is fluid, with analytic goals shifting as the study matures. At an early stage, the goal is primarily to focus the questions that themselves will be the targets of further data elicitation. At the later stages of study, the primary goal is to organize data that address these questions into specific themes, interpretations, and explanations. Conclusions derive credibility from a process of "triangulation", which is the degree to which information from different independent sources generate the same theme or point to the same conclusion. Subjectivist analysis also employs a strategy known as "member checking" whereby investigators take preliminary conclusions back to the persons in the setting under study, asking if these conclusions make sense, and if not, why not. In subjectivist investigation, unlike objectivist studies, the agenda is never completely closed. The investigator is constantly on the alert for new information that can require a significant reorganization of the findings and conclusions that have been drawn to date.

13.6 Communicating Evaluation Results

Once any study, qualitative or quantitative, is complete, the results need to be communicated to the stakeholders and others who might be interested. In many ways, communication of evaluation results, a term we prefer over "reporting", is the most challenging aspect of evaluation. Elementary theory tells us that, in general, successful communication requires a sender, one or more recipients, and a channel linking them, along with a message that travels along this channel (Ong and Coiera 2011).

Seen from this perspective, successful communication of evaluation results is challenging in several respects. It requires that the recipient of the message actually receive it. That is, for evaluations, the recipient must read the written report or attend the meeting intended to convey evaluation results.

6 Examples include: Atlas.ti: ► https://atlasti.com/ (Accessed November 18, 2019) and NVivo ► https://www.qsrinternational.com/nvivo/home (Accessed November 18, 2019).

For this reason, the investigator is invariably challenged to create a report the stakeholders will want to read or to choreograph a meeting they will be motivated to attend. Successful communication also requires that the recipient understand the message, which challenges investigators to draft written documents at the right reading level, with audience-appropriate technical detail. Sometimes there must be several different forms of the written report to match several different audiences. Overall, we encourage investigators to recognize that their obligation to communicate does not end with the submission of a written document comprising their technical evaluation report. The report is one channel for communication, not an end in itself.

Depending on the nature, number, and location of the recipients—and permissions which have been obtained or written into evaluation agreements--many options exist for communicating the results of a study, including:

- Written reports
  - Document(s) prepared for specific audience(s)
  - Internal newsletter article
  - Published journal article, with appropriate permissions
  - Monograph, picture album, or book
- One-to-one or small group meetings
  - With stakeholders or specific stakeholder groups
  - With the general public, if appropriate
- Formal oral presentations
  - To groups of project stakeholders
  - Conference presentation with a poster or published paper in proceedings
  - To external meetings or seminars
- Internet
  - Project Web site or blog
  - Web "chat", forum or Twitter feed to socialize results
  - Online preprint
  - Internet based journal
- Other
  - Video or podcast describing the study and information resource
  - Interview with a journalist on newspaper, TV, radio

A written, textual report is not the sole medium for communicating evaluation results. Verbal, graphical, or multimedia approaches can be helpful as ways to enhance communication with specific audiences. Another useful strategy is to hold a "town meeting" to discuss a traditional written report after it has been released. Photographs or videos can portray the work setting for a study, the people in the setting, and the people using the resource. If appropriate permissions are obtained, these images—whether included as part of a written report, shown at a town meeting, or placed on a Web site—can be worth many thousands of words. The same may be true for recorded statements of resource users. If made available, with permission, as part of a multimedia report, the voices of the participants can convey a feeling behind the words that can enhance the credibility of the investigator's conclusions (Fig. 13.4).

In addition to the varying formats for communication described above, investigators have other decisions to make after the data collection and analysis phases of a study are complete. One key decision is what personal role they will adopt after the formal investigative aspects of the work are complete. They may elect only to communicate the results, but they may also choose to persuade stakeholders to take specific actions in response to the study results, and perhaps even assist in the implementation of these actions. This raises a key question: Is the role of an evaluator simply to record and communicate study findings and then to move on to the next study, or is it also to engage with the study stakeholders and help them change how they work as a result of the study?

To answer this question about the role of an evaluator, we need to understand that an evaluation study, particularly a successful one, has the potential to trigger a series of events, starting with the communication of study results, but then including interpretation, recommendation, and even implementation. Some evaluators—perhaps enthused by the clarity of their results and an opportunity to use them to improve health care, biomedical research, or education—prefer to go beyond

Fig. 13.4 A picture is worth 1000 words: in the report of a study to establish the need for an electronic patient record, a casual photograph like this may prove much more persuasive than a table of data or paragraphs of prose

reporting the results and conclusions to making recommendations. The dilemma often faced by evaluators is whether to retain their scientific detachment and merely report the study results, or to stay engaged somewhat longer. Investigators who choose to remain may become engaged in helping the stakeholders interpret what the results mean, guiding them in reaching decisions and perhaps even in implementing the actions decided upon. The longer they stay, the greater the extent to which evaluators must leave behind their scientific detachment and take on a role more commonly associated with change agents (Lunenburg 2010). Some confounding of these roles is inevitable when the evaluation is performed by individuals within the organization that developed the information resource under study. There is no hard-and-fast rule for deciding on the most appropriate role for the evaluator; the most important initial realization for investigators is that the different options exist and that a decision among them must inevitably be made.

13.7 Conclusion: Evaluation as an Ethical and Scientific Imperative

Evaluation takes place, either formally or informally, throughout the resource development cycle: from defining the need to monitoring the continuing impact of a resource once it is deployed (Stead et al. 1994). We have seen in this chapter that different issues are explored, at different degrees of intensity, at each stage of resource development. For meaningful evaluation to occur, adequate amounts must be allocated for these studies when time and money are budgeted for a development effort. Evaluation cannot be left to the end of a project. While formal evaluations, as we have described them here, are still seen as optional for resources of the types that are the foci of biomedical and health informatics, the increasing complexity and prevalence of these resources have raised concerns about their safety and effectiveness when used in the real world (e.g., Koppel et al. 2005). For the moment, we would argue that formal evaluations, using the range of methods described in this chapter, are mandated by the professional ethics of biomedical informatics as an applied scientific discipline (see ▶ Chap. 12).

Formal evaluations of biomedical information resources may someday be a statutory or regulatory requirement in many or all parts of the world, as they are already for new drugs or medical devices. If and when that day comes, the wide variety of questions to be addressed and the diversity of legitimate methods available to address those questions, as described in this chapter, will make it difficult to describe with exactitude how these studies should be done. There have been some published academic checklists or guidelines

describing things to study and report in such studies (Talmon et al. 2009), but this is a bridge to be crossed in the future. We express the hope that writers of such guidelines and regulations will not overprescribe the methods to be used, while insisting on rigor in drawing conclusions from data collected using study designs thoughtfully matched to carefully identified questions. We hope the reader has learned from this chapter that rigor in evaluation is achievable in many ways, that information resources raise unique challenges when they are the foci for evaluation, and that overly rigid prescription of evaluation methods, however well intentioned, could defeat their well-intentioned purpose. However, it is also clear that the intensity of the evaluation effort should be closely matched to the resource's maturity (Stead et al. 1994). The UK Medical Research Council's Framework for Complex Interventions (Campbell et al. 2000), or a more recent variation intended for digital interventions (Murray et al. 2016) point out that it is unwise to conduct an expensive user-effect field trial of an information resource that is barely complete, is still in prototype form, may evolve considerably before taking its final shape, or is so early in its development that it may fail because programming bugs have not been eliminated.

We believe that readers of this chapter will to varying degrees be critical appraisers of, participants in, and/or conductors of evaluation studies. In playing any or all of these roles, it is important to recognize that evaluation sits at the junction where the art of the possible, given the complexity of informatics interventions, meets the rigor of scientific method drawn from the objectivist and subjectivist traditions.

Acknowledgment The authors wish to acknowledge Nikolas Koscielniak for his multiple important contributions to this chapter. This chapter is adapted from material in an earlier edition of the textbook that was also co-authored by Douglas K. Owens.

Appendices

Appendix A: Two Evaluation Scenarios

Here we introduce two scenarios that collectively capture many of the dilemmas facing those planning and conducting evaluations in biomedical informatics:

1. A prototype information resource has been developed, but its usability and potential for benefit need to be assessed prior to deployment;
2. A commercial resource has been deployed across a large enterprise, and there is need to understand its impact on users as well as on the organization.

These scenarios do not address the full scope of evaluations in biomedical informatics, but they cover a lot of what people do. For each, we introduce sets of evaluation questions that frequently arise and examine the dilemmas that investigators face in the design and execution of evaluation studies.

Scenario 1: A Prototype Information Resource Has Been Developed, but Its Usability and Potential for Benefit Need to Be Assessed Prior to Deployment

The primary evaluation issue here is the upcoming decision to continue with the development of the prototype information resource. Validation of the design and structure of the resource will have been conducted, either formally or informally, but not yet a usability study. If this looks promising, a laboratory evaluation of key functions is also advised before making the substantial investment required to turn a promising prototype into a system that is stable and likely to bring more benefits than problems to users in the field. Here, typical questions will include:

- Who are the target users, and what are their background skills and knowledge?
- Does the resource make sense to target users?

- Following a brief introduction, can target users navigate themselves around important parts of the resource?
- Can target users carry out a selection of relevant tasks using the resource, in reasonable time and with reasonable accuracy?
- What user characteristics correlate with the ability to use the resource and achieve fast, accurate performance with it?
- What other kinds of people can use it safely?
- How to improve the layout, design, wording, menus etc.
- Is there a long learning curve? What user training needs are there?
- How much on-going help will users require once they are initially trained?
- What concerns do users have about the system – e.g., accuracy, privacy, effect on their jobs, other side effects
- Based on the performance of prototypes in users' hands, does the resource have the potential to meet user needs?

These questions fall within the scope of the usability and laboratory function testing approaches listed in Table 15.1. A wide range of techniques–borrowed from the human-computer interaction field and employing both objectivist and subjectivist approaches–can be used, including:

- Seeking the views of potential users after both a demonstration of the resource and a hands-on exploration. Methods such as focus groups may be very useful to identify not only immediate problems with the software and how it might be improved, but also potential broader concerns and unexpected issues that may include user privacy and long term issues around user training and working relationships.
- Studying users while they carry out a list of pre-designed tasks using the information resource. Methods for studying users includes watching over their shoulder, video observation (sometimes with several video cameras per user); think aloud protocols (asking the user to verbalize their impressions as they navigate and use the system); and automatic logging of keystrokes, navigation paths, and time to complete tasks.
- Use of validated questionnaires to capture user impressions, often before and after an experience with the system, one example being the Telemedicine Preparedness questionnaire (Demiris et al. 2000).
- Specific techniques to explore how users might improve the layout or design of the software. For example, to help understand what users think of as a "logical" menu structure for an information resource, investigators can use a card sorting technique. This entails listing each function available on all the menus on a separate card and then asking users to sort these cards into several piles according to which function seems to go with which [► www.useit.com].

Depending on the aim of a usability study, it may suffice to employ a small number of potential users. Nielsen has shown that, if the aim is to identify only major software faults, the proportion identified rises quickly up to about 5 or 6 users then much more slowly to plateau at about 15–20 users (Nielsen 1994). Five users will often identify 80% of software problems. However, investigators conducting such small studies, useful though they may be for software development, cannot then expect to publish them in a scientific journal. The achievement in this case is having found answers to a very specific question about a specific software prototype. This kind of local reality test is unlikely to appeal to the editors or readers of a journal. By contrast, the results of formal laboratory function studies, that typically employ more users, are more amenable to journal publication.

Scenario 2: A Commercial Resource Has Been Deployed Across a Large Enterprise, and There Is Need to Understand its

Impact on Users as Well as on the Organization

The type of evaluation questions that arise here include:

- In what fraction of occasions when the resource could have been used, was it actually used?
- Who uses it, why, are these the intended users, and are they satisfied with it?
- Does using the resource improve influence information/communication flows?
- Does using the resource influence their knowledge or skills?
- Does using the resource improve their work?
- For clinical information resources, does using the resource change outcomes for patients?
- How does the resource influence the whole organization and relevant sub units?
- Do the overall benefits and costs or risks differ for specific groups of users, departments, the whole organization?
- How much does the resource really cost the organization?
- Should the organization keep the resource as it is, improve it or replace it?
- How can the resource be improved, at what cost, and what benefits would result?

To each of the above questions, one can add: "Why, or why not?", to get a broader understanding of what is happening as a result of use of the resource.

This evaluation scenario, suggesting a problem impact study, is often what people think of first when the concept of evaluation is introduced. However, we have seen in this chapter that it is one of many evaluation scenarios, arising relatively late in the life cycle of an information resource. When these impact-oriented evaluations are undertaken, they usually result from a realization by stakeholders, who have invested significantly in an information resource, that the benefits of the resource are uncertain and there is need to justify recurring costs. These stakeholders usually vary in the kind of evaluation methods that will convince them about the impacts that the resource is or is not having. Many such stakeholders will wish to see quantified indices of benefits or harms from the resource, for example the number of users and daily uses, the amount the resource improves productivity or reduces costs, or perhaps other benefits such as reduced waiting times to perform key tasks or procedures, lengths of hospital stay or occurrence of adverse events. Such data are collected through objectivist studies as discussed earlier. Other stakeholders may prefer to see evidence of perceived benefit and positive views of staff, in which case staff surveys, focus groups and unstructured interviews may prove the best evaluation methods. Often, a combination of many methods is necessary to extend the investigation from understanding what impact the resource has to why this impact occurs – or fails to occur.

If the investigator is pursuing objectivist methods, deciding which of the possible effect variables to include in an impact study and developing ways to measure them can be the most challenging aspect of an evaluation study design. (These and related issues receive the attention of five full chapters of a textbook by the authors of this chapter (Friedman and Wyatt 2005).) Investigators usually wish to limit the number of effect measures employed in a study for many reasons: limited evaluation resources, to minimize manipulation of the practice environment, and to avoid statistical analytical problems that result from a large number of measures.

Effect or impact studies can also use subjectivist approaches to allow the most relevant "effect" issues to emerge over time and with increasingly deep immersion into the study environment. This emergent feature of subjectivist work obviates the need to decide in advance which effect variables to explore, and is considered by proponents of subjectivist approaches to be among their major advantages.

In health care particularly, every intervention carries some risk, which must be judged in comparison to the risks of doing nothing or of providing an alternative intervention. It is difficult to decide whether an information resource is an improvement unless the performance of the current decision-takers is also measured in a comparison-based evaluation. For example, if physicians' decisions are to

become more accurate following introduction of a decision-support tool, the resource needs to be "right" when the user would usually be "wrong" This could mean that the tool's error rate is lower than that of the physician, or its errors are in different cases, or they should be of a different kind or less serious than those of the clinician, so as not to introduce new errors caused by the clinician following resource advice even when that advice is incorrect – "automation bias" (Goddard et al. 2012).

For effect studies, it is often important to know something about how the practitioners carry out their work prior to the introduction of the information resource. Suitable measures include the accuracy, timing, and confidence level of their decisions and the amount of information they require before making a decision. Although data for such a study can sometimes be collected by using abstracts of cases or problems in a laboratory setting (Fig. 15.2), these studies inevitably raise questions of generalization to the real world. We observe here one of many trade-offs that occur in the design of evaluation studies. Although control over the mix of cases possible in a laboratory study can lead to a more precise estimate of practitioner decision making, ultimately it may prove better to conduct a baseline study while the individuals are doing real work in a real practice setting. Often this audit of current decisions and actions provides useful input to the design of the information resource, and a reference against which resource performance may later be compared.

When conducting problem impact studies in health care settings, investigators can sometimes save themselves much time and effort without sacrificing validity by measuring effect in terms of certain health care processes rather than patient outcomes, in other words by employing a user effect study as a proxy for a problem impact study. For example, measuring the mortality or complication rate in patients with heart attacks requires data collection from hundreds of patients, as complications and death are (fortunately) rare events. However, as long as large, rigorous trials or meta-analyses have determined that a certain procedure (e.g., giving heart attack patients streptokinase within 24 h) correlates closely with the desired patient outcome, it is perfectly valid to measure the rate of performing this procedure as a valid "surrogate" for the desired outcome. Mant and Hicks demonstrated that measuring the quality of care by quantifying a key process in this way may require one tenth as many patients as measuring outcomes (Mant and Hicks 1995).

Appendix B: Exemplary Evaluation Studies

In this appendix, we briefly summarize studies that align with many of the study types described in Tables 13.1 and 13.2.

Usability Study *Assessing Performance of an Electronic Health Record Using Cognitive Task Analysis.*

Saitwal et al. (2010) is a pure usability testing study that evaluates the Armed Forces Health Longitudinal Technology Application EHR using a cognitive task analysis approach, referred to as Goals, Operators, Methods, and Selection rules (GOMS). Specifically, authors evaluated the system response time and the complexity of the graphical user interface (GUI) when completing a set of 14 prototypical tasks using the EHR. Authors paid special attention to inter-rater reliability of the two evaluators using GOMS to analyze the GUI of the system through task completion. Each task was broken down into a series of steps, with the intent to determine the percent of steps classified as "mental operators". Execution time was then calculated for each step and summed to obtain a total time for task completion.

Lab Function Study *Diagnostic inaccuracy of smartphone applications for melanoma detection.*

Wolf et al. (2013) conducted an evaluation study of smartphone applications capable of detecting melanoma and sought to determine the diagnostic inaccuracy. The study is exemplary of a lab function study and complements the Beaudoin et al. (2016) study

described below because study authors paid special attention to measuring application function in a lab setting using digital clinical images with a previous diagnosis obtained via histologic analysis by a dermatopathologist. Authors employed a comparative analysis between four different smartphone applications and assessed the sensitivity, positive predictive value, and negative predictive value of each application compared to histologic diagnosis. Rather than focus on the function in a real health care setting with real users, authors were interested in facilitating decision-making as to which applications performed best under controlled conditions.

Field Function Study *Evaluation of a machine learning capability for a clinical decision support system to enhance antimicrobial stewardship programs.*

Beaudoin et al. (2016) conducted an observational study to evaluate the function of a combined clinical decision support system (antimicrobial prescription surveillance system (APSS)) and a learning module for antimicrobial stewardship pharmacists in a Canadian university hospital system. Authors developed a rule-based machine learning module designed from expert pharmacist recommendations which triggers alerts for inappropriate prescribing of piperacillin–tazobactam. The combined system was deployed to pharmacists and outputs were studied prospectively over a five-week period within the hospital system. Analyses assessed accuracy, positive predictive value, and sensitivity of the combined system, the individual learning module, and the APSS compared to the pharmacist opinion. This is an exemplary field function study because authors are evaluating the ability of the combined rule-based learning module and APSS to detect inappropriate prescribing in the field with real patients.

Lab User Effect Study *Applying human factors principles to alert design increases efficiency and reduces prescribing errors in a scenario-based simulation.*

Russ et al. (2014) describe a study evaluating the redesign of alerts using human factors principles and their influence on prescribing by providers. The study is exemplary of a lab user effect study because it analyzed frequency of prescribing errors by providers, and it was conducted in a simulated environment (the Human-Computer Interaction and Simulation Laboratory in a Veterans Affairs Medical Center). Authors were particularly interested in three types of alerts: drug-drug interactions, drug-allergy, and drug disease. Three scenarios were developed for this study that included 19 possible alerts. These alerts were intended to be familiar and unfamiliar to prescribers. Authors used a crossover design with a two-week "washout period" for participants to complete both original and redesigned alerts to reduce contamination in repeated measures. Special attention was paid to a repeated measures comparative analysis of the influence of original versus redesigned alerts on outcomes of perceived workload and prescribing errors. Authors also employed elements of usability testing during this study, such as assessing learnability, efficiency, satisfaction and usability errors.

Field User Effect Study *Reminders to physicians from an introspective computer medical record: A two-year randomized trial.*

McDonald et al. (1984) conducted a two-year randomized controlled trial to evaluate the effects of a computer-stored medical record system which reminds physicians about actions needed for patients prior to a patient encounter. This study most closely aligns with a field user effect study for the attention to behavior change in preventive care delivery associated with use of the information resource, and is exemplary because its rigorous design accounts for the hierarchical nature of clinicians working in teams without having to manipulate the practice environment. Randomization occurs at the clustered team level and analyses were performed at both the cluster and individual levels. The study did include problem impact metrics, however no significant changes were observed in these outcomes during the study.

Field User Effect Study *Electronic health records and health care quality over time in a federally qualified health center.*

Kern et al. (2015) conducted a three-year comparative study across six sites of a federally qualified health center in New York to analyze the association between post-implementation of an electronic health record (EHR) and quality of care delivery as measured by change in compliance with Stage 1 Meaningful Use quality measures. This study is an exemplary field user effect study for its attention to measures of clinician behavior in care delivery through test/screening ordering using the EHR and explicit use of statistical analysis techniques to account for repeated measures on patients over time. The study also includes two problem impact metrics (change in HbA1c and LDL cholesterol) analyzed over the study period; however, the study intent was primarily focused on clinician ordering behavior.

Problem Impact Study *Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): A randomised trial.*

Lester et al. (2010) is an exemplar for problem impact studies. Authors conducted a randomized controlled trial to measure improvement in patient adherence to antiretroviral therapy (ART) and suppression of viral load following receipt of mobile phone communications with health care workers. The study randomized patients to the intervention group (receiving mobile phone messages from healthcare workers) or to the control group (standard care). Outcomes were clearly identified and focused on behavioral effects (drug adherence) and an overall intent to measure the extent that improvements in adherence influenced patient health status (viral load). The special attention to randomization and use of effect size metrics for analysis are critical components to measuring the overall impact of mobile phone communications on patient health.

Suggested Reading

Ammenwerth, E., & Rigby, M. (Eds.). (2016). *Evidence-based health informatics*. Amsterdam: IOS Press. This work includes an extensive exploration of evaluation methods pertinent to health informatics.

Anderson, J. G., & Aydin, C. E. (2005). *Evaluating the organizational impact of health care information systems*. New York: Springer. This is an excellent edited volume that covers a wide range of methodological and substantive approaches to evaluation in informatics.

Brender, J. (2006). *Handbook for evaluation for health informatics*. Burlington: Elsevier Academic Press. Along with the Friedman and Wyatt text cited below, one of few textbooks available that focuses on evaluation in health informatics.

Cohen, P. R. (1995). *Empirical methods for artificial intelligence*. Cambridge, MA: MIT Press. This is a nicely written, detailed book that is focused on evaluation of artificial intelligence applications, not necessarily those operating in medical domains. It emphasizes objectivist methods and could serve as a basic statistics course for computer science students.

Fink, A. (2004). *Evaluation fundamentals: Insights into the outcomes, effectiveness, and quality of health programs* (2nd ed.). Thousand Oaks: Sage Publications. A popular text that discusses evaluation in the general domain of health.

Friedman, C. P., & Wyatt, J. C. (2006). *Evaluation methods in biomedical informatics*. New York: Springer. This is the book on which the current chapter is based. It offers expanded discussion of almost all issues and concepts raised in the current chapter.

Jain, R. (1991). *The art of computer systems performance analysis: Techniques for experimental design, measurement, simulation, and modelling*. New York: Wiley. This work offers a technical discussion of a range of objectivist methods used to study computer systems. The scope is broader than Cohen's book (1995) described earlier. It contains many case studies and examples and assumes knowledge of basic statistics.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Thousand Oaks: Sage Publications. This is a classic book on subjectivist methods. The work is very rigorous but also very easy to read. Because it does not focus on medical

domains or information systems, readers must make their own extrapolations.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks: Sage Publications. This is a valuable textbook on evaluation, emphasizing objectivist methods, and is very well written. It is generic in scope, and the reader must relate the content to biomedical informatics. There are several excellent chapters addressing pragmatic issues of evaluation. These nicely complement the chapters on statistics and formal study designs.

Questions for Discussion

1. Associate each of the following hypothetical evaluation scenarios with one or more of the nine types of studies listed in Table 13.1. Note that some scenarios may include more than one type of study.
   (a) An order communication system is implemented in a small hospital. Changes in laboratory workload are assessed.
   (b) The developers of the order communication system recruit five potential users to help them assess how readily each of the main functions can be accessed from the opening screen and how long it takes users to complete them.
   (c) A study team performs a thorough analysis of the information required by psychiatrists to whom patients are referred by a community social worker.
   (d) A biomedical informatics expert is asked for her opinion about a PhD project on a new bioinformatics algorithm. She requests copies of the student's code and documentation for review.
   (e) A new intensive care unit system is implemented alongside manual paper charting for a month. At the end of this time, the quality of the computer-derived data and data recorded on the paper charts is compared. A panel of intensive care experts is asked to identify, independently, episodes of hypotension from each data set.
   (f) A biomedical informatics professor is invited to join the steering group for a series of apps to support people living with diabetes. The only documentation available to critique at the first meeting is a statement of the project goal, description of the planned development method, and the advertisements and job descriptions for team members.
   (g) Developers invite educationalists to test a prototype of a computer-aided learning system as part of a user-centered design workshop
   (h) A program is devised that generates a predicted 24-h blood glucose profile using seven clinical parameters. Another program uses this profile and other patient data to advise on insulin dosages. Diabetologists are asked to prescribe insulin for a series of "paper patients" given the 24-h profile alone, and then again after seeing the computer-generated advice. They are also asked their opinion of the advice.
   (i) A program to generate alerts to prevent drug interactions is installed in a geriatric clinic that already has a computer-based medical record system. Rates of clinically significant drug interactions are compared before and after installation of the alerting program.
2. Choose any alternative area of biomedicine (e.g., drug trials) as a point of comparison, and list at least four factors that make studies in biomedical informatics more difficult to conduct successfully than in that area. Given these difficulties, discuss whether it is worthwhile to conduct empirical studies in biomedical informatics or whether we should use intuition or the

marketplace as the primary indicators of the value of an information resource.

3. Assume that you run a philanthropic organization that supports biomedical informatics. In investing the scarce resources of your organization, you have to choose between funding a new system or resource development, or funding empirical studies of resources already developed. What would you choose? How would you justify your decision?
4. To what extent is it possible to be certain how effective a medical informatics resource really is? What are the most important criteria of effectiveness?
5. Do you believe that independent, unbiased observers of the same behavior or outcome should agree on the quality of that outcome?
6. Many of the evaluation approaches assert that a single unbiased observer is a legitimate source of information in an evaluation, even if that observer's data or judgments are unsubstantiated by other people. Give examples drawn from our society where we vest important decisions in a single experienced and presumed impartial individual.
7. Do you agree with the statement that all evaluations appear equivocal when subjected to serious scrutiny? Explain your answer.

References

Ammenwerth, E. (2015). Evidence-based health informatics: How do we know what we know? *Methods of Information in Medicine, 54*(04), 298–307.

Anderson, J. G., Aydin, C. E., & Jay, S. J. (Eds.). (1994). *Evaluating health care information systems*. Thousand Oaks: Sage Publications Inc.

Beaudoin, M., Kabanza, F., Nault, V., & Valiquette, L. (2016). Evaluation of a machine learning capability for a clinical decision support system to enhance antimicrobial stewardship programs. *Artificial Intelligence in Medicine, 68*, 29–36.

Black, A. D., Car, J., Pagliari, C., Anandan, C., Cresswell, K., Bokun, T., et al. (2011). The impact of eHealth on the quality and safety of health care: A systematic overview. *PLoS Medicine, 8*(1), e1000387.

Brender, J. (2005). *Handbook of evaluation methods for health informatics*. Burlington: Academic Press.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi experimental designs for research*. Boston: Houghton Mifflin, reprinted often since.

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *BMJ, 321*(7262), 694–696.

Friedman, C. P., Abbas, U. L. (2003), Is medical informatics a mature science? A review of measurement practice in outcome studies of clinical systems, *International Journal of Medical Informatics, 69*; (2–3), Pages 261–272, ISSN 1386-5056, https://doi.org/10.1016/S1386-5056(02)00109-0.

Davey Smith, D. (2007). Capitalizing on Mendelian randomization to assess the effects of treatments. *Journal of the Royal Society of Medicine, 100*(9), 432–435. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1963388/.

Demiris, G., Speedie, S., & Finkelstein, S. (2000). A questionnaire for the assessment of patients' impressions of the risks and benefits of home telecare. *Journal of Telemedicine and Telecare, 6*(5), 278–284.

Elizabeth Murray, Eric B. Hekler, Gerhard Andersson, Linda M. Collins, Aiden Doherty, Chris Hollis, Daniel E. Rivera, Robert West, Jeremy C. Wyatt, (2016) Evaluating Digital Health Interventions. American Journal of Preventive Medicine 51 (5):843–851.

Eminovic, N., Wyatt, J. C., Tarpey, A. M., Murray, G., & Ingrams, G. J. (2004, June 02). First evaluation of the NHS direct online clinical enquiry service: A nurse-led Web chat triage service for the public. *Journal of Medical Internet Research, 6*(2), E17.

European Union Medical Devices Regulatory Framework. (2018). https://ec.europa.eu/growth/sectors/medical-devices/regulatory-framework_en. Accessed 24 Oct 2018.

Forsythe, D. E. (1992). Using ethnography to build a working system: Rethinking basic design assumptions. In *Proceedings annual symposium computer applications in medical care* (pp. 505–509).

Forsythe, D. E., Buchanan, B. G., Osheroff, J. A., & Miller, R. A. (1992). Expanding the concept of medical information: An observational study of physicians' information needs. *Computers and Biomedical Research, 25*, 181–200.

Fox, J. (1993). Decision support systems as safety-critical components: Towards a safety culture for medical informatics. *Methods of Information in Medicine, 32*, 345–348.

Friedman, C. P., & Wyatt, J. C. (2005). *Evaluation methods in biomedical informatics* (2nd ed., p. 386). New York: Springer-Publishing. ISBN 0-387-25889-2.

Gaschnig, J., Klahr, P., Pople, H., Shortliffe, E., & Terry, A. (1983). Evaluation of expert systems: Issues and case studies. In F. Hayes-Roth, D. A. Waterman, &

D. Lenat (Eds.), *Building expert systems*. Reading: Addison Wesley.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association: JAMIA, 19*, 121–127.

Gray, E., Marti, J., Brewster, D. H., Wyatt, J. C., Piaguet-Rossel R., & Hall P. S. (2019). Feasibility and results of four real-world evidence methods for estimating the effectiveness of adjuvant chemotherapy in early stage breast cancer. *J Clin Epidemiol*.

Haddow, G., Bruce, A., Sathanandam, S., & Wyatt, J. C. (2011). 'Nothing is really safe': A focus group study on the processes of anonymizing and sharing of health data for research purposes. *Journal of Evaluation in Clinical Practice, 17*, 1140–1146.

Herasevich, V., & Pickering, B. W. (2017). *Health information technology evaluation handbook: From meaningful use to meaningful outcome*. Boca Raton: CRC Press.

House, E. (1980). *Evaluating with validity*. San Francisco: Sage.

Kern, L. M., Edwards, A. M., Pichardo, M., & Kaushal, R. (2015). Electronic health records and health care quality over time in a federally qualified health center. *Journal of the American Medical Informatics Association, 22*(2), 453–458.

Koppel, R., Metlay, J. P., Cohen, A., Abaluck, B., Localio, A. R., Kimmel, S. E., & Strom, B. L. (2005). Role of computerized physician order entry systems in facilitating medication errors. *JAMA: The Journal of the American Medical Association, 293*(10), 1197–1203.

Lester, R. T., Ritvo, P., Mills, E. J., Kariri, A., Karanja, S., Chung, M. H., Jack, W., Habyarimana, J., Sadatsafavi, M., Najafzadeh, M., Marra, C. A., Estambale, B., Ngugi, E., Ball, T. B., Thabane, L., Gelmon, L. J., Kimani, J., Ackers, M., & Plummer, F. A. (2010). Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): A randomised trial. *The Lancet, 376*(9755), 1838–1845.

Littlejohns, P., Wyatt, J. C., & Garvican, L. (2003, April 19). Evaluating computerised health information systems: Hard lessons still to be learnt. *BMJ, 326*(7394), 860–863.

Liu, J. L. Y., & Wyatt, J. C. (2011). The case for randomized controlled trials to assess the impact of clinical information systems. *Journal of the American Medical Informatics Association: JAMIA, 18*(2), 173–180.

Liu, Y. I., Kamaya, A., et al. (2011). A Bayesian network for differentiating benign from malignant thyroid nodules using sonographic and demographic features. *AJR: American Journal of Roentgenology, 196*(5), W598–W605.

Lundsgaarde, H. P. (1987). Evaluating medical expert systems. *Social Science & Medicine, 24*, 805–819.

Lunenburg, F. C. (2010). Managing change: The role of the change agent. *International Journal of Management, Business and Administration, 13*(1), 1–6.

Mant, J., & Hicks, N. (1995). Detecting differences in quality of care: The sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ, 311*, 793–796.

McDonald, C. J., Hui, S. L., Smith, D. M., Tierney, W. M., Cohen, S. J., Weinberger, M., & McCabe, G. P. (1984). Reminders to physicians from an introspective computer medical record: A two-year randomized trial. *Annals of Internal Medicine, 100*(1), 130–138.

McMurry, T. L., Hu, Y., Blackstone, E. H., & Kozower, B. D. (2015, July). Propensity scores: Methods, considerations, and applications in the Journal of Thoracic and Cardiovascular Surgery. *The Journal of Thoracic and Cardiovascular Surgery, 150*(1), 14–19. https://doi.org/10.1016/j.jtcvs.2015.03.057. Epub 2015 Apr 2.

Michaelis, J., Wellek, S., & Willems, J. L. (1990). Reference standards for software evaluation. *Methods of Information in Medicine, 29*, 289–297.

Murray, M. D., Harris, L. E., Overhage, J. M., Zhou, X. H., Eckert, G. J., Smith, F. E., Buchanan, N. N., Wolinsky, F. D., McDonald, C. J., & Tierney, W. M. (2004, March). Failure of computerized treatment suggestions to improve health outcomes of outpatients with uncomplicated hypertension: Results of a randomized controlled trial. *Pharmacotherapy, 24*(3), 324–337.

Nielsen, J. (1994). *Usability inspection methods*. Paper presented at the conference companion on human factors in computing systems, Boston.

Office of the National Coordinator for Health Information Technology (ONC) website. (2014). FDASIA Committee Report. https://www.healthit.gov/sites/default/files/fdasiahealthitreport_final.pdf. Accessed 25 Oct 2018.

Ong, M. S., & Coiera, E. (2011, June). A systematic review of failures in handoff communication during intrahospital transfers. *Joint Commission Journal on Quality and Patient Safety, 37*(6), 274–284.

Pinsky, P. F., Miller, A., Kramer, B. S., Church, T., Reding, D., Prorok, P., Gelmann, E., Schoen, R. E., Buys, S., Hayes, R. B., & Berg, C. D. (2007, April 15). Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. *American Journal of Epidemiology, 165*(8), 874–881.

Pope, C., Halford, S., Turnbull, J., Prichard, J., Calestani, M., & May, C. (2013). Using computer decision support systems in NHS emergency and urgent care: Ethnographic study using normalisation process theory. *BMC Health Services Research, 13*(1), 111.

Ramnarayan, P., Kapoor, R. R., Coren, M., Nanduri, V., Tomlinson, A. L., Taylor, P. M., Wyatt, J. C., & Britto, J. F. (2003, November–December). Measuring the impact of diagnostic decision support on the quality of clinical decision mak-

ing: Development of a reliable and valid composite score. *Journal of the American Medical Informatics Association: JAMIA, 10*(6), 563–572.

Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2017). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics, 21*(1), 4–21.

Rigby, M., Forsström, J., Ruth, R., & Wyatt, J. (2001). Verifying quality and safety in health informatics services. *BMJ, 323*, 552–556.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd ed.). Philadephia: Lippincott Williams & Wilkins.

Russ, A. L., Zillich, A. J., Melton, B. L., Russell, S. A., Chen, S., Spina, J. R., et al. (2014). Applying human factors principles to alert design increases efficiency and reduces prescribing errors in a scenario-based simulation. *Journal of the American Medical Informatics Association, 21*(e2), e287–e296.

Saitwal, H., Feng, X., Walji, M., Patel, V., & Zhang, J. (2010). Assessing performance of an electronic health record (EHR) using cognitive task analysis. *International Journal of Medical Informatics, 79*(7), 501–506.

Scott, G. P., Shah, P., Wyatt, J. C., Makubate, B., & Cross, F. W. (2011, August 11). Making electronic prescribing alerts more effective: Scenario-based experimental study in junior doctors. *Journal of the American Medical Informatics Association: JAMIA, 18*(6), 789–798.

Scott, P. J., Brown, A. W., Adedeji, T., Wyatt, J. C., Georgiou, A., Eisenstein, E. L., & Friedman, C. P. (2019). A review of measurement practice in studies of clinical decision support systems 1998–2017. *Journal of the American Medical Informatics Association, 26*(10), 1120–1128.

Sheikh, A., Cornford, T., Barber, N., Avery, A., Takian, A., Lichtner, V., et al. (2011). Implementation and adoption of nationwide electronic health records in secondary care in England: Final qualitative results from prospective national evaluation in "early adopter" hospitals. *BMJ, 343*, d6054.

Sherman, R. E., Anderson, S. A., Dal Pan, G. J., Gray, G. W., Gross, T., Hunter, N. L., LaVange, L., Marinac-Dabic, D., Marks, P. W., Robb, M. A., Shuren, J., Temple, R., Woodcock, J., Yue, L. Q., & Califf, R. M. (2016, December 8). Real-world evidence – what is it and what can it tell us? *The New England Journal of Medicine, 375*(23), 2293–2297.

Slight, S. P., & Bates, D. W. (2014). A risk-based regulatory framework for health IT: Recommendations of the FDASIA working group. *Journal of the American Medical Informatics Association, 21*(e2), e181–e184.

Somerville I. (2015). Software Engineering (10th edition). Pearson, 816 pages. ISBN-10: 0133943038.

Spiegelhalter, D. J. (1983). Evaluation of medical decision-aids, with an application to a system for dyspepsia. *Statistics in Medicine, 2*, 207–216.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929–1958.

Stead, W., Haynes, R. B., Fuller, S., et al. (1994). Designing medical informatics research and library projects to increase what is learned. *Journal of theAmerican Medical Informatics Association, 1*, 28–34.

Streeter, A. J., Lin, N. X., Crathorne, L., Haasova, M., Hyde, C., Melzer, D., & Henley, W. E. (2017, July). Adjusting for unmeasured confounding in nonrandomized longitudinal studies: A methodological review. *Journal of Clinical Epidemiology, 87*, 23–34. https://doi.org/10.1016/j.jclinepi.2017.04.022. Epub 2017 Apr 28.

Szczepura, A., & Kankaanpaa, J. (1996). *Assessment of health care technologies*. London: Wiley.

Talmon, J., Ammenwerth, E., Brender, J., de Keizer, N., Nykänen, P., & Rigby, M. (2009). STARE-HI—statement on reporting of evaluation studies in health informatics. *International Journal of Medical Informatics, 7*, 1–9.

van Gennip, E. M., & Talmon, J. L. (Eds.). (1995). *Assessment and evaluation of information technologies in medicine*. Amsterdam: IOS Press.

Van Way, C. W., Murphy, J. R., Dunn, E. L., & Elerding, S. C. (1982). A feasibility study of computer-aided diagnosis in appendicitis. *Surgery, Gynecology & Obstetrics, 155*, 685–688.

Ventres, W., Kooienga, S., Vuckovic, N., Marlin, R., Nygren, P., & Stewart, V. (2006) *The Annals of Family Medicine, 4*(2) 124-131; https://doi.org/10.1016/10.1370/afm.425.

Wasson, J. H., Sox, H. C., Neff, R. K., & Goldman, L. (1985). Clinical prediction rules: Applications and methodological standards. *The New England Journal of Medicine, 313*, 793–799.

Wolf, J. A., Moreau, J. F., Akilov, O., Patton, T., English, J. C., Ho, J., & Ferris, L. K. (2013). Diagnostic inaccuracy of smartphone applications for melanoma detection. *JAMA Dermatology, 149*(4), 422–426.

Wright, A., Sittig, D. F., Ash, J. S., Erickson, J. L., Hickman, T. T., Paterno, M., et al. (2015). Lessons learned from implementing service-oriented clinical decision support at four sites: A qualitative study. *International Journal of Medical Informatics, 84*(11), 901–911.

Wyatt, J., & Spiegelhalter, D. (1990). Evaluating medical expert systems: What to test and how? *Medical Informatics (London), 15*, 205–217.

Wyatt, J., & Wyatt, S. (2003). When and how to evaluate clinical information systems ? *International Journal of Medical Informatics, 69*, 251–259.

Wyatt, J. C., Batley, R. P., & Keen, J. (2010, October). GP preferences for information systems: Conjoint analysis of speed, reliability, access and users. *Journal of Evaluation in Clinical Practice, 16*(5), 911–915.

Zhang, J., Johnson, T. R., Patel, V. L., Paige, D. L., & Kubose, T. (2003). Using usability heuristics to evaluate patient safety of medical devices. *Journal of Biomedical Informatics, 36*(1–2), 23–30.

Biomedical Informatics Applications

Contents

Electronic Health Records

Genevieve B. Melton, Clement J. McDonald, Paul C. Tang, and George Hripcsak

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_14

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What is the definition of an EHR?
- What are the functional components of an EHR?
- What are the benefits of an EHR?
- What are the some of the impediments to development, configuration, and use of an EHR?

14.1 What Is an Electronic Health Record?

The preceding chapters introduced the conceptual basis for the field of biomedical informatics, including the use of patient data in clinical practice and research. The chapters in this section cover the various technologies, systems, and approaches of biomedical informatics in practice. This chapter focuses on the **patient record** and associated systems, commonly referred to as the patient's chart, medical record, or health record. In particular, we define and examine the use of **electronic health records (EHRs)**,[1] discuss their purpose and functional components, potential benefits and costs, and describe current challenges and opportunities in their dissemination, optimal use, and innovation.

1 The terms "electronic health record" (EHR) and "electronic health record system" (EHRS or EHR system) are often used interchangeably, with no generally agreed upon distinction. The terms "electronic medical record" (EMR) and "electronic medical record system" (EMRS or EMR system) have also been used but many have moved towards using the term "health" versus "medical" as these systems are being used broadly across the continuum of care by a multitude of roles; again, the distinctions between EMR and EMRS are not generally agreed upon. The term "computer-based record system" has also been used in the past. In this chapter, and throughout the book, we will use the term "electronic health record", with the acronyms "EHR" and "EHRs" to designate the singular and plural forms, respectively.

14.1.1 Purpose of a Patient Record

Stanley Reiser (1991) wrote that the purpose of a patient record is "to recall observations, to inform others, to instruct students, to gain knowledge, to monitor performance, and to justify interventions." The many uses described in this statement, although diverse, have a single goal—to leverage patient data and information within the record to care for patients and to further health sciences, including the conduct of research and public health activities that address population health. Traditionally, the patient record was on paper and almost exclusively used by providers, nurses, and other care team members to document and facilitate the care of patients. The paper record contained clinical notes documenting assessments, decision-making, and care rendered; paper charts also often contained diagnostic results (e.g., laboratory and imaging results); physiologic information (e.g., vital signs); and patient care orders. With the increased digitization of health care, modern EHRs have become increasingly ubiquitous; they are designed not only for patient care but also to facilitate broader value-added functions and views of patient information, providing much more than a static view of events.

14.1.2 EHR Overview

An EHR is an electronic repository of maintained information about an individual's health status and health care with functionality to enable provision of care and information stored in order to serve the multiple legitimate uses and users of the record. While traditional patient records have been illness-focused, health care is evolving to encourage health care providers to focus on the continuum of health and health care from wellness to illness and recovery.

As a result of this shift and increasing system inter-operability of EHRs, many anticipate that EHRs will increasingly carry a much

greater portion of a person's health-related information from a wide range of sources over their lifetime (e.g., diagnostic images, intensive care electrophysiologic monitoring, patient recorded biometrics, genomic information). Today, the Department of Veterans Affairs (VA) has already committed to keeping existing electronic patient data for 75 years. In many cases, EHRs incorporate or integrate with a range of additional multimedia sources, such as radiology images and echocardiographic video loops.

EHRs also often include active tools used to manage patient with a wide range of features. This includes information management tools to provide clinical reminders and alerts, dynamic tracking and trending which can often be personalized, linkages with knowledge sources for health care clinical decision support (CDS; see ▶ Chap. 26), and analysis of aggregate data both for care management and for research. The EHR also helps users organize, interpret, and react to patient data. Examples of tools provided in current EHRs are discussed in ▶ Sect. 14.3. As such, EHRs need to be able to provide different views and presentations of patient data to meet the needs of various user types and patient care contexts, along with associated functionality to serve patient care and various secondary EHR uses as described in ▶ Chap. 2.

EHRs can also analyze a patient's record, call attention to trends and dangerous conditions, and suggest corrective actions much like an airplane flight control information system. Another powerful aspect of EHRs is their ability to organize data at both the individual and population level such as a view to facilitate care for one patient or one for a population of patients to assist with care management decisions or answer epidemiologic questions.

One advantage of EHRs is the availability of information entry controls and capabilities. In addition to increased legibility compared to paper, EHRs can also increase the quality of data by applying validity checks as data is being entered like typographical errors checks and other checks (e.g., dosing ranges for medications). EHRs can require data entry in specified fields, conditional on the value of other fields. As such, EHRs not only store data, but can also conditionally enforce the capture of certain data elements. This enforcement power should be used judiciously, however, and not *require* the entry of unavailable data (e.g., the age of an unidentified patient receiving care during an emergency trauma) especially during order entry, and potentially prevent the clinician from completing an important order needed for clinical care (Strom et al. 2010).

The degree to which a particular EHR achieves its intended value depends on several factors:

Comprehensiveness of information Does the system include information from all organizations and clinicians who participated in a patient's care and from all settings where care was delivered (e.g., office practice, hospital, homecare, care coordination, virtual care)? Does it include the full spectrum of clinical data, including clinical notes, laboratory test results, medication details, images, and patient reported outcomes, including those collected by validate patient self-assessments. Increasingly, genetic data (including both germline and somatic tumor data) will become key to clinical care and EHRs. Incorporation of genetic data (including its interpretation(s) and analysis model(s)) will bring important data management and knowledge management issues including data storage (e.g., 1100 patients for 1 year requiring 2 terabytes (Burykin 2011)) and evolving interpretation(s) of genetic data and computational analyses of this data (see ▶ Chap. 30).

Duration of use and retention of data EHRs gain value over time through accumulation of a greater proportion of the patients' medical history. A record system with 5 years of patient data will be more valuable than one with only the last month's records. Retention of medical records follow, at minimum, the statute of limitations for medical malpractice, which are state-based laws. EHR records can be archived on various systems or maintained for ongoing access.

Degree of structure of data Narrative notes stored in EHRs have the advantage over their paper predecessors as they can be searched by word or patterns. The success of such searches is limited by the sophistication of the EHR's text mining and search capabilities along with the quality of the user's search criteria particularly since there can often be language variability in expression of medical terms including the use of abbreviations.

A great deal of valuable information is contained in clinical notes mostly recorded in narrative. Though often time consuming and cumbersome, one way to obtain structured data with clinical notes is to ask the clinician to enter key information through structured forms which restrict data entry with a controlled vocabulary that is standards-based. Powerful automated techniques like natural language processing can also be increasingly leveraged to extract clinical data from notes (see ► Chap. 9).

Ubiquity of access With today's secure networks and other distributed technologies, clinicians and patients can access a patient's EHR from geographically distributed sites. Paper records have significant **inaccessibility** issues as they can only be in one place and with at most one user at any point in time. Previously, completing discharge summaries and signing orders with paper records or borrowing records for administrative or research purposes from medical record departments was logistically challenging. In contrast, EHRs are ubiquitously available to users for these and other purposes.

In some cases, the collective data about one patient from independent care systems can also be accessed through health information exchanges (see ► Chap. 17). Such availability can also support health care continuity during disasters. Brown et al. (2007) found a "stark contrast" between the care VA versus non-VA patients received after Hurricane Katrina, because appropriate and uninterrupted care were supported by nationwide access to the comprehensive VA EHRs. EHRs not only make data more accessible to authorized users, but they also provide the benefit of greater control over data and user access and improved enforcement of applicable privacy regulations as required by the Health Insurance Portability and Accountability Act (HIPAA) (see ► Chap. 31).

System challenges can be experienced from several perspectives:

Challenges with system use and training Physicians and other key personnel have to take time from their work to learn how to use the system. Furthermore, clinical workflows have to be re-designed in order to utilize the EHR effectively. It is increasingly appreciated that EHR usability needs to be improved so that clinicians can easily document and provide patient care.

Readability of clinical notes One of the advantages of EHRs is the ease with which text can be entered compared to paper notes. Unfortunately, this can result in something referred to as "note bloat", resulting from cutting and pasting text and inserting other results (► Sect. 14.3.1.2), resulting in voluminous documentation compared to paper records, making EHR notes often more difficult to review efficiently.

System failures and ensuring adequate redundancy and security Computer-based systems have the potential for catastrophic failures that could cause extended unavailability of patients' computer records. To combat this, EHRs today often run across distributed technologies offering redundancy such as parallel operation at geographically separate sites and **hot fail over** with separate computer systems running synchronously with the primary system that can take over near instantaneously from the primary system if it were to fail. Yet, nothing provides complete protection; contingency plans with **downtime procedures** must be developed for handling brief (even planned) or longer system outages. Also, cybersecurity is increasingly a concern that health care systems are working to improve as attacks are of increasing sophistication and frequency (see ► Chap. 18).

14.2 Historical Perspective: Development of EHRs

The initial development of automated systems in health care was stimulated by regulatory and reimbursement requirements. Early health care systems in the inpatient setting provided charge capture functionality to meet billing requirements in a fee-for-service environment.

The Flexner report on medical education was the first formal statement made about the function and contents of the medical record (Flexner 1910). In advocating a scientific approach to medical education, the Flexner report also encouraged physicians to keep a patient-oriented medical record. Three years earlier, Dr. Henry Plummer initiated the "unit record" for the Mayo Clinic (including its St. Mary's Hospital), placing all the patient's visits and types of information in a single folder. This innovation represented the first longitudinal medical record (Melton 3rd 1996). The Presbyterian Hospital (New York) adopted the unit record for its inpatient and outpatient care in 1916, studying the effect of the unit record on length of stay and quality of care (Openchowski 1925) and writing a series of letters and books about the unit record that disseminated the approach around the nation (Lamb 1955).

The first record we could find of a computer-based medical record was a short newspaper article describing a new "electronic brain" – to replace punched and file index cards and to track hospital and medical records by the Michigan Hospital Service (Brain 1956). The first, operational EHRs emerged in the early 1970's. Some started an out patient systems, including Costar from Massachusetts General Hospital (Grossman et al. 1973; Barnett et al. 1979; Barnett 1984), RMRS from the Regenstrief Institute (McDonald 1973; McDonald et al. 1975; McDonald et al. 1999), Duke University (Stead 1977; Stead & Hammond 1988), STOR (Simborg and Whiting-O'Keefe III 1981), and others (see outpatient EHR review by Kuhn et al. 1984). Other systems began on the inpatient side including HELP (Warner 1972) and Lockheed's hospital information system (HIS) at El Camino Hospital, which became operational in 1971 (Coffey 1979).

Weed's problem-oriented medical record book (POMR) (1968) shaped medical thinking about both manual and automated medical records. His computer-based inpatient system followed (Schultz et al. 1971). Morris Collen, who also pioneered the multiphasic screening system (1969), wrote a readable 500-page history of medical informatics (1995) that provides rich details about these early medical records systems, as does a three-decade summary of computer-based medical record research projects from the U.S. Agency for Health Care Policy and Research (AHCPR) (Fitzmaurice et al. 2002).

EHRs can provide Clinical Decision Support (CDS) by suggesting needed action based on the patient data it carries. A few early systems: HELP (Warner 1972; Pryor 1988) the RMRS (McDonald 1973, 1976) offered CDS as part of their initial design. Other early EHRs added CDS capability as they grew: the Columbia University system (Johnson et al. 1991; Hripcsak et al. 1999), the CCC (Center for Clinical Computing) system at Beth Israel Deaconess Medical Center (Rind et al. 1994; Slack and Bleich 1999; Bleich et al. 1985; Halamka and Safran 1998), and others (Giuse and Mickish 1996; Teich et al. 1999; Cheung et al. 2001; Duncan et al. 2001; Brown et al. 2003).

Since those early years, hundreds of commercial venders have emerged to supply outpatient practice computer systems and a few dozen offered for inpatient systems. However, as hospital systems merged into ever-larger aggregations and pulled office practices through acquisitions, the boundaries between outpatient and inpatient computer systems have blurred and the number of health care system EHR vendors has shrunk considerably. However, specialized clinical information systems that cover the special needs within large and complex health care systems including medical and imaging areas continue. Today, a majority of health systems in the US use a select few EHRs provided by major health information technology (IT) vendors.

14.3 Functional Components of an EHR

An EHR is not simply a recording of the patient's clinical state but also has linkages and functional tools to facilitate communication and decision-making. We summarize the components of a comprehensive EHR and illustrate functionality with examples from systems currently in use. The functional components are:

1. Patient data capture, aggregation, and review
2. Computerized provider order entry
3. Clinical decision support
4. Access to knowledge resources
5. Care team and patient communication
6. Billing and coding

Increasingly, requirements around certified Health IT for EHRs are resulting in systems with specific functionality and system behavior. Certified Health IT system requirements have led to EHRs being more standards-based, and interoperable (2015 Edition Certification Regulations – 170.315[2]).

14.3.1 Patient Data Capture, Aggregation, and Review

Providing an integrated view of relevant patient data and having functionality to enter and supplement patient data are overarching EHR goals. However, EHRs may miss certain patient data, including (1) patient data existing only on old paper records, (2) data from care provided outside of the current organization (e.g., unconnected office practices, free-standing radiology centers, home-health agencies, nursing homes), and (3) differences in data representation despite electronic and organizational links, the latter of which can be a result of different EHR vendors, different implementations of a given vendor's system at different institutions, and by unwillingness to share data. Though some progress has been made in sharing EHR data among institutions especially when institutions have a common EHR vendor, sharing of EHR data remains challenging. For example, institutions may choose different sets of functionality from vendors and employ different business rules and use different codes for identifying tests, measurements and treatment.

14.3.1.1 Data Integration and Standards

An integrated, mature EHR accommodates a broad spectrum of data types ranging from text to numbers and from signals (e.g., EKG waveform) to images as well as increasingly audio and video. More complex data such as radiology images are usually delivered for human viewing — via the DICOM[3] standard or general commercial imaging standards such as JPEG[4] or motion JPEG used for cardiac echocardiograms (see ▶ Chap. 12). ◘ Figure 14.1 shows an example screenshot of WorldVistA CPRS EHR, which integrates a variety of text data and images into a patient report data screen including: demographics, a detailed list of the patient's procedures, a DICOM chest x-ray image, and JPG photo of a skin lesion. Other tabs in the system provide links to problems, medications, orders, notes, consults, discharge summary, and labs.

In addition to challenges with EHR clinical data exchange, another important challenge in the US to the construction of an integrated view of patient data between systems is the lack of a national patient identifier. Because each organization assigns its own medical record number, a receiving organization cannot directly map a local medical record number from an external care organization to its own. Linking algorithms for identity management of patients are typically

2 Certification of Health IT, Testing Process & Test Methods, 2015 Edition Test Method. ▶ https://www.healthit.gov/topic/certification-ehrs/2015-edition-test-method (Accessed 6/4/2020).

3 Digital Imaging and Communications in Medicine, ▶ https://www.dicomstandard.org/ (Accessed 6/4/2020).

4 JPEG from Wikipedia, the free encyclopedia, ▶ http://en.wikipedia.org/wiki/JPEG (Accessed 6/4/2020).

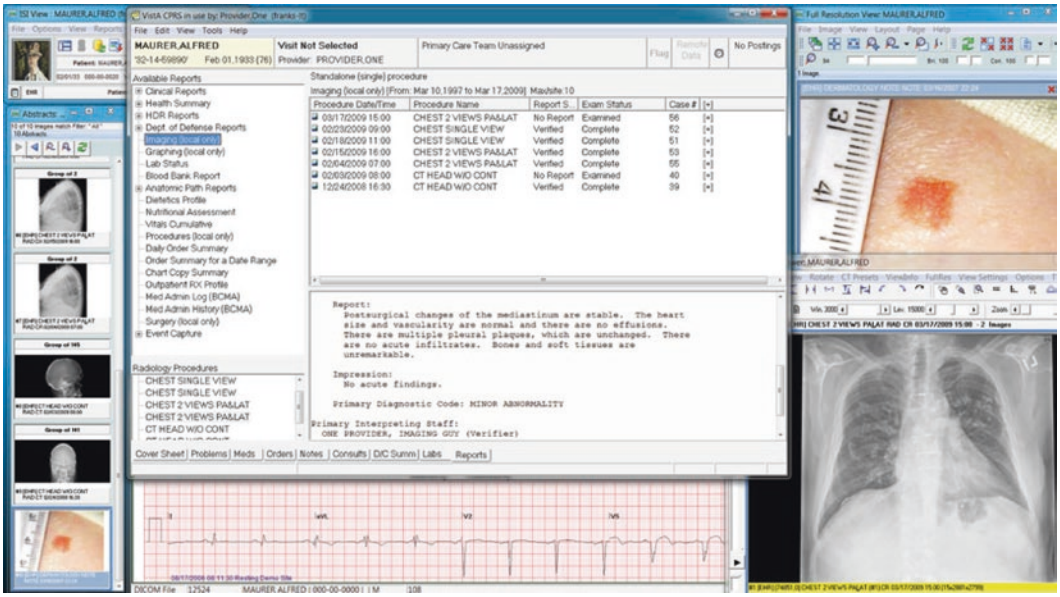


Fig. 14.1 A screenshot of the combined WorldVistA Computer Based Patient Record System (CPRS) and ISI Imaging system. These systems are derived from the Department of Veterans Affairs VistA and VistA Imaging systems (► http://www.va.gov/vista_monograph/). The figure shows how clinical images can be presented with laboratory test results, medications, notes and other relevant clinical information in a single longitudinal medical record. (Source: Courtesy of WorldVistA (► worldvista.org) and II Group (► www.isigp.com), 2012)

based on name, birth date, and other patient characteristics such as address and employment. The performance of these algorithms must be monitored for data integrity issues and errors and associated processes created to manage patient identity for cases where algorithms are not sufficient to adjudicate potential matches (Just et al. 2016; Zech et al. 2016).

One of the more significant barriers today to the integration of health record data from different organizations are the local and idiosyncratic identifiers used to label observations and coded observation values — recapitulating the Babel story. However, those barriers are shrinking as Health IT regulations[5] and institutions adopt terminology standards, including LOINC[6] for observations, questions, variables, and assessments (McDonald et al. 2003; Vreeman et al. 2010); SNOMED CT[7] (Wang et al. 2002) for diagnoses, symptoms, findings, organisms and answers; UCUM[8] for computable units of measure; and RxNorm[9,10] for clinical drug names, ingredients, and orderable drug names for various purposes (see also ► Chaps. 8 and 31). Supporting this trend are laboratory instrument vendors, which are beginning to specify LOINC codes to use for each of the tests results that their instruments generate.[11]

As healthcare providers consolidate and bring together EHR data or implement a new

5 Certification of Health IT, Testing Process & Test Methods, 2015 Edition Test Method. ► https://www.healthit.gov/topic/certification-ehrs/2015-edition-test-method (Accessed 6/4/2020).

6 Logical Observation Identifiers Names and Codes (LOINC®) from Regenstrief. ► http://loinc.org/ (Accessed 6/4/2020).

7 SNOMED Clinical Terms® (SNOMED CT®) Five-step briefing. ► https://www.snomed.org/snomed-ct/ (Accessed 6/4/2020).

8 The Unified Code for Units of Measure. ► http://unitsofmeasure.org/ (Accessed 6/4/2020).

9 RxNorm Overview. ► http://www.nlm.nih.gov/research/umls/rxnorm/overview.html (Accessed 6/4/2020).

10 RxTerms. ► https://wwwcf.nlm.nih.gov/umlslicense/rxtermApp/rxTerm.cfm (Accessed 6/4/2020).

11 ► https://ivdconnectivity.org/fda-encourages-livd (Accessed 6/4/2020).

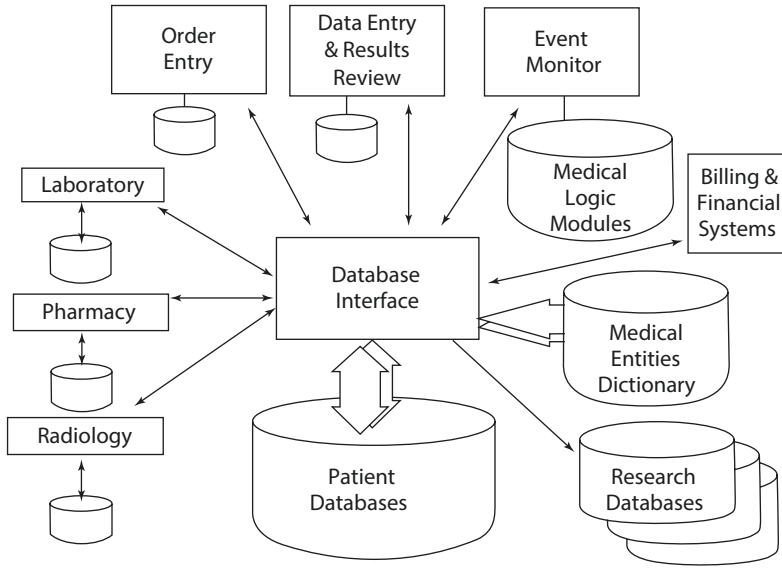


Fig. 14.2 A block diagram of multiple-source-data systems that contribute patient data, which ultimately reside in a computerized patient record (CPR). The database interface, commonly called an interface engine, performs a number of functions. It may simply be a router of information to the central database. It may also provide more intelligent filtering, translating, and alerting functions, as it does at Columbia University Medical Center. (Source: Courtesy of Columbia University Medical Center, New York)

EHR, organizations have used a number of approaches to load EHRs with pre-existing patient data. One approach is to interface the EHR from the available electronic source (e.g., dictation service, pharmacy system, and laboratory information system) and load data from these sources for a pre-specified length of time (e.g., 12 months). A second approach is to abstract select data (e.g., key laboratory results, the problem lists, and active medications) and either automatically or hand enter those data into the new EHR prior to each patient's visit for a period of time. The third approach is to scan and store 1–2 years of the old paper records or to produce electronic "printout" versions (e.g., as Portable Document Format [PDF]) of content stored in by preceding EHR. This approach can be applied to any kind of document, including handwritten records that predate the EHR installation. **Optical Character Recognition** (**OCR**) capability is built into most document scanners today, and converts typed text within scanned documents to computer understandable text with 98–99% character accuracy, which can make this content potentially searchable.

Today, most clinical data sources and EHRs can send and receive clinical content as version 2.x **Health Level 7** (**HL7**)[12] messages. Most organizations use interface engines for HL7 messages and integration platforms (either part of the integration engine or separate technology platform) that can support other data formats to send, receive, and, when necessary, translate the format of and the codes within exchanged data (see ▶ Chap. 17); ◘ Fig. 14.2 shows an example of architecture to integrate data from multiple source systems. The Columbia University Medical Center computerized patient record (CPR) interface depicted in this diagram not only provides message-handling capability but can also automatically translate codes from the external source to the preferred codes of the receiving EHR. And although many vendors now offer single systems that serve "all" needs, they never escape the need for standards-based data exchange from ancillary systems (e.g., EKG carts, cardiol-

12 Health Level Seven International, ▶ http://www.hl7.org/ (Accessed 6/4/2020).

ogy systems, radiology imaging systems, anesthesia systems, off-site laboratories, community pharmacies and external collaborating health systems). At least one high-capability open-source interface engine, NextGen Connect (formerly Mirth Connect),[13,14] is available and used relatively widely for data exchange.

HL7 **Fast Healthcare Interoperability Resources (FHIR)** is an elegant application-programming interface for exchanging clinical data (see ► Chap. 17) with a recognizable heritage from V2. In 2018, it was embraced by a surge of large organizations including Apple, (Apple Health), Microsoft and Google (Mandl et al. 2019), many federal agencies (CMS, ONC, the Veterans Administration), major EHR vendors, and health insurance companies. Most of them are also using the related, SMART on FHIR App specification (Mandel et al. 2016), with which users can develop Apps designed to access data within EHRs from outside of that EHR.

14.3.1.2 Clinician Data Entry

Clinical data may be entered as narrative free-text, as codes, or as a combination of the two. Trade-offs exist between the use of codes and narrative text. The major advantage of structured data is that it makes the data "understandable" to the computer and thus enables selective retrieval, clinical research, quality improvement, and clinical operations management. The coding of diagnoses, allergies, problems, orders, and medications is of particular importance for these purposes.

Because of the chance of errors with the hand entry of data, EHRs apply **validity checks** scrupulously. A number of different kinds of checks apply to clinical data (Schwartz et al. 1985). **Range checks** can detect or prevent entry of values that are out of range (e.g., a serum potassium level of 50.0 mmol/L—which is impossibly outside the normal range of 3.5–5.0 mol/L). **Pattern checks** for including regular expressions can verify that the entered data have a required pattern (e.g., the three digits, hyphen, and four digits of a local telephone number). Range and pattern checks (among others can be implemented using standard browser features).**Computed checks** can verify that values have the correct mathematical relationship (e.g., white blood cell differential counts, reported as percentages, must sum to 100). **Consistency checks** can detect errors by comparing entered data (e.g., the recording of cancer of the prostate as the diagnosis for a female patient). Software for accomplishing this is embedded in standard web browser). **Delta checks** warn of large and unlikely differences between the values of a new result and of the previous observations (e.g., a recorded weight that changes by 100 lbs. in 2 weeks). **Spelling checks** verify the spelling of individual words.

Clinician-gathered patient information requires special comment because it presents one of the most difficult challenge to EHR developers and users. Physicians spend at least 20% of their time documenting the clinical encounter (Gottschalk and Flocke 2005; Hollingsworth et al. 1998). The burden has risen over time for several reasons (Poissant et al. 2005). EHRs tend to require far more data entry than the pre-existing manual systems. Many studies suggests that the EHR functions taken together may consume up to 1–2 hours of the physician's free time per clinic day (Sinsky et al. 2016; McDonald et al. 2014). In one study, the computer system was a primary cause of clinician dissatisfaction (Edgar 2009) and their reason for leaving military medicine. In addition, EHR documentation requirements have been repeatedly been cited as a significant cause of physician burnout (Gardner et al. 2019). EHRs tend to forbid simple narrative text responses; so providers have to dig through menus to find the coded term that expresses their intended meaning. Billing requirements and fear of malpractice have fueled the demand for ever more data entry. Adding to provider' data entry pain is the fact that EHR user inter faces can be clumsy and non-intuitive.

The requirement that providers enter all of their findings, impressions and plans into their note and orders into the computer comes from a dictum that says the person who cre-

13 ► https://github.com/nextgenhealthcare (Accessed 6/4/2020).

14 NextGen Connect Integration Engine. ► https://www.nextgen.com/products-and-services/integration-engine and ► https://github.com/nextgenhealthcare (Both accessed 6/4/2020).

ates the content should enter it. This dictum often makes sense for prescriptions, orders, and perhaps diagnoses and procedure orders, because immediate provider entry during the course of care makes diagnostic testing, treatments and check out more efficient and provides crucial grist for CDS. The justification for direct entry of visit notes by clinicians is weaker because of its time cost to physicians is high and the information is not a prerequisite to the checkout process.

Today, clinical notes can be entered into the EHR via one of three general mechanisms: (1) transcription or use of speech recognition systems to convert spoken word to dictated or written notes, (2) clinic staff (scribes) who transfer text or codes of providers written or spoken content into the computer system, and (3) direct data entry by physicians themselves (potentially facilitated by electronic templates or macros).

Free-form narrative entry—by typing, dictation, or speech recognition—allows clinicians to express important clinical information in the most natural manner. When clinicians communicate in narrative, they naturally prioritize findings and leave much information implicit. For example, an experienced clinician often leaves out "pertinent negatives" (i.e., findings that the patient does not have but that nevertheless inform the decision-making process) knowing that the clinician who reads the record will interpret them properly to be absent. The result is usually a more concise history with a high signal-to-noise ratio that not only shortens the data capture time but also lessens the cognitive burden on the reading clinician. Weir and colleagues present compelling evidence about these advantages, especially when narrative is focused and vivid, and emphasize that too much information interferes with inter-provider communication (Weir et al. 2011).

Most EHRs let physicians cut and paste notes from previous visits and other sources or even have automated functionality to "bring forward" content from previous visits. For example, a physician can cut and paste parts of a visit note into a letter to a referring physician and into an admission note, a most appropriate use of this capability. However, when over-used, this can cause 'note bloat.' In addition, without proper attention to detail, users may copy information that is no longer pertinent or true or possibly lose context especially with time expressions (e.g., "yesterday"). Studies note high rates of text duplicated from previous notes of over 50% (Wrenn et al. 2010; Zhang et al. 2014).

Dictation with **transcription** has been common historically for entering narrative information into EHRs. Transcriptionists are often able to maintain a degree of structure in the transcribed document via section headers, and the structure can also be delivered as an HL7 CDA document (Ferranti et al. 2006). **Speech recognition** software offers an approach to "dictating" without the cost or delay of transcription by translating clinician speech to text automatically. Historically, these systems resulted in errors that required significant time to find and correct misunderstood words. Increasingly, these solutions have improved speech recognition algorithms, and companies can reach accuracies better than 99% without training. Skeptical readers can try it themselves.[15]

In addition, some dictation services use speech recognition to generate a draft transcription, which the transcriptionist corrects while listening to the audio dictation, thus saving transcriptionist time. Natural-language processing (NLP) (see ► Chap. 9) offers hope for automatic encoding of narrative text (Nadkarni et al. 2011). Some companies are exploring the use of NLP to auto-encode transcribed text, and employ the transcriptionist to correct any NLP coding errors (see ► Chap. 9).

Some practices have **scribes** (a variant on the stenographers of old) to do much of the physicians' data entry work (Koshy et al. 2010; Misra-Hebert et al. 2016). Scribes typically work alongside the care provider in the examination room, or remotely through an audiovisual connection or recording. The Joint Commission is agnostic about the use of scribes, but provides guidelines for their use (The Joint Commission 2018[16]).

15 ► https://cloud.google.com/speech-to-text/

16 The Joint Commission: Documentation assistance provided by scribes. ► https://www.jointcommission.org/en/standards/standard-faqs/nursing-care-center/record-of-care-treatment-and-services-rc/000002210/ (Accessed 6/4/2020).

Another data-entry method is to have clinicians record information on a **structured form**, from which data and associated documentation are created (Downs et al. 2006; Hagen et al. 1998). One system, called CHICA (Anand et al. 2018), originally generated a patient specific and scannable paper document and used optical character and mark recognition to capture the recoded data in a two-step process. Today, data can be data captured from a handheld electronic tablet or via a webpage (Anand et al. 2017). In addition to a child-specific data-capture form completed by the child's family in the waiting room (Fig. 14.3a), the CHICA computer uses the entered data to generate a physician encounter form with a tailored agenda for the encounter (Fig. 14.3b). The CHICA system generates a prose version of associated form responses which can be incorporated into and used to help generate a clinical note.

The third alternative is the structured, coded entry of data by clinicians. A major issue associated with direct physician entry is

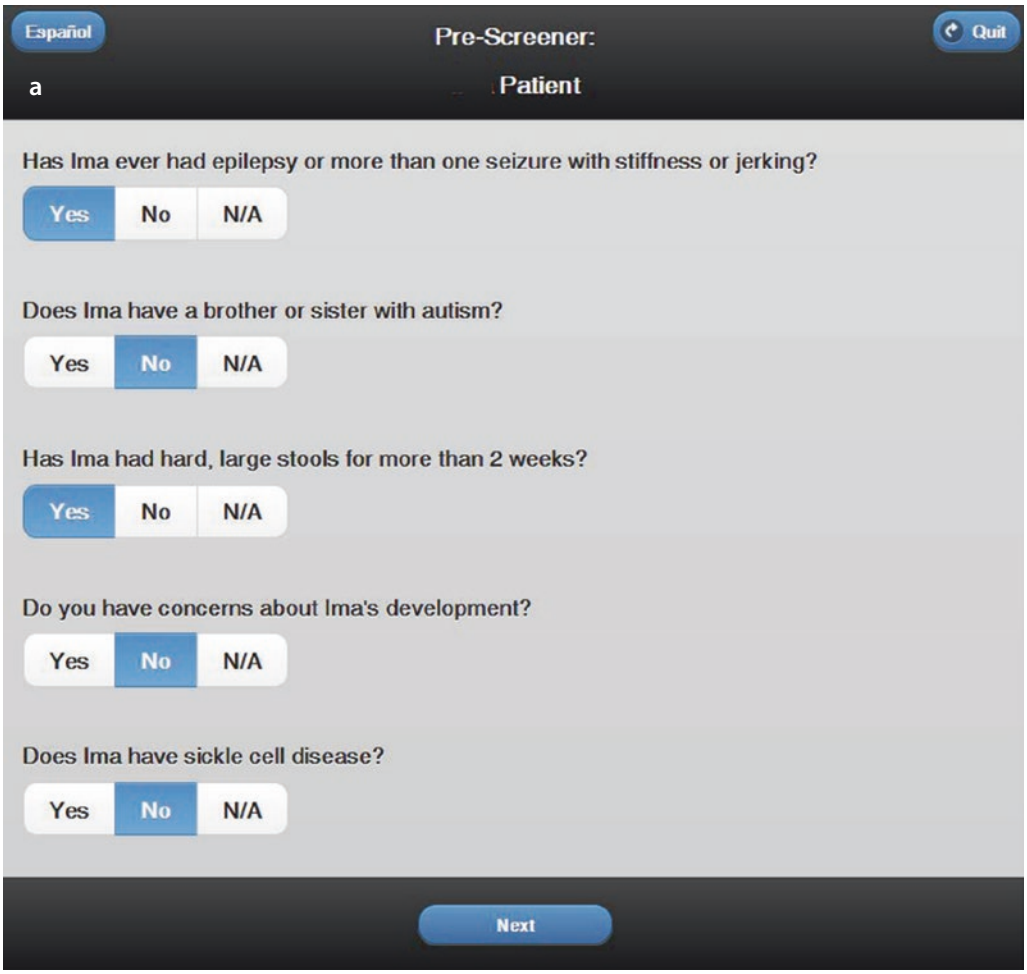


Fig. 14.3 **a** The family completes the first form with questions tailored to patient's age and other factors. Form can be displayed on a tablet or printed on a tailored paper form that is scanned by an OCR system that passes the content to the EHR. **b** The computer generates a physician encounter form based on the contents of the first form and adds reminders. The form is displayed as a web form in the EHR or printed on paper that an OCR system interprets. Coded results are stored in the computer and a prose version is returned by the system to be incorporated in the physician's note. (Source: Courtesy of Prof Stephen M Downs, Indiana University, Indianapolis, IN)

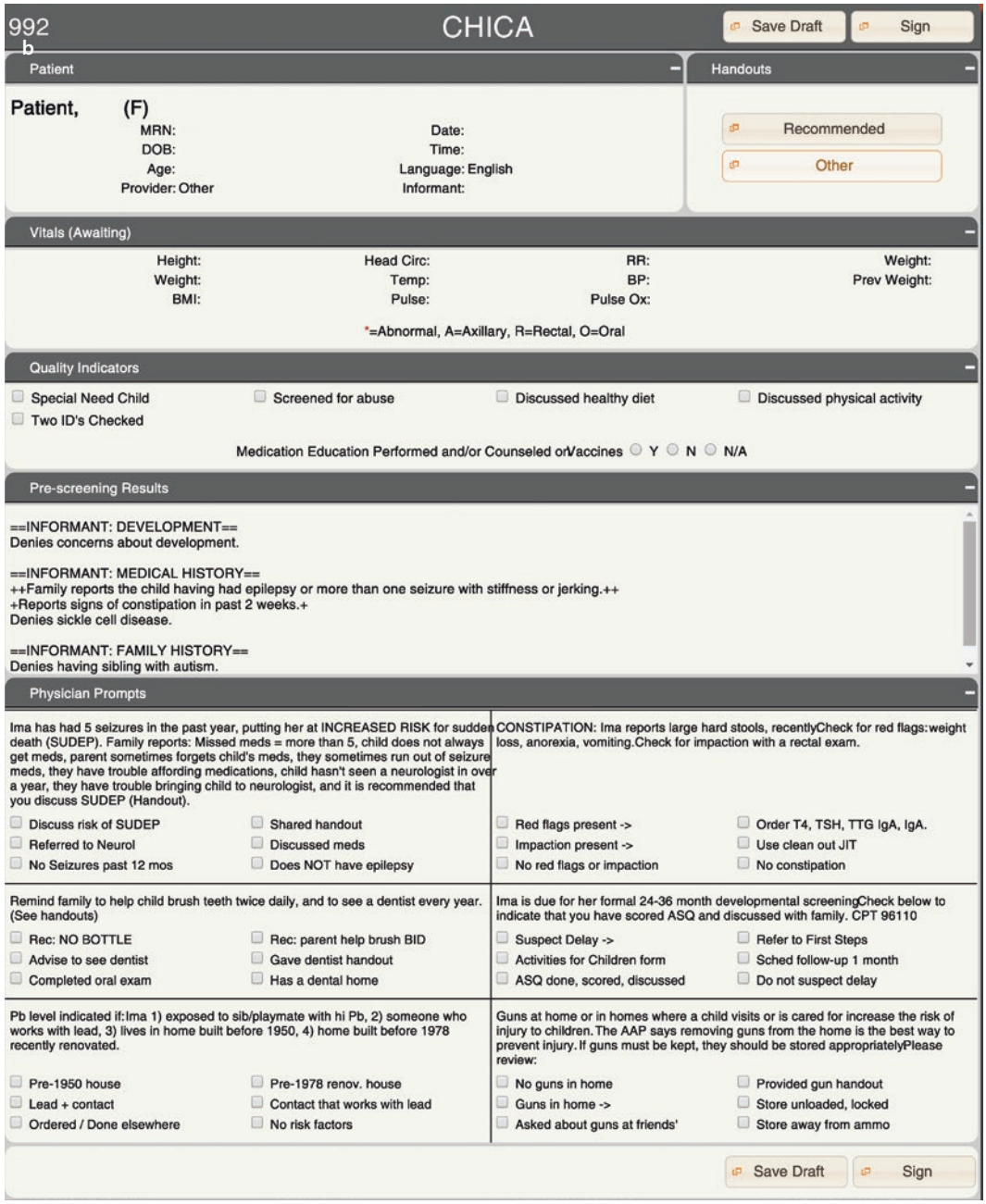


Fig. 14.3 (continued)

the physician time cost. Studies document that structured data entry consumes more clinician time than the traditional record keeping (Chaudhry et al. 2006), as much as 20 seconds per SNOMED CT coded diagnosis (Fung et al. 2011). On the other hand, this option has the advantage that the computer can immediately check the entry for consistency with previously stored information and can ask for additional detail or dimensions conditional on the information just entered. Some of these data will be entered into fields, with menu selection. For ease of entry, such menus should not be very long, require scrolling, or impose a rigid hierarchy (Kuhn et al. 1984). Using a process called auto complete,

clinicians can code items by typing in a few letters of an item name, then choose the item they need from the list of items that match the string they entered. In some cases this process can be fast and efficient, but critics have described it as "death by a thousand clicks" (Fry and Schulte 2019).

The use of templates and menus can speed note entry, but they can also generate excessive boilerplate and discourages specificity (i.e., it is easier to pick an available menu option than to describe a finding or event in detail). Notes written via templates may not convey as clear, or as accurate, a picture of the patient's state as a provider note written in narrative. Developers can develop separate data capture forms using APIs specific to the manufacturer resulting in forms that are relatively easier to use and often can adapt to various form factors.

Among its many other capabilities, FHIR has developed specifications for web data capture forms. Questionnaire[17] is FHIR's data capture resource. Questionnaire supports skip logic, nesting and repeating groups of questions and many simple data validation checks. It also supports nested, and repeating, groups of questions and thus can accommodate complicated forms such as the surgeon general is family history (see ◘ Fig. 14.4a). FHIR has developed an enhanced version of Questionnaire called Structured Data Capture (SDC)[18] in STU4 trial use. It adds many capabilities to Questionnaire, including arithmetic and logical calculations (see ◘ Fig. 14.4b – BMI and ◘ Fig. 14.4c – Apgar score) and mechanisms for pre-populating forms with existing patient data from an EHR. It also includes regular expressions for validation of data entry along with mechanisms for storing form content data into designated FHIR resources. Finally, SDC supports adaptive (CAT) survey instruments such as PROMIS[19] patient reported outcomes.

The long-term solution for effective data capture of information generated by clinicians is still evolving. Semi-structured data entry combines the use of narrative text fields amenable to natural language processing combined with structured data entry fields where needed. With time and better input devices, acquisition of data will become faster and easier. In addition, direct entry of some data by patients could reduce the data entry burden for clinicians (Janamanchi et al. 2009).

14.3.1.3 Data Display

Once stored in the computer, EHRs can present patient data in different formats for different purposes. These systems can also present content in novel formats. Clinicians need more than just integrated access to patient data; they also need various views of these data: in chronologic order as flowsheets or graphs to highlight changes over time, and as snapshots that show a computer view of the patients' current status and their most important observations.

Timeline Graphs

A graphical presentation can help the physician to assimilate and draw conclusions from the information quickly and draw conclusions (Fafchamps et al. 1991; Tang and Patel 1994; Starren and Johnson 2000). An anesthesia system vendor provides an especially good example of the use of numbers and graphics in a timeline to convey the patient's state in form that can be digested at a glance (Vigoda and Lubarsky 2006). Sparklines – "small, high resolution graphics embedded in a context of words, numbers, images" (Tufte 2006), which today's browsers and spreadsheets can easily generate – provide a way to embed graphic timelines into any report. One study found that with sparklines, "physicians were able to assess laboratory data faster". Sparklines enable more information to be presented more compactly in a single view and thus reduce the need to scroll or flip between screens" (Bauer et al. 2010).

17 ► https://www.hl7.org/fhir/questionnaire.html (Accessed 6/4/2020).

18 ► http://hl7.org/fhir/uv/sdc/2019May/ (Accessed 6/4/2020).

19 ► http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis (Accessed 6/4/2020).

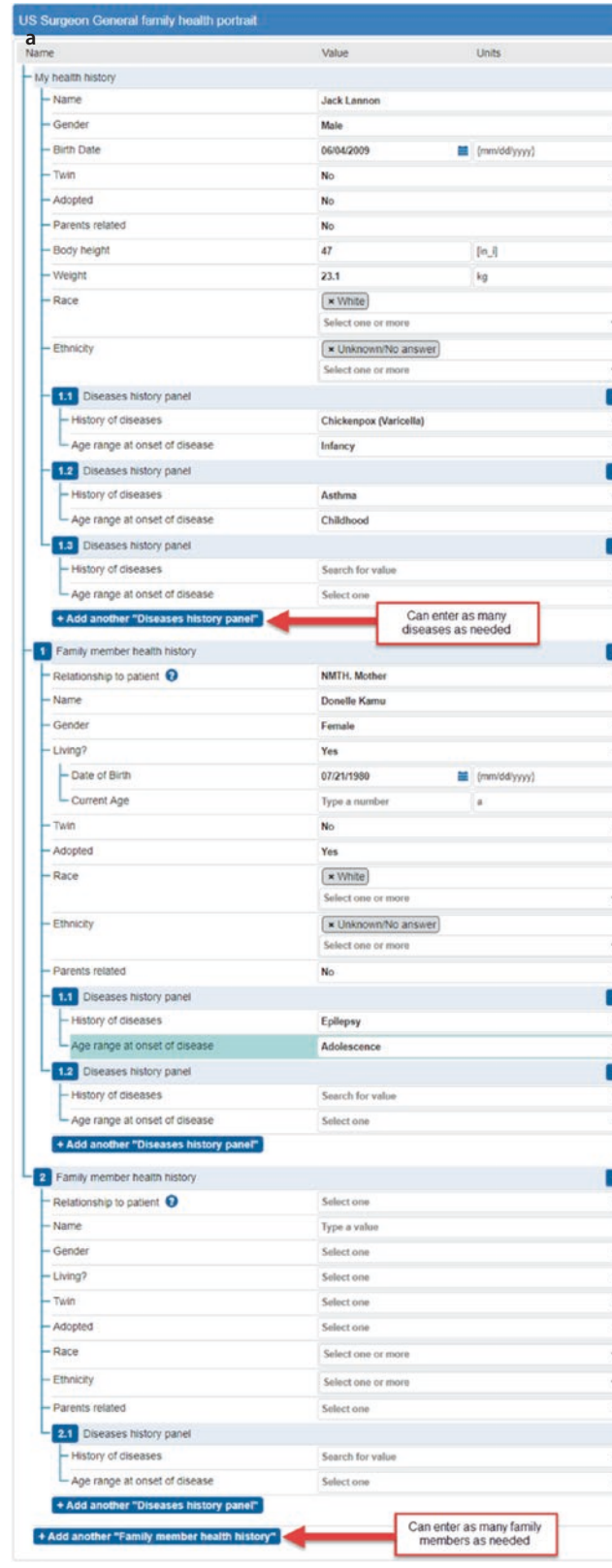


Fig. 14.4 **a** A FHIR questionnaire with the same content as the Surgeon General's Family History, rendered by the NLM FHIR Questionnaire App. The questionnaire asks questions about the proband and about each disease the proband experienced and when it occurred. These two questions repeat for each such disease the proband has experienced. Then it asks almost the same set of questions about each of the proband's relatives as it asked about the proband, and this whole set of questions repeats for as many relatives as the user wants to enter. The same pair of questions about disease and age range also repeat within each set of questions about relatives. See FHIR definition for such forms at ► http://hl7.org/fhir/uv/sdc/2019May/. **b** An SDC form rendered by the NLM FHIR Questionnaire App that captures height and weight (among other things) and automatically computes BMI as soon as these data elements are available via a FHIRpath expression. More information about FHIRpath can be found here: ► https://github.com/lhncbc/fhirpath.js. Try it live on the Demo URL (► https://lhcforms.nlm.nih.gov/sdc). **c** A rendered SDC form for the Apgar instrument to rate the health of a newborn. The answer to each question has an associated pre-defined score. This SDC form computes the overall Apgar score by adding the answers' scores as the user selects them. The overall score appears at the bottom of the form. See demo URL (► https://lhcforms.nlm.nih.gov/48334-7). You can click on the gears to change the input control

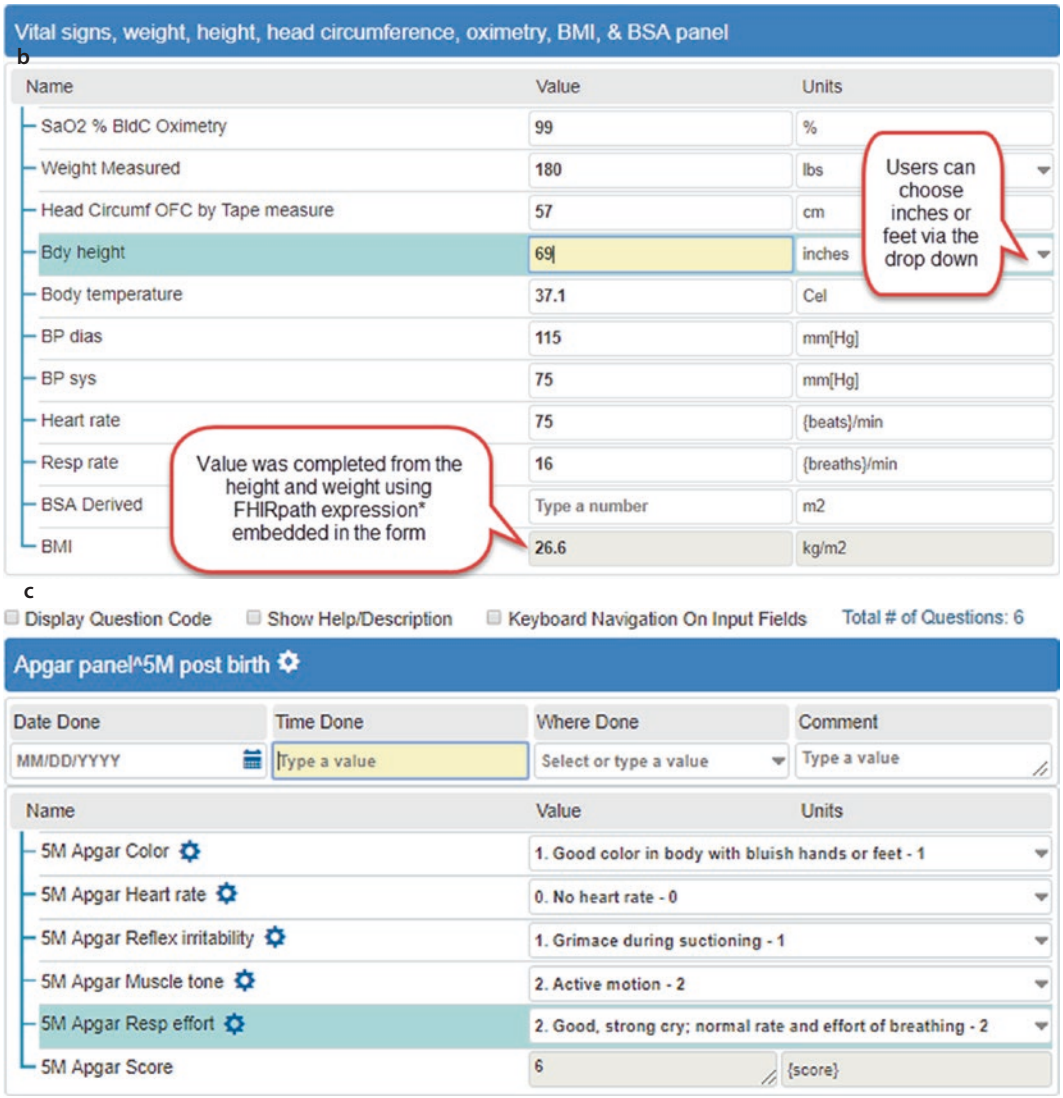


Fig. 14.4 (continued)

Timeline Flowsheets

Figure 14.5 shows an integrated view of a flowsheet of the radiology impressions with the rows representing different kinds of radiology examinations and the columns representing study dates. Clicking on the radiology image icon brings up the radiology images. Figure 14.6 shows the previously highly popular pocket rounds report that provides laboratory and nursing measurements as a very compact flowsheet that fits in a white coat pocket (Simonaitis et al. 2006).

Flowsheets and other formats can be specialized for management of a particular problem. For example, a flowsheet used to monitor patients who have hypertension (high blood pressure) and might contain values for weight, blood pressure, heart rate, and medications that control hypertension with doses as well as results of laboratory tests that monitor

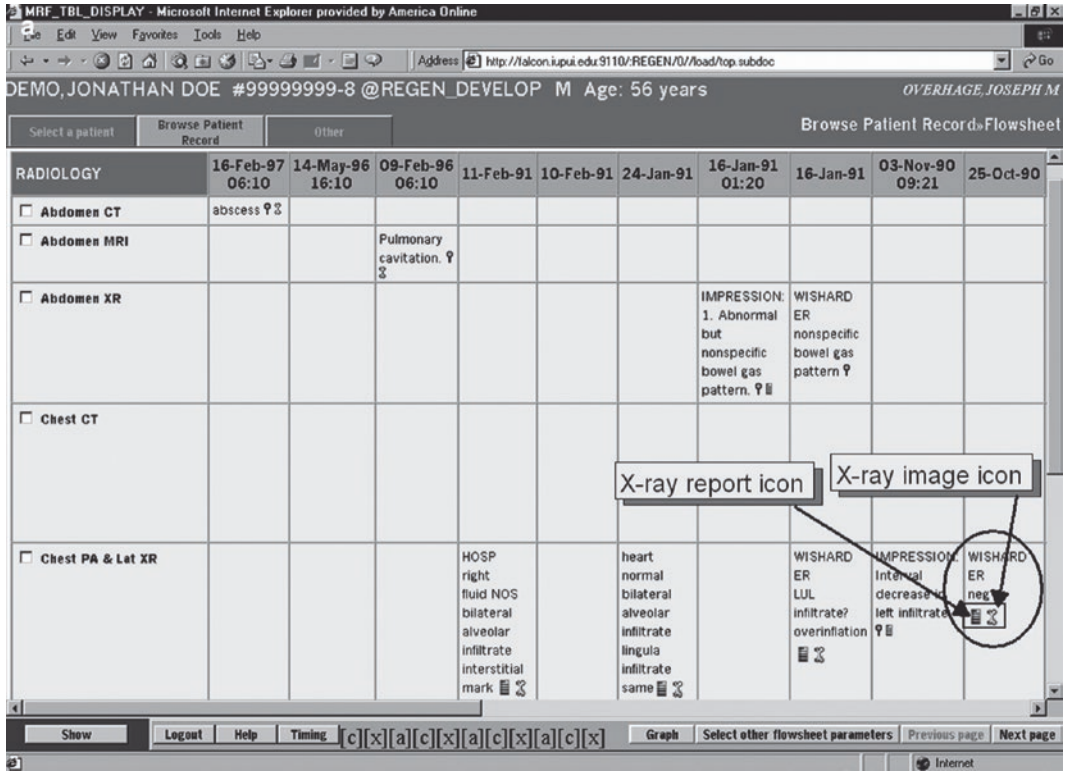


Fig. 14.5 A flow sheet of radiology reports. The rows all report one kind of study and the columns report one date. Each cell shows the impression part of the radiology report as a quick summary of the content of that report. The cells include two icons. Clicking on the report icon provides the full radiology report. Clicking on the radiology image icon provides the images. (Source: Courtesy of Regenstrief Institute, Indianapolis, IN). The CareWeb program from Regenstrief, which generated this flowsheet, presents cross-institutional patient flowsheets from the Indiana HIE to office practices today

complications of hypertension, or the medications used to treat it. Physicians at the University of Wisconsin and University of Texas-Southwestern have developed tables of mappings from problem categories (e.g., renal failure, ischemic heart disease) to observations, and to medications, with mappings developed through a multi-institutional consensus process with universal code specifications that use SNOMED CT or ICD (for defining problem classes), LOINC for observations, and RxNorm for medications (Fig. 14.7) (Buchanan 2017; Willett et al. 2018). Physicians at the University of Wisconsin and University of Texas-Southwestern have developed Problem Concept Maps that will be available for free download[20,21] under a LOINC-like agreement (Fig. 14.7). In this example of a *Problem Oriented View*, displays medications and lab results relevant to the problem of acute systolic heart failure. If a user chose a different problem in the panel on the left, it would display content relevant to that problem.

20 125 SNOMED CT groupers published as online supplement to open-access article above in reference #3. ▶ https://www.thieme-connect.de/media/10.1055-s-00035026/201803/supmat/10-1055-s-0038-1668090-s180031ra.pdf (Accessed 6/4/2020).

21 See Problem List MD at ▶ https://problemlist.org (Accessed 6/4/2020).

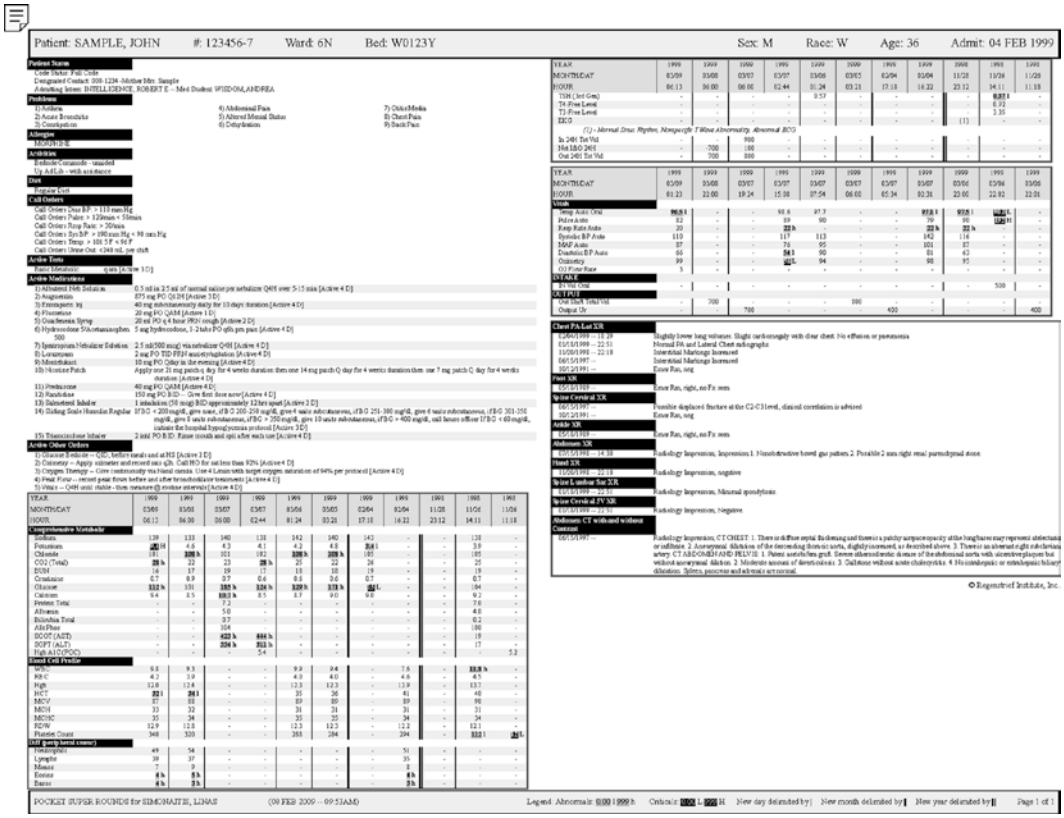


Fig. 14.6 The Pocket rounds report—so called because when folded from top to bottom, it fits in the clinician's white coat pocket as a booklet. It is a dense report (12 lines per inch, 36 characters per inch), printed in landscape mode on one 8 1/2 × 11 in. page), and includes the all active orders (including medications), recent laboratory results, vital signs and the summary impressions of radiology, endoscopy, and cardiology reports. (Source: Courtesy of L. Simonaitis, Regenstrief Institute, Indianapolis, IN)

The LHC Flowsheet on FHIR app (named for the NLM's Lister Hill Center), is an open source web app (► https://lhcflowsheet.nlm.nih.gov/) that generates clinical flowsheets from any FHIR server with LOINC coded observation identifiers (Fig. 14.8a, b). It can display very large datasets (the example presents a patient with more than 20 thousand observations), can scroll quickly in the X and Y-axis and show or hide selected groups of variables. This app can also provide a problem-focused flowsheet as well as time axis (column) compression, variable (row) axis compression and units' conversion according to user preferences. It can also convert observations with mixed units of measure into a preferred unit of measure using LHC's unit validation/converter (UCUM).[22]

Summaries and Snapshots

EHRs can highlight important components (e.g., active allergies, active problems, active treatments, and recent observations) in clinical summaries or snapshots (Tang et al. 1999). Figure 14.9 shows an example from the Epic EHR where active patient problems, active medications, allergies, health maintenance reminders, and other relevant sum-

22 ► https://ucum.nlm.nih.gov/ucum-lhc/demo.html (Accessed 6/4/2020).

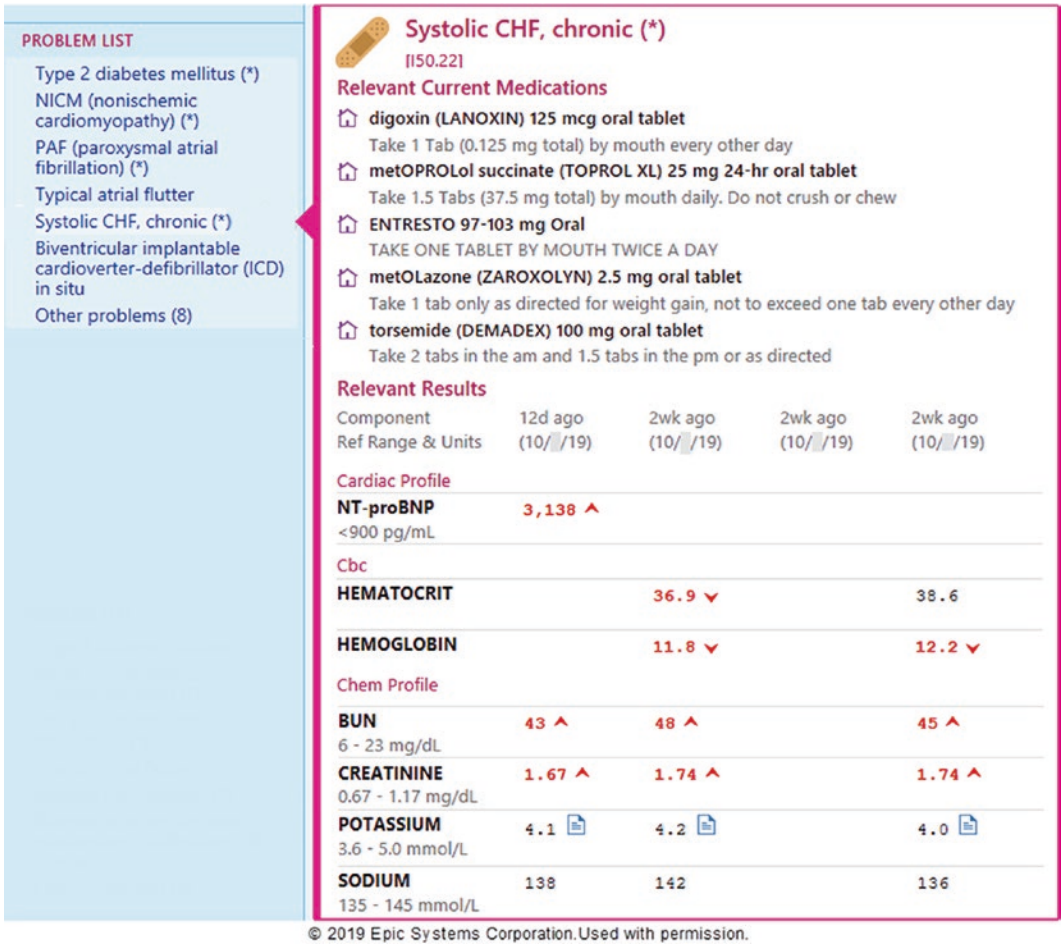


Fig. 14.7 In this example of a *Problem Oriented View*, displays medications and lab results relevant to the problem of acute systolic heart failure If a user chose a different problems in the panel on the left, it would display content relevant to that problem. (© 2019 Epic Systems Corporation. Used with permission)

mary information are summarized. These views are automatically updated and kept current as new data arrives. In the future, we can expect more sophisticated summarizing and surveillance strategies, such as automated detection of adverse events (Bates et al. 2003) or automated time-series events (e.g., cancer chemotherapy cycles). We may also see patient data views that distinguish abnormal changes that have been explained, or treated, from those that have not, and displays that dynamically organize the supporting evidence for existing problems (Tang and Patel 1994; Tang et al. 1994a; Buchanan 2017). Ultimately, computers should be able to produce concise and flowing summary reports that are like an experienced physician's hospital hand crafted discharge summary.

Researchers are developing increasingly sophisticated summaries. The HARVEST system (Hirsch et al. 2015), for example, processes all of a patient's notes, extracts unique concepts, ranks them in importance for the patient, and displays them in a word cloud. Clicking a word reveals the notes that support the concept, and clicking the individual note reveals the relevant snippet(s) of text. The user can pick a subset of the patient's timeline. The system reveals information that the user may not know to ask for and has been found to be especially useful for emergency department doctors and quality assurance nurses.

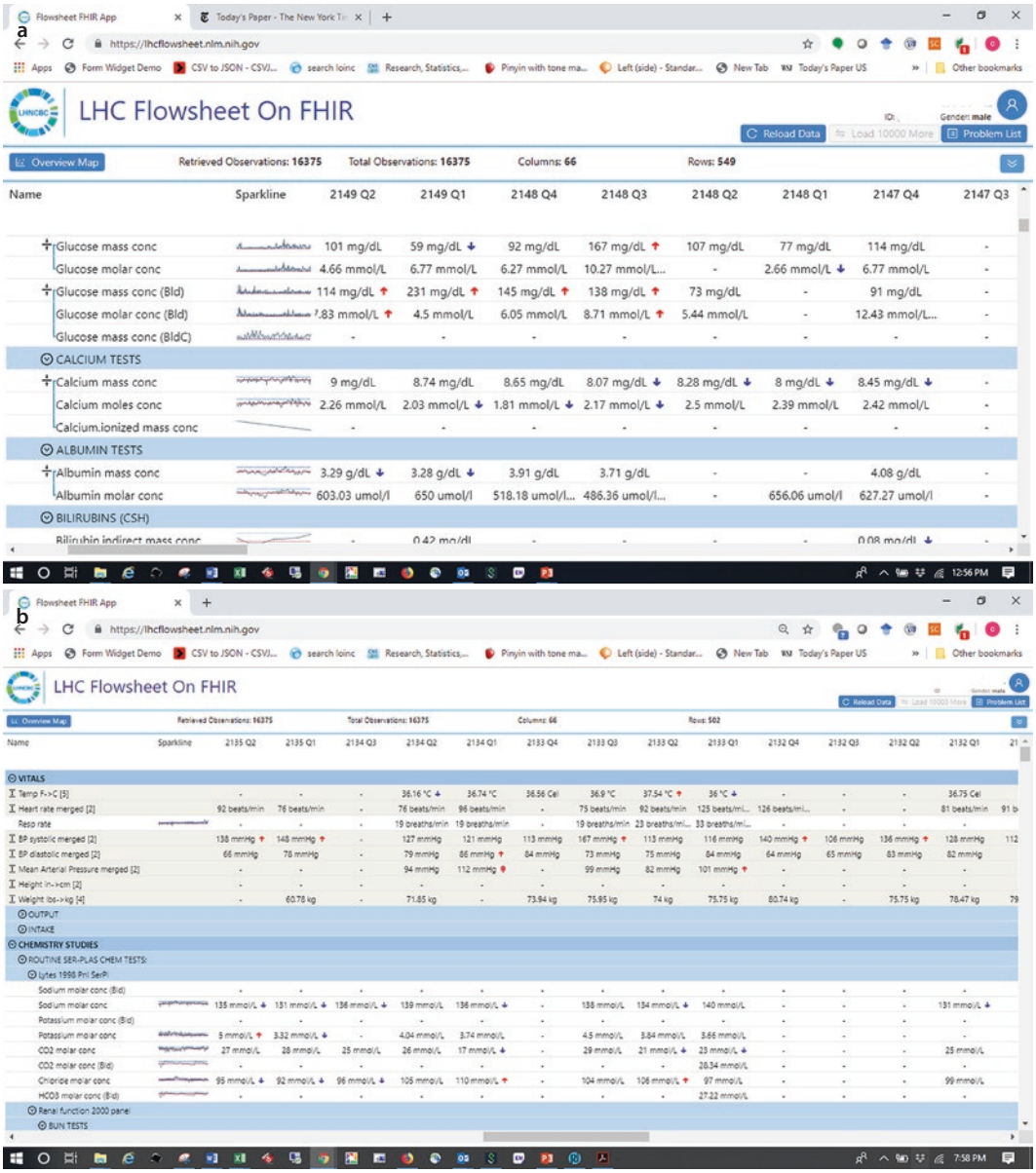


Fig. 14.8 LHC Flowsheet on FHIR app presenting the content of a large, (more than 15,000 observations) de-identified medical records. The dates and values within the dataset have been slightly shifted from their original values and salted with additional data using random methods. In this display, the user may choose to collapse columns to one value per quarter to facilitate user interpretation. Out of range results are flagged as high (red up arrow) or low (blue down arrow) using HL7 interpretation codes and normal value ranges. A spark line appears graphing all of the values in a row, appears to the right of each observation name. In **a**, values for each observation have separate rows, whereas **b** all of the values for within one equivalence class are folded into a single row after converting all values with any molar or mass unit into a common, but configurable, mass unit

Dynamic On-Demand Data Views

Anyone who has reviewed a patient's chart knows how hard it can be to find a particular piece of information. From 10% (Fries 1974) to 81% (Tang et al. 1994) of the time, when looking in the paper record, physicians did not find important patient information that was present in that record. Furthermore, the ques-

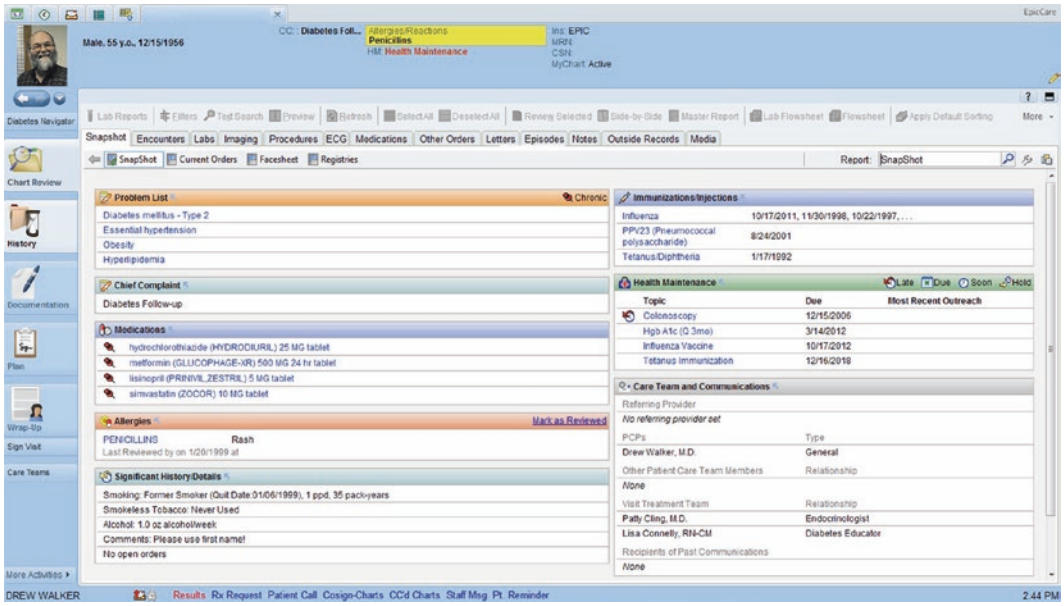


Fig. 14.9 Summary record. The patient's active medical problems, current medications, and drug allergies are among the core data that physicians must keep in mind when making any decision on patient care. This one-page screen provides an instant display of core clinical data elements as well as reminders about required preventive care. (Source: Courtesy of Epic Systems, Madison, WI)

tions clinicians routinely ask are often the ones that are difficult to answer from perusal of a typical medical record. Common questions include whether a specific test has ever been performed, what kinds of medications have been tried, and how the patient has responded to particular treatments (e.g., a class of medications) in the past. Physicians constantly ask these questions as they flip back and forth in the chart searching for the facts needed to support or refute a given hypotheses as their thinking about the patient evolves. On demand search tools help clinicians locate and then organize relevant patient data, into flowsheets, problem oriented displays (see section "► Timeline Flowsheets") or graphs (Fafchamps et al. 1991; Tang et al. 1994a; Starren and Johnson 2000) to facilitate provider assimilation of the relevant facts.

14

14.3.2 Computerized Provider Order Entry

One of the most important components of an EHR is computerized provider order entry (CPOE), where clinicians make and act upon therapeutic and diagnostic decisions by entering orders. CPOE systems can reduce errors and costs compared to paper systems, in which orders are transcribed manually from one paper form (e.g., the orders section of the paper chart) to another (e.g., the nurse's work list, a laboratory request form), or faxed to a receiving area for fulfillment (e.g., transportation services, pharmacy). CPOE orders pass electronically from the decision-maker to the order filler with minimal to no manual labor. Order entry systems also provide opportunities to deliver CDS when providers are entering orders and making clinical decisions. Most existing CPOEs provide alerts about drug interactions, allergies, and dosing adjustments for renal insufficiency when new drug orders are entered. However, EHR implementers should be selective about which alerts to evoke and be parsimonious with the use of interruptive alerts to avoid wasting provider time on trivial or low-likelihood outcomes (Miller et al. 2005a; Phansalkar et al. 2012a, b). We discuss this capability in more fully in the next section.

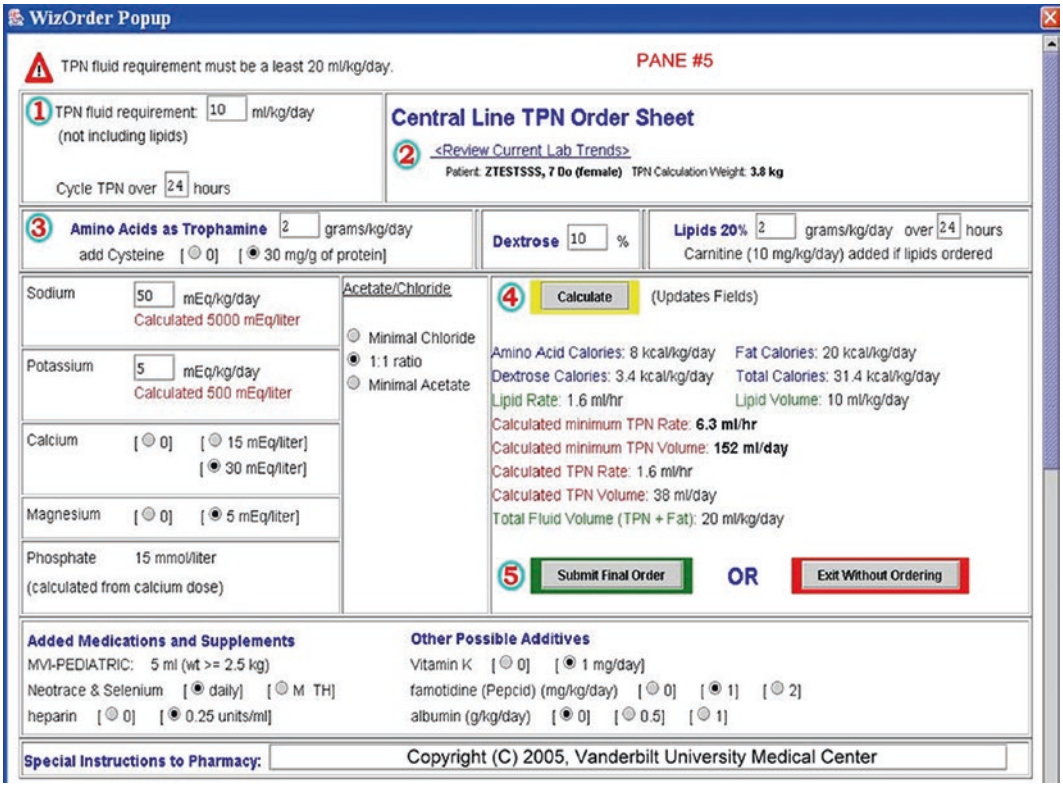


Fig. 14.10 Neonatal Intensive Care Unit (NICU) Total Parenteral Nutrition (TPN) Advisor provides information about complex interactive advice and performs various calculations in response to the provider's prescribed goal for amount of fluid, calories, nutrition, and special additives. (Source: Miller et al. (2005b). Elsevier Reprint License No. 2800411402464)

Order entry systems can also remind providers about important orders, which might otherwise be forgotten. Very intelligently designed order entry systems can shrink the work of entering complicated orders such as declining dosing of prednisone, intravenous fluid orders, and total parenteral nutrition (TPN) orders the last of which require entry of many additives and calculations to avoid dangerous mixtures and to reach specified targets for calories each additives. Figure 14.10 shows an example of a TPN order entry screen from Vanderbilt (Miller et al. 2005b). However, with some order entry systems, entry of intravenous and declining dose orders can be more difficult than with the manual alternative.

When a CPOE system is operational, simply changing the default drug or dosing based on the latest scientific evidence can shift ordering behavior toward the optimum standard of care, with benefits to quality and costs. Because of these advantages, health care organizations have adopted CPOE widely and federal regulations for certified health IT require core CPOE functionality for medications, laboratories, diagnostic imaging, and EHRs, which include checking for drug-drug and drug-allergy interactions[23]–though some question the wisdom of those interaction requirements (Khajouei and Jaspers 2010) and, in one study, the use of these checks with alerts had no effect in the rate of adverse drug reactions (Nebeker et al. 2005).

23 Certification of Health IT, Testing Process & Test Methods, 2015 Edition Test Method. 2015 Edition Certification Regulations – 170.315(a) (1, 2, 3, 4): Computerized Provider Order Entry and Drug-drug and Drug-allergy checks for CPOE ► https://www.healthit.gov/topic/certification-ehrs/2015-edition-test-method (Accessed 6/4/2020).

14.3.3 Clinical Decision Support

Clinical trials have shown that certain reminders from CDS can improve care processes (McDonald 1976; Haynes 2011; Damiani et al. 2010; Schedlbauer et al. 2009; Ranta et al. 2015; Clyne et al. 2012; Tajmir et al. 2017) but the efficacy of CDS broadly is mixed (Delvaux et al. 2017; Parshuram et al. 2018; Muth et al. 2018; Fried et al. 2017). EHRs can deliver CDS in batch mode at intervals across a whole practice population to identify patients who are not reaching to treatment targets or are past due for immunizations, cancer screening, or have missed their recent appointments, to cite a few examples. In batch mode, clinical practices can utilize lists of patients generated by CDS to contact the patient and encourage him or her to reach a goal or to schedule an appointment for the delivery of suggested care and can reach patients who have not kept scheduled appointments.

Decision support—especially for preventive care—is most efficiently delivered in the course of routine care while the patient and provider are together. Suggestions can be delivered during the physician order entry process, which in some cases can be the best point in the workflow at which to discourage or countermand an order that might be dangerous or wasteful. It is also a convenient point to offer reminders about needed tests or treatments, which can easily be initiated during that order session.

One of the best ways for CDS systems to remind providers about tests or treatments is by presenting pre-constructed order(s) to the provider who can confirm or reject the order(s) with a single keystroke or mouse click. It is best to annotate such suggestions with their rationale (e.g., "the patient is due for his pneumonia vaccine because he has emphysema and is over 65") so that the provider understands the rationale for the suggestion (Mamlin et al. 2007).

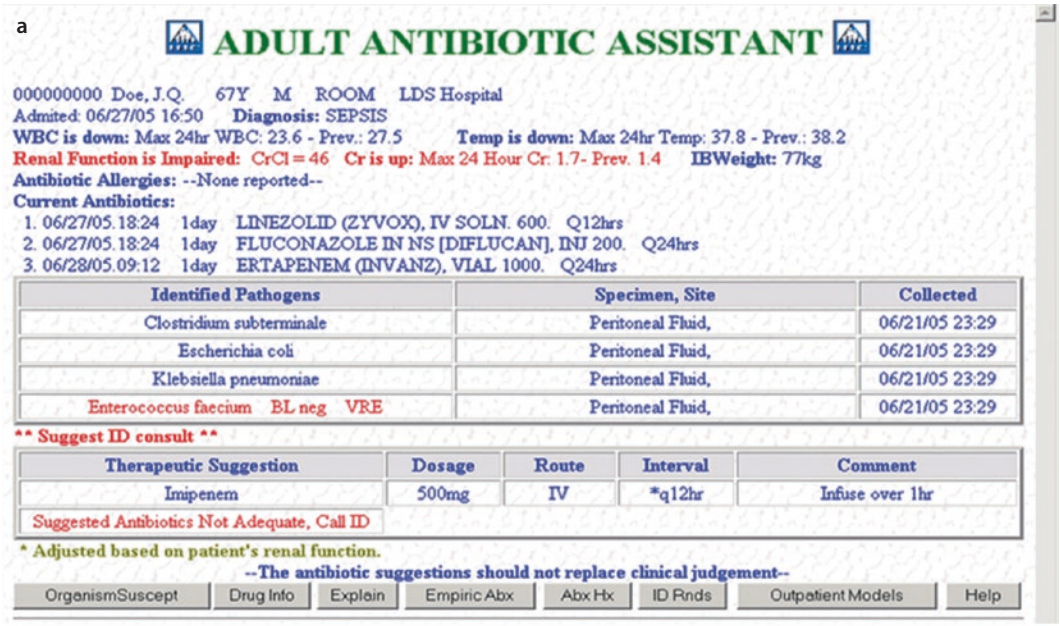
Figure 14.11 shows some suggestions from a sophisticated inpatient CDS system developed by Intermountain Health Care. This system used a wide range of clinical information to recommend antibiotic choice, dose, and duration of treatment from the system improved clinical outcomes and reduced costs of infections among patients managed with the assistance of this system (Evans et al. 1998; Pestotnik 2005). Vanderbilt's inpatient "WizOrder" CPOE system also addressed antibiotic orders, as shown in Fig. 14.12; it suggests the use of Cefepine rather than Ceftazidine, and provides choices of dosing by indication.

Clinical alerts attached to a laboratory test result can include suggestions for appropriate follow up or treatments for some abnormalities (Ozdas et al. 2008; Rosenbloom et al. 2005). Also, CPOE functionality can warn the physician about allergies (Fig. 14.13a) and drug interactions (Fig. 14.13b) before the provider completes a medication order, as exemplified by screenshots from Partner's outpatient medical record orders.

Reminders and alerts are employed widely in outpatient care. Indeed, the outpatient setting is where the first study of clinical reminders and the first randomized trial of medical informatics systems, was performed (McDonald 1976) and remains the setting for the majority of such studies (Garg et al. 2005). Reminders to physicians in outpatient settings quadrupled the use of recommended vaccines in eligible patients compared with those who did not receive reminders (McDonald et al. 2014; McPhee et al. 1991; Hunt et al. 1998; Teich et al. 2000). Reminder systems can also suggest needed tests and treatments for eligible patients (Overhage 1997). Figure 14.14 shows an Epic system screen with reminders to consider ordering a cardiac echocardiogram and starting an ACE inhibitor—in an outpatient patient with a diagnosis of heart failure but no record of a cardiac echocardiogram or treatment with one of the most beneficial drugs for heart failure.

Though the outpatient setting is the primary setting for preventive care reminders, preventive reminders have also been applied effectively in the hospital setting (Dexter et al. 2001). Furthermore, reminders directed to inpatient nurses improve preventive care even more than reminders directed to physicians (Dexter et al. 2004).

a



b

- Patient should receive IV antibiotics.
- Renal function dictates that dosage should be adjusted.
- Cultures show fungi or yeast that were not considered pathogens.
- The suggested antibiotic(s) will treat the identified anaerobes.
- Patient's vitals (Temp, WBC, Bands) do not support chest Xray: Wed Jun 22 06:14:00 MDT 2005)
- Suggest vancomycin & an aminoglycoside to empirically treat the Dx of sepsis.
- Suggest ticar/clav or imipenem due to the site of Clostridium infection.
- Prophylactic antibiotics are not suggested for this patient at this time.
- Suggest ID consult based on the complexity of this patient's condition.

--The antibiotic suggestions should not replace clinical judgement.--
The electronic medical record may not contain all patient information.

Fig. 14.11 Example of the main screen **a** from the Intermountain Health Care Antibiotic Assistant program needed. The program displays evidence of an infection-relevant patient data (e.g., kidney function, temperature), recommendations for antibiotics based on the culture results, and **b** disclaimers. (Source: Courtesy of R. Scott Evans, Robert A. Larsen, Stanley L. Pestotnik, David C. Classen, Reed M. Gardner, and John P. Burke, LDS Hospital, Salt Lake City, UT (Larsen et al. 1989) © Cambridge University Press)

14.3.4 Access to Knowledge Resources

Many clinical questions, whether addressed to a colleague or answered by searching through textbooks and published papers, are asked in the context of a specific patient (Covell et al. 1985). Thus, one appropriate time to offer knowledge resources to clinicians is while they are writing notes or entering orders for a specific patient. Clinicians typically have access to a selection of knowledge sources, which can be accessed from a web browser at any point in time today. Some are from public sources, such as the National Library of Medicine's (NLM) PubMed and MedlinePlus, Centers for Disease Control and Prevention's (CDC) vaccines and international travel information, and Agency for Healthcare Research and Quality's (AHRQ) National Guideline Clearinghouse. Others come from commercial vendors like

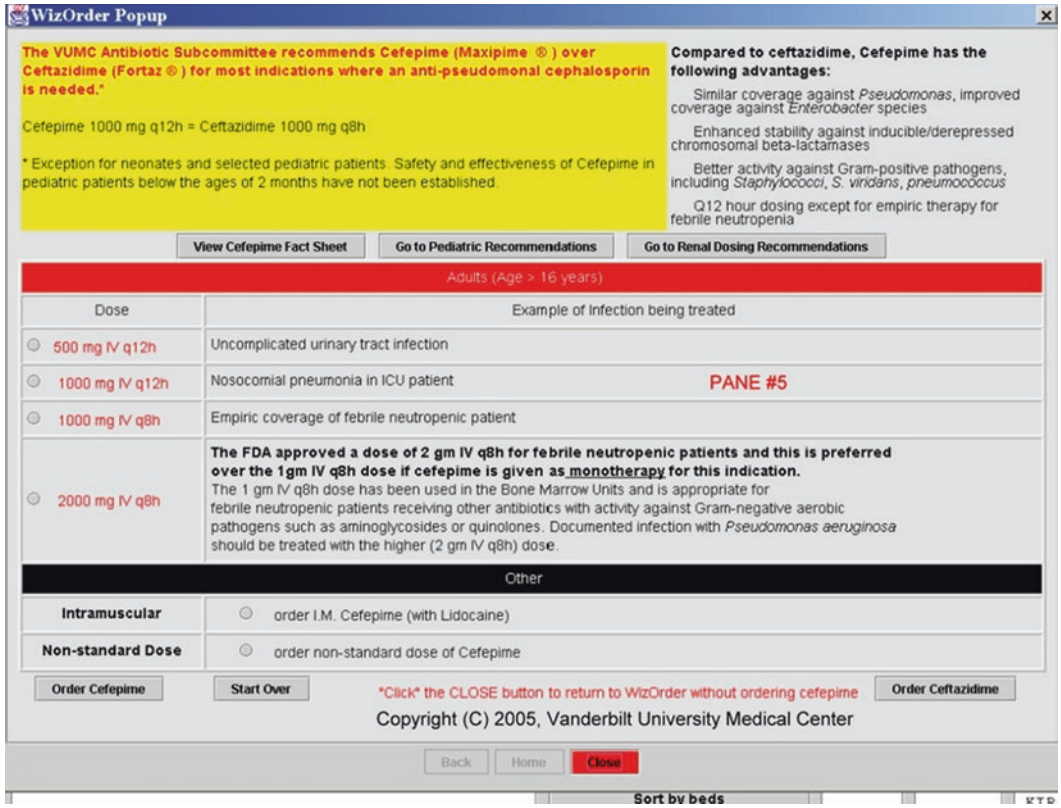


Fig. 14.12 User ordered an antibiotic for which the Vanderbilt's former inpatient "WizOrder" CPOE system, based on their Pharmaceuticals and Therapeutics (P and T) Committee input, recommended a substitution. This educational advisor guided clinician through ordering an alternative antibiotic. Links to "package inserts" (via buttons) detailed how to prescribe recommended drug under various circumstances. (Source: Miller et al. (2005b). Elsevier Reprint License No. 2800411402464)

14

UpToDate, Micromedex, and a variety of electronic textbooks. Some EHRs are proactive and routinely present short informational nuggets adjacent to the order item that the clinician has chosen. Through an **Infobutton** designed to pull context-specific information, EHRs can also pull literature, textbook or other sources of information relevant to a particular clinical situation, and present that information to the clinician on the fly (Del Fiol et al. 2012). The Infobutton standard is now a core CDS functionality within certified Health IT requirements[24] (see Fig. 14.15). To support this function, HL7 Version 3 standard has produced the Context Aware Knowledge Retrieval Application

24 Certification of Health IT, Testing Process & Test Methods, 2015 Edition Test Method, Clinical Decision Support. 2015 Edition Certification Regulations - 170.315(a)(9): Clinical Decision Support ► https://www.healthit.gov/test-method/clinical-decision-support-cds#ccg (Accessed 6/4/2020).

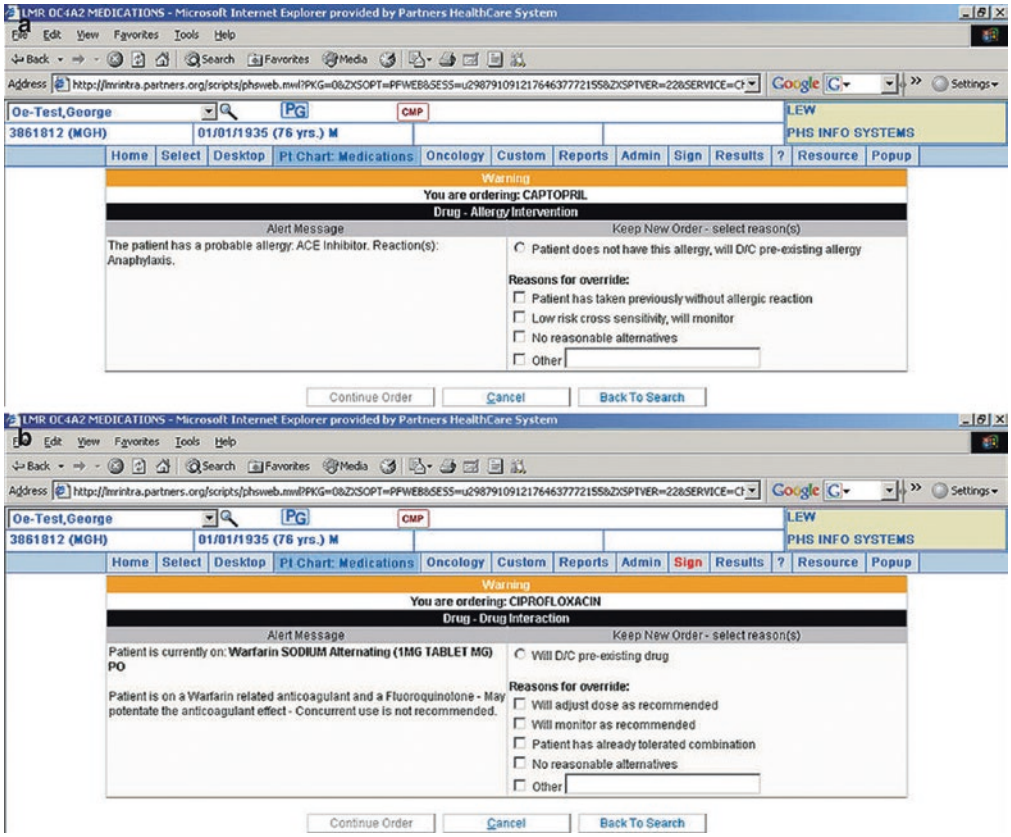


Fig. 14.13 Drug-alert display screens from Partners former outpatient medical record application (Longitudinal Medical Record, LMR). The screens show **a** a drug-allergy alert for captopril, and **b** a drug-drug interaction between ciprofloxacin and warfarin. (Source: Courtesy of Partners Health Care System, Chestnut Hill, MA)

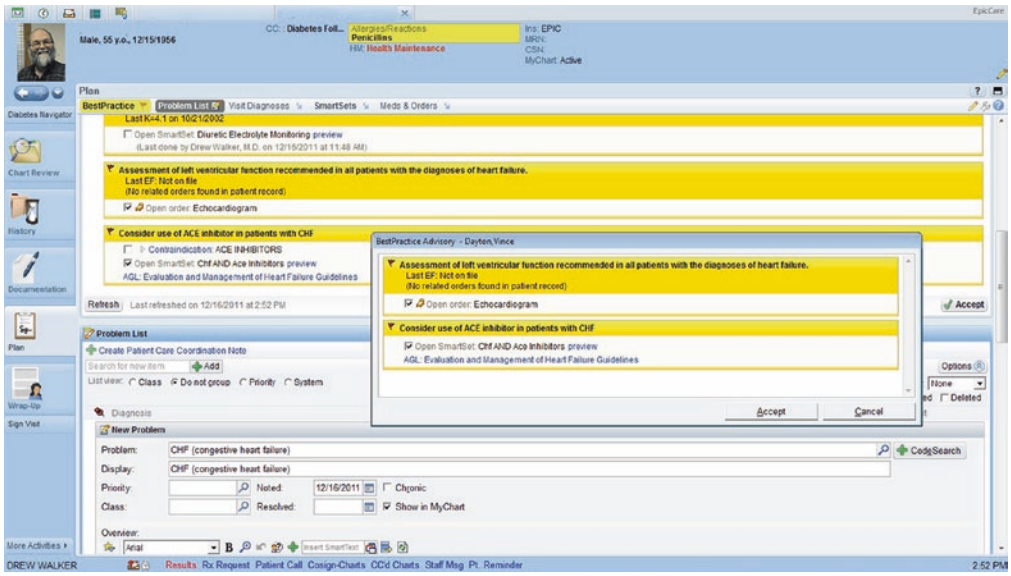


Fig. 14.14 Example of CDS alerts to order an echocardiogram and to start an ACE inhibitor in a patient with diagnosed congestive heart failure. (Source: Courtesy of Epic Systems, Madison, WI)

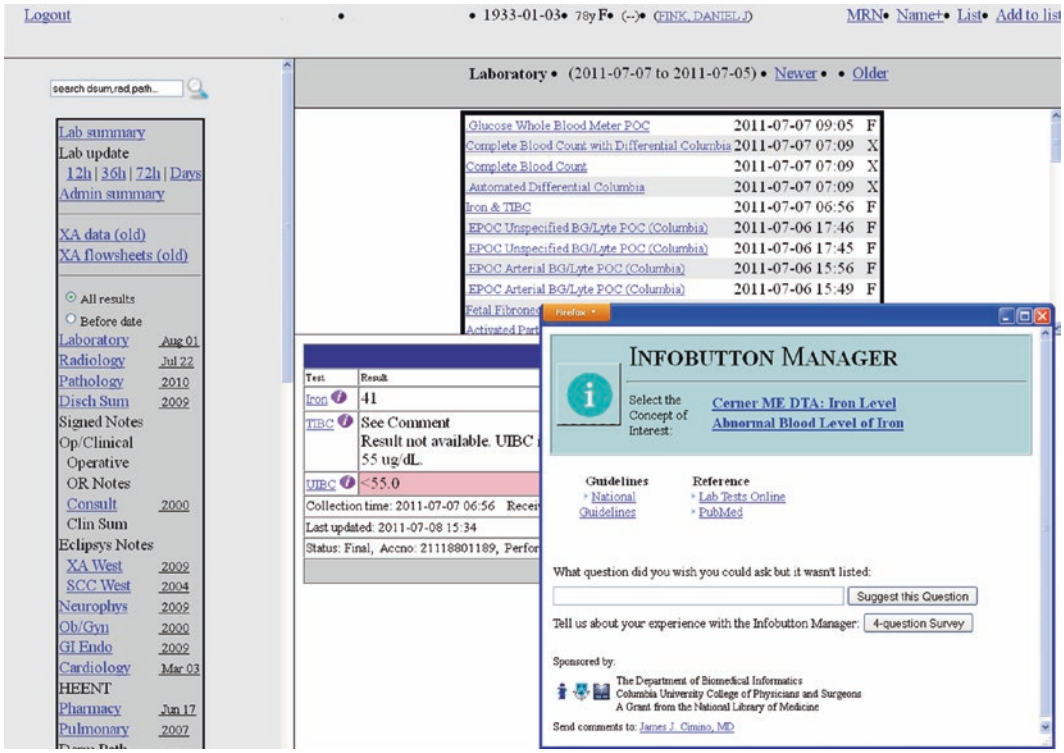


Fig. 14.15 This figure shows the use of Columbia University Medical Center's Infobuttons during results review. Clicking on the Infobutton adjacent to the Iron result generates a window (image) with a menu of questions. When the user clicks on one of the questions, the Infobutton delivers the answers. (Source: Courtesy of Columbia University Medical Center, New York)

("Infobutton" standard).[25] HL7 FHIR also has powerful CDS capabilities; namely, support for CDS Hooks and two scripting languages: FHIRPath[26] and CQL[27] for algebraic and logical calculations (See ► Chap 8).

One idea increasingly of interest is to provide to clinicians at the point of care optimized treatment plans or customized information, derived from patients "just like" the one in front of the clinician. Conceptually, this "green button" (Longhurst et al. 2014), like other CDS, would be accessible to clinicians at the point of care and provide aggregate patient data (e.g., outcomes to a particular medication for disease treatment according to similar patient characteristics) to help support treatment decisions in the absence of high quality evidence.

14

25 HL7 Version 3 Standard: Context Aware Knowledge Retrieval Application ("Infobutton"), Knowledge Request, Release 2. ► http://www.hl7.org/implement/standards/product_brief.cfm?product_id=208 (Accessed 6/4/2020).

26 FHIRPath STU1 Release. [Internet]. 2019 [cited 01/29/2019]. Available from: ► http://hl7.org/fhirpath/ (Accessed 6/4/2020).

27 Health Level 7. Clinical quality language (CQL) standard. [Internet]. 2018 [cited 01/29/2019]. Available from: ► http://www.hl7.org/implement/standards/product_brief.cfm?product_id=400 (Accessed 6/4/2020).

14.3.5 Care Team and Patient Communication

Communication tools, that support timely and efficient communication between patients and the health care team and amongst team members, can enhance coordination of care and disease management. Patients are provided secure online access to their EHR and integrated communication tools to ask medical questions or conveniently perform other

clinical (e.g., renew a prescription) or administrative tasks (e.g., schedule an appointment) (Tang 2003). Increasingly, the delivery of optimal patient care also requires multiple health care professionals that may cross several organizations; thus, it is important that communication among team members and organizations is delivered effectively, efficiently, and on time. Such communications usually focus on a single patient and may require a care provider to assess or interchange information from several systems and providers in order to coordinate relevant care.

Direct connectivity to patients is and will be increasingly important to patient-provider communication. It will permit direct-to-patient reminders (Sherifali et al. 2011) and deliver home health monitoring data (such as home blood pressure measurements and glucose testing results) to the EHR and other information systems (Earle 2011; Green et al. 2008). The patient's personal health record (PHR) will also become an important destination for clinical messages and test results (see ▶ Chap. 13). Relevant information can be "pushed" to the patient or their PHR via e-mail, pager services, or other secure texting or closed loop communication (Major et al. 2002; Poon et al. 2002; Gulacti and Lok 2017; Rief et al. 2017; Przvbvlo et al. 2014) or "pulled" by users on demand during their routine interactions with the computer.

EHRs can also provide electronic functionality to assist in the transfer (or hand-off) of care responsibility from one clinician to another. When the transfer is between settings (e.g., from hospital to nursing home), the sending clinician usually provides a brief verbal or written turnover note to the receiving clinician(s) summarizing the patient's problems, treatments, and other relevant clinical issues. ◘ Figure 14.16 shows an example of a "turn-over report" that includes instructions from the "sending" physician, as well as relevant recent laboratory test results and other data pulled from the patient's EHR and a "to-do" list, that ensures that critical tasks are complete (Stein et al. 2010). Such reports facilitate communication among team members, and can improve both coordination and patient safety. However, keeping the contents of such reports up to date and accurate can be a challenge (Arsoniadis et al. 2017).

Although most patient encounters are defined by scheduled face-to-face visits (e.g., outpatient visit, home health visit), provider decision-making also occurs during non-face-to-face and nonscheduled events (e.g., patient telephone calls, prescription renewal requests, and the arrival of new test results). Recently, CMS has indicated they will pay for another kind of non-face-to-face event, namely virtual visits.[28] EHRs and other Health IT will support and facilitate these non-face-to-face events, and video-visit capabilities will become part of many EHR vendor system offerings.

EHRs are traditionally bounded by the institution in which they reside. The National Health Information Infrastructure (NHII) (NCVHS 2001) has proposed a future in which providers caring for a patient could reach beyond his or her local institution to automatically obtain patient information from all relevant sources (see ▶ Chap. 13). Today, examples of such regional "EHRs," often referred to as **Health Information Exchanges** (**HIE**), serve routine and emergency care, public health and other functions. The first HIE was the IHIE (Indiana Health Information Exchange) (McDonald et al. 2005) which started in 1994 with 3 Indianapolis hospitals and now includes hospitals from most of Indiana. Other early HIEs include Ontario, Canada (electronic Child Health Network),[29] Kentucky (Kentucky Health Information Exchange),[30] and Memphis (Frisse et al. 2008) Today, scores of HIEs are in operation ▶ https://strategichie.com/membership/member-list/. A study from the Memphis HIE showed that the extra patient information provided by this HIE saves resource use and costs (Frisse et al. 2011).

28 ▶ https://www.cms.gov/outreach-and-education/medicare-learning-network-mln/mlnproducts/downloads/telehealthsrvcsfctsht.pdf (Accessed 6/4/2020).

29 eCHN electronic Child Health Network. ▶ http://www.echn.ca/ (Accessed 6/4/2020).

30 Kentucky Health Information Exchange Frequently Asked Questions. ▶ http://khie.ky.gov/Pages/faq.aspx?fc=010 (Accessed 6/4/2020).

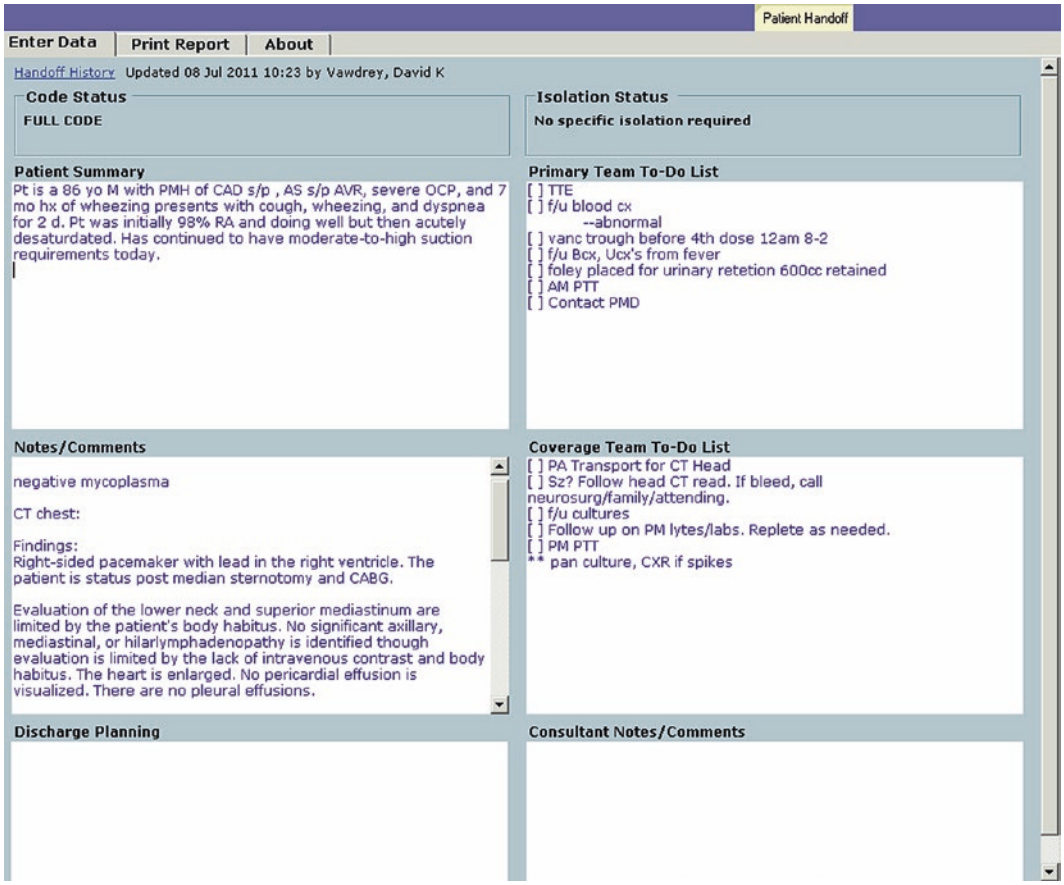


Fig. 14.16 Patient handoff report—a user-customizable hard copy report with automatic inclusion of patient allergies, active medications, 24-h vital signs, recent common laboratory test results, isolation requirements, code status, and other EHR data. This system was developed by a customer within a vendor EHR product (Sunrise Clinical Manager, Allscripts, Chicago, IL) and was disseminated among other customers around the nation. (Source: Courtesy of Columbia University Medical Center, New York)

According to a December 2018 ONC report, half of all US hospitals now share some data through one or more HIEs.[31]

In 2010, the **Office of the National Coordinator** (**ONC**) proposed the Nationwide Health Information Network (NwHIN) to connect regional HIEs and promote health data exchange (see ► Chaps. 13 and 31).[32] It included NHIN Connect and NHIN Direct,[33] the former with a special focus on large enterprises and government organizations, the latter with a focus on simpler or local networks. The Direct protocol[34] is an email (SMTP)-based protocol designed to deliver encrypted messages and attachments securely among pre-arranged groups of individuals or organizations. ONC has nurtured the development of the Direct Protocol and required

31 ► https://www.ruralcenter.org/resource-library/methods-used-to-enable-interoperability-among-us-non-federal-acute-care-hospitals (Accessed 6/4/2020).

32 eHealth Exchange. ► https://ehealthexchange.org/ (Accessed 6/4/2020).

33 ► https://www.healthdatamanagement.com/news/connect-nhin-direct-what-are-they (Accessed 6/4/2020).

34 ► http://wiki.directproject.org/Main_Page (Accessed 6/4/2020).

EHR vendors to support it, and they have, but minimally in some cases.

A least seven organizations, provide general tools and trust policies to create new secure health care networks or connect existing ones. Some of them are direct descendants of NHIN Connect. These seven are described and compared in an excellent 2017 ONC report.[35] Among the seven, Direct Trust and National Association for Trusted Exchange (NATE) use the Direct Protocol exclusively, and Surescripts and eHealth Exchange use it optionally. The other three use a mix of non-Direct Protocols. Some of these organizations are discussing partnerships or mergers. Unfortunately, support for delivering structured data is limited in many of these systems, and EHR vendors have not chosen a common approach.

With the goal of one national network of networks, ONC has proposed an overarching set of policies and protocols, called "The trusted exchange framework and common agreement" (TEFCA).[36] It has been met with both praise and criticism. FHIR offers additional mechanisms for linking independent networks and is highlighted in the TEFCA proposal. To see how this all evolves, stay tuned.

14.3.6 Billing and Coding

While originally billing and coding systems were most often separate from the main EHR which served as a clinical system of record, over time this has changed. Today, the majority of vendor EHRs have billing and coding functionality, as well as other aspects of revenue cycle (e.g., prior authorization, accounts receivable). Tying the various aspects of revenue cycle functionality into EHRs has proven to have several efficiencies for health care organizations. For instance, the workflow for coders is more streamlined. The same system can be used to review records, code, and provide feedback to clinicians on documentation, as well as allow clinicians to drop charges directly into the EHR. In some cases, EHR coding and billing can also be augmented with 3rd party systems, such as those leveraging computer-assisted coding with clinical NLP technologies (see ► Chap. 9).

14.4 EHRs for Secondary and Population-Based Uses

This section considers secondary uses of EHRs, which are increasing greatly. Some of this functionality can be done by or in concert with other systems and platforms. We expect that EHRs will continue to evolve and transform as information systems over time with greater analysis and secondary use functions. Medical personnel, quality and patient safety professionals, and administrators use these capabilities to find particular patterns and events that predict patient outcomes. Public health professionals can use reporting functions of computer-stored records for surveillance, including looking for emergence of new diseases or other health threats that warrant medical attention.

14.4.1 Population-Based Clinical Care

Although the functions of CDS for a single patient on the one hand and across a larger patient population on the other are different, their internal logic is similar. In both, the central procedure is to determine if a single patient at hand or which across the full set or subset of patients satisfy pre-specified criteria and to act appropriately when the patient meets those criteria. Surveillance queries generally address a large subset, or all, of a patient population; the output is often a tabular report of selected raw data on all the patient records retrieved or a statistical summary of the values contained in the records. Decision support systems usually address patients who are under active care and

35 ► https://www.healthit.gov/sites/default/files/analysis_of_existing_trust_arrangements_printable.pdf (Accessed 6/4/2020).

36 ► https://www.healthit.gov/sites/default/files/page/2019-04/FINALTEFCAQTF41719508version.pdf (Accessed 6/4/2020).

generate an **alert** or **reminder message** (McDonald 1976) to that patient. Organizations can use these systems at a population level for care coordination, patient empanelment for primary care providers, and other tasks.

For example, a cross-population query can be used to identify patients who are due for periodic screening examinations such as immunizations, mammograms, and cervical Pap tests and then can generate letters to patients or call lists for office staff to encourage the preventive care. This can also be especially useful for conducting ad hoc searches such as those required to identify and notify patients who have been receiving a recalled drug. Such systems can also facilitate quality management and patient safety activities, identify candidate patients for concurrent review and gather many of the data required to complete such audits.

14.4.2 Clinical Research

Researchers can use EHRs particularly associated cross-patient queries and alerting capabilities to identify patients who meet or have a high chance of meeting eligibility requirements for a prospective clinical trial. For example, an investigator could identify all patients seen in a medical clinic who have a particular diagnosis and satisfy the eligibility requirements specified in a given study protocol (Kho et al. 2007). These approaches can sometimes be applied in real time. At one institution, the physician's workstation was programmed to ask permission to invite the patient into a study, when that physician entered a problem that suggested the patient might be a candidate for a local clinical trial. If the physician gave permission, an automated electronic page could then be triggered and sent to the nurse recruiter who would then invite the patient to participate in the study. One early such study was for patients with back pain (Damush et al. 2002).

Randomized **prospective studies** are the gold standard for clinical investigations, but **retrospective studies** of existing data have contributed much to medical progress (see ▶ Chap. 29). Retrospective studies can also obtain answers at a small fraction of the time and cost of comparable prospective studies. EHRs can often provide much of the data required for a retrospective study. They can, for example, identify study cases and comparable control cases, and provide data needed for statistical analysis of the comparison cases (Brownstein et al. 2007). Combined with access to discarded specimens, they also offer powerful approaches to retrospective genome association studies that researcher can do much faster and at fraction of the comparable prospective studies (Kohane 2011; Roden et al. 2008).

Computer-stored records do not eliminate all the work required to complete an epidemiologic study; chart reviews and patient interviews may still be necessary. Computer-stored records are likely to be most complete and accurate with respect to visit diagnoses that are carefully coded for administrative purpose, as well as to prescribed drugs, and laboratory tests, because the latter two usually come directly from automated laboratory and pharmacy systems, respectively. Consequently, computer-stored records are likely to contribute to research on a physician's practice patterns, on the efficacy of tests and treatments, and on the toxicity of drugs. The research opportunities will only improve with FHIR API and coding standards required by proposed CMS and ONC rules. NIH and AHRQ have both encouraged research interest in FHIR[37,38] Also, improvements in NLP techniques may make the content of narrative text more accessible to automatic searches (see ▶ Chap. 9).

37 ▶ https://grants.nih.gov/grants/guide/notice-files/NOT-HS-19-020.html (Accessed 6/4/2020).

38 ▶ https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-122.html (Accessed 6/4/2020).

14.4.3 Quality Reporting

EHRs are also increasingly important in the production or autonomation of quality reports that are used for both internal quality improvement activities and for external regulatory or public reporting. Although it is difficult for paper-based records to incorporate patient-generated input and it requires careful tagging of data sources, EHRs can increasingly include data contributed by patients (e.g., patient-reported outcomes such as functional status, pain scores, and symptom; review of systems). Patient-reported data is also being incorporated in future quality measures particularly for disease specific or episodic specific conditions, e.g., use of the Seattle Angina Questionnaire Short Form (SAQ-7) and Rose Dyspnea Scale (RDS) following Non-Emergent Percutaneous Coronary Intervention (PCI).[39]

With changing reimbursement payment models focusing more on outcomes measures instead of volume of transactions, generating efficient and timely reports of clinical quality measures will play an increasingly important role in management and payment. FHIR will likely play a role here, as well. FHIR supports the Clinical Quality Language (CQL)[40] and FHIRPath,[41] a subset CQL. CQL was developed for CMS quality reporting but has wider applicability to other initiatives.

14.4.4 Administration

In the past, administrators had to rely on data from billing systems to understand practice patterns and resource utilization. However, claims data have their limits, including their delayed and retrospective nature. From direct comparisons between medical record content and diagnoses coded from that record, we know that the accuracy of coding varies by kind of diagnosis, setting and hospital size, and variability in granularity. Claims data can also provide structured and coded records of ambulatory prescriptions, but generally provide no test results or clinical measurement values. So, considering only claims-based diagnostic codes can lead to inappropriate policymaking and conclusions (Tang et al. 2007).

EHRs complement claim- and administrative-based data and can provide information about the relationships among diagnoses, severity of illness indication, and resource consumption. Thus, these systems are important tools for administrators who wish to make informed decisions in the increasingly value-based world of health care. On the other hand, the use of EHR data for billing and administrative purposes can incentivize clinicians to bias their documentation for maximal payment, and possibly reducing the clinical accuracy of the diagnoses. It may therefore be best to base financial decisions on variables that are not open to interpretation.

Despite Reiser's (1991) clinically oriented goals for EHRs, much of what these systems are currently is driven by complex and prescriptive medical-legal, reimbursement, and regulatory requirements (Cusack et al. 2013). These requirements may lead to redundant data capture, cumbersome documentation processes, and information that is biased towards optimized billing. One potential solution would be policy changes such that the tie between payment and documentation elements are less emphasized, including the proposed CMS rule "Patients Over Paperwork"[42] which features decreased documentation burden for office visits.

39 ► https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=3516 (Accessed 6/4/2020).

40 ► https://ecqi.healthit.gov/cql-clinical-quality-language (Accessed 6/4/2020).

41 ► https://www.hl7.org/fhir/fhirpath.html (Accessed 6/4/2020).

42 ► CMS.gov. Patients Over Paperwork. ► https://www.cms.gov/About-CMS/story-page/patients-over-paperwork.html (Accessed 6/4/2020).

14.5 Challenges Ahead

Although many commercial products are labeled as EHRs, some do not satisfy all the criteria that we defined at the beginning of this chapter. Even beyond matters of definition, however, it is important to recognize that the concept of an EHR is neither unified nor static. As the capability of technology evolves, the function of the EHR will expand. Certified health IT specifications appear also to have pushed forward certain types of core EHR functionality.

Greater involvement of patients in their own care, for example, means that **personal health records** (**PHRs**) will increasingly incorporate data captured at home and also support two-way communication between patients and their health care team (see also ▶ Chap. 13). The potential for patient-entered data includes history, symptoms, and outcomes entered by patients as well as data uploaded automatically by home monitoring devices such as scales, blood pressure monitors, glucose meters, pulmonary function devices, smart phones and Fitbits. By integrating these patient-generated data into the EHR, either by uploading the data into the EHR or by linking the EHR and the PHR, a number of long-term objectives can be achieved: patient-generated data may in some circumstances be more accurate or complete, the time spent entering data during an office visit by both the provider and the patient may be reduced, and the information may allow the production of outcomes measures that are better attuned to patients' goals. Patient-delivered data will be welcome when this information has been requested by the practice (e.g. initial visit history check list), and a mutual understanding exists about the types and volumes of data that can be accepted and delays between receipt and review.[43]

The future of EHRs depends on both technical and nontechnical considerations. Computing technology will continue to advance, with processing power doubling every 1.5 years according to Moore's law (see ▶ Chap. 1). Software will improve with more powerful applications, better user interfaces, and more integrated CDS, including CDS using third party solutions and CDS integration specifications (e.g., FHIR®©, CDS hooks[44]). New kinds of software that support collaboration will continue to improve; social media are growing rapidly both inside and outside of health care. For example, as both providers and patients engage increasingly in social media, new ways to capture data, share data, collaborate, and share expertise may emerge. Perhaps the greater need for leadership and action will be in the social and organizational foundations that must be laid if EHRs are to serve as the information infrastructure for health care. We touch briefly on some of these challenges in this final section.

14.5.1 Usability

An intuitive and efficient user interface is an important desired characteristic of an EHR. Designers must understand the cognitive aspects of the human computer interaction (HCI) and each of the various workflows if they are to build user interfaces that are easy-to-learn and easy-to-use (see ▶ Chap. 5). Improving these systems using best practices and principles of HCI will require changes not only in how the system behaves but also in how humans interact with the system.

User interface requirements of a nurse entering patient data in the inpatient setting are different from the requirements of a clerk entering patient charges. Usability for clinicians means fast computer response times, and the fewest possible data input fields. *A system that is slow or requires too much input is not usable by clinicians, particularly in the time-constrained setting of clinical care*. The menus and vocabularies that constrain input must

43 We have included examples from various systems in this chapter, both developed by users and commercially available, to illustrate a portion of the functionality of EHRs currently in use.

44 CDS hooks HL7 group. ▶ http://wiki.hl7.org/index.php?title=201809_CDS_Hooks (Accessed 6/4/2020).

include synonyms for all the ways health professionals name the items, and the system must have keyboard options for all inputs and actions because switching from mouse to keyboard steals user time.

What information the provider needs and what tasks the provider performs should influence what information the EHR presents and how the system presents it. Development of technology that matches the data-processing power of computers with the cognitive capability of human beings to formulate insightful questions and to interpret data is still a rate-limiting step (Tang and Patel 1994). For example, one can imagine an interface in which speech input, typed narrative, and mouse-based structured data entry are accepted and seamlessly stored into a single data structure within the EHR, with a hybrid user display that shows both a narrative version of the information and a structured version of the same information that highlights missing fields or inconsistent values. Along these lines, ◘ Fig. 14.17 shows a historical example of order generation by the Gopher 3 system in operation at Eskanazi Hospital (a.k.a. Wishard hospital) clinics. Physicians would write their problem in narrative text. Using NLP methods in parallel, the computer would generate a list of code orders that it inferred from these notes. Users could then confirm order(s) with simple click(s) and add any further details required to complete each order (Duke et al. 2014). This same functionality is now provided through the Regenstrief Clinical Learning system – a realistic medical record system with rich sample patient data available for teaching medical students about EHR functionality.[45]

14.5.2 Standards

We alluded to the importance of standards earlier in this chapter, when we discussed the architectural requirements of integrating data from multiple sources. Standards are the focus of ▶ Chap. 8. Here, we stress the importance of national standards to the development, implementation, and use of EHRs (Miller and Gardner 1997). Standards are especially important for integrating clinical data from different organizations. Health information exchanges (HIEs) continue to expand in size and numbers but the healthcare systems that feed them will have adopt meaningful use coding and messages/API structure standards more fully than before. HIEs will be able to efficiently import and integrate structured data about one patient from many organizations. Messaging and API standards are increasingly well developed and in widespread use for laboratory data[46] (HL7s LRI), prescriptions sent to pharmacies[47] (NCPDPs SCRIPT stands), many kinds of diagnostic (DICOM) images.[48] FHIR®© is now supported by major federal agencies (ONC, CMS, NIH, CDC, AHRQ and increasingly by the FDA) as well as by the high-tech industry (Apple, Amazon, Google and Microsoft) and health care software developers. It is now mainstream for many in healthcare applications and communications. The incomplete adoption of standard coding systems for observation identifiers, however, remains a major obstacle to the integration of patient data from independent care providers and large care delivery systems alike.

The HIPAA legislation[49] includes mandated standards for administrative messages (×12) privacy, security, and clinical data. Federal agencies have already promulgated regulations based on this legislation for the first three of these categories.[50] A series of legislative measures, notably with the 2009

45 ▶ https://www.regenstrief.org/resources/clinical-learning/ (Accessed 6/4/2020).

46 ▶ https://www.lri.fr/presentation_en.php (Accessed 6/4/2020).

47 ▶ https://www.ncpdp.org/NCPDP/media/pdf/NCPDPEprescribingBasics.pdf (Accessed 6/4/2020).

48 ▶ https://searchhealthit.techtarget.com/definition/DICOM-Digital-Imaging-and-Communications-in-Medicine (Accessed 6/4/2020).

49 ▶ https://www.edibasics.com/edi-resources/document-standards/hipaa (Accessed 6/4/2020).

50 HIPAA for professionals. ▶ https://www.hhs.gov/hipaa/for-professionals/index.html (Accessed 6/4/2020).

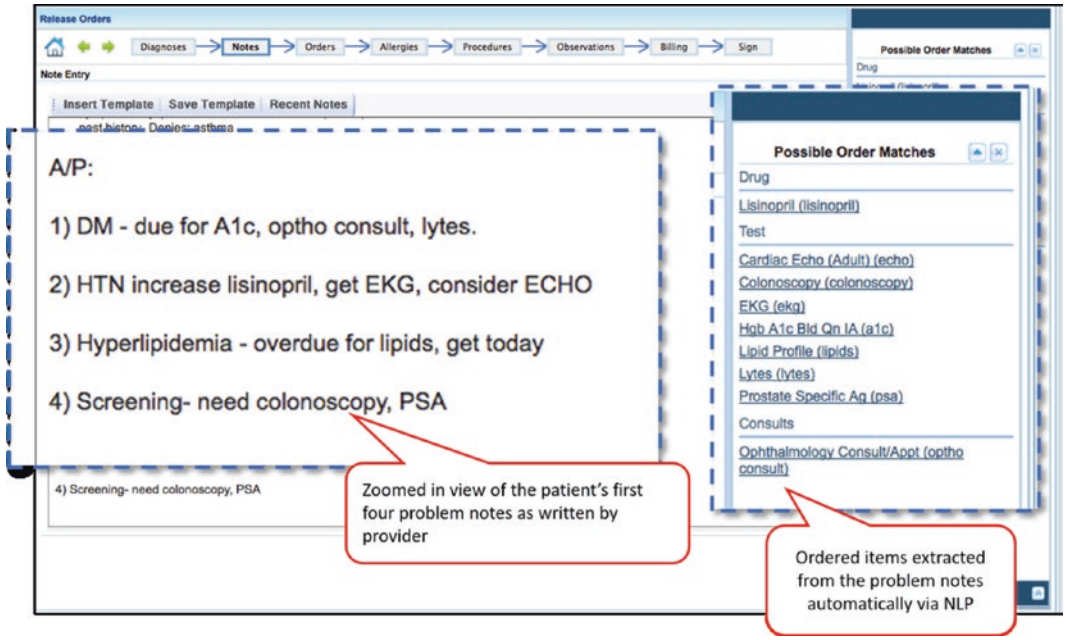


Fig. 14.17 An example of NLP and rule-based conversion of provider notes to order items from Blain Takasu. Notes written in free text in the A/P section are inferred by NLP analysis to suggest possible matches in the order items box

HITECH Act[51] (see ▶ Chaps. 8 and 31) and subsequently with the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA),[52] have stimulated significant efforts to increase the adoption and functionality of EHRs, as well as leverage these systems for quality reporting. More recently, the twenty-first Century Cure Act has significant provisions around delivery of Health IT usability (Sect. 4001), Conditions of Certification (Sect. 4002(a)), Trusted Exchange Framework and Common Agreement (Sect. 4003(b)), and guidelines around reasonable and necessary activities that do not constitute information blocking (Sect. 4004)[53] and these positions are present in proposed ONC and CMS rules.

14.5.3 Costs and Benefits

The National Academy of Medicine (formerly Institute of Medicine) declared the EHR an essential infrastructure for the delivery of health care, and the protection of patient safety (IOM Committee on Improving the Patient Record 2001). Like any infrastructure project, the benefits specifically attributable to infrastructure are not immediate and sometimes difficult to establish; an infrastructure plays an enabling role in all projects that take advantage of it. Early randomized controlled clinical studies showed that computer-based decision-support systems reduce costs and improve quality compared with usual care supported with a paper medical record

51 ▶ https://www.healthit.gov/sites/default/files/hitech_act_excerpt_from_arra_with_index.pdf (Accessed 6/4/2020).

52 ▶ CMS.gov: MACRA. ▶ https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-MIPS-and-APMs.html (Accessed 6/4/2020).

53 H.R.34 – 114th Congress. ▶ https://www.congress.gov/bill/114th-congress/house-bill/34 (Accessed 6/4/2020).

(Tierney et al. 1993; Bates et al. 1997, 2003; Classen et al. 1997), and meta-analyses of Health IT have demonstrated quality benefits (Buntin et al. 2011; Lau et al. 2010; Clyne et al. 2012). However, others have not found consistent associations between EHRs and CDS and better quality (Romano and Stafford 2011; Delvaux et al. 2017; Parshuram et al. 2018; Muth et al. 2018).

Because of the significant resources needed and the significant broad-based potential benefits of these systems, the decision to implement an EHR is a strategic one for most healthcare organizations. Hence, the evaluation of the costs and benefits must consider the effects on the organization's strategic goals, as well as the objectives for individual health care (Samantaray et al. 2011). Today, there are a number of **Open Source** options for EHR software with a range of capabilities (Syzdykkova et al. 2017).

The cost of installing an EHR in a large health system can exceed $100 million and even $1 billion for the largest implementations.[54] The cost of the system itself in license fees and related items is usually only a portion of that number. Other costs include configuration, training, and lost revenue as care providers learn to use the system. The benefits of such an investment are often related to the integration of a health system's diverse components into a single, coordinated enterprise.

14.5.4 Leadership

Leaders from all segments of the health care industry must work together to articulate the needs, to continue to define and expand upon the standards, to fund the development, to implement the social change, and to write the laws to accelerate the development and routine use of EHRs in health care. Because of the prominent role of the federal government in health care—as a payer, provider, policymaker, and regulator—federal leadership to create incentives for developing and adopting standards and for promoting the implementation and use of EHRs remains crucial. Technological change will continue to occur at a rapid pace, driven by consumer demand for entertainment, retail, games, and business tools. Nurturing the use of IT in health care requires leaders including informatician leaders who promote the use of EHRs and work to overcome the obstacles that impede widespread use of computers for the benefit of health care.

54 EHR Intelligence. Top 5 Most Expensive Implementations of 2017. ► https://ehrintelligence.com/news/top-5-most-expensive-ehr-implementations-of-2017 (Accessed 6/5/2020).

Suggested Reading

Barnett, G. O. (1984). The application of computer-based medical-record systems in ambulatory practice. *New England Journal of Medicine, 310*(25), 1643–1650. This seminal article compares the characteristics of manual and automated ambulatory patient record systems, discusses implementation issues, and predicts future developments in technology.

Bates, D. W., Kuperman, G. J., Wang, S., et al. (2003). Ten commandments for effective clinical decision support: Making the practice of evidence-based medicine a reality. *Journal of the American Medical Informatics Association, 10*(6), 523–530. The authors present ten very practical tips to designers and installers of clinical decision support systems.

Berner, E. S. (Ed.). (2010). *Clinical decision support systems, theory and practice: Health informatics series* (3rd ed.). New York: Springer. This text focuses on the design, evaluation, and application of Clinical Decision Support systems, and examines the impact of computer-based diagnostic tools both from the practitioner's and the patient's perspectives. It is designed for informatics specialists, teachers or students in health informatics, and clinicians.

Collen, M. F. (1995). *A history of medical informatics in the United States, 1950–1990*. Indianapolis: American Medical Informatics Association, Hartman Publishing. This rich history of medical informatics from the late 1960s to the late 1980s includes an extremely detailed set of references.

Collen, M. F., & Ball, M. J. (Eds.). (2015). *The history of medical informatics in the United States* (2nd ed.). London: Springer-Verlag, Springer Nature. This rich history of medical informatics from the 1990s to mid 2010s (25 years) provides an updated medical informatics perspective. It includes an extremely detailed set of references.

Hartley, C. P., & Jones, E. D. (2011). *EHR implementation: A step-by-step guide for the medical practice* (2nd ed.). Chicago: American Medical Association. This book provides rich details for implementing an EHR. It is a great resource for anyone trying to learn about EHR deployments, covering topics related to preparation, support, and implementation.

Institute of Medicine (IOM) Roundtable on Value and Science-Driven Health Care. (2011). *Digital infrastructure for the learning health system: The foundation for continuous improvement in health and health care – workshop series summary*. Washington, DC: National Academy Press. This report summarizes three workshops that presented new approaches to the construction of advanced medical record system that would gather the crucial data needed to improve the health care system.

Kuperman, G. J., Gardner, R. M., & Pryor, T. A. (1991). *The HELP system*. Berlin/Heidelberg: Springer-Verlag GmbH and Co. K. The HELP (Health Evaluation through Logical Processing) system was a computerized hospital information system developed by the authors at the LDS Hospital at the University of Utah, USA. It provided clinical, hospital administration and financial services through the use of a modular, integrated design. This book thoroughly documents the HELP system. Chapters discuss the use of the HELP system in intensive care units, the use of APACHE and APACHE II on the HELP system, various clinical applications and inactive or experimental HELP system modules. Although the HELP system has now been retired from routine use, it remains an important example of several key issues in EHR implementation and use that continue in the commercial systems of today..

Osheroff, J., Teich, J., Levick, D., et al. (2012). *Improving outcomes with clinical decision support: An implementers guide* (2nd ed.). Scottsdale: Scottsdale Institute, AMIA, AMDIS and SHM. This text provides guidance on using clinical decision support interventions to improve care delivery and outcomes in a hospital, health system or physician practice. The book also presents considerations for health IT software suppliers to effectively support their CDS implementer clients.

Sittig, D. F., & Ash, J. S. (2011). *Clinical information systems: Overcoming adverse consequences (Jones and Bartlett series in biomedical informatics)* (1st ed.). Burlington: Jones and Bartlett Learning. This book explores the challenges and obstacles with implementation of clinical information systems including the nine categories of unintended adverse consequences with implementation and optimization of these systems as well as best practices.

Weed, L. L. (1969). *Medical records, medical evaluation and patient care: The problem-oriented record as a basic tool*. Chicago: Year Book Medical Publishers. In this classic book, Weed presents his plan for collecting and structuring patient data to produce a problem-oriented medical record.

Questions for Discussion

1. What is the definition of an EHR? What, then, is an EHR? What are five advantages of an EHR over a paper-based record? Name three limitations of an EHR.
2. What are the five functional components of an EHR? Think of the information systems used in health care institutions in which you work or that you have seen. Which of the components that you named do those systems have? Which are missing? How do the missing elements limit the value to the clinicians or patients?
3. Discuss three ways in which a computer system can facilitate information transfer between hospitals and ambulatory care facilities, thus enhancing continuity of care for previously hospitalized patients who have been discharged and are now being followed up by their primary physicians.

4. Much of medical care today is practiced in teams, and coordinating the care delivered by teams is a major challenge. Thinking in terms of the EHR functional components, describe four ways that EHRs can facilitate care coordination. Describe two ways in which EHRs are likely to create additional challenges in care coordination.
5. How does the health care financing environment affect the use, costs, and benefits of an EHR? How has the financing environment affected the functionality of information systems? How has it affected the user population?
6. Would a computer scan of a paper-based record be an EHR? What are two advantages and two limitations of this approach?
7. Among the key issues for designing an EHR are what information should be captured and how can it be entered into the system. Physicians may enter data directly or may record data on a paper worksheet (encounter form) for later transcription by a data-entry worker. What are two advantages and two disadvantages of each method? Discuss the relative advantages and disadvantages of entry of free text instead of entry of fully coded information. Describe an intermediate or compromise method.
8. EHR data may be used in clinical research, quality improvement, and monitoring the health of populations. Describe three ways that the design of the EHR may affect how the data may be used for other purposes.
9. Identify four locations where clinicians need access to the information contained in an EHR. What are the major costs or risks of providing access from each of these locations?
10. What are three important reasons to have physicians enter orders directly into an EHR? What are three challenges in implementing such a system?
11. Consider the task of creating a summary report for clinical data collected over time and stored in an EHR. Clinical laboratories traditionally provide summary test results in flowsheet format, thus highlighting clinically important changes over time. A medical record system that contains information for patients who have chronic diseases must present serial clinical observations, history information, and medications, as well as laboratory test results. Suggest a suitable format for presenting the information collected during a series of ambulatory-care patient visits.
12. The public demands that the confidentiality of patient data must be maintained in any patient record system. Describe three protections and auditing methods that can be applied to paper-based systems. Describe three technical and three nontechnical measures you would like to see applied to ensure the confidentiality of patient data in an EHR. How do the risks of privacy breaches differ for the two systems?

References

Anand, V., McKee, S., Dugan, T. M., & Downs, S. M. (2017). Leveraging electronic tablets for general pediatric care: A pilot study. *Applied Clinical Informatics, 6*(1), 1–15.

Anand, V., Carroll, A. E., Biondich, P. G., Dugan, T. M., & Downs, S. M. (2018). Pediatric decision support using adapted Arden Syntax. *Artificial Intelligence in Medicine, 92*, 15–23.

Arsoniadis, E. G., Skube, S. J., Bjerke, T. M., Jarabek, B., & Melton, G. B. (2017). Assessing provider-generated free-text quality in EHR-integrated handoff notes. *Studies in Health Technology and Informatics, 245*, 999–1003.

Barnett, G. O. (1984). The application of computer-based medical-record systems in ambulatory practice. *The New England Journal of Medicine, 310*(25), 1643–1650.

Barnett, G. O., Justice, N. S., Somand, M. E., et al. (1979). COSTAR – a computer-based medical information system for ambulatory care. *Proceedings of the Institute of Electrical and Electronics Engineers (IEEE), 67*, 1226–1237.

Bates, D. W., Spell, N., Cullen, D. J., et al. (1997). The costs of adverse drug events in hospitalized patients. *Journal of the American Medical Association, 277*(4), 307–311.

Bates, D. W., Evans, R. S., Murff, H., et al. (2003). Detecting adverse events using information technology. *Journal of the American Medical Informatics Association: JAMIA, 10*(2), 115–128.

Bauer, D. T., Guerlain, S., & Brown, P. J. (2010). The design and evaluation of a graphical display for laboratory data. *Journal of the American Medical Informatics Association: JAMIA, 17*(4), 416–424.

Bleich, H. L., Beckley, R. F., Horowitz, G. L., Jackson, J. D., Moody, E. S., Franklin, C., Goodman, S. R., McKay, M. W., Pope, R. A., Walden, T., et al. (1985). Clinical computing in a teaching hospital. *The New England Journal of Medicine, 312*(12), 756–764.

'Brain' to Store Medical Data. (1956, October 24). *The New York Times*.

Brown, S. H., Lincoln, M. J., Groen, P. J., et al. (2003). VistA–U.S. Department of Veterans Affairs national-scale HIS. *International Journal of Medical Informatics, 69*(2–3), 135–156.

Brown, S. H., Fischietti, L. F., Graham, G., Bates, J., et al. (2007). Use of electronic health records in disaster response: The experience of department of veterans affairs after hurricane Katrina. *American Journal of Public Health, 97*, S136–S141.

Brownstein, J. S., Sordo, M., Kohane, I. S., & Mandl, K. D. (2007). The tell-tale heart: Population-based surveillance reveals an association of rofecoxib and celecoxib with myocardial infarction. *PLoS One, 2*(9), e840.

Buchanan, J. (2017). Accelerating the benefits of the problem oriented medical record. *Applied Clinical Informatics, 8*(1), 180.

Buntin, M. B., Burke, M. F., Hoaglin, M. C., & Blumenthal, D. (2011). The benefits of health information technology: A review of the recent literature shows predominantly positive results. *Health Affairs (Millwood), 30*(3), 464–471.

Burykin, A., Peck, T., & Buchman, T.G. (2011). Using "off-the-shelf" tools for terabyte-scale waveform recording in intensive care: computer system design, database description and lessons learned. In *Computational Methods Programs Biomed. 103*(3) (pp. 151–60). https://doi.org/10.1016/j.cmpb.2010.10.004.

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S. C., et al. (2006). Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine, 144*(10), 742–752.

Cheung, N. T., Fung, K. W., Wong, K. C., et al. (2001). Medical informatics – the state of the art in the hospital authority. *International Journal of Medical Informatics, 62*, 113–119.

Classen, D. C., Pestotnik, S. L., Evans, R. S., Lloyd, J. F., & Burke, J. P. (1997). Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. *Journal of the American Medical Association, 277*(4), 301–306.

Clyne, B., Bradley, M. C., Hughes, C., et al. (2012). Electronic prescribing and other forms of technology to reduce inappropriate medication use and polypharmacy in older people: A review of current evidence. *Clinical Geriatric Medicine, 28*(2), 301–322.

Coffey, R. (1979). *How medical information systems affect costs: The El Camino experience*. Washington, DC: Public Health Service. Publication DHEW-PHS-80-3626.

Collen, M. F. (1969). Value of multiphasic health checkups. *The New England Journal of Medicine, 280*(19), 1072–1073.

Collen, M. F. (1995). *A history of medical informatics in the United States, 1950 to 1990*. Bethesda: American Medical Informatics Association.

Covell, D., Uman, G., & Manning, P. (1985). Information needs in office practice: Are they being met? *Annals of Internal Medicine, 103*, 596–599.

Cusack, C. M., Hripcsak, G., Bloomrosen, M., Rosenbloom, T., Weaver, C., Wright, A., Vawdrey, D., Walker, J., & Mamykina, L. (2013). The future state of clinical data capture and documentation: A report from AMIA's 2011 Policy Meeting. *Journal of the American Medical Informatics Association: JAMIA, 20*, 134–140.

Damiani, G., Pinnarelli, L., Colosimo, S. C., Almiento, R., Sicuro, L., Galasso, R., Sommello, L., & Ricciardi, W. (2010). The effectiveness of computerized clinical guidelines in the process of care: A systematic review. *BMC Health Services Research, 10*, 2.

Damush, T. M., Weinberger, M., Clark, D. O., Tierney, W. M., Rao, J. K., Perkins, S. M., & Verel, K. (2002). Acute low back pain self management intervention for urban primary care patients: Rationale, design, and predictors of participation. *Arthritis and Rheumatism, 47*, 372–379.

Davis, L. S., Collen, M. F., Rubin, F. L., & Van Brunt, E. E. (1968). Computer stored medical record. *Computer Biomedical Research, 1*, 452–469.

Del Fiol, G., Huser, V., Strasberg, H. R., et al. (2012). Implementations of the HL7 context-aware knowledge retrieval ("infobutton") standard: Challenges, strengths, limitations, and uptake. *Journal of Biomedical Informatics, 45*(4), 726–735.

Delvaux, N., Van Thienen, K., Heselmans, A., et al. (2017). The effects of computerized clinical decision support systems on laboratory test ordering: A systematic review. *Archives of Pathology & Laboratory Medicine, 141*(4), 585–595.

Dexter, P. R., Perkins, S., Overhage, J. M., et al. (2001). A computerized reminder system to increase the use of preventive care for hospitalized patients. *The New England Journal of Medicine, 345*(13), 965–970.

Dexter, P. R., Perkins, S. M., Maharry, K. S., Jones, K., & McDonald, C. J. (2004). Inpatient computer-based standing orders vs. physician reminders to increase influenza and pneumococcal vaccination rates. *Journal of the American Medical Association, 292*(19), 2366–2371.

Downs, S. M., Biondich, P. G., Anand, V., Zore, M., & Carroll, A. E. (2006). Using Arden syntax and

adaptive turnaround documents to evaluate clinical guidelines. In *AMIA annual symposium proceedings, 2006* (pp. 214–218).

Duke, J. D., Morea, J., Mamlin, B., Martin, D. K., Simonaitis, L., Takesue, B. Y., Dixon, B. E., & Dexter, P. R. (2014). Regenstrief Institute's medical gopher: A next-generation homegrown electronic medical record system. *International Journal of Medical Informatic, 83*(3), 170–179.

Duncan, R. G., Saperia, D., Dulbandzhyan, R., et al. (2001). Integrated web-based viewing and secure remote access to a clinical data repository and diverse clinical systems. In *Proceedings of the AMIA fall symposium, 2001* (pp. 149–153).

Earle, K. (2011). In people with poorly controlled hypertension, self-management including telemonitoring is more effective than usual care for reducing systolic blood pressure at 6 and 12 months. *Evidence-Based Medicine, 16*(1), 17–18.

Edgar, E. P. (2009). *Physician retention in Army Medical Department. Strategic Research Project, U.S. Army War College, Carlisle Barracks PA* 17013–5050. Available at: http://handle.dtic.mil/100.2/ADA499087

Evans, R. S., Pestotnik, S. L., Classen, D. C., Clemmer, T. P., Weaver, L. K., Orme, J. F., Lloyd, J. F., & Burke, J. P. (1998). A computer-assisted management program for antibiotics and other antinfective agents. *The New England Journal of Medicine, 338*(4), 232–238.

Fafchamps, D., Young, C. Y., & Tang, P. C. (1991). Modelling work practices: Input to the design of a physician's workstation. In *Proceedings of the 15th annual symposium on computer applications in medical care* (pp. 788–792). Washington, DC: CAMC.

Ferranti, J. M., Musser, R. C., Kawamoto, K., & Hammond, W. E. (2006). The clinical document architecture and the continuity of care record: A critical analysis. *Journal of the American Medical Informatics Association: JAMIA, 13*(3), 245–252.

Fitzmaurice, J. M., Adams, K., & Eisenberg, J. M. (2002). Three decades of research on computer applications in health care: Medical informatics support at the agency for healthcare research and quality. *Journal of the American Medical Informatics Association: JAMIA, 9*(2), 144–160.

Flexner, A. (1910). *Medical education in the united states and Canada: A report to the Carnegie foundation for the advancement of teaching*. New York: The Carnegie Foundation for the Advancement of Teaching. OCLC 9795002.

Fried, T. R., Niehoff, K. M., Street, R. L., et al. (2017). Effect of the tool to reduce inappropriate medications on medication communication and deprescribing. *Journal of the American Geriatrics Society, 65*(10), 2265–2271.

Fries, J. F. (1974). Alternatives in medical record formats. *Medical Care, 12*(10), 871–881.

Frisse, M. E., King, J. K., Rice, W. B., et al. (2008). A regional health information exchange: Architecture and implementation. In *American Medical Informatics Association annual symposium proceedings, 2008* (pp. 212–216).

Frisse, M. E., Johnson, K., Nian, H., et al. (2011). The financial impact of health information exchange on emergency department care. *Journal of the American Medical Informatics Association: JAMIA, 19*(3), 328–333.

Fry, E., & Schulte, F. (2019, March 18). Death by a thousand clicks: Where electronic health records went wrong. *Fortune Magazine*.

Fung, K. W., Xu, J., Rosenbloom, S. T., et al. (2011). Testing three problem list terminologies in a simulated data entry environment. In *American Medical Informatics Association annual symposium proceeds, 2011* (pp. 445–454).

Gardner, R. L., Cooper, E., Haskell, J., et al. (2019). Physician stress and burnout: The impact of health information technology. *JAMIA, 26*(2), 106–114.

Garg, A. X., Adhikari, N. K. J., McDonald, H., et al. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *Journal of the American Medical Association, 293*(10), 1223–1238.

Giuse, D. A., & Mickish, A. (1996). Increasing the availability of the computerized patient record. In *Proceedings of the AMIA annual fall symposium, 1996* (pp. 633–637).

Gottschalk, A., & Flocke, S. A. (2005). Time spent in face-to-face patient care and work outside the examination room. *Annals of Family Medicine, 3*(6), 488–493.

Green, B. B., Cook, A. J., Ralston, J. D., et al. (2008). Effectiveness of home blood pressure monitoring, web communication, and pharmacist care on hypertension control: A randomized controlled trial. *Journal of the American Medical Association, 299*(24), 2857–2867.

Grossman, J. H., Barnett, G. O., Koepsell, T. D., Nesson, H. R., Dorsey, J. L., & Phillips, R. R. (1973). An automated medical record system. *Journal of the American Medical Association, 224*(12), 1616–1621.

Gulacti, U., & Lok, U. (2017). Comparison of secure messaging application (WhatsApp) and standard telephone usage for consultations on Length of Stay in the ED. A prospective randomized controlled study. *Applied Clinical Informatics, 8*(3), 742–753.

Hagen, P. T., Turner, D., Daniels, L., & Joyce, D. (1998). Very large-scale distributed scanning solution for automated entry of patient information. *TEPR Proceedings (Toward an ElectronicPatient Record), 1*, 228–232.

Halamka, J. D., & Safran, C. (1998). CareWeb: A web-based medical record for an integrated healthcare delivery system. In *Proceedings of Medinfo 1998, Part 1* (pp. 36–39).

Haynes, R. B. (Ed.). (2011). Computerized clinical decision support systems: How effective are they? [article collection] *Implementation Science*, *6*, 87–108. Articles available at: http://www.implementationscience.com/series/CCDSS. Accessed 2 Jan 2012.

Health IT Standards Committee. (2011). Recommendations to ONC on the assignment of code sets to clinical concepts [data elements] for use in quality measures. [Letter]. Retrieved from http://www.healthit.gov/sites/default/files/standards-certification/HITSC_CQMWG_VTF_Transmit_090911.pdf. Accessed 3 Jan 2012.

Hirsch, J. S., Tanenbaum, J. S., Lipsky Gorman, S., Liu, C., Schmitz, E., Hashorva, D., Ervits, A., Vawdrey, D., Sturm, M., & Elhadad, N. (2015). HARVEST, a longitudinal patient record summarizer. *Journal of the American Medical Informatics Association: JAMIA, 22*, 263–274.

Hollingsworth, J. C., Chisholm, C. D., Giles, B. K., Cordell, W. H., & Nelson, D. R. (1998). How do physicians and nurses spend their time in the emergency department? *Annals of Emergency Medicine, 31*, 87–91.

Hripcsak, G., Cimino, J. J., & Sengupta, S. (1999). WebCIS: Large scale deployment of a web-based clinical information system. In *Proceedings of the annual AMIA symposium, 1999* (pp. 804–808).

Hunt, D. L., Haynes, R. B., et al. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes: A systematic review. *JAMA : The Journal of the American Medical Association, 280*(15), 1339–1346.

Institute of Medicine Committee on Improving the Patient Record. (2001). Crossing the quality chasm: A new health system for the 21st century. Washington, DC: National Academy Press. Available at: http://www.iom.edu/Reports/2001/Crossing-the-Quality-Chasm-A-New-Health-System-for-the-21st-Century.aspx. Accessed 2 Jan 2012.

Janamanchi, B., Katsamakas, E., Raghupathi, W., & Gao, W. (2009). The state and profile of open source software projects in health and medical informatics. *International Journal of Medical Informatics, 78*(2009), 457–472.

Johnson, S. B., Friedman, C., Cimino, J. J., Clark, T., Hripcsak, G., & Clayton, P. D. (1991). Conceptual data model for a central patient database. In *Proceedings of the symposium on computer applications in medical care* (pp. 381–385).

Just, B. H., Marc, D., Munns, M., et al. (2016). Why patient matching is a challenge: Research on master patient index (MPI) data discrepancies in key identifying fields. *Perspectives in Health Information Management, 13*(Spring), 1e.

Khajouei, R., & Jaspers, M. W. (2010). The impact of CPOE medication systems' design aspects on usability, workflow and medication orders: A systematic review. *Methods of Information in Medicine, 49*(1), 3–19.

Kho, A., Zafar, A., & Tierney, W. (2007). Information technology in PBRNs: The Indiana University Medical Group Research Network (IUMG ResNet) experience. *The Journal of the American Board of Family Medicine, 20*(2), 196–203.

Kohane, I. S. (2011). Using electronic health records to drive discovery in disease genomics. *Nature Reviews Genetics, 12*(6), 417–428.

Koshy, S., Feustel, P. J., Hong, M., & Kogan, B. A. (2010). Scribes in an ambulatory urology practice: Patient and physician satisfaction. *Journal of Urology, 184*(1), 258–262.

Kuhn, I. M., Wiederhold, G., Rodnick, J.E., et al. (1984). Automated ambulatory medical record systems in the U.S. In Blum, B. (Ed.), Information systems for patient care (pp.199–217), New York: Springer. Original 1982 report available at: http://infolab.stanford.edu/TR/CS-TR-82-928.html. Accessed 2 Jan 2012.

Lamb, A. (1955). *The Presbyterian Hospital and the Columbia-Presbyterian Medical Center, 1868–1943: A history of a great medical adventure*. New York: Columbia University Press.

Larsen, R. A., Evans, R. S., Burke, J. P., Pestotnik, S. L., Gardner, R. M., & Classen, D. C. (1989). Improved perioperative antibiotic use and reduced surgical wound infections through use of computer decision analysis. *Infection Control and Hospital Epidemiology, 10*(7), 316–320.

Lau, F., Kuziemsky, C., Price, M., & Gardner, J. (2010). A review on systematic reviews of health information system studies. *Journal of the American Medical Informatics Association: JAMIA, 17*(6), 637–645.

Longhurst, C. A., Harrington, R. A., & Shah, N. H. (2014). A 'green button' for using aggregate patient data at the point of care. *Health Affairs, 33*(7), 1229–1235.

Major, K., Shabot, M. M., & Cunneen, S. (2002). Wireless clinical alerts and patient outcomes in the surgical intensive care unit. *American Surgery, 68*, 1057–1060.

Mamlin BW, Overhage JM, Tierney WM, Dexter PR, McDonald CJ. Clinical Decision Support Within the Regenstrief Medical Record System. Book chapter - Clinical Decision Support Systems -Theory and Practice - Series: Health Informatics, Berner, Eta S. (Ed.) 2nd ed., 2007, 190–21.

Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association, 23*(5), 899–908.

Mandl, K. D., Gottlieb, D., & Ellis, A. (2019). Beyond one-off integrations: A commercial, substitutable, reusable, standards-based, electronic health record–connected app. *Journal of Medical Internet Research, 21*(2), e12902. https://doi.org/10.2196/12902.

McDonald, C. J. (1973, November) The clinical automaton—An ambulatory care record with control capabilities. In *Proceedings of the IEEE Conf Cyber Soc*. (pp. 27–31).

McDonald, C. J. (1976). Protocol-based computer reminders. The quality of care and the non-perfectibility of man. *The New England Journal of Medicine, 295*(24), 1351–1355. https://doi.org/10.1056/NEJM197612092952405.

McDonald, C. J., Bhargava, B., & Jeris, D. W. (1975). A clinical information system (CIS) for ambulatory care. In *Proceedings AFIPS national computing conference*, Anaheim.

McDonald, C. J., Overhage, J. M., Tierney, W. M., et al. (1999). The regenstrief medical record system: A quarter century experience. *International Journal of Medical Informatics, 54*(3), 225–253.

McDonald, C. J., Huff, S. M., Suico, J. G., et al. (2003). LOINC, a universal standard for identifying laboratory observations: A 5-year update. *Clinical Chemistry, 49*(4), 624–633.

McDonald, C. J., Overhage, J. M., Barnes, M., Schadow, G., Blevins, L., Dexter, P. R., & Mamlin, B. W. (2005). The Indiana network for patient care: A working local health information infrastructure (LHII). *Health Affairs (Millwood), 24*(5), 1214–1220.

McDonald CJ, Callaghan FM, Weissman A, Goodwin RM, Mundkur M, Kuhn T. Use of Internist's Free Time by Ambulatory Care Electronic Medical Record Systems. JAMA Intern Med. 2014;174(11):1860–1863. https://doi.org/10.1001/jamainternmed.2014.4506.

McPhee, S. J., Bird, J. A., Fordham, D., Rodnick, J. E., & Osborn, E. H. (1991). Promoting cancer prevention activities by primary care physicians: Results of a randomized, controlled trial. *Journal of the American Medical Association, 266*(4), 538–544.

Melton, L. J., 3rd. (1996). History of the Rochester epidemiology project. *Mayo Clinic Proceedings, 71*, 266–274.

Miller, R. A., & Gardner, R. M. (1997). Recommendations for responsible monitoring and regulation of clinical software systems. *Journal of the American Medical Informatics Association: JAMIA, 4*, 442–457.

Miller, R. A., Gardner, R. M., Johnson, K. B., & Hripcsak, G. (2005a). Clinical decision support and electronic prescribing systems: A time for responsible thought and action. *Journal of the American Medical Informatics Association: JAMIA, 12*(4), 403–409.

Miller, R. A., Waitman, L. R., Chen, S., & Rosenbloom, S. T. (2005b). The anatomy of decision support during inpatient care provider order entry (CPOE): Empirical observations from a decade of CPOE experience at Vanderbilt. *Journal of Biomedical Informatics, 38*(6), 469–485.

Misra-Hebert, A. D., et al. (2016, March). Medical scribes: How do their notes stack up? *Journal of Family Practice*, 155+. Gale Academic Onefile, Accessed 14 Oct 2019.

Muth, C., Ulhmann, L., Haefeli, W. E., et al. (2018). Effectiveness of a complex intervention on Prioritising Multimedication in Multimorbidity (PRIMUM) in primary care: Results of a pragmatics cluster randomised controlled trial. *BMJ Open, 8*(2), e017740.

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association: JAMIA, 18*(5), 544–551.

National Committee on Vital and Health Statistics. (2001). Information for health: A strategy for building the National Health Information Infrastructure. Report and recommendations from the National Committee on Vital and Health Statistics. Available at http://www.ncvhs.hhs.gov/nhiilayo.pdf. Accessed 17 Dec 2012.

Nebeker, J. R., Hoffman, J. M., Weir, C. R., Bennett, C. L., & Hurdle, J. F. (2005). High rates of adverse drug events in a highly computerized hospital. *Archives of Internal Medicine, 165*(10), 1111–1116.

Openchowski, M. W. (1925). The effect of the unit record system and improved organization on hospital economy and efficiency. *Archives of Surgery, 10*(3), 925–934.

Overhage JM, Tierney WM, Zhou XH, McDonald CJ. A Randomized Trial of "Corollary Orders" to Prevent Errors of Omission. J Am Med Inform Assoc. 1997;4(5):364–75. PMID: 9292842.

Ozdas, A., Miller, R. A., & Waitman, L. R. (2008). Care Provider Order Entry (CPOE): A perspective on factors leading to success or to failure. *Yearbook of Medical Informatics, 2007*, 128–137. Review. Erratum in: *Yearbook of Medical Informatics*, *2008*, 19.

Parshuram, C. S., Dryden-Palmer, K., Farrell, C., et al. (2018). Effect of a pediatric early warning system on all-cause mortality in hospitalized pediatric patients: The EPOCH randomized clinical trial. *JAMA, 319*(10), 1002–1012.

Pestotnik, S. L. (2005). Expert clinical decision support systems to enhance antimicrobial stewardship programs: Insights from the Society of Infectious Diseases Pharmacists. *Pharmacotherapy, 25*(8), 1116–1125.

Phansalkar, S., Desai, A. A., Bell, D., et al. (2012a). High-priority drug-drug interactions for use in electronic health records. *Journal of the American Medical Informatics Association: JAMIA, 19*(5), 735–743.

Phansalkar, S., van der Sijs, H., Tucker, A. D., et al. (2012b). Drug-drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records. *Journal of the American Medical Informatics Association: JAMIA, 20*(3), 489–493.

Poissant, L., Pereira, J., Tamblyn, R., & Kawasumi, Y. (2005, September 1). The impact of electronic health records on time efficiency of physicians and nurses: A systematic review. *Journal of the American Medical Informatics Association, 12*(5), 505–516.

Poon, E. G., Kuperman, G. J., Fiskio, J., & Bates, D. W. (2002). Real-time notification of laboratory data requested by users through alphanumeric pagers. *Journal of the American Medical Informatics Association: JAMIA, 9*, 217–222.

Pryor, T.A. (1988) The HELP medical record system. In *MD Computing. 5*(5). (pp. 22–33). PMID: 3231033.

Przvbvlo, J. A., Wang, A., Loftus, P., et al. (2014). Smarter hospital communication: Secure smartphone text messaging improves provider satisfaction and perception of efficacy, workflow. *Journal of Hospital Medicine, 9*(9), 573–578.

Ranta, A., Dovey, S., Weatherall, M., et al. (2015). Cluster randomized controlled trial of TIA electronic decision support in primary care. *Neurology, 84*(15), 1545–1551.

Reiser, S. (1991). The clinical record in medicine. Part 1: Learning from cases. *Annals of Internal Medicine, 114*(10), 902–907.

Rief, J. J., Hamm, M. E., Zickmund, S. L., et al. (2017). Using health information technology to Foster engagement: Patients' experiences with an active patient health record. *Health Communications, 32*(3), 310–319.

Rind, D. M., Safran, C., Phillips, R. S., Wang, Q., Calkins, D. R., Delbanco, T. L., Bleich, H. L., & Slack, W. V. (1994). Effect of computer-based alerts on the treatment and outcomes of hospitalized patients. *Archives of Internal Medicine, 154*(13), 1511–1517.

Roden, D. M., Pulley, J. M., Basford, M. A., Bernard, G. R., Clayton, E. W., Balser, J. R., & Masys, D. R. (2008, September). Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clinical Pharmacology and Therapeutics, 84*(3), 362–369.

Romano, M. J., & Stafford, R. S. (2011). Electronic health records and clinical decision support systems: Impact on national ambulatory care quality. *Archives of Internal Medicine, 171*(10), 897–903.

Rosenbloom, S. T., Geissbuhler, A. J., Dupont, W. D., Giuse, D. A., Talbert, D. A., Tierney, W. M., Plummer, W. D., Stead, W. W., & Miller, R. A. (2005). Effect of CPOE user interface design on user-initiated access to educational and patient information during clinical care. *Journal of the American Medical Informatics Association: JAMIA, 12*(4), 458–473.

Samantaray, R., Njoku, V. O., Brunner, J. W., Raghavan, V., Kendall, M. L., & Shih, S. C. (2011). Promoting electronic health record adoption among small independent primary care practices. *The American Journal of Managed Care, 17*(5), 353–358.

Schedlbauer, A., Prasad, V., Mulvaney, C., Phansalkar, S., Stanton, W., Bates, D. W., & Avery, A. J. (2009). What evidence supports the use of computerized alerts and prompts to improve clinicians' prescribing behavior? *Journal of the American Medical Informatics Association: JAMIA, 16*(4), 531–538.

Schultz, J. R., Cantrill, S. V., & Morgan, K. G. (1971). *AFIPS Conference Proceedings, 38*, 239–264.

Schwartz, R. J., Weiss, K. M., & Buchanan, A. V. (1985). Error control in medical data. *MD Computing, 2*(2), 19–25.

Sherifali, D., Greb, J. L., Amirthavasar, G., Hunt, D., Haynes, R. B., Harper, W., Holbrook, A., Capes, S., Goeree, R., O'Reilly, D., Pullenayegum, E., & Gerstein, H. C. (2011). Effect of computer-generated tailored feedback on glycemic control in people with diabetes in the community: A randomized controlled trial. *Diabetes Care, 34*(8), 1794–1798.

Simborg, D. W., & Whiting-O'Keefe, Q. E. III. (1981, November 4). The Physician's Office and Group practice—ambulatory medical record systems: A automated medical record systems: Summary time oriented record (STOR)—a progress report. In *Proceedings of the annual symposium on computer application in Medical Care* (p. 100).

Simonaitis, L., Belsito, A., Warvel, J., Hui, S., & McDonald, C. J. (2006). Extensible stylesheet language formatting objects (XSL-FO): A tool to transform patient data into attractive clinical reports. In *AMIA annual symposium proceedings, 2006* (pp. 719–723).

Sinsky, C., Colligan, L., Li, L., Prgomet, M., Reynolds, S., Goeders, L., Westbrook, J., Tutty, M., & Blike, G. (2016). Allocation of physician time in ambulatory practice: A time and motion study in 4 specialties. *Annals of Internal Medicine, 2165*(11), 753–760.

Slack, W. V., & Bleich, H. L. (1999). The CCC system in two teaching hospitals: A progress report. *International Journal of Medical Informatics, 54*(3), 183–196.

Starren, J., & Johnson, S. B. (2000). An object-oriented taxonomy of medical data presentations. *Journal of the American Medical Informatics Association: JAMIA, 7*(1), 1–20.

Stead, W.W., Brame, R.G., Hammond, W.E., et. al. (1977). A computerized obstetric medical record. In *Obstetrics & Gynecology. 49*(4). (pp. 502-9). PMID 854253.

Stead, W. W., & Hammond, W. E. (1988). Computer-based medical records: The centerpiece of TMR. *MD Computing, 5*(5), 48–62.

Stein, D. M., Vawdrey, D. K., Stetson, P. D., & Bakken, S. (2010). An analysis of team checklists in physician signout notes. In *American Medical Informatics Association annual symposium proceedings* (pp. 767–771).

Strom, B. L., Schinnar, R., Aberra, F., Bilker, W., Hennessy, S., Leonard, C. E., & Pifer, E. (2010). Unintended effects of a computerized physician order entry nearly hard-stop alert to prevent a drug interaction: A randomized controlled trial. *Archives of Internal Medicine, 170*(17), 1578–1583.

Syzdykkova, A., Malta, A., Zolfo, M., et al. (2017). Open-source electronic health record systems for low-resource settings: Systematic review. *JMIR Medical Informatics, 5*(4), e44.

Tajmir, S., Raja, A. S., Ip, I. K., et al. (2017). Impact of clinical decision support on radiography for acute ankle injuries: A randomized trial. *The Western Journal of Emergency Medicine, 18*(3), 487–495.

Tang, P. C. (2003). *Key capabilities of an electronic health record system* (letter report). Committee on Data Standards for Patient Safety. Board on Health Care Services, Institute of Medicine.

Tang, P. C., & Patel, V. L. (1994). Major issues in user interface design for health professional workstations: Summary and recommendations. *International Journal of Biomedical Computing, 34*(104), 130–148.

Tang, P. C., Annevelink, J., Suermondt, H. J., & Young, C. Y. (1994a). Semantic integration in a physician's workstation. *International Journal of Bio-Medical Computing, 35*(1), 47–60.

Tang, P. C., Fafchamps, D., & Shortliffe, E. H. (1994b). Traditional medical records as a source of clinical data in the outpatient setting. In *Proceedings of the annual symposium computer applications in Medical Care* (pp. 575–579).

Tang, P. C., Marquardt, W. C., Boggs, B., et al. (1999). NetReach: Building a clinical infrastructure for the enterprise. In J. M. Overhage (Ed.), *Fourth annual proceedings of the Davies CPR recognition symposium* (pp. 25–68). Chicago: McGraw-Hill.

Tang, P. C., Ralston, M., Arrigotti, M. F., Qureshi, L., & Graham, J. (2007). Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: Implications for performance measures. *Journal of the American Medical Informatics Association: JAMIA, 14*, 10–15.

Teich, J. M., Glaser, J. P., Beckley, R. F., et al. (1999). The Brigham integrated computing system (BICS): Advanced clinical systems in an academic hospital environment. *International Journal of Medical Informatics, 54*(3), 197–208.

Teich, J. M., Merchia, P. R., Schmiz, J. L., et al. (2000). Effects of computerized physician order entry on prescribing practices. *Archives of Internal Medicine, 160*(18), 2741–2747.

Tierney, W. M., Miller, M. E., Overhage, J. M., & McDonald, C. J. (1993). Physician inpatient order writing on microcomputer workstations: Effects on resource utilization. *Journal of the American Medical Association, 269*(3), 379–383.

Tufte, E. (2006). *Beautiful evidence*. Cheshire: Graphics Press. ISBN 978-0-9613921-7-8..

Vigoda, M. M., & Lubarsky, D. A. (2006). Failure to recognize loss of incoming data in an anesthesia record-keeping system may have increased medical liability. *Anesthesia and Analgesia, 102*, 1798–1802.

Vreeman, D. J., McDonald, C. J., & Huff, S. M. (2010). Representing patient assessments in LOINC®. In *American Medical Informatics Association annual symposium proceedings, 2010* (pp. 832–836).

Wang, A.Y., Sable, J.H., & Spackman, K.A. (2002). The SNOMED clinical terms development process: refinement and analysis of content. In *American Medical Informatics Association annual symposium proceedings*, 2002 (pp. 845–9). PMCID: PMC2244575.

Warner, H. R. (1972). A computer-based patient information system for patient care. In G. A. Bekey & M. D. Schwartz (Eds.), *Hospital information systems* (pp. 293–332). New York: Marcel Dekker.

Weir, C. R., Hammond, K. W., Embi, P. J., et al. (2011). An exploration of the impact of computerized patient documentation on clinical collaboration. *International Journal of Medical Informatics, 80*(8), e62–e71.

Willett, D. L., Kannan, V., Chu, L., Buchanan, J. R., Velasco, F. T., Clark, J. D., Fish, J. S., Ortuzar, A. R., Youngblood, J. E., Bhat, D. G., & Basit, M. A. (2018). SNOMET CT concept hierarching for sharing definitions of clinical conditions using electronic health record data. *Applied Clinical Informatics, 9*(3), 667.

Wrenn, J. O., Stein, D. M., Bakken, S., & Stetson, P. D. (2010). Quantifying clinical narrative redundancy in an electronic health record. *Journal of the American Medical Informatics Association: JAMIA, 17*(1), 49–53.

Zech, J., Husk, G., Moore, T., et al. (2016). Measuring the degree of unmatched patient records in a health information exchange using exact matching. *Applied Clinical Informatics, 7*(2), 330–340.

Zhang, R., Pakhomov, S. V., Lee, J. T., & Melton, G. B. (2014). Using language models to identify relevant new information in inpatient clinical notes. In *American Medical Informatics Association annual symposium proceedings* (pp. 1268–1276).

Health Information Infrastructure

William A. Yasnoff

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_15

Learning Objectives

After reading this chapter you should know the answers to these questions:

- What is the vision and purpose of Health Information Infrastructure (HII)?
- What kinds of impacts will HII have, and over what time periods?
- How do HII requirements lead to effective architectural specifications?
- What are the political and technical barriers to HII implementation?
- How can HII progress be effectively evaluated?

15.1 Introduction

This chapter addresses **health information infrastructure (HII)**, community level informatics systems designed to make comprehensive electronic patient records available when and where needed for the entire population. There are numerous difficult and highly interdependent challenges that HII systems must overcome, including privacy, stakeholder cooperation, assuring all-digital information, and providing financial sustainability. As a result, while HII has been pursued for years with myriad approaches in many countries, progress has been slow and no proven formula for success has yet been identified.

While the discussion here is focused on the development of the HII in the United States, many other countries are involved in similar activities and in fact have progressed further along this road. Canada, Australia, and a number of European nations have devoted considerable time and resources to their own national HIIs. A few countries, such as Finland, Estonia, and Brazil have actually succeeded in developing effective HII systems that have been working nationwide for a number of years. It should be noted, however, that all of these nations have centralized, government-controlled healthcare systems. This organizational difference from the multifaceted, mainly private healthcare system in the U.S. results in a somewhat different set of issues and problems. One can hope that the lessons learned from HII development activities across the globe can be effectively shared to ease the difficulties of everyone who is working toward these important goals.

HII at first seems like a vague term – what does it really mean? This is not a trivial question – as with all information systems, if we don't understand clearly what we are trying to accomplish, as well as how we will measure whether we've achieved our goals, success will be elusive. The overall goal may be stated as "comprehensive electronic patient information when and where needed." This includes both immediate access to comprehensive records for individuals (for care) and the ability to search and aggregate information across the population (for public health, medical research, quality improvement, and policy).

We know that patients in hospitals always have a unified chart (be it paper or electronic) that contains all their hospital records from all sources. However, there is no equivalent "outpatient chart" with comprehensive records from all providers in a single place. The lack of this information is a serious problem: a survey of doctor visits in 2015 found that 55% of patients reported that their medical history was missing or incomplete, while 49% indicated that their physician was not aware of which prescription medications they were taking.[1] Naturally, the result of this lack of information is undertreatment, overtreatment, and medical errors.

15.2 Vision & Benefits of HII

The vision of HII is comprehensive electronic patient information when and where needed, allowing providers to have complete and current information upon which to base clinical decisions. In addition, clinical decision sup-

1 Surescripts Survey Finds Patients Prefer Digitally Savvy Doctors and Demand a Connected Healthcare Experience. Released 28 Sept 2015. Retrieval 28 Aug 2018: ▶ https://surescripts.com/news-center/press-releases/!content/surescripts-survey-finds-patients-prefer-digitally-savvy-doctors-and-demand-a-connected-healthcare-experience

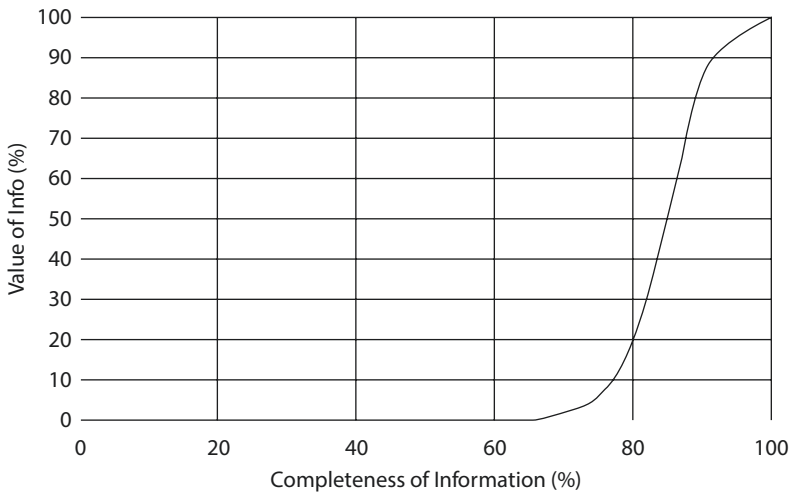
port (see ▶ Chap. 22) would be integrated with information delivery. In this way, both clinicians and patients could receive reminders of the most recent **clinical guidelines** and research results. This would avoid the need for clinicians to have superhuman memory capabilities to assure the effective practice of medicine, and enable patients more easily to adhere to complex treatment protocols and to be better informed. Patients could also review and add information to their record and thereby become more active participants in their care. In addition, the availability of comprehensive records for each patient would enable value-added services, such as immediate electronic notifications to patients' family members about emergency care, as well as authorized queries in support of medical research, public health, and public policy decisions.

15.2.1 Value Versus Completeness of Information

In considering HII, it is extremely important to appreciate that medical information for a given patient must, in general, be relatively complete before it is truly valuable for clinical use (see ◘ Fig. 15.1). For example, if a physician had access to an electronic information system that could retrieve half of each patient's list of medications, it is unlikely such a system would be actively used. Knowing that the information was incomplete, the physician would still need to rely on other traditional sources of information to fill in the missing data (including questioning the patient). So there would be little added benefit for investing the time to obtain the partial information from the new system. Similarly, applying clinical decision support to incomplete patient data may produce erroneous, misleading, or even potentially dangerous results. Therefore, HII systems must reliably provide reasonably complete information to be valuable to clinicians for patient care, and therefore to make their use worthwhile.

Besides their limited value, incomplete records are potentially dangerous. The conclusions that providers draw from a partial picture of a patient's history may often prove to be incorrect. For example, missing contraindications could result in the prescription of a medication with serious adverse effects. Incomplete records also are a source of unnecessary costs. When test and/or procedure results are not available, but are needed for care, they are likely to be repeated.

Because of the above factors, the cost of obtaining incomplete records is not typically accompanied by substantial benefits. As a result, organizations compiling such records find themselves under great financial stress. Many such "health information exchange"



◘ **Fig. 15.1** Estimated value vs. completeness of health information. Medical information of any given type for a patient typically needs to be over 85% complete before it starts being truly valuable to clinicians

(HIE) organizations have failed in the past few years, e.g., Washington (DC), Kansas, Tennessee, CalRHIO, and CareSpark (Kingsport, TN).

Importantly, today no patients in the U.S. can be assured that wherever they seek care, their comprehensive records from all sources will be available to their provider.

The U.S. Congress has recognized the importance of comprehensive records and has mandated that the problem be solved in the 21st Century Cures Act, enacted in late 2016:

- *"The Secretary shall use existing authorities to encourage partnerships … with the goal of offering patients access to their electronic health information in a single, longitudinal format that is easy to understand, secure, and may be updated automatically."* […and …]
- *"… promote policies that ensure that a patient's electronic health information is accessible to that patient and the patient's designees, in a manner that facilitates communication with the patient's health care providers and other individuals, including researchers, consistent with such patient's consent."*[2]

The Final Rule implementing the 21st Century Cures Act,[3] reinforces these goals by requiring application programming interfaces (APIs) that allow all data elements of a patient's electronic health record to be accessed, exchanged, and used without special effort and a more granular approach to consent management. It also requires that patients be allowed to access all of their electronic health record information at no cost and prohibits "information blocking" in response to legitimate requests for patient records.

15.2.2 Value in Patient Care

The potential benefits of HII are both numerous and substantial. Perhaps most important are error reduction and improved quality of care. Many studies have shown that the complexity of present-day medical care results in very frequent errors of both omission and commission (IOM 1999). The source of this problem was clearly articulated by Masys, who observed that current medical practice depends upon the "clinical decision-making capacity and reliability of autonomous practitioners for classes of problems that routinely exceed the bounds of unaided human cognition" (Masys 2002). Electronic health information systems can contribute significantly to alleviating this problem by reminding practitioners about recommended actions at the point of care. This can include both notifications of actions that may have been missed and warnings about planned treatments or procedures that may be harmful or unnecessary. Literally dozens of research studies have shown that such reminders improve safety and reduce costs (Bates 2000). In one such study, medication errors were reduced by 55% (Bates et al. 1998). Another study by the Rand Corporation showed that only 55% of U.S. adults were receiving recommended care (McGlynn et al. 2003). The same techniques used to reduce medical errors with electronic health information systems also contribute substantially to ensuring that recommended care is provided. This is becoming increasingly important as the population ages and the prevalence of chronic disease increases.

Guidelines and reminders also can improve the effectiveness of dissemination of new research results. Widespread dissemination of new research to the clinical setting is very slow; one study showed an average of 17 years (Balas and Boren 2000). Patient-specific reminders delivered at the point of care, high-

2 21st Century Cures Act, H.R.34 — 114th Congress (2015–2016), Section 4006. Retrieval 29 Oct 2018: ▶ https://www.congress.gov/bill/114th-congress/house-bill/34

3 21st Century Cures Act: Interoperability, Information Blocking and the ONC Health IT Certification Program. Released May 1, 2020. Retrieval 23 May 2020: ▶ https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification

lighting important new research results, could substantially accelerate this adoption rate.

Another important contribution of HII to the research domain is improving the efficiency of clinical trials. At present, most such trials require the creation of a unique information infrastructure to ensure protocol compliance and to collect essential research data. With an effective HII, every practitioner would have access to a fully functional and comprehensive **electronic health record (EHR)** for each patient, so clinical trials could routinely be implemented through the dissemination of guidelines that specify the research protocol. Data collection could occur automatically in the course of administering the protocol, reducing time and costs. In addition, there would be substantial value in analyzing **de-identified aggregate data** from routine patient care to assess the outcomes of various treatments and monitor the health of the population.

Another critical function for HII is early detection of patterns of disease, particularly early detection of outbreaks from newly-virulent microorganisms or possible bioterrorism. Our current system of disease **surveillance**, which primarily depends on alert clinicians diagnosing and reporting unusual conditions, is both slow and potentially unreliable. These problems are illustrated by delayed detection of the anthrax attacks in the Fall of 2001, when seven cases of cutaneous anthrax in the New York City area 2 weeks before the so-called "index" case in Florida went unreported (Lipton and Johnson 2001). Since all the patients were seen by different clinicians, the pattern could not have been evident to any of them even if the correct diagnosis had immediately been made in every case. Wagner et al. described nine categories of requirements for surveillance systems for potential bioterrorism outbreaks—several categories must have immediate electronic reporting to ensure early detection (Wagner et al. 2003).

HII would allow immediate electronic reporting of both relevant clinical events and laboratory results to public health (see ▶ Chap. 18). Not only would this be an invaluable aid in early detection of bioterrorism, it would also serve to improve the detection of the much more common naturally occurring disease outbreaks. In fact, early results from a number of electronic reporting demonstration projects show that disease outbreaks can routinely be detected sooner than was ever possible using the current system (Overhage et al. 2001). While early detection has been shown to be a key factor in reducing morbidity and mortality from bioterrorism (Kaufmann et al. 1997), it will also be extremely helpful in reducing the negative consequences from other disease outbreaks.

Although the U.S. Congress mandated the creation of a national public health situational awareness network in the Pandemic and All-Hazards Preparedness Acts of both 2006 and 2013, the General Accounting Office has documented that such a system has yet to be deployed in two separate reports[4,5] and has issued letters in both 2019 and 2020[6] highlighting the failure of DHHS to implement the required capabilities. HII can in fact function as an effective public health situational awareness network by reporting and/or providing access to relevant disease events in near real time to public health authorities without the necessity of creating a separate, duplicative, and expensive system solely for public health (Sittig and Singh 2020). For example,

4 General Accounting Office Report 11-99 (2010) Public Health information Technology: Additional Strategic Planning Needed to Guide HHS's Efforts to Establish Situational Awareness Capabilities. Retrieval 23 May 2020: ▶ https://www.gao.gov/products/GAO-11-99

5 General Accounting Office Report 17-377 (2017) Public Health Information Technology: HHS Has Made Little Progress toward Implementing Enhanced Situational Awareness Network Capabilities. Retrieval 23 May 2020: ▶ https://www.gao.gov/products/GAO-17-377

6 General Accounting Office (2020) Priority Open Recommendations: Department of Health and Human Services (April 23, 2020). Retrieval 23 May 2020: ▶ https://www.gao.gov/assets/710/706568.pdf

imagine how valuable it would be in tracking the Covid-19 pandemic if timely information on patients reporting relevant symptoms to their physicians were available nationwide on a daily basis.

Finally, HII can substantially reduce healthcare costs. The inefficiencies and duplication in our present paper-based healthcare system are enormous. One study showed that the anticipated nationwide savings from implementing advanced **computerized physician order entry (CPOE)** systems in the outpatient environment would be $44 billion per year (Johnston et al. 2003), while a related study (Walker et al. 2004) estimated $78 billion more in savings from **health information exchange (HIE)** (for a total of $112 billion per year). Substantial additional savings are possible in the inpatient setting—numerous hospitals have reported large net savings from implementation of EHRs. Another analysis concluded that the total efficiency and patient safety savings from HII would be in range of $142–371 billion each year (Hillestad et al. 2005), and a survey of the recent literature found predominantly positive benefits from HII (Menachemi et al. 2018). It is important to note that much of the savings depends not just on the widespread implementation of EHRs, but the effective interchange of this information to ensure that the *complete* medical record for every patient is immediately available in every care setting.

Inasmuch as the current cost trend of healthcare is unsustainable, particularly in the face of our aging population, this issue is both important and urgent. Without comprehensive electronic patient information, any healthcare reform is largely guesswork in our current "black box" healthcare environment where the results of interventions often take years to understand. We do not currently have mechanisms for timely monitoring of healthcare outcomes to inform needed course corrections in any proposed reform. In essence, healthcare must be "informed" before it can be effectively "reformed."

15.3 History

In the U.S., the first major report to address HII was issued in 1991 by the Institute of Medicine (now known, since 2015, as the National Academy of Medicine, a part of the National Academies of Sciences, Engineering, and Medicine or NASEM). This report, "The Computer-Based Patient Record" (IOM 1991), was the first in a series of national expert panel reports recommending transformation of the healthcare system from reliance on paper to electronic information management (see ▶ Chap. 14). In response to the IOM report, the Computer-based Patient Record Institute (CPRI), a private not-for-profit corporation, was formed for the purpose of facilitating the transition to computer-based records. A number of **community health information networks (CHINs)** were established around the country in an effort to coalesce the multiple community stakeholders in common efforts towards electronic information exchange. The Institute of Medicine updated its original report in 1997 (IOM 1997), again emphasizing the urgency to apply information technology to the information intensive field of health care.

However, most of the CHINs were not successful. Perhaps the primary reason for this was that the standards and technology were not yet ready for cost-effective community-based electronic HIE. Another problem was the focus on availability of aggregated health information for secondary uses (e.g., policy development), rather than individual information for the direct provision of patient care. Also, there was neither a sense of extreme urgency nor were there substantial funds available to pursue these endeavors. However, at least one community (Indianapolis, Indiana) continued to move forward throughout this period and has now emerged as a national example of the application of information technology to health care both in individual healthcare settings and throughout the community (McDonald et al. 2005).

Widespread attention was focused on this issue with the IOM report "To Err is Human" (IOM 1999). This landmark study documented the accumulated evidence of the high error rate in the medical care system, including an estimated 44,000–98,000 preventable deaths each year in hospitals alone. It has proven to be a milestone in terms of public awareness of the negative consequences of paper-based information management in healthcare. Along with the follow-up report, "Crossing the Quality Chasm" (IOM 2001), the systematic inability of the healthcare system to operate at a high degree of reliability has been thoroughly elucidated. A more recent analysis estimated a much larger number of preventable deaths due to medical errors – over 400,000 (Makary and Daniel 2016). These reports clearly place the blame on the system, not on the dedicated healthcare professionals who work in an environment without effective tools to promote quality and to minimize errors.

Several additional national expert panel reports have emphasized the IOM findings. In 2001, the President's Information Technology Advisory Committee (PITAC) issued a report entitled "Transforming Health Care Through Information Technology" (PITAC 2001). That same year, the Computer Science and Telecommunications Board of the National Research Council (NRC) released "Networking Health: Prescriptions for the Internet" (NRC 2001), which emphasized the potential for using the Internet to improve electronic exchange of healthcare information. That same year, the National Committee on Vital and Health Statistics (NCVHS) outlined a vision for building a National HII in its report, "Information for Health" (NCVHS 2001). NCVHS, a statutory advisory body to the U.S. Department of Health and Human Services (DHHS), indicated that Federal government leadership was needed to facilitate further development of HII. In response, DHHS began an HII initiative, organizing a large national conference in 2003 to develop a consensus agenda to guide progress.[7] (Yasnoff et al. 2004)

In April, 2004, a Presidential Executive Order created the Office of the National Coordinator for Health Information Technology (ONC) in DHHS (see also ▶ Chap. 29). The initial efforts of ONC focused on promoting standards and certification to support adoption of EHRs by physicians and hospitals. It also promoted implementation of an "institution centric" model for HIE by **Regional Health Information Organizations (RHIOs)**, wherein electronic records for a given patient stored at sites of past care episodes are located, assembled, and delivered in real time when needed for patient care. Four demonstration projects implementing this model were funded, but did not lead to sustainable systems.

In 2008, ONC was codified in law by the Health Information Technology for Economic and Clinical Health (HITECH) portion of the ARRA statute (▶ Chap. 29). In addition, $20+ billion was appropriated including $2 billion for ONC and the remainder for payment of EHR incentives through Medicare and Medicaid to providers who achieved "Meaningful Use" of these systems. The ONC used its resources to establish **regional extension centers (RECs)** to subsidize assistance to providers adopting and using EHRs ($677 million), fund states to establish HIEs ($564 million) and initiate several research programs.

In December, 2010, the President's Council of Advisors on Science and Technology (PCAST) issued a report expressing concern about ONC strategy, specifically indicating that its HIE efforts through the states "*will not solve the fundamental need for data to be universally accessed, integrated, and understood while also being protected*" (PCAST 2010).

7 Department of Health and Human Services. (2003) *The National Health Information Infrastructure*. Retrieval 29 Oct 2018: ▶ http://aspe.hhs.gov/sp/nhii/

Findings of a 2011 survey of HIEs "*call into question whether RHIOs in their current form can be self-sustaining and effective*" (Adler-Milstein et al. 2011).

Over the past few years, these concerns have been proven correct; HIE efforts have largely floundered. In the 2013 "reboot" report, a group of U.S. Senators criticized lack of a clear path to interoperability and sustainability, inadequate privacy and security protections, and failure to achieve health care cost reductions.[8] That same year, DHHS itself admitted that "[current policy] alone will not be enough to achieve the widespread interoperability and electronic exchange of information necessary for delivery reform where information will routinely follow the patient regardless of where they receive care" in an RFI requesting ideas for "accelerating HIE progress."[9] At least one experienced former DHHS official bemoaned the lack of an HII architecture to guide efforts.[10] Even the mainstream media noticed the continuing lack of progress (Creswell 2014).

A systematic review found that HIE use through 2015 in the U.S. was "growing but still limited" (Devine et al. 2017). The official evaluation of ONC's state HIE program tried to put a positive spin on these activities with the statistically meaningless conclusion that "about half the states [were] performing at or above the national average," a result that by definition would always be the case. The report also observed that "sustainability was a persistent concern among grantees."[11] The latter challenge has since resulted in the shutdown of multiple projects.

Although those few surviving HIEs have nearly all abandoned their original institution-centric architectures in favor of patient-centric repositories (more on HII architecture in ▶ Sect. 15.5 below), the predicted reductions in health care costs from the widespread use of EHRs have not materialized primarily due to the incomplete nature of patient records available through current HIE systems. A 2017 study observed "To date, only a small number of HIE studies have demonstrated benefits to patients, providers, public health, or payers" (Yeager et al. 2017).

Overall, it is clear that three decades after the 1991 IOM report urging universal adoption of EHRs, the U.S. still lacks a clear and feasible roadmap leading to the ultimate goal of widespread availability of comprehensive electronic patient information when and where needed. Despite much progress, no one in the U.S. as yet receives their medical care with the assured, immediate availability of all their records across multiple providers and provider organizations.

15.4 Requirements for HII

As with any informatics system development project, it is critical at the outset to understand the desired end result. In the case of a large, extremely complex system such as HII, this is especially important because there are many stakeholders with conflicting incentives and agendas, as well as challenging policy and operational issues. The ultimate goal is the "universal availability of comprehensive electronic patient records when and where needed." In transforming this goal into a design specification, it is critical to understand the issues

15

8 Reboot: Re-examining the strategies needed to successfully adopt health IT. Retrieval 29 Oct 2018: ▶ https://www.thune.senate.gov/public/_cache/files/0cf0490e-76af-4934-b534-83f5613c7370/C60F25439BE1CEC36DF9E3834942D908.ehr-white-paper-april-15.pdf

9 CMS. Advancing Interoperability and Health Information Exchange (Request for Information). Retrieval 29 Oct 2018: ▶ https://www.federalregister.gov/documents/2013/03/07/2013-05266/advancing-interoperability-and-health-information-exchange

10 Loonsk JW. Where's the plan for interoperability? Healthcare IT News (22 Sept 2014). Retrieval 29 Oct 2018: ▶ https://www.healthcareitnews.com/news/wheres-plan-interoperability

11 Dullabh PM, Parashuram S, Hovey L, Ubri P, Fischer K. (2016) Evaluation of the State HIE Cooperative Agreement Program: Final Report. Retrieval 29 Oct 2018: ▶ https://www.healthit.gov/sites/default/files/reports/finalsummativereportmarch_2016.pdf

and constraints that must be addressed. Then any proposed system design must first demonstrate on paper how the objectives will be achieved within those limitations.

15.4.1 Privacy and Trust

The most important and overriding requirement of HII is privacy. Clearly, health records are very sensitive – perhaps the most sensitive personal information that exists. In addition to our natural desire to keep our medical information private, improper disclosure can lead to employment and other types of discrimination. Furthermore, failure to assure the privacy of records will naturally result in patients being unwilling to disclose important personal details to their providers – or even to avoid seeking care at all. In addition to the contents of the records, the very existence of certain records (e.g., a visit to psychiatric hospital) is sensitive even if no details are available. Therefore, extraordinary care must be taken to ensure that information is protected from unauthorized disclosure and use.

In general, U.S. Federal law (the HIPAA Privacy Rule as introduced in ▶ Chap. 12) requires patient consent for disclosure and use of medical records. However, consent is not required for record release for treatment, payment, and healthcare operations. These "TPO" exceptions have, as a practical matter, allowed healthcare organizations to utilize medical records extensively while bypassing patient consent. The organization that holds medical information has sole discretion to make the decision whether a proposed disclosure is or is not a TPO exception. Until recently, TPO disclosures did not even need to be recorded, effectively preventing discovery of improper disclosures. Even under the HITECH legislation that requires records of TPO disclosures, such records are not automatically available to the subjects of the disclosures. The net effect is that individuals not only lack control over the dissemination of their medical records, but are not even informed when they are disclosed beyond where they were created.

It seems appropriate to question whether this disclosure regime is adequate for electronic health records. The general public understands that making electronic patient records available for good and laudable purposes simultaneously makes them more available for evil and nefarious purposes, thereby necessitating higher levels of protection to avoid abuses. Assigning decision-making for disclosure of personal medical records to anyone other than the patient or the patient's representative inherently erodes trust. In essence, the patient is being told, "we are going to decide for you where your medical records should go because we know what's in your interest better than you do." A patient may wonder why, if a given disclosure is in their interest, their consent would not be sought. Furthermore, failing to seek such consent inevitably leads to suspicion that the disclosure is in fact not in the patient's interest, but rather in the interest of the organization deciding that the records should be released.

The concern about privacy of medical records is not at all theoretical or insignificant. In at least two consumer surveys, 13–17% of consumers indicated that they already employ "information hiding" behaviors with respect to their medical records[12] (Shortliffe 2011). This includes activities such as obtaining laboratory tests under an assumed name or seeking out-of-state treatment to conceal an illness from their primary care provider. Even assuming that everyone engaged in such behaviors was willing to admit to them in such a survey, this represents a substantial proportion of consumers who would, at a minimum, refuse to participate in an electronic medical information system that did not provide them with control over their own records. Of even greater concern, such a large percentage of consumers would likely organize and use their political power to halt the deployment and operation of such a system. Indeed, it was a much smaller percentage of concerned citizens that, citing the threat to privacy, convinced Congress to repeal the pro-

12 California Health Care Foundation. (2005) *National Consumer Health Privacy Survey*. Retrieval 29 Oct 2018: ▶ http://www.chcf.org/publications/2005/11/national-consumer-health-privacy-survey-2005

visions in the original HIPAA legislation calling for a unique medical identifier for all U.S. residents (see ► Chap. 12).

In view of this, there are those who argue that all decisions about release of patient records need to be entrusted to the patient (with rare exceptions, such as mental incompetence). They also suggest that attention to these concerns may be especially important for enabling HII, because patients must trust that their records are not being misused in such a system. Some argue that patients are not sufficiently informed to make such decisions and may make mistakes that are harmful to them, whereas others believe that the negative consequences of delegating this decision-making to others than the patient could be much greater. Advocates of patient control of medical information argue, by analogy, that society has accepted that individuals retain the right to make decisions about how their own money is spent, even though this can lead to adverse consequences when those decisions prove to be unwise. In considering these issues, it should be noted that prior to the 2002 HIPAA Privacy Rule that established the TPO exceptions, both law and practice had always required patient consent for all access to medical records. While acknowledging the need for consumer education about decisions relating to release of medical records, patient-control advocates believe that the same freedom and personal responsibility that applies to an individual's financial decisions should be applied to the medical records domain. These medical information privacy policy issues may be even more urgent in the context of the enhanced trust necessary when seeking to implement an effective and widely accepted HII.

15

15.4.2 Stakeholder Cooperation

To ensure the availability of comprehensive patient records, all healthcare stakeholders that generate such records must consistently make them available. While it would be ideal if such cooperation were voluntary and universal, assuring long-term collaboration of competing healthcare stakeholders is problematic. Indeed, only a handful of communities have succeeded in developing and maintaining an organization that includes the active participation of the majority of healthcare providers. Even in these communities, the system could be disrupted at any time by the arbitrary withdrawal of one or more participants. The unfortunate reality is that healthcare stakeholders are often quite reluctant to share patient records, fearing loss of competitive advantage.

Therefore, some would argue that mandating healthcare stakeholder participation in a system for sharing electronic patient records is highly desirable, since it would result in consistently more comprehensive individual records. Since imposing a new requirement on healthcare stakeholders would be a daunting political challenge, such an approach would be most easily accomplished as part of an existing mandate. Proponents of this approach have noted that one such mandate that could be utilized is the HIPAA Privacy Rule itself, which requires all providers to respond to patient requests for their own records (U.S. 45 CFR 164.524(a)). Furthermore, if patients request their records in electronic form, and they are available in electronic form, this regulation also requires that they be delivered in electronic form. Although not well known, this latter provision is included in the original HIPAA Privacy Rule (U.S. 45 CFR 164.524(c)(2)), and has been reinforced by HITECH. It is also being promoted by the "blue button" initiative that seeks to allow patients to retrieve their own records electronically[13] and the growing movement by patients advocating guaranteed access to their own data.[14]

Advocates argue that patient control, in addition to being an effective approach to pri-

13 Chopra A, Park T, Levin PL. (2010) '*Blue Button*' *Provides Access to Downloadable Personal Health Data*. Retrieval 29 Oct 2018: ► http://www.whitehouse.gov/blog/2010/10/07/blue-button-provides-access-downloadable-personal-health-data

14 Miliard M. (2014) Patients want online access to records. Healthcare IT News (5 May 2014). Retrieval 29 Oct 2018: ► https://www.healthcareitnews.com/news/patients-want-online-access-records

vacy, could also serve to ensure ongoing, consistent healthcare stakeholder participation. Of course, in order for this approach to be practical, the rights of patients to electronic copies of their records under HIPAA would need to be universally enforced. Such enforcement has to date been inconsistent, and, until recently, exclusively dependent on the Office of Civil Rights at DHHS (since patients do not have a private right of action). Under HITECH, state attorneys general may also bring legal action, which provides another legal avenue for improving compliance.

Another option to ensure that providers make their electronic records available to patients for use in compiling comprehensive records would be to create a linkage to reimbursement for care. In this scenario, each provider would be required to offer to deposit the new information generated from an encounter in a place of the patient's choice. In cases where the patients designated a destination for their information, payment for the care received would be contingent on the deposit of the required data. While this may seem somewhat coercive and even radical, it is consistent with practices in other service industries. Whether it be car repair, plumbing, or legal services, payment is nearly always contingent on the client receiving detailed justifications and descriptions of the services provided. While a growing number of health care providers are increasingly providing such "visit summaries," at least in paper form, this remains the exception in the medical domain.

15.4.3 Ensuring Information in Standard Electronic Form

It is self-evident that the electronic exchange of health information cannot occur if the information itself is not in electronic form. Over the past decade, nearly all U.S. hospitals and most U.S. physicians have adopted EHRs as a result of the Federal government's "Meaningful Use" financial incentives.

However, the major obstacle for physician adoption of EHRs has not been merely cost, as is often cited, but the very unfavorable ongoing cost/benefit ratio. Most of the benefits of EHRs in physician offices accrue not to the physician, but to other stakeholders. In one study, 89% of the economic benefit was attributed to other stakeholders (Hersh 2004). It is unreasonable to expect physicians to shoulder 100% of the cost of systems while accruing only 11% of the benefits. Even as physician EHR adoption levels have increased, there have been increasing complaints about the burden that EHRs impose on physicians (Sinsky et al. 2016).

It is important to note that EHRs alone, even if adopted by all healthcare providers, are a necessary but not sufficient condition for achieving HII. Indeed, each EHR simply converts an existing paper "silo" of information to electronic form. These provider-based systems manage the *provider* information on the patient in question, but do not have *all* the information for each patient. To achieve the goal of availability of comprehensive patient information, there must also be an efficient and cost-effective mechanism to aggregate the scattered records of each patient from all their various providers. Such aggregation also requires effective standards for encoding and communicating EHR information. Major gains in quality and efficiency of care will be attainable only through HII that ensures the availability of every patient's comprehensive record when and where needed.

15.4.4 Financial Sustainability

There are three fundamental approaches that can be used individually or in combination to provide long-term financial sustainability for HII: (1) public subsidy; (2) leveraging anticipated future healthcare cost savings; and/or (3) leveraging new value created. The first approach has been advocated by those who assert, with some justification, that HII represents a public good that benefits everyone. They compare HII to other publicly available infrastructure, such as roads, and suggest that taxation is an appropriate funding mechanism. Of course, new taxes are consistently unpopular and politically undesirable, and

other key infrastructures such as public utilities and the Internet, although regulated, are funded through user fees rather than taxation. Note, however, that at least two U.S. states (Maryland and Vermont) are using this mechanism to help fund their HII.

The most common approach suggested for long-term HII sustainability is leveraging anticipated healthcare cost savings. This is based on the substantial and growing body of evidence that the availability of more comprehensive electronic patient records to providers results in higher quality and lower cost care (AHRQ 2006; Menachemi et al. 2018). Some of the best examples include large, mostly closed healthcare systems such as Kaiser, Group Health and the Veterans Administration, where the availability of more complete patient records in electronic form over time has been consistently associated with both cost savings and better care. While the case for HII reducing healthcare costs is compelling, the distribution and timing of those savings is difficult to predict. In addition, cost savings to the healthcare system means revenue losses to one or more stakeholders – clearly an undesirable result from their perspective. Finally, the allocation of savings for a given population of patients is unknown, with the result that organizations are reluctant to make specific financial commitments that could be larger than their own expected benefits.

The final but least frequently mentioned path to financial sustainability of HII is utilizing the new value created by the availability of comprehensive electronic information. While it is widely recognized that this information will be extremely valuable for a wide variety of purposes, this option has remained largely unexplored. One example of such new value is the potential reduction in cost for delivering laboratory results to ordering physicians. The expenses borne by individual laboratories for their own infrastructure providing this essential service can be greatly reduced by a single uniform community infrastructure providing electronic delivery to physicians through one mechanism. Another example is availability of medical information for research – both to find eligible subjects for clinical trials and to utilize the data itself for research queries. While this latter application has the potential to defray a substantial portion of the costs of HII, it requires efficient mechanisms for both searching data and recording and maintaining patient consent that have not generally been incorporated into HII systems.

Perhaps the most lucrative HII revenue source lies in the development of innovative applications that rely on the underlying information to deliver compelling value to consumers and other healthcare stakeholders. For example, HII allows the delivery of timely and accurate reminders and alerts to patients for recommended preventive services, needed medication refills, and other medically related events of immediate interest to patients and their families. It also would allow deployment of applications that assist consumers automatically with management of their chronic diseases. Utilizing new value to finance HII avoids the prediction and allocation problems inherent in attempts to leverage expected healthcare cost savings, with the added incentive that any such savings would fully accrue to whoever achieves them.

15.4.5 Community Focus

Most observers believe that successful HII must be focused on the community. An essential element in HII is trust, which is inherently local. Furthermore, health care itself is predominantly local, since the vast majority of medical care for residents of a given community is provided in that community. Indeed, people traveling away from home who are injured or become ill inevitably will return home at their earliest opportunity if their condition permits (and does not resolve quickly). Since medical care is predominantly local, creating a system that delivers comprehensive electronic patient information in a community solves the overwhelming majority of information needs in that community. While move-

ment of health information over long distances has some value and ultimately must be addressed to assure completeness of records, its contribution to a total solution is marginal.

The lack of any examples of working HII in communities larger than about 10 million people provides additional evidence of the need for local focus. Keeping the scope of such projects relatively small also increases their likelihood of success by reducing complexity, thereby avoiding the huge increases in failure rates of extremely large-scale IT projects. This rule of thumb is reinforced by the relatively small populations of countries that have successfully implemented effective HII such as Finland (5.5 million) and Estonia (1.3 million).

In thinking about HII, analogies are often made to the international financial system that efficiently transfers and makes funds available to individuals anywhere in the world. However, it is often forgotten that these financial institutions, that also are heavily dependent on trust, began as "building and loan funds" in very small communities designed to share financial resources among close neighbors. It took many decades of building trust before large-scale national and international financial institutions emerged.

15.4.6 Governance and Organizational Issues

Trust is arguably the most important element in considering the appropriate governance for HII. Even in a system where patients exert full control over their own records, the organization that operates the HII must earn the full faith and confidence of consumers for the security, integrity, and protection of the records, as well as ensuring that records are appropriately available only for purposes that consumers specify. Furthermore, the organization ideally must be devoid of any biases or hidden agendas that would favor one category of healthcare stakeholders over another, or favor specific stakeholders within a given category.

None of the existing healthcare stakeholders seem well suited to meet the trust requirement. Many argue that government cannot operate an HII because it is inherently not trusted with sensitive personal records, and furthermore needs to assume the role of providing regulatory oversight for whatever organization does take the HII responsibility. Similarly, it seems problematic for employers to be responsible for the HII since one of the primary concerns of consumers is to avoid disclosing sensitive medical information to their employers. Health plans and insurers are typically not trusted by consumers because their incentives are not aligned – they have a financial incentive to deny care, which is a natural concern to consumers. Hospitals are in competition with each other and therefore are not in a good position to cooperate in a long-term HII effort. Physicians are the most trusted healthcare stakeholders, from a consumer perspective, but are not organized in a way to facilitate the creation of HII. Furthermore, they are also in competition with each other and, most importantly, do not generally have the informatics capabilities necessary for such a complex endeavor.

Therefore, many believe that an independent (perhaps entirely new) organization is needed to operate HII in communities. This organization would have no direct connections to existing healthcare stakeholders and therefore would be unbiased. Its sole function would be to protect and make available comprehensive electronic patient records on behalf of consumers. Such an independent organization would also ideally facilitate cooperation among all existing stakeholders, who would know that the HII activity was completely neutral and designed primarily to serve consumers.

15.5 Architecture for HII

15.5.1 Institution-Centric Architecture

Initially, most developing HII systems chose an institution-centric approach to data storage, leaving patient records wherever they are created (Fig. 15.2). Although records are not stored centrally in this model, there is a need to maintain at least a central index of where information can be found for a particular patient; without such an index, finding information about each patient would require queries to every possible source of medical information worldwide -- clearly an impractical approach. When a given patient's record is requested, the index is used to generate queries to the locations where information is stored. The responses to those queries are then aggregated (in real time) to produce the patient's complete record. After the patient encounter, the new data is entered into the clinician's EHR system and another pointer (to that system) is added to the index so it will be queried (in addition to all the other prior locations) next time that patient's record is requested.

While this architecture appeals to healthcare stakeholders because they continue to "control" the records they generate, one can argue that it fails to meet several key requirements, does not scale effectively, and is complex and expensive to operate. The most critical requirement that is not addressed by this architecture is searching the data, e.g., to find all patients with a cholesterol level above 300. To do such a search, the records of every patient must be assembled from their various locations and examined one at a time. Known as a sequential search, it has a very long completion time that increases linearly with the size of the population. For example, in a modest-sized HIE with 500,000 patients, assuming retrieval and processing of each patient's records requires just 2 seconds (a very low estimate), each such search would take at least 12 days (1 million seconds). Furthermore, every such search would require that each provider record system connected to the HIE retrieve and transmit all its information – a very substantial computing and communications burden (that also increases the risk of interception of information). In standard database systems, impractical sequential search times are reduced by pre-indexing the contents of the records. However, such pre-indexing would in essence create a central repository of indices that could be used to reconstruct most of the original data, and therefore is inconsistent with this architectural approach.

It may be argued that the searches could themselves be distributed to the provider systems, and then the results aggregated into a

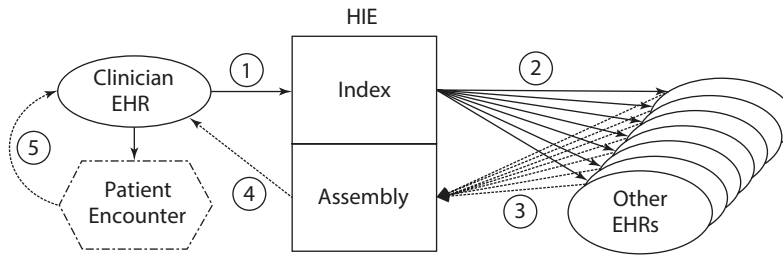


Fig. 15.2 Institution-centric HII architecture. (1) The clinician EHR requests prior patient records from the HIE; this clinician's EHR is added to the index for future queries for this patient (if not already present). (2) Queries are sent to EHRs at all sites of prior care recorded in the HIE Index. (3) EHRs at each prior site of care return records for that patient to the HIE; the HIE must wait for all responses. (4) The returned records are assembled and sent to the clinician EHR; any inconsistencies or incompatibilities between records must be resolved in real time. (5) After the care episode, the new information is stored in the clinician EHR only. (Used with permission of Health Record Banking Alliance (HRBA))

coherent result. However, this approach also fails because individual patient records are incomplete in each system. Therefore, searches that require multiple items of patient data (e.g., patients with chest pain who have taken a certain medication in the past year), will produce anomalous results unless all the instances of the relevant data for a given patient are in a single provider system (i.e., if one system finds a patient with chest pain, but without any indication of the medication of interest [which is in another provider's system], that patient will not be reported as satisfying the conditions) (Weber 2013). It is possible to launch multiple searches each limited to a single criterion and then combine the results from each to generate a correct result. However, this would require multiples of the completion time for a single criterion (e.g., 12 days × 2 = 24 days for the two criteria example), making the retrieval times and processing burdens even more untenable.

In addition to the scaling issues for this architecture related to searching, there is also a problem with response time for assembling a patient record. When a given patient record is requested, the locations where the patient has available records are found using the central index. Then, a **query-response cycle** is required for each location where patient records are available. Following completion of the query-response cycles, all the information obtained must be integrated into a comprehensive record and made available to the requestor. While the query-response cycles can all be done in parallel, the final integration of results must wait for the slowest response. As the number of connected systems increases, so does the probability of a slow (or absent) response from one of them when queried for patient records. In addition, more systems mean more processing time to integrate multiple sources of information into one coherent record. Thus, the response time will become slower as the number of connected systems increases (Lapsia et al. 2012).

The institution-centric architecture also introduces high levels of operational complexity. Since the completeness of retrieval of a given patient's records is dependent on the availability of all the systems that contain information about that patient, ongoing real-time monitoring of all connected information sources is essential. This translates into a requirement for a 24 × 7 **network operations center (NOC)**, that constantly monitors the operational status of every medical information system and is staffed with senior IT personnel who can immediately troubleshoot and correct any problems detected (Fig. 15.3). Even with modest system failure rates (e.g., one per thousand), a community with thousands of EHRs will typically have a handful of systems that are unresponsive to queries for patient records and require immediate expert attention to restore to full operation.

Fig. 15.3 Example of a Network Operations Center (NOC). (Image used by permission of NTT Ltd (available at ▶ https://www.gin.ntt.net/support-center/noc/))

The cost of this around-the-clock monitoring is very substantial, since a staff of at least five full-time network engineers is required to assure that at least one person is always available for every shift 7 days a week.

Adding to the cost of the NOC, every EHR system in an institution-centric model must incur additional expenses to always be able to respond to queries in real-time. This will be extremely problematic for physician offices, since their EHR systems will need to operate 24 × 7 and include additional hardware, software, and telecommunications capabilities to simultaneously support such queries while also serving its local users. Clearly, the transaction volumes generated will be substantial, since each patient's records will be queried whenever they receive care at any location. Contrast this to a central repository model where the information from a care episode is transmitted once to the repository and no further queries to the source systems are ever needed. This analysis has been confirmed by a simulation study of the institution-centric architecture demonstrating that both the transaction volume and probability of incomplete records (from missing data due to a malfunctioning network node) increase exponentially with the average number of sites where each patient's data is located (Lapsia et al. 2012).

15.5.2 Patient-Centric Architecture (Health Record Banking)

Health record banking is a patient-centric approach to developing community HII that both addresses the key requirements and can overcome the challenges that have stymied current efforts[15] (Yasnoff et al. 2013). A **health record bank (HRB)** is defined as "an independent organization that provides a secure electronic repository for storing and maintaining an individual's lifetime health and medical records from multiple sources and assuring that the individual always has complete control over who accesses their information" (Detmer et al. 2008).

Using a community HRB to provide patient information for medical care is straightforward (Fig. 15.4). Prior to seeking care (or at the time of care in an emergency), the patient gives permission for the caregiver to access his/her HRB account records (either all or part) through a secure Internet portal. The provider then accesses (and optionally, downloads) the records through a similar secure web site. When the care episode is completed, the caregiver then transmits any new information generated to the HRB to be added to the account-holder's lifetime health record. The updated record is then immediately available for subsequent care.

History of HRBs

The health record banking concept has been evolving for over two decades since it was initially proposed (Szolovits et al. 1994). The

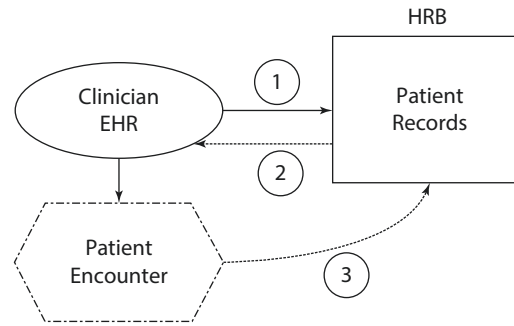


Fig. 15.4 Patient-centric HII architecture. (1) The clinician EHR requests prior patient records from the HRB. (2) The prior patient records are immediately sent to the clinician EHR. (3) After the care episode, the new information is stored in the clinician EHR and sent to the HRB; any inconsistencies or incompatibilities with prior records in the HRB need to be resolved before that patient's records are requested again (but not in real time). (Used with permission of Health Record Banking Alliance (HRBA))

15 Yasnoff WA. (2006) *Health Record Banking: A Practical Approach to the National Health Information Infrastructure.* Retrieval 29 Oct 2018: ▶ http://nhiiadvisors.com/slides/Health%20Rec%20Banking.html

term "health information bank" was introduced in 1997 in the U.K. (Dodd 1997), and was subsequently described as the "bank of health" (Ramsaroop and Ball 2000). A legal analysis of the implications of a "health record trust" was published in 2002 (Kostyack 2002), an Italian system known as the "health current account" was described in 2004 (Saccavini and Greco 2004), and the "health record bank" concept was described by Dyson in 2005 (Dyson 2005). In 2006, a Heritage Foundation policy paper endorsed health record banking,[16] additional papers described HRBs in more detail (Ball and Gold 2006; Shabo 2006), the non-profit Health Record Banking Alliance was formed,[17] the State of Washington endorsed the concept after a 16-month study,[18] and the non-profit Dossia consortium was formed by several large employers to implement and operate an HRB for their employees. In 2007, the Information Technology and Innovation Foundation recommended that the health record banking approach be used to build the U.S. HII,[19] while Gold and Ball described the "health record banking imperative" (Gold and Ball 2007). That same year, both Microsoft and Google introduced patient-controlled medical record repositories. In 2009, three pilot HRBs were funded by the State of Washington and the role of HRBs in protecting privacy was described (Kendall 2009). The HRB concept, although not always named as such, started appearing with greater frequency in articles discussing the need for comprehensive EHRs (Steinbrook 2008; Mandl and Kohane 2008; Kidd 2008; Miller et al. 2009; Krist and Woolf 2011).

More recently, discussion and activities related to HRBs have accelerated and expanded even more. Multiple articles have been published advocating for patient control of their own records and considering the issues involved[20] (Kish and Topol 2015; Haun and Topol 2017). An entire supplement to the Journal of General Internal Medicne was devoted exclusively to this topic (JGIM 2015). Even a White House Senior Advisor and the Administrator of CMS have joined the chorus touting the advantages of patient control of their own records.[21] In addition, at least two publications have discussed using the value of the information to facilitate sustainability[22] (Porter 2018).

There are also a continuing stream of articles describing HRBs and advocating for their establishment and use. These include the rationale for HRBs (Yasnoff et al. 2013), potential HRB use in public health (Yasnoff et al. 2014), and a description of lessons

16 Haislmaier EF. (2006) *Health Care Information Technology: Getting the Policy Right*. Retrieval 29 Oct 2018: ▶ http://www.heritage.org/Research/Reports/2006/06/Health-Care-Information-Technology-Getting-the-Policy-Right

17 Health Record Banking Alliance. (2006) Retrieval 29 Oct 2018: ▶ http://www.healthbanking.org

18 State of Washington Health Care Authority. (2006) *Washington State Health Information Infrastructure: Final Report and Roadmap for State Action*. Retrieval 29 Oct 2018: ▶ http://providersedge.com/ehdocs/ehr_articles/Washington_State_Health_Information_Infrastructure-Final_Report_and_Roadmap_for_State_Action.pdf

19 Castro D. (2007) *Improving Health Care: Why a Dose of IT May Be Just What the Doctor Ordered.* Information Technology and Innovation Foundation. Retrieval 29 Oct 2018: ▶ http://www.itif.org/publications/improving-health-care-why-dose-it-may-be-just-what-doctor-ordered

20 Yaraghi N. (2016) You Should Control Your Own Health Care Data. *U.S. News* (12 Feb 2016). Retrieval 29 Oct 2018: ▶ https://www.usnews.com/opinion/blogs/policy-dose/articles/2016-02-12/to-protect-patients-privacy-give-them-control-of-their-health-data

21 Kushner J and Verma S. (2018) Giving patients control of their health information will help give patients control of their health. Recode (15 Mar 2018). Retrieval 29 Oct 2018: ▶ https://www.recode.net/2018/3/15/17114684/health-care-digital-information-jared-kushner-seema-verma-president-trump

22 Yaraghi N. (2015) A Sustinable Busines Model for Health Information Exchange Platforms: The Solution to Interoperability in Health Care IT. Brookings Institution (30 Jan 2015). Retrieval 29 Oct 2018: ▶ https://www.brookings.edu/research/a-sustainable-business-model-for-health-information-exchange-platforms-the-solution-to-interoperability-in-health-care-it/

learned from an early HRB startup (Yasnoff and Shortliffe 2014). Endorsement of the HRB concept (albeit referenced under various different terms) has come from a variety of sources including a prominent observer of the digital transformation of health care (Mikk et al. 2017), a policy think tank,[23] a former ONC Director,[24] and a leading market research firm.[25] Even the current CMS Administrator has been openly advocating for lifetime, longitudinal patient records accessible to and controlled by patients.[26]

Several countries have established successful HRBs, including Finland,[27] Estonia,[28] and Brazil.[29] In addition, a number of startup companies such as Project Hugo,[30] CareDox,[31] Patients Know Best,[32] Betterpath,[33] and Ciitizen[34] (as well as a number of others) are pursuing the challenge of developing HRBs. Even the tech giant Apple appears to be moving towards HRB implementation with its Apple Health Kit serving as the infrastructure for a patient record repository controlled by individuals on their own smartphones.[35]

How Requirements Lead to HRB Architecture

Figure 15.5 shows how the HII requirements for complete records, low cost, and high benefits lead directly to the HRB architecture. In order to ensure complete records, information from each patient encounter must be available. The only currently available mandate for this is for the patient to request records (in digital form) from their provider, invoking their rights under the HIPAA privacy rule. In order for patients to feel comfortable doing that, they must trust the system. The first element of trust is the architectural end point of security; the patients must know that the records being sent from their provider will be protected from improper use. This also leads to the architectural end point of patient control of all access to the records; how can it be justified to patients for some other entity to

23 Kendall D and Quill E. (2015) A Lifetime Electronic Health Record for Every American. Third Way (28 May 2015). Retrieval 29 Oct 2018: ▶ https://www.thirdway.org/report/a-lifetime-electronic-health-record-for-every-american

24 Blumenthal D. (2016) The Biggest Obstacle to the Health-Care Revolution. Wall Street Journal (28 Jun 2016) Retrieval 29 Oct 2018: ▶ https://blogs.wsj.com/experts/2016/06/28/how-to-make-health-care-records-as-mobile-as-patients/

25 Frost & Sullivan. (2015) Moving Beyond the Limitations of Fragmented Solutions: Empowering Patients with Integrated, Mobile On-Demand Access to the Health Information Continuum. Retrieval 29 Oct 2018: ▶ http://coranetsolutions.com/wp-content/uploads/Moving%20Beyond%20the%20Limitations%20of%20Fragmented%20Solutions%20Whitepaper.pdf

26 Coombs B. (2018) Medicare chief says it's time health care caught up to other industries to benefit consumers. CNBC (30 Apr 2018). Retrieval 29 Oct 2018: ▶ https://www.cnbc.com/2018/04/30/cms-verma-says-its-time-health-care-caught-up-to-other-industries.html

27 Kunnamo I. (2018) National health record bank (the eArchive) in Finland. Retrieval 29 Oct 2018: ▶ https://www.dropbox.com/s/yfpck5hpzkd8lh0/HRBA%20Kunnamo.pdf

28 e-Estonia Healthcare. Retrieval 29 Oct 2018: ▶ https://e-estonia.com/solutions/healthcare/e-health-record/

29 Hummel GS. (2016) Brazil eHealth – Overview, Trends, and Opportunities. Consulate General of the Kingdom of the Netherlands. Retrieval 29 Oct 2018: ▶ https://www.rvo.nl/sites/default/files/2017/01/Brazil%20Healthcare%20-%20Guilherme%20Hummel.pdf

30 Retrieval 29 Oct 2018: ▶ http://hugophr.com

31 Retrieval 29 Oct 2018: ▶ https://caredox.com

32 Retrieval 29 Oct 2018: ▶ https://www.patientsknowbest.com/patients.html

33 Retrieval 29 Oct 2018: ▶ https://www.betterpath.com

34 Retrieval 29 Oct 2018: ▶ https://www.ciitizen.com

35 Bell K. (2018) Apple wants to put medical records on your iPhone. Here's how it works. Mashable (24 Jan 2018). Retrieval 29 Oct 2018: ▶ https://mashable.com/2018/01/24/apple-iphone-medical-records-how-to/

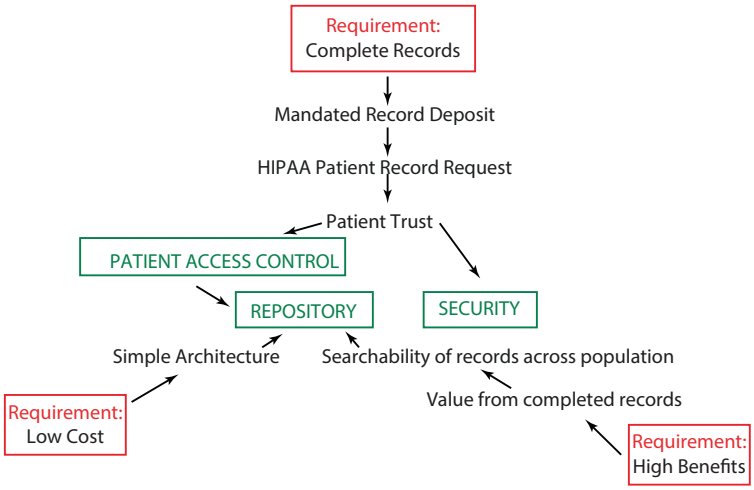


Fig. 15.5 How requirements lead to HRB architecture. This diagram shows how the key requirements of complete records, low cost, and high benefits lead directly to the architectural specifications of patient access control, a repository for data storage, and trustworthy security that are characteristic of health record banks

decide which records to release to whom and when? In order to assure feasible patient access control, the records must be stored in a repository (the third and final architectural end point) so that patients have a single point for indicating and revising their record access permissions.

The requirement for low cost means that the architecture must be simple. The most straightforward architecture for compiling and accessing complete records for each person is a repository – it is easy (and therefore inexpensive) to implement and operate.

Finally, the requirement for high value leads to the intermediate requirement to be able to search records across the population. It is those searching operations that provide critical information of great value for public health, medical research, quality improvement, and policy. Such searchability is much more easily achieved when the records are in a repository.

Therefore, the architecture needed for a low cost, high benefit HII with complete records is a secure repository with record access controlled by patients. These are the exact characteristics of health record banks, which were conceived and designed to specifically address these key HII requirements.

Patient Control Ensures Privacy and Stakeholder Cooperation

In an HRB, everything is done with *consumer consent*, with account-holders controlling their copy of all their records and deciding who gets to see any or all of it. This protects privacy (since each consumer sets their own customized privacy policy), promotes trust, and ensures stakeholder cooperation since all holders of medical information must provide it when requested by the patient (Kendall 2009). Of course, the operations of an HRB must be open and transparent with independent auditing of privacy practices. World-class state-of-the-art computer security is needed to protect the HRB, which will be a natural target for hackers. At least one new security approach is now available that absolutely prevents large-scale data loss from a repository of medical records (Yasnoff 2016), the key security requirement for an HRB (see subsection on HRB Security below).

Natural concerns arise from the ability of the patient to suppress any or all of their HRB account information, which could lead to misdiagnosis and dangerous treatment. This capability could be abused by patients who, for example, may seek multiple prescriptions for controlled substances for the pur-

pose of diversion for illegal sale. With respect to the possibility of medical errors resulting from incomplete information, the patient would be clearly and unmistakably warned about this when choosing not to disclose any specific information (e.g., "Failure to disclose any of your medical information may lead to serious medical problems, including your death"). The expectation is that few people will choose to do this, particularly after such a warning. However, as noted earlier, 13–17% of patients already engage in this practice, leading many observers to conclude that the general public may not be comfortable with a system that provides easy access to their records unless they are in control of such access. This issue ultimately becomes one of public policy and may also be a subject of discussion between the doctor and the patient (i.e., the doctor will want to be assured by the patient that all information is being provided). Clearly, physicians should not be liable for the consequences of the patient's choice to withhold information.

With respect to patients who use their power to hide information as a way to facilitate improper or illegal activity, there is clearly an overriding public policy concern. For example, in the case of controlled substances, it may be necessary to report to the physician (or, if legislatively mandated, to regulatory authorities) whenever a patient suppresses any information about controlled substance prescriptions. The information itself would still be under the patient's control, but the physician would be alerted with a notice such as "some controlled substance prescription information has been withheld at the patient's request." There may be other situations where such warnings are needed.

Assuring that Standardized, Encoded Electronic Patient Information Is Consistently Deposited

HRBs can provide ongoing incentives for deposits of EHR records by clinicians, which would help to offset the unfavorable cost/benefit of EHRs for office-based providers. These incentives could be paid on a per-encounter or per-month basis. In addition, those few physicians who do not currently have EHRs could receive no-cost Internet-accessible EHR systems (at HRB expense) with the understanding that information from patient encounters would be automatically transferred to the HRB. Another option is to link reimbursement for medical services to HRB deposits – i.e., providers would not be paid unless the medical record information generated from those services is transmitted to an HRB. This makes sense economically, as the value of medical services is greatly limited if the information about patients is not readily available for their ongoing care.

Incentives for HRB deposits also serve to ensure compliance with data standards, both initially and on an ongoing basis. Clearly, any EHR provided through the HRB would, by definition, transmit information back to the HRB in a standard format (since the HRB would only provide systems that can do so). For physicians who already have EHRs, HRB reimbursements for deposits from those systems would naturally require complete and fully encoded encounter data using established standards to be sent to the HRB. Over time, higher levels of encoding of medical information can be promoted through the gradual introduction of more stringent standards requirements (with plenty of lead time to allow for system upgrades). Compliance with such changes in standards could also be further assured through a direct relationship to reimbursement.

HRB Business Model

Health record banking has advantages on both the cost and revenue sides of the business model; the cost is lower and the revenue opportunities greater. Because of the lower operating costs and additional functionality

for searching records, one can envision a variety of business models for HRBs that do not depend on public subsidies or attempt to capture any healthcare savings, but are solely funded through new value created for consumers and other stakeholders.[36]

Due to the simplicity of HRB operations, the cost is substantially less than an equivalent institution-centric architecture. For an HRB, providing access at the point of care only involves a single retrieval from the bank's repository of records. In an institution-centric model, the records for a given patient are located at an arbitrary number of dispersed sites, and must be assembled in real-time and integrated into a comprehensive record before they can be used for patient care. Not only is this process of assembly complex, time-consuming, and prone to error, it necessitates, as noted above, the creation of a fully staffed 24 × 7 NOC to monitor the availability of all information sources as well as troubleshoot and correct those that are malfunctioning.

The estimated cost for the NOC in an institution-centric model is substantial. For example, given a population of 1,000,000, at least 1,000 systems would need monitoring (one for every 1,000 patients). Assuming a reasonable failure rate for fully functional query connectivity to each system of once/year (representing a **mean time between failures [MTBF]** of over 8,700 hours), there would be an average of 2.73 failures/day or 0.91 failures per 8-hour shift that would need troubleshooting attention. A minimum staff for the NOC would be one person 24 × 7; given 21 shifts/week plus leeway for vacations and sick leave, this would require at least 5 full-time equivalent staff costing about $200,000 each including equipment, overhead and fringe benefits. Assuming an additional $500,000/year for hardware and software to operate the institution-centric system (over and above the data repository needed for an HRB) yields an annual cost of $1.5 million or $1.50/person/year. This would add nearly 20% for the institution-centric model to the estimated $8/person/year needed to operate an HRB (Kaelber et al. 2008).

Beyond this, the additional costs imposed in the institution-centric model for each connected EHR for additional hardware, software, telecommunications capability, and additional operational expenses to maintain 24 × 7 system availability must also be included. Even if such costs were only a very modest $1,000/year/system (less than $100/month), this would result in an additional $1,000,000 or $1/person/year. Adding this to the $1.50/person/year for the NOC gives a total estimated cost of $2.50/person/year, resulting in over 30% higher costs for the institution-centric model than a basic HRB. Added to this would be the costs and complexity of establishing and maintaining data sharing agreements among all the entities, which would be substantial.

An HRB with comprehensive electronic medical records for individuals in a community can generate substantial value (Yasnoff and Shortliffe 2014). In addition to empowering physicians to provide safer, more effective, and more efficient care, the availability of such records readily enables many types of heretofore infeasible yet very desirable services for patients, providers, and other stakeholders.

Perhaps the most compelling example of a service for patients is the "peace of mind" alert that immediately notifies the patient's loved ones about a critical event, such as an emergency room visit. As soon as the emergency provider accesses the patient's HRB record, the alert, which is delivered electronically in a manner chosen by the recipient, is delivered. Such a service is particularly valuable for children of seniors who may not be

36 Health Record Banking Alliance. (2012) Health Record Banking: A Foundation for Myriad Health Information Sharing Business Models. White Paper (12 Dec 2012). Retrieval 29 Oct 2018: ▶ http://www.healthbanking.org/uploads/9/6/9/4/9694117/hrba_white_paper_business_model_dec_2012.pdf

immediately notified that their parent is undergoing emergency care.

Another example of a patient service is the "preventive care reminder" that indicates to the patient recommended preventive interventions, e.g. a flu shot, customized based on demographics and history. Since the reminder is based on comprehensive records, it is highly likely to be correct, and therefore relevant, to the patient. Once the service is obtained, further reminders would cease. In this way, patients would not be distracted by irrelevant or redundant reminders.

A third example of a patient service is medication refill reminders. These could be sent via text message, allowing the patient to respond briefly to indicate approval of the refill. While a number of pharmacies currently provide such services, the HRB reminders would be independent of pharmacy, and the record of the patient's permission for such communications would only need to be maintained in one place.

An example of a service for providers is the "normal or unchanged lab result" message to patients. Every day, providers review results for previously ordered tests and communicate those results with their recommendations to their patients. With an HRB, deposits of new lab results that are either normal, or within a specified range of a previous result for that patient, could be automatically sent to the patient over the provider's signature indicating that all is well. It would no longer be necessary for the provider to take the time to review and comment manually on such results, which would be a significant productivity enhancement.

There are many other potentially valuable benefits of comprehensive HRB patient records, such as the ability to comprehensively monitor clinical care for public health, query clinical records for research and policy purposes (with patient permission), and eliminate the need for separate registries of clinical information (e.g. diabetes, asthma, cancer, immunization) since all the data normally stored in such registries would already be available in the HRB.

HRB Security

It has long been known that centralized data storage is the best way to ensure security (Turn et al. 1976). The reason for this is clear: distributed data is inherently less secure since it must be transmitted multiple times for each use. However, the single source of all data provided by a central database is also an inherent weakness; if all the data is accessible all-at-once for good purposes, it must necessarily also be available for misappropriation and misuse. Multiple recent large-scale healthcare security breaches (Associated Press 2015; Abelson and Goldstein 2015; Reuters 2015) validate the risk of centralizing substantial amounts of sensitive medical information. How can data for each person in a community be readily available while ensuring its security?

The **personal grid architecture** (Yasnoff 2016) addresses this difficult problem by storing each patient's information in a separate file, separately encrypted with its own strong password. The clear advantage of this approach is that there is no longer a single access point for multiple patient records. Indeed, if an adversary somehow obtained a complete copy of all the data, it would be of very limited value since access to each individual record would require breaking a strong encryption key, a very costly and time-consuming process. Thus, not only is the data protected, but the incentive for hackers to obtain the data is largely eliminated.

However, this approach also has a serious drawback with respect to searching records across a population. To do this, each record must be retrieved, decrypted, and searched in sequence, which is prohibitively slow. Indeed, the reason relational and other databases are used for storage is that these systems "pre-index" the data to allow rapid retrieval across records. Such "pre-indexing" is not compatible with the personal grid architecture as it would create the single access point to all the data that the architecture has deliberately eliminated to ensure security, thereby defeating the core purpose of the approach.

Fortunately, the development of cloud computing has provided a convenient, acceptable solution to searching across population

data in the personal grid. Although sequential searching, which is inherently slow, is still required, the searching can be distributed over hundreds or even thousands of servers from a cloud computing provider. While the amount of computing needed remains the same, this parallel approach reduces the search time by orders of magnitude. It is estimated, for example, that the search time for a population of 5 million using 500 parallel search servers would be just 7 minutes. In the medical environment, there are no use cases where response times for population searches need to be less than an hour (in fact, overnight is usually adequate), so this searching methodology can effectively overcome the requirement for sequential search that the personal grid imposes to ensure security.

Some observers have suggested that blockchain, the increasingly popular secure distributed ledger methodology[37,38] could be a useful alternative for securing medical records. However, blockchain has a number of serious drawbacks for this application: (1) Medical records are massively larger than financial ledger entries, so the storage requirements for the required copies of the entire dataset at each of the multiple blockchain nodes would be prohibitive; (2) Every blockchain node would receive and have a copy of everyone's medical records for that dataset, which unnecessarily increases the risk of a security breach even if the records are encrypted both in transit and at rest; (3) Adding records to the blockchain requires extensive computing resources for which there is no clear source of compensation; and (4) Security for medical records must be trusted by the general public, which requires that they understand why the records are secure. The complexity of blockchain and its dependence on advanced encryption and game theory methods makes it extremely difficult and time consuming to fully explain to non-technical audiences, precluding an informed basis for widespread trust. In view of these issues, the effective, inexpensive, and easy-to-understand personal grid architecture is a better choice to meet the security requirements for medical record storage.

Summary Comparison of Institution-Centric and Patient-Centric Architectures

Table 15.1 summarizes the characteristics of the institution-centric approach to HII compared to the patient-centric (HRB) model. The patient-centric model is simpler and more straightforward, and deals directly with the issue of privacy by putting patients in control of their own information. Interoperability is much more easily accomplished in the patient-centric model since standards compliance can be reinforced with financial incentives, and reconciliation of inconsistencies between records need not be real-time. The patient-centric approach is financially sustainable with a variety of business models, and can provide powerful incentives to clinicians to deposit EHR data for their patients. Finally, the patient-centric model avoids the substantial processing burden on clinician EHRs from queries each time any patient whose record is stored is seen anywhere. In each of the categories of requirements, organizational issues, cost, operations, and incentives, the patient-centric approach has substantial advantages.

15.6 Progress Towards HRBs

15.6.1 HRB Opposition

If, as has been clearly elucidated in this chapter, HRBs are the most effective and efficient solution for HII architecture, why hasn't this approach been widely adopted? The short answer is that key healthcare stakeholders, such as insurers, health plans, and hospitals often oppose it (although typically not

37 Brakeville S and Perepa B. (2018) Blockchain basics: Introduction to distributed ledgers. IBM Developer Tutorial (31 Oct 2018). Retrieval at: ► https://developer.ibm.com/tutorials/cl-blockchain-basics-intro-bluemix-trs/

38 Marr B. (2017) A Complete Beginner's Guide to Blockchain. Forbes (24 Jan 2017). Retrieval at: ► https://www.forbes.com/sites/bernardmarr/2017/01/24/a-complete-beginners-guide-to-blockchain/?sh=187ac78a6e60

Table 15.1 Comparison of the institution-centric and patient-centric approaches to HII

Issue	Institution-centric	Patient-centric
Requirements		
Privacy	Patient consent difficult to implement; many complex data sharing agreements needed	Simple; patients in control of all access to their own records; consent easy to implement
Security	Inherently weak because records transmitted multiple times for each use	Very strong using new security techniques that eliminate large-scale data loss from repositories
Searchability	Impractical to search population data	Search feasible using parallel processing in the cloud
Completeness	Requires queries to all data sources each time a patient's records are requested; all must respond for completeness	Comprehensive data available at all times for each patient
Organizational issues		
Cooperation needed	Extensive; community-wide	Unifying; HIPAA mandates records on patient request
Organizational complexity	High; ongoing collaboration of multiple competing stakeholders necessary	Low; HRB is neutral and independent of all stakeholders
Cost		
Startup cost	Substantial (due to high complexity)	Moderate
Operational cost	Inefficient/expensive	Efficient/inexpensive
Business model	Complex; no clear approach has emerged; typically requires ongoing subsidies from health care stakeholders	Sustainable using new value of complete information; many options possible funded by patients/payers/purchasers
Operations		
IT design	Complex, based on "fetch and show"; requires queries to multiple entities, real-time reconciliation of inconsistencies, and NOC	Simple, based on "deposit once to account"; no secondary queries or real-time reconciliation needed; NOC unnecessary
Reliability	Prone to error (record sources unavailable)	Reliable; one operation to retrieve individual patient data
Interoperability	Compliance voluntary	Compliance can be assured with financial incentives
Incentives		
Clinician incentives	Not included	Easy to include
Clinician burden	Extensive; incoming query each time current patients seen anywhere increases EHR costs	Minimal; information deposited once in HRB; no incoming queries

openly). However, this is an oversimplification of the complex incentives involved.

For insurers, their claims data provides the most comprehensive picture of individual health care service usage. While this information is not sufficiently detailed to facilitate clinical management of patients, it gives the insurers the "informational advantage" in negotiations with their customers, employers that purchase insurance for their employees, and their payees, physicians, hospitals, and health plans. HRBs would make much more detailed information about each patient potentially available to all healthcare stakeholders, eliminating this informational advantage. In addition, HRBs have the potential to reduce health care costs. While this would seem to be positive for insurers (making their offerings less expensive to employers), it also would reduce their profitability, which is typically a percentage of those health care costs. For these reasons (and perhaps others), insurers have worked actively against HRB development, for example, by legislatively derailing promising efforts to jump start HRB development in Washington State. Of course, since comprehensive patient-controlled electronic records are so appealing to the public, the opposition of insurers has been behind the scenes. The latter is a theme that is common to all the stakeholder opposition, as it is difficult to openly justify opposing the safer, more effective, more efficient care likely to result from HRBs.

Health plans and hospitals have somewhat different reasons for opposing HRBs. Mostly, this involves concerns about empowering their competitors. Typically, the largest hospitals and health plans in each region have more complete information than their smaller competitors, and therefore see HRBs as weakening or eliminating that perceived market advantage. Sadly, this is true despite the fact that HRBs would allow all providers to deliver better care. Another issue, particularly for hospitals, is that they are concerned about the idea of sharing their electronic patient information after spending hundreds of millions of dollars (or more) to install and operate EHR systems. The financial interests of health plans and hospitals also appear to conflict with the deployment of HRBs, since their adoption would likely reduce their fee-for-service incomes by eliminating unnecessary and duplicative care. Again, this opposition is behind the scenes to avoid having to openly advocate for higher institutional income over patient safety and effective, efficient care.

In a situation such as this, where progress that would benefit everyone (i.e., HRBs) is opposed by specific groups, one would hope that the government, representing the interests of all the people, would help ensure that the good of everyone prevails. However, in the U.S., the ability of the government to impose major changes is quite limited (due to the system of "checks and balances"), and is especially so when the proposed changes are opposed by many key stakeholders. It has therefore been relatively easy for major healthcare stakeholders to delay and/or redirect HRB efforts by, for example, advocating for other solutions, raising seemingly valid objections to reasonable efforts to move forward, or creating organizations that appear to be working towards a solution while in reality being dedicated to maintaining the status quo.

15.6.2 Factors Accelerating HRB Progress

Perhaps the most important factor accelerating HRB progress is the recent and ongoing change in the reimbursement system for care. The move from "fee for service" to "pay for value" is changing the incentives for all health care stakeholders. Under "fee for service," the more efficient care enabled by HRBs would directly translate to lower reimbursement, a clearly undesirable outcome for providers. But a "pay for value" system financially rewards efficiency, thereby creating a strong incentive for the more complete and timely patient information that HRBs can make available. While the transition from "fee for service" to "pay for value" is still in its relatively early stages, it is reasonable to expect that ongoing

progress in this direction will be accompanied by growing support for the HRB architecture.

In addition, the growing recognition of an "individual right" to personal data ownership and control is also an enabling force for HRB adoption. The adoption of the EU General Data Protection Rule[39] as well as the widespread backlash against use of individual data for profit without permission of the data subjects[40,41,42] is creating a higher level of awareness and support for personal data control. This is likely to accelerate the demand for organizations such as HRBs that can provide individuals with the ability to capture and control their sensitive medical record information.

There are at least four additional factors that are currently facilitating and even accelerating the development of health record banks (HRBs):

- Patient records are now largely electronic.
  - This is a necessary prerequisite for any HII approach. Thanks largely to the subsidies to providers and hospitals to acquire EHR systems, most patient records are now digital.
- Effective standards are available.
  - Standards for transmission of EHR data have evolved. While not perfect, the HL7 CCDA standard is an effective methodology for transmitting and receiving patient data. Other potentially even more effective standards, such as FHIR (Fast Healthcare Interoperability Resources), are rapidly evolving.
- Smart phones are nearly ubiquitous
  - Smart phones provide a convenient and readily accessible mechanism for individuals to access and control their digital health records.
- New computer security methods can prevent large-scale breaches
  - Community-based repositories of digital health records must be able to reliably prevent large-scale breaches. The new "personal health grid" architecture does this in an easy-to-understand way.

Finally, as mentioned earlier, there are now several successful examples of HRB implementations outside the U.S. As the benefits of these systems are documented and become more widely known, this should also increase the demand for HRBs throughout the world.

15.7 Evaluation

The last element in the strategy for promoting a complex and lengthy project such as the HII is evaluation to both gauge progress and define a complete system. Evaluation measures should have several key features. First, they should be sufficiently sensitive so that their values change at a reasonable rate (a measure that only changes value after 5 years will not be particularly helpful). Second, the measures must be comprehensive enough to reflect activities that affect most of the stakeholders and activities needing change. This ensures that efforts in every area will be reflected in improved measures. Third, the measures must be meaningful to policymakers. Fourth, periodic determinations of the current values of the measures should be easy so that the measurement process does not detract from the actual work. Finally, the totality of the measures must reflect the desired end state so that when the goals for all the measures are attained, the project is complete.

A number of different types or dimensions of measures for HII progress are possi-

39 Retrieval 29 Oct 2018: ▶ https://gdpr-info.eu

40 Breland A. (2017) Tech faces public anger over internet privacy repeal. The Hill (2 Apr 2017). Retrieval 29 Oct 2018: ▶ https://thehill.com/policy/technology/326816-tech-faces-public-anger-over-internet-privacy-repeal

41 King C. (2018) Tech Industry Pursues a Federal Privacy Law, on Its Own Terms. New York Times (26 Aug 2018). Retrieval 29 Oct 2018: ▶ https://www.nytimes.com/2018/08/26/technology/tech-industry-federal-privacy-law.html

42 Wakabayashi D. (2018) California Passes Sweeping Law to Protect Online Privacy. New York Times (28 Jun 2018). Retrieval 29 Oct 2018: ▶ https://www.nytimes.com/2018/06/28/technology/california-online-privacy-law.html

ble. Aggregate measures assess HII progress over the entire nation. Examples include the percentage of the population covered by an HII and the percentage of healthcare personnel who utilize EHRs. Another type of measure is based on the setting of care. Progress in implementation of EHR systems in the inpatient, outpatient, long-term care, home, and community environments could clearly be part of an HII measurement program. Yet another dimension is healthcare functions performed using information systems support, including, for example, registration systems, decision support, and CPOE. Finally, it is also important to assess progress with respect to the semantic encoding of EHRs. Clearly, there is a progression from the electronic exchange of images of documents, where the content is only readable by the end user viewing the image, to fully standardized and encoded EHRs where all the information is indexed and accessible in machine-readable form.

Sadly, the evidence is now overwhelming that U.S. HIEs in their current form are, with rare exceptions, not succeeding. Labkoff and Yasnoff described four criteria for the quantitative evaluation of HII progress in communities: (1) completeness of information, (2) degree of usage, (3) types of usage, and (4) financial sustainability (Labkoff and Yasnoff 2007). Using these criteria, four of the most advanced community HII projects in the U.S. achieved scores of 60–78% (on a 0–100 scale), indicating substantial additional work was required before the HII could be viewed as complete.

The 2010 PCAST report stated, "*HIEs have drawbacks that make them ill-suited as the basis for a national health information architecture*" (PCAST 2010). Among those drawbacks, PCAST cited administrative burdens (data sharing agreements to ensure stakeholder cooperation), financial sustainability, interoperability, and an architecture that cannot be scaled effectively. A recent survey of HIEs (Adler-Milstein et al. 2011) found only 13 HIEs in the U.S. (covering 3% of hospitals and 0.9% of physician practices) capable of meeting Stage 1 Meaningful Use criteria, and even those metrics by no means ensure the availability of comprehensive electronic patient information when and where needed. Of those, only 6 were reported to be financially viable. More importantly, *none* of the HIEs surveyed had the capabilities of a comprehensive system as specified by an expert panel.

Overall, the current approaches to building HII consistently fail to meet one or more of the requirements described above: privacy, stakeholder cooperation, ensuring fully electronic information, financial sustainability, and independent governance. While these problems are highly interdependent, it is useful to consider them in the context of the decisions that communities have made about HII architecture, privacy, and business model that, while appearing attractive to stakeholders in the short term, have so far been largely unsuccessful. Exploration and large-scale testing of alternative approaches that directly address the requirements, such as health record banking, seem both necessary and increasingly urgent.

15.8 Conclusions

While progress has been made and efforts are continuing, successful development and operation of comprehensive HII systems remains a largely unsolved problem in the U.S. Happily, we are now seeing successful HII implementations in other countries that can provide important examples of feasible and effective systems. The extensive focus on building HII systems has greatly improved our understanding of the requirements, barriers, and challenges, as well as potential solutions. Despite the daunting obstacles, the benefits of HII are sufficiently urgent and compelling to ensure major ongoing work in this domain. Through these activities, the HII path to comprehensive electronic patient records when and where needed is becoming clearer, and substantial

additional progress is highly likely over the next few years.

Suggested Readings

Castro, D. (2007). *Improving health care: Why a dose of IT may be just what the doctor ordered.* Information Technology and Innovation Foundation. Retrieval 31 Oct 2018: http://www.itif.org/publications/improving-health-care-why-dose-it-may-be-just-what-doctor-ordered. This is the first independent report that endorsed patient-centric architecture (HRBs) as an effective approach to HII. It describes clearly the problems and challenges of HIEs.

Kendall, D., & Quill, E. (2015, May 28). A lifetime electronic health record for every American. *Third Way*. Retrieval 29 Oct 2018: https://www.thirdway.org/report/a-lifetime-electronic-health-record-for-every-american. This paper is a more recent endorsement of the HRB concept by an independent think tank.

Krist, A. H., & Woolf, S. H. (2011) A vision for patient-centered health information systems. *Journal of the American Medical Association, 305*(3):300–301. A vision of how fully functional patient-centric electronic medical record systems could be the basis for an effective HII.

Lapsia, V., Lamb, K., & Yasnoff, W. A. (2012). Where should electronic records for patients be stored? *International Journal of Medical Informatics, 81*(12), 821–827. This paper elucidates clearly the advantages of patient-centric architecture by comparing it via simulation to an institution-centric approach.

Miller, R. H., & Miller, B. S. (2007). The Santa Barbara County Care Data Exchange: What happened? *Health Affairs, 26*(5), w568–w580. This paper describes the history of one of the earliest HIEs, including details about the factors leading to its failure.

Mikk, K. A., Sleeper, H. A., & Topol, E. J. (2017). The pathway to patient data ownership and better health. *JAMA, 318*(15), 1433–1434. This paper describes how patient data ownership and records stored in repositories can result in an effective HII.

National Committee on Vital and Health Statistics. (2001). *Information for health: A strategy for building the National Health Information Infrastructure. Report and recommendations from the National Committee on Vital and Health Statistics.* Retrieval 31 Oct 2018: https://aspe.hhs.gov/report/information-health-strategy-building-national-health-information-infrastructure. This seminal work was the first to call for a national HII, coining the term. It comprehensively describes the need for HII, the problems it would solve, and the necessity for government investment to incentivize its development.

Yasnoff, W. A. (2016). A secure and efficiently searchable health information architecture. *Journal of Biomedical Informatics, 61*, 237–246. This paper describes a new and innovative personal grid architecture for digital health records that absolutely prevents large-scale data loss, an essential capability to ensure user trust in large central repositories of health records.

Yasnoff, W. A., & Shortliffe, E. H. (2014). Lessons learned from a health record bank start-up. *Methods of Information in Medicine, 53*, 66–72. This paper gives a detailed post-mortem of a health record bank startup and includes specific recommendations for future success.

Yasnoff, W. A., Humphreys, B. L., Overhage, J. M., Detmer, D. E., Brennan, P. F., Morris, R. W., Middleton, B., Bates, D. W., & Fanning, J. P. (2004). A consensus action agenda for achieving the national health information infrastructure. *Journal of the American Medical Informatics Association, 11*(4), 332–338. This paper describes the results of the first national consensus conference on HII held in Washington, DC, in 2003. This was the meeting that led to the creation of ONC in 2004.

Yasnoff, W. A., Sweeney, L., & Shortliffe, E. H. (2013). Putting health IT on the path to success. *Journal of the American Medical Association, 309*(10), 989–990. This paper provides a concise overview of the problems with HIE and the rationale for HRBs.

Questions for Discussion

1. Make the case for and against investing $billions in the HII. How successful have the HITECH Meaningful Use

incentives been in promoting HII development? What could have been done differently to make them more effective?

2. What organizational options would you consider if you were beginning the development of HII? What are the pros and cons of each? How would you proceed with making a decision about which one to use?
3. Estimate the required bandwidth and transaction rate for patient-centric (HRB) vs. institution-centric HII architecture.
4. Consider the policy implications of universal availability of comprehensive electronic patient records. What are the risks and how could they be mitigated?
5. Given the architectural and other advantages of HRBs, why have most communities adopted institution-centric architectures up to now? What are some steps that might be helpful in encouraging communities to evaluate alternative architectures such as HRBs?
6. Show specifically the potential locations where patient consent functionality could be added to the institution-centric and patient-centric HII architectures in Figs. 15.2 and 15.4 and describe the granularity of consent that would be possible at each proposed location. After eliminating any redundant functionality, compare and contrast the consent implementation issues for the two alternative architectures, describing the advantages and disadvantages of each. Which architecture more efficiently addresses the issue of patient consent? Why?

References

Abelson, R., & Goldstein, M. (2015, May 2). Millions of anthem customers targeted in cyberattack. *New York Times*. Retrieval 29 Oct 2018: http://www.nytimes.com/2015/02/05/business/hackers-breached-data-of-millions-insurer-says.html

Adler-Milstein, J., Bates, D. W., & Jha, A. K. (2011). A survey of health information exchange organizations in the United States: Implications for meaningful use. *Annals of Internal Medicine, 154*, 666–671.

Agency for Healthcare Research and Quality. (2006). *Costs and benefits of health information technology.* Evidence Report/Technology Assessment 132, publication 06-E006. Retrieval 29 Oct 2018: https://www.ahrq.gov/downloads/pub/evidence/pdf/hitsys-costs/hitsys.pdf

Associated Press. (2015, July 17). Online attacks at UCLA Health exposed 4.5 million. *New York Times*. Retrieval 29 Oct 2018: http://www.nytimes.com/2015/07/18/business/online-attacks-at-ucla-health-exposed-4-5-million.html

Balas, E. A., & Boren, S. A. (2000). Managing clinical knowledge for health care improvement. In *Yearbook of medical informatics 2000: Patient-centered systems* (pp. 65–70). Schattauer: Stuttgart.

Ball, M., & Gold, J. (2006). Banking on health: Personal records and information exchange. *Journal of Healthcare Information Management, 20*(2), 71–83.

Bates, D. W. (2000). Using information technology to reduce rates of medication errors in hospitals. *British Medical Journal, 320*, 788–791.

Bates, D. W., Leape, L. L., Cullen, D. J., et al. (1998). Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *Journal of theAmerican Medical Association, 280*(15), 1311–1316.

Creswell, J. (2014, September 30). Doctors find barriers to sharing digital medical records. *New York Times*. Retrieval 29 Oct 2018: https://www.nytimes.com/2014/10/01/business/digital-medical-records-become-common-but-sharing-remains-challenging.html

Detmer, D., Bloomrosen, M., Raymond, B., & Tang, P. (2008). Integrated personal health records: Transformative tools for consumer-centric care. *BMC Medical Informatics and Decision Making, 8*, 45. Retrieval 29 Oct 2018: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2596104/

Devine, E. B., Totten, A. M., Gorman, P., Eden, K. B., Kassakian, S., Woods, S., Daeges, M., Pappas, M., McDonagh, M., & Hersh, W. R. (2017, December 7). Health information exchange use (1990–2015): A systematic review. *EGEMS (Washington, DC), 5*(1), 27. Retrieval 29 Oct 2018: https://egems.academyhealth.org/articles/10.5334/egems.249/

Dodd, B. (1997, October). An independent "health information bank" could solve health data security issues. *British Journal of Healthcare Computing and Information Management, 14*(8), 2.

Dyson, E. (2005, September). Personal health information: Data comes alive! *Release, 1.0*(24), 1.

Gold, J. D., & Ball, M. J. (2007). The health record banking imperative: A conceptual model. *IBM Systems Journal, 46*(1), 43–55.

Haun, K., & Topol, E. J. (2017, January 2). The health data conundrum. *New York Times*. Retrieval 29 Oct 2018: https://www.nytimes.com/2017/01/02/opinion/the-health-data-conundrum.html

Hersh, W. (2004). Healthcare information technology: Progress and barriers. *Journal of the American Medical Association, 292*, 2273–2274.

Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., & Taylor, R. (2005). Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs, 24*, 1103–1117.

Institute of Medicine. (1991). *The computer-based patient record: An essential technology for patient care*. Washington, DC: National Academy Press.

Institute of Medicine. (1997). *The computer-based patient record: An essential technology for patient care* (Revised ed.). Washington, DC: National Academy Press.

Institute of Medicine Committee on Quality of Health Care in America. (1999). *To err is human: Building a safer health care system*. Washington, DC: National Academy Press.

Insitute of Medicine Committee on Quality of Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press.

JGIM. (2015). Special supplement on patient control of medical records (8 articles). *The Journal of General Internal Medicine, 30*(1 Supplement). Retrieval 29 Oct 2018: https://link.springer.com/journal/11606/30/1/suppl/page/1

Johnston, D., Pan, E., Walker, J., Bates, D. W., & Middleton, B. (2003). *The value of computerized provider order entry in ambulatory settings*. Boston: Center for Information Technology Leadership, Partners Healthcare.

Kaelber, D. C., et al. (2008). *The value of personal health records* (p. 52). Boston/Chicago: Center for Information Technology Leadership, Partners Healthcare/Health Information Management and Systems Society.

Kaufmann, A., Meltzer, M., & Schmid, G. (1997). The economic impact of a bioterrorist attack: Are prevention and post-attack intervention programs justifiable? *Emerging Infectious Diseases, 3*(2), 83–94.

Kendall, D. B. (2009). Protecting patient privacy through health record trusts. *Health Affairs, 28*(2), 444–446.

Kidd, M. R. (2008). Personal electronic health records: MySpace or HealthSpace? *British Medical Journal, 336*, 1029–1030.

Kish, L. J., & Topol, E. J. (2015). Unpatients – why patients should own their medical data. *Nature Biotechnology, 33*, 921–924. Retrieval 29 Oct 2018: http://www.nature.com/nbt/journal/v33/n9/full/nbt.3340.html

Kostyack, P. (2002). The emergence of the healthcare information trust. *Matrix: Journal of Law-Medicine, 12*(293), 339–447.

Krist, A. H., & Woolf, S. H. (2011). A vision for patient-centered health information systems. *Journal of the American Medical Association, 305*(3), 300–301.

Labkoff, S. E., & Yasnoff, W. A. (2007). A framework for systematic evaluation of health information infrastructure progress in communities. *Journal of Biomedical Informatics, 40*(2), 100–105.

Lapsia, V., Lamb, K., & Yasnoff, W. A. (2012). Where should electronic records for patients be stored? *International Journal of Medical Informatics, 81*(12), 821–827.

Lipton, E., & Johnson, K. (2001, December 26). The anthrax trail: Tracking bioterror's tangled course. *New York Times*, Section A, p. 1.

Makary, M. A., & Daniel, M. (2016). Medical error – the third leading cause of death in the US. *British Medical Journal, 353*, i2139.

Mandl, K. D., & Kohane, I. S. (2008). Tectonic shifts in the health information economy. *The New England Journal of Medicine, 358*(16), 1732–1737.

Masys, D. R. (2002). Effects of current and future information technologies on the health care workforce. *Health Affairs, 21*(5), 33–41.

McDonald, C. J., Overhage, J. M., Barnes, M., et al. (2005). The Indiana network for patient care: A working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. *Health Affairs, 24*(5), 1214–1220.

McGlynn, E. A., Asch, S. M., Adams, J., et al. (2003). The quality of health care delivered to adults in the United States. *The New England Journal of Medicine, 348*, 2635–2645.

Menachemi, N., Rahurkar, S., Harle, C. A., et al. (2018). The benefits of health information exchange: An updated systematic review. *Journal of the American Medical Informatics Association, 25*(9), 1259–1265.

Mikk, K. A., Sleeper, H. A., & Topol, E. J. (2017). The pathway to patient data ownership and better health. *Journal of the American Medical Association, 318*(15), 1433–1434.

Miller, H., Yasnoff, W., & Burde, H. (2009). *Personal health records: The essential missing element in twenty-first century healthcare*. Chicago: Health Information and Management Systems Society.

National Committee on Vital and Health Statistics. (2001). *Information for health: A strategy for building the National Health Information Infrastructure. Report and recommendations from the National Committee on Vital and Health Statistics.* Retrieval 31 Oct 2018: https://aspe.hhs.gov/report/information-health-strategy-building-national-health-information-infrastructure

National Research Council, Computer Science and Telecommunications Board. (2001). *Networking health: Prescriptions for the Internet*. Washington, DC: National Academy Press.

Overhage, J. M., Suico, J., & McDonald, C. J. (2001, November). Electronic laboratory reporting: Barriers, solutions and findings. *Journal of Public Health Management and Practice, 7*(6), 60–66.

Porter, E. (2018, March 6). Your data is critical to a robotic age. Shouldn't you be paid for it? *New York Times*. Retrieval 29 Oct 2018: https://www.nytimes.

com/2018/03/06/business/economy/user-data-pay.html

President's Council of Advisors on Science and Technology. (2010). *Realizing the full potential of health information technology to improve healthcare for Americans: The path forward.* Washington, DC: Executive Office of the President (40). Retrieval 31 Oct 2018: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf

President's Information Technology Advisory Committee. (2001, February). *Transforming health care through information technology*. President's Information Technology Advisory Committee: Panel on Transforming Health Care. Retrieval 31 Oct 2018: https://www.nitrd.gov/pubs/pitac/pitac-hc-9feb01.pdf

Ramsaroop, P., & Ball, M. (2000). The "bank of health": A model for more useful patient health records. *MD Computing, 17*, 45–48.

Reuters. (2015, March 17). Premara blue cross says data breach exposed medical data. *New York Times*. Retrieval 29 Oct 2018: http://www.nytimes.com/2015/03/18/business/premera-blue-cross-says-data-breach-exposed-medical-data.html

Saccavini, C., & Greco, F. (2004). *Health current account project*. Presented at EuroPAC MIR (Session 7.19).

Shabo, A. (2006). A global socio-economic-medico-legal model for the sustainability of longitudinal health records. *Methods of information in medicine, 45*, 240–245 (Part 1), 498–505 (Part 2).

Shortliffe, E. H. (2011). Demonstrating value to secure trust. In C. Grossmann, B. Powers, & J. M. McGinnis (Eds.), *Digital infrastructure for the learning health system: 5. weaving a strong trust fabric*. Institute of Medicine. Retrieval 29 Oct 2018: https://www.ncbi.nlm.nih.gov/books/NBK83558/.

Sinsky, C., Colligan, L., Ling, L., et al. (2016). Allocation of physician time in ambulatory practice: A time and motion study in 4 specialties. *Annals of Internal Medicine, 165*(11), 753–760. Retrieval 29 Oct 2018: http://annals.org/aim/article-abstract/2546704/allocation-physician-time-ambulatory-practice-time-motion-study-4-specialties.

Sittig, D. F., & Singh, H. (2020, May 18). COVID-19 and the need for a national health information technology infrastructure. *Journal of the American Medical Association, 323*(23), 2373–2374. https://doi.org/10.1001/jama.2020.7239.

Steinbrook, R. (2008). Personally controlled online health data – The next big thing in medical care? *The New England Journal of Medicine, 358*(16), 1653–1656.

Szolovits, P., Doyle, J., Long, W. J., Kohane, I., & Pauker, S. G. (1994). *Guardian angel: Patient-centered health information systems.* Technical Report MIT/LCS/TR-604, Massachusetts Institute of Technology Laboratory for Computer Science.

Turn, R., Shapiro, N. Z., & Juncosa, M. L. (1976). Privacy and security in centralized vs. decentralized database systems. *Policy Sciences, 7*, 17–29.

Wagner, M. M., Dato, V., Dowling, J. N., & Allswede, M. (2003). Representative threats for research in public health surveillance. *Journal of Biomedical Informatics, 36*(3), 177–188.

Walker, J., Pan, E., Johnston, D., Adler-Milstein, J., Bates, D. W., & Middleton, B. (2004). *The value of healthcare information exchange and lnteroperability*. Boston: Center for Information Technology Leadership, Partners Healthcare.

Weber, G. M. (2013). Federated queries of clinical data repositories: the sum of the parts does not equal the whole. *Journal of the American Medical Informatics Association, 20*, e155–e161. Retrieval 29 Oct 2018: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3715334/

Yasnoff, W. A. (2016). A secure and efficiently searchable health information architecture. *Journal of Biomedical Informatics, 61*, 237–246.

Yasnoff, W. A., & Shortliffe, E. H. (2014). Lessons learned from a health record bank start-up. *Methods of Information in Medicine, 53*, 66–72.

Yasnoff, W. A., Humphreys, B. L., Overhage, J. M., Detmer, D. E., Brennan, P. F., Morris, R. W., Middleton, B., Bates, D. W., & Fanning, J. P. (2004). A consensus action agenda for achieving the national health information infrastructure. *Journal of the American Medical Informatics Association, 11*(4), 332–338.

Yasnoff, W. A., Sweeney, L., & Shortliffe, E. H. (2013). Putting health IT on the path to success. *Journal of the American Medical Association, 309*(10), 989–990.

Yasnoff, W. A., Shortliffe, E. H., & Shortell, S. M. (2014). A proposal for financially sustainable population health organizations. *Population Health Management, 17*(5), 255–256.

Yeager, V. A., Vest, J. R., Walker, D., Diana, M. L., & Menachemi, N. (2017, September 4). Challenges to conducting health information exchange research and evaluation: Reflections and recommendations for examining the value of HIE. *EGEMS (Washington, DC), 5*(1), 15. Retrieval 29 Oct 2018: https://egems.academyhealth.org/articles/10.5334/egems.217/

Management of Information in Health Care Organizations

Lynn Harold Vogel and William C. Reed

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_16

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What are the primary information requirements of health care organizations (HCOs)?
- What are the clinical, financial, and administrative functions provided by health care information systems (HCISs), and what are the potential benefits of implementing such systems?
- How have changes in health care delivery models changed the scope and requirements of HCISs over time?
- How do differences among business strategies and organizational structures influence information systems choices?
- What are the major challenges to implementing and managing HCISs?
- How are ongoing health care reforms, technological advances, and changing social norms likely to affect HCIS requirements in the future?

16.1 Overview

Health care organizations (HCOs), and Integrated Delivery Networks **(IDNs)** like many other business entities, are information-intensive enterprises. Health care personnel require sufficient data and information management tools to make appropriate decisions. At the same time, they need to care for patients and manage and run the enterprise; they also need to document and communicate plans and activities, and to meet the requirements of numerous regulatory and accrediting organizations. *Clinicians* assess patient status, plan patient care, administer appropriate treatments, and educate patients and families regarding clinical management of various conditions. They are also concerned about evaluating the clinical outcomes, quality, and increasingly, the cost of health services provided. *Administrators* determine appropriate staffing levels, manage inventories of drugs and supplies, and negotiate payment contracts for services. *Governing boards* make decisions about whether to invest in new business lines, how to partner with other organizations, and how to eliminate underutilized services. Collectively, health care professionals comprise a heterogeneous group with diverse objectives and information requirements, and in the end, all are expected to be focused on the *patients* who are, after all, the reason for all of this.

The purpose of **health care information systems** (HCISs) is to support the access, processing, and management of the information that health care professionals need to perform their jobs effectively and efficiently. HCISs facilitate communication, integrate information, and coordinate action among multiple health care professionals—and increasingly, patients. In addition, HCISs organize and store substantial amounts of data and they support record-keeping and reporting functions. Many of the clinical information functions of an HCIS were detailed in our discussion of the computer-based patient record (CPR) in ▶ Chap. 14; systems to support nurses and other care providers are discussed in ▶ Chap. 17. Furthermore, HCISs are key elements that interface with the health information infrastructure (HII), as discussed in ▶ Chap. 15. An HCIS also supports the financial and administrative functions of a health organization and associated operating units, including the operations of ancillary and other clinical-support departments. The evolving complexities of HCOs place great demands on an HCIS. Many HCOs are broadening their scope of activities to cover the care continuum, partially in response to **Accountable Care Organization** (**ACO**), **Value-based Purchasing** and **Bundled Payment** initiatives from the federal government. HCISs must organize, manage, and integrate large amounts of clinical and financial data collected by diverse users in a variety of organizational settings (from patient homes to physicians' offices to hospitals to health care systems) and must provide health care workers (and, increasingly, patients) with timely access to complete, accurate, and up-to-date information presented in a useful format. The diversity and extent of the modern IDN is illustrated in ◘ Fig. 16.1.

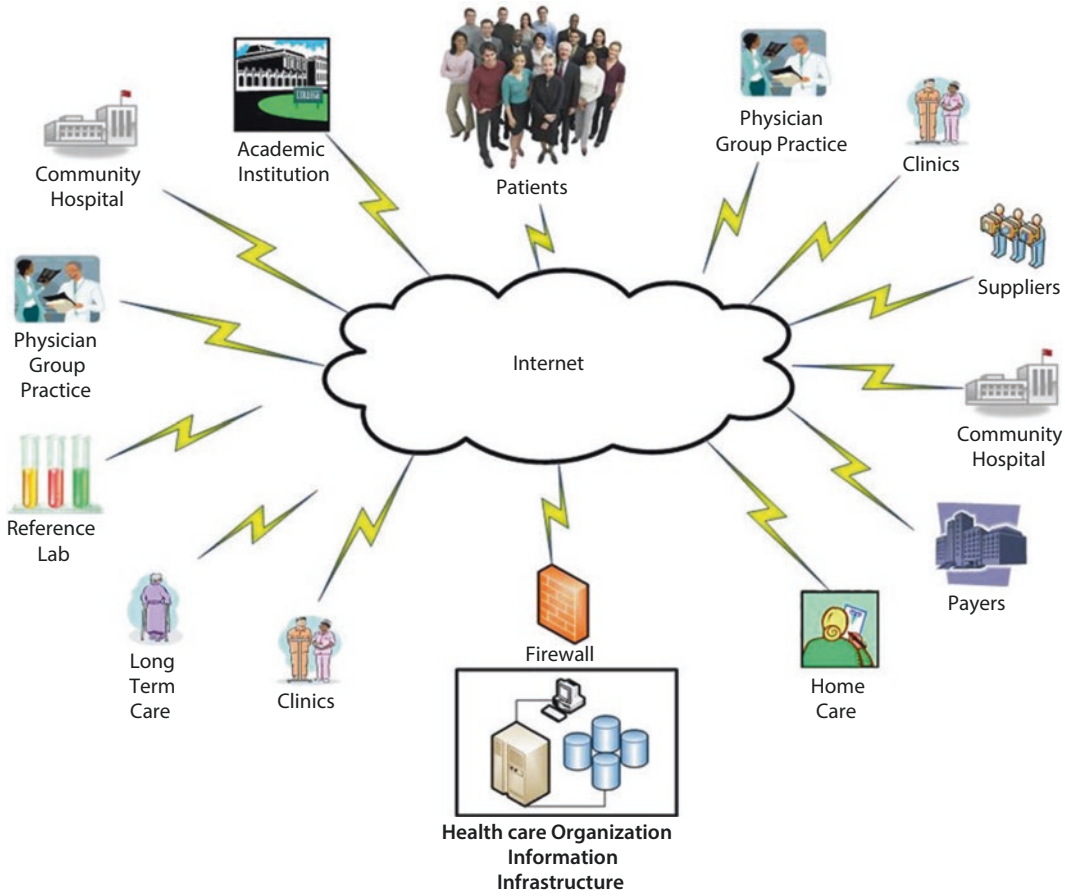


Fig. 16.1 Major organizational components of an integrated delivery network (IDN). A typical IDN might include several components of the same type (e.g., clinics, community hospitals, physician group practices, etc.). Components within the same geographic area may have direct data connections, but increasingly the Internet is the preferred way to connect organizational components

16.2 Historical Evolution of the Technology of Health Care Information Systems (HCISs)

Technological advances as well as changes in the information and organizational requirements of HCOs, have driven many of the changes in system architecture, hardware, software, and functionality of HCISs over time. The tradeoff between functionality and ease of integration is an important factor that influences the choices that vendors have made in systems design.

16.2.1 Central and Mainframe-Based Systems

The earliest HCISs (typically found in hospitals) were designed according to the philosophy that a single comprehensive or **central computer system** could best meet an HCO's information processing requirements. Advocates of the centralized approach emphasized the importance of first identifying all the hospital's information needs and then designing a single, unified framework to meet these needs. Patient management and billing functions were the initial focus of such efforts. One

result of this design goal was the development of systems in which a single large computer performed all information processing and managed all the data files using application-independent file-management programs—initially focusing almost exclusively on financial and billing data. Users accessed these systems via general-purpose **video-display terminals (VDTs)** affectionately known as "green screens" because the displayed numbers and text were often green on a dark background.

Central systems integrated and communicated information well because they provided users with a centralized data store and a single, standardized method to access information simply and rapidly. On the other hand, the biggest limitation of central systems was their inability to accommodate the diverse needs of individual departments. There is a tradeoff between the uniformity (and relative simplicity) of a generalizable system and the nonuniformity and greater responsiveness of custom-designed systems that solve specific problems for specialized departments. Generality—a characteristic that enhances communication and data integration in a homogeneous environment—can be a drawback in an HCO because of the complexity and heterogeneity of the information-management tasks. As a result, central systems have proved too unwieldy and inflexible to support evolving HCO requirements, except in smaller facilities.

16.2.2 Departmental Systems

By the late 1970s, departmental systems had begun to emerge. Advances in technology resulted in decreases in the price of hardware and improvements in software, making it feasible for individual departments within a hospital to acquire and operate their own computers. In a **departmental system**, one or a few computers can be dedicated to processing specific functional tasks within the department. Distinct software application modules carry out specific tasks, and a common framework, which is specified initially, defines the interfaces that will allow data to be shared among the modules. Radiology (▶ Chap. 22) and Laboratory systems are examples of these types of systems. As HCOs increasingly selected the best functionality for their departmental systems, their IT strategy became known as **best of breed**.

The departmental approach responded to many of the challenges of central systems. Although individual departmental systems are constrained to function with predefined interfaces, they do not have to conform to the general standards of an overall system, so they can be designed to accommodate the special needs of specific areas. For example, the processing capabilities and file structures suitable for managing the data acquired from a patient-monitoring system in the intensive-care unit (analog and digital signals acquired in real time) differ from the features that are appropriate for a system that reports radiology results (image and text storage and processing). Furthermore, modification of departmental systems, although laborious with any approach, is simpler because of the smaller scope of the system. The price for this greater flexibility is increased difficulty in integrating data and communicating among modules of the HCISs. In reality, installing a system is never as easy as simply plugging in the connections.

The challenge of sharing data among many different information systems that emerged in the 1980s and 1990s was daunting. As noted earlier, the various components of the HCISs were in most cases developed by different vendors, using different hardware (e.g., DEC, IBM), operating systems (e.g., PICK, Altos, DOS, VMS, MUMPS on minicomputers, and IBM's 360 OS on mainframes) and programming languages (e.g., BASIC, PL/I, COBOL, MUMPS, and even Assembler). Sharing data among two different systems typically required a two-way interface—one to send data from System A to System B, the other to send data or acknowledge receipt from B back to A. Adding a third system didn't require simply one additional interface because the new system would in many cases have to be interfaced to both of the original systems, resulting in the possibility of six interfaces. Introducing a fourth system into the HCIS environment

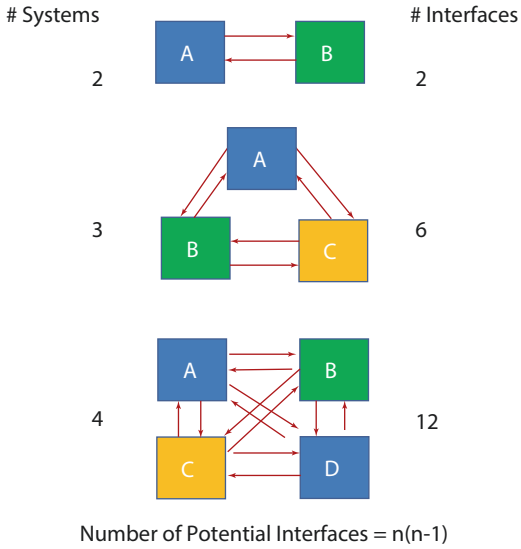


Fig. 16.2 The challenge of moving data from one system to another becomes complicated with the addition of each new system. Considering that even small size hospitals may have several hundred applications, interfacing is a major challenge. While not all systems need to have two-way interfaces to every other system, this figure illustrates the challenges that even small numbers of systems bring

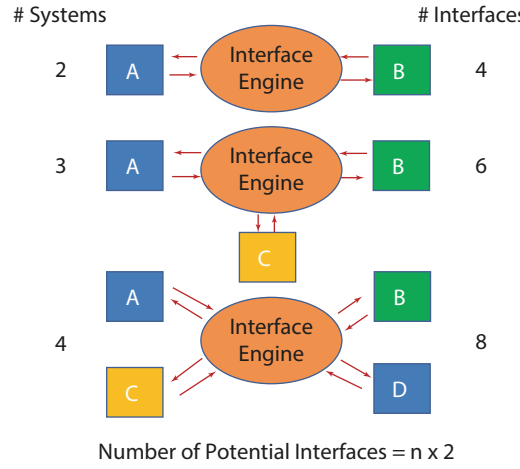


Fig. 16.3 The introduction of the Interface Engine **(IE)** made system interfaces much more manageable, particularly so with the implementation of **HL7** data messaging standards. With an IE, each additional system only added two additional interfaces to the mix, one to send data and one to acknowledge receipt of the data

increased the complexity further, since it often meant the need for two-way interfaces to each of the original three systems, for a total of 12 (Fig. 16.2). With the prospect of interfaces increasing exponentially as new systems were added (represented by the formula, $I = n^{(n-1)}$ where I represents the number of interfaces needed and n represents the number of systems), it was clear that a new solution was needed to address the complexity and cost of interfacing.

In response, an industry niche was born which focused on creating a hardware and software application combination designed to manage the interfacing challenges among disparate systems in the HCIS environment. Instead of each system having to interface to every other system independently, an interface engine served as the central connecting point for all interfaces (Fig. 16.3). Each system had only to connect to the interface engine; the engine would then manage the sending of data to and from any other system that needed it. The interface engine concept, which originated in health care, has given rise to a whole series of strategies for managing multiple systems. Many of the vendors who got their start in health care interfacing subsequently found new markets in financial services as well as other industries.

Continued advancements in departmental systems were made possible as computing power increased and corresponding costs decreased through the advent of microcomputers. Even smaller ancillary departments such as Respiratory Therapy, which previously could not justify a major computer acquisition, could now purchase dedicated file servers and workstations and participate in the HCIS environment. Health care providers in nursing units or at the bedside, physicians in their offices or homes, and managers in the administrative offices could eventually access and analyze data locally using what were initially termed microcomputers (later known as desktop personal computers or PCs).

16.2.3 Integration Challenges

In the early 1980s, researchers at the University of California, San Francisco (UCSF) Hospital successfully implemented one of the first **Local Area Networks (LANs)**

to support communication among several of the hospital's standalone systems. Using technology developed at the Johns Hopkins University, they connected minicomputers that supported patient registration, medical records, radiology, the clinical laboratory, and the outpatient pharmacy. Interestingly, each of the five computer systems was different from the others: the computers were made by different manufacturers and ran different operating systems but were able to communicate with each other through standardized communications protocols.

By the late 1980s, HCISs based on evolving network-communications standards were being developed and implemented in HCOs. As **distributed computer systems**, connected through electronic networks, these HCISs consisted of a federation of independent systems that had been tailored for specific application areas. The computers operated autonomously and shared data (and sometimes programs and other resources, such as printers) by exchanging information over a local area network (LAN) using standard protocols such as **TCP/IP** and **Health Level 7(HL7)** for communication and in many cases utilizing the interface engine strategy we discussed earlier in ▶ Sect. 16.2.2. One advantage of LAN-connected distributed systems was that individual departments could have greater flexibility in choosing hardware and software that optimally suited their specific needs. On the downside, the distribution of information processing capabilities and responsibility for data among diverse systems made the tasks of data integration, communication, and security more difficult—a fact that continues to the present day. Development of industry-wide standard network and interface protocols such as TCP/IP and HL7 has eased the technical problems of electronic communication considerably.

Still, there are problems to overcome in managing and controlling access to a patient database that is fragmented over multiple computers, each with its own file structure and method of file management. Furthermore, when no global architecture or vocabulary standards are imposed on the HCISs, individual departments and entities may encode data values in ways that are incompatible with the definitions chosen by other areas of the organization. The promise of sharing among independent departments, entities, and even independent institutions has increased the importance of defining clinical data standards (see ▶ Chap. 7) As noted earlier, some HCOs pursued a best of breed strategy in which they chose the best system, regardless of vendor and technology, then worked to integrate that system into their overall HCIS environment. Some HCOs modified this strategy by choosing suites of related applications, e.g., selecting all ancillary systems from a single vendor (also known as **best of cluster**), thereby reducing the overall number of vendors they work with and, in theory, reducing the costs and difficulty of integration.

Commercial software vendors have supported this strategy by broadening their offerings of application suites and managing the integration at the suite level rather than at the level of individual applications. For example, Epic, one the of major enterprise-wide HCIS vendors, started in the late 1960s with a focus on physician billing systems, and over the subsequent 40 years evolved into a fully integrated product suite encompassing ambulatory, inpatient, ancillary and billing functionality. Cerner, another major enterprise-wide HCIS vendor, started with a suite of products for ancillary systems such as clinical laboratories, pharmacy and radiology, and over a similar time period evolved into a full-service product suite including inpatient, outpatient and billing functionality.

With the increasing availability of single vendor systems, the need to develop and implement applications that support the specific functionality has diminished significantly. PC-based universal workstations are now the norm and some HCOs and IDNs support thousands of PCs in enterprise-wide networked environments. The requirement for direct access to independent ancillary systems has been largely eliminated not only by enterprise data networks, but by vendors offering integrated product suites which include support for general clinical as well as ancillary service functionality. Where specialized systems remain, interfaces join such systems to

a core clinical system or to a centralized clinical data repository. These permit the ability to access patient databases (by clinicians), human resources documents (by administrators and employees), financial information (by administrators) and basic information about facilities, departments, and staff (by the public) through a single enterprise-wide data network (See ◘ Fig. 16.1).

Historically, as more information systems were added to the HCIS environment, the challenge of moving data from one system to another became overwhelming. In response, beginning in the 1980s, two unique developments occurred: (1) the interface engine; and (2) Health Level Seven (HL7), a standard for the structure of the data messages that were being sent from one information system to another (see ▶ Chap. 7).

The creation of HL7 was yet another response to the challenge of moving data among disparate health care systems. HL7 is a health care-based initiative focused on the development of data messaging standards for the sharing of data among the many individual systems that comprise an HCIS. The basic idea was to use messaging standards so that data could be sent back and forth using standard formats within the HCIS environment. Most of the departmental systems that were introduced at this time were the products of companies focused on specific niche markets, including laboratories, pharmacies and radiology departments. Consequently, there was strong support for both the interface engine and the HL7 efforts as mechanisms to permit smaller vendors to compete successfully in the marketplace. In recent years, many of these pioneering vendors have been purchased and their products included as components of larger single vendor product families.

In hospitals, clinical and administrative personnel have traditionally had distinct areas of responsibility and performed many of their functions separately. Thus, it is not surprising that administrative and clinical data have often been managed separately—administrative data in business offices and clinical data in medical-records departments. When computers were first introduced, the hospital's information processing was often performed on separate computers with separate databases, thus minimizing conflicts about priorities in services and investment. In addition, information systems to support hospital functions and ambulatory care historically have, due to organizational boundaries, developed independently. Many hospitals, for example, have rich databases for inpatient data but maintain less information for outpatients. As **fee-for-service** reimbursement models continue to be challenged for their focus on activity-driven care, alternatives such as **ACOs, bundled payments** for services, and **pay for performance** proposals will stimulate efforts toward greater data integration.

By the late 1980s, **clinical information system** (CIS) components of HISs offered clinically oriented capabilities, such as order writing and results communications. During the same period, **ambulatory medical record systems (AMRSs)** and **practice management systems (PMSs)** were being developed to support large outpatient clinics and physician offices, respectively. These systems performed functions analogous to those of hospital systems, but were generally less complex, reflecting the complexity of patient care delivered in outpatient settings. However, there has historically been little or no systems integration between hospital and ambulatory settings.

The historical lack of integration of data from diverse sources creates a host of problems. If clinical and administrative data are stored on separate systems, then data needed by both must either be entered separately into each system, be copied from one system to another, or data from both sources transferred to yet another location to be analyzed. In addition to the expense of redundant data entry and data maintenance incurred by these approaches (see also the related discussion for the health information infrastructure in ▶ Chap. 15), the consistency of information tends to be poor because data may be updated in one place and not in the other, or information may be copied incorrectly from one place to another. In the extreme example, the same data may be represented differently in different settings. As we noted earlier within the hospital setting, many of these issues have been addressed through the development of

automated interfaces to transfer demographic data, orders, results, and charges between clinical systems and billing systems. Even with an interface engine managing data among disparate systems, however, an organization still must solve the thorny issues of synchronization of data and comparability of similar data types.

With the development of IDNs and other complex HCOs, the sharing of data elements among operating units becomes both more critical and more problematic. Data integration issues are further compounded in IDNs by the acquisition of previously independent organizations that have clinical and administrative information systems incompatible with those of the rest of the IDN. It is still not unusual to encounter minimal automated information exchange among organizations even within an IDN. Patients register and reregister at the physician's office, diagnostic imaging center, ambulatory surgery facility, and acute-care hospital—and sometimes face multiple registrations even within a single facility. Each facility may continue to keep its own clinical records, and shadow files may be established at multiple locations with copies of critical information such as operative reports and hospital discharge summaries. Inconsistencies in these multiple electronic and manual databases can result in inappropriate patient management and inappropriate resource allocation. For example, medications that are first given to a patient while she is a hospital inpatient may inadvertently be discontinued when she is transported to a rehabilitation hospital or nursing home. Also, information about a patient's known allergies and medication history may be unavailable to physicians treating an unconscious patient in an emergency department.

The objectives of coordinated, high-quality, and cost-effective health care cannot be completely satisfied if an organization's multiple computer systems operate in isolation. Unfortunately, free-standing systems within HCOs are still common, although HCOs and IDNs are increasingly investing in the implementation of new more consistent systems across all their facilities or in integrating existing systems to allow data sharing. The capital investment required to pursue a strategy of system-wide data integration can be significant, and with ongoing challenges to reimbursement rates for both hospitals and physicians, the funding to pursue this strategy is often limited either due to competing investment requirements (e.g., acquiring or maintaining buildings and equipment) or the continued downward trend in reimbursement for services.

16.2.4 Evolution to Enterprise-Wide Health Care System Information Systems

If an HCO or IDN is to manage patient care effectively, project a focused market identity, and control its operating costs, it must perform in a unified and consistent manner. For these reasons, information technologies to support data and process integration are recognized as critical to an IDN's or HCO's operations. From an organizational perspective, information should be available when and where it is needed; users must have an integrated view, regardless of system or geographic boundaries; data must have a consistent interpretation; and adequate security must be in place to ensure access only by authorized personnel and only for appropriate uses. Unfortunately, these criteria are much easier to describe than to meet.

Over time, changes in the health care economic and regulatory environments have radically transformed the structure, strategic goals, and operational processes of health care organizations through a gradual shifting of financial risk from third party payers (e.g., traditional insurance companies such as Blue Cross and Blue Shield, Medicare and Medicaid programs that emerged in the 1960s and 1970s) to the providers themselves and in many cases directly to the patient through higher deductibles, co-pays, benefit caps, etc. This shifting of risk initially brought about a consolidation of health care providers into **integrated delivery networks (IDNs)** in the 1990s.

Mergers and acquisitions (M&A) have been an important part of corporate life for close to 200 years; in health care, the M&A experience has been largely felt over the past 20 years. One major difference is that HCOs often form "affiliations" rather than participate in outright merger and acquisition activity—except in situations in which HCOs actually acquire physician practices. So while "M&A" is an appropriate reference for corporate activity outside of health care, "MA&A" (Mergers, Affiliations and Acquisitions) may be a more appropriate term for the health care industry. Cost savings are often touted as the rationale for M&A activity, but that goal is seldom able to be documented and more typical the result is to drive excess capacity from the system (e.g., an oversupply of hospital beds) and to secure market share.

IDNs are still a prominent feature in many health care markets, often driven by new regulatory requirements aimed at improved efficiency while emphasizing greater patient privacy and safety. While the most successful of IDNs have achieved a measure of structural and operational integration, gains from the integration of clinical activities and from the consolidation of HCISs have been much more difficult. Many IDNs scaled back their original goals for integrating clinical activities and began to shed home care services, health plans and managed care entities. Most recently, the pendulum has swung back as IDNs acquire both physician practices and hospitals while shifting their focus to becoming identified as an ACO, as reimbursement constraints and federal **ACO** initiatives strive to improve both the efficiency and effectiveness of HCOs. All these changes have tremendous implications for HCISs.

The expertise gained from managing an inpatient-driven organization producing a relatively large amount of revenue from a relatively small set of events (e.g., a hospital) does not readily translate to the successful management of other organizational activities that in many cases required many more events to produce a similar level of revenue (e.g., from outpatient clinics). In some cases, it was even a challenge to translate management processes from inpatient operations to outpatient clinics, or one hospital to another. Attempts to apply hospital management principles to ambulatory clinics have been challenged because inpatient-based hospitals generate a relatively small number of patient bills with high dollar amounts whereas ambulatory clinics do just the opposite—generate a relatively large number of patient bills, each with a relatively small dollar amount. To date, it is fair to say that few IDNs have gained the degree of cost savings and efficiencies they had originally projected. The immense upfront costs of implementing (or integrating) the required HCISs have contributed to this limited success. Regardless of organizational structure, all health care organizations are striving toward greater information access and integration, including improved information linkages with physicians and patients. The "typical" IDN is a melding of diverse organizations, and the associated information systems infrastructure is still far from integrated; rather, it remains in many cases an amalgam of heterogeneous systems, processes, and data stores (Blackstone & Fuhr Jr. 2003; Kastor 2001; Shortell et al. 2000).

16.2.5 Information Requirements

The most important function of any HCIS is to present data to decision makers so that they can improve the quality and timeliness of the decisions they need to make. From a clinical perspective, the most important function of an HCIS is to present patient-specific data to care givers so that they can easily interpret the data for diagnostic and treatment planning purposes and support the necessary communication among the many health care workers who cooperate in providing health services to patients. From an administrative perspective, the most pressing information needs are those related to the daily operation and management of the organization—bills must be generated accurately and rapidly, employees and vendors must be paid, supplies must be ordered, and so on. In addition, administrators need information to make short-term and long-term planning decisions.

Since clinical system information requirements are discussed in ► Chaps. 14, 17, 21,

and 24, we focus here on operational information requirements, and specifically on four broad categories: daily operations, planning, communication, and documentation and reporting.

- *Operational requirements*. Health care workers—both care givers and administrators—require detailed and up-to-date information to perform the daily tasks that keep a hospital, clinic, or physician practice running—the bread-and-butter tasks of the institution. These include not only clinical activities, but financial management, acquisition and management of supplies, posting charges, sending bills and receiving payments. Queries for operational purposes can include: In what room is patient John Smith? What drugs is he receiving? What tests are scheduled for Mr. Smith during his stay and after his discharge? What insurance coverage does he have? Is the staffing skill mix sufficient to handle the current volume and special needs of patients in Care Center 3 West? What are the names and telephone numbers of patients who have appointments for tomorrow and need to be called for a reminder? What authorization is needed to perform an ultrasound procedure on Jane Blue under the terms of her health insurance coverage? Are the daily, monthly and annual financial reports accurate? Are the charges for services and supplies accurately collected and transferred to the billing system? Are bills sent out in a timely manner and remittances received and posted to the correct accounts? HCISs can support these operational requirements for information by organizing data for prompt and easy access. Because the HCO may have developed product-line specialization within a particular facility (e.g., a diagnostic imaging center or women's health center), however, answering even a simple request may require accessing information stored in different systems at several different facilities.
- *Planning requirements*. Health professionals also require information to make short-term and long-term decisions about patient care and organizational management. The importance of appropriate clinical decision-making is obvious—we devote all of ▶ Chaps. 3 and 24 to explaining methods to help clinicians select diagnostic tests, interpret test results, and choose treatments for their patients. The decisions made by administrators and managers are no less important to their choices concerning the acquisition and use of health care resources. In fact, clinicians and administrators alike must choose wisely in their use of resources to provide high-quality care and excellent service at a competitive price. In addition, HCOs live in a highly regulated world and must report to local, state, federal and private entities on how they manage the care they provide and on the safety of that care. HCISs should help health care personnel (including auditors) to answer queries such as these: What are the organization's clinical guidelines for managing the care of patients with this condition? Have similar patients experienced better clinical outcomes with medical treatment or with surgical intervention? What are the financial and medical implications of closing the maternity service? If we added six care managers to the outpatient-clinic staff, can we improve patient outcomes and reduce emergency admissions? Will the proposed contract to provide health services to Medicaid patients be profitable given the current cost structure and current utilization patterns? How many incidents occurred during the last month, including patient falls, medications that were given to the wrong patient or administered with the wrong dose? How often were supplies such as fire extinguishers or oxygen sources inspected? Often, the data necessary for planning and meeting regulation requirements are generated from many sources. HCISs can assist by aggregating, analyzing, and summarizing the information relevant to decision-making and compliance.
- *Communication requirements*. Communication and coordination of patient care and operations across multiple personnel, multiple business units, and far-flung geography are not possible without investment in an underlying technology infrastruc-

ture. For example, the routing of paper medical records, a cumbersome process even within a single hospital, is an impossibility for a regional network of providers trying to act in coordination. Similarly, it is neither timely nor cost effective to copy and distribute hard copy documents to all participants in a regionally distributed organization. An HCO's technology infrastructure can enable information exchange via web-based access to shared databases and documents, collaborative platforms, electronic mail, document-management systems, and on-line calendaring systems, as well as providing and controlling access for authorized users at the place and time that information is required.

- *Documentation and reporting requirements*. The need to maintain records for future reference or analysis and reporting makes up the fourth category of informational requirements. Some requirements are internally imposed. For example, a complete record of each patient's health status and treatment history is necessary to ensure continuity of care across multiple providers and over time. External requirements create a large demand for data collection and record keeping in HCOs (as with mandated reporting of vaccination records to public health agencies). As discussed in ▶ Chap. 14, the medical record is a legal document. If necessary, the courts can refer to the record to determine whether a patient received proper care. Insurance companies require itemized billing statements, and medical records substantiate the clinical justification of services provided and the charges submitted to them. The **Joint Commission** (**JC**), which certifies the qualifications and performance of many health care organizations, has specific requirements concerning the content and quality of medical records, as well as requirements for organization-wide information-management processes. Furthermore, to qualify for participation in the Medicare and Medicaid programs, the JC requires that hospitals follow standardized procedures for auditing the medical staff and monitoring the quality of patient care, and they must be able to show that HCOs meet the safety requirements for infectious disease management, buildings, and equipment. Employer and consumer groups are also joining the list of external monitors.

16.2.6 Process Integration

To be truly effective, information systems must mesh smoothly with the people who use them and with the specific operational workflows of the organization. But **process integration** poses a significant challenge for HCOs and for the HCIS's as well. Today's health care-delivery models represent a radical departure from historical models of care delivery. Changes in reimbursement and documentation requirements often lead, for example, to changes in the responsibilities and work patterns of physicians, nurses, and other care providers; the development of entirely new job categories (such as care managers who coordinate a patient's care across facilities or between encounters); and the more active participation of patients in their own personal health management (◘ Table 16.1). Process integration is further complicated in that component entities typically have evolved different operational policies and procedures, which can reflect different historical and leadership experiences from one office to another, or in the extreme example, from one floor to another within a single hospital.

The most progressive HCOs are developing new enterprise-wide processes for providing easy and uniform access to health services, for deploying consistent clinical guidelines, and for coordinating and managing patient care across multiple care settings throughout the organization. Integrated information technologies are essential to supporting such enterprise-wide processes. Mechanisms for information management aimed at integrating operations across entities must address not only the migration from legacy systems but also the migration from legacy work processes to new, more consistent and more standardized policies and processes within and across entities.

Table 16.1 The changing health care environment and its implications for an IDN's core competencies. Columns 1 and 2 are used with permission with CSC. Column 3 is the authors' addition

Characteristics	Old care model	Twentieth Century care model	Twenty-first Century care model
Goal of care	Manage sickness	Manage wellness	Prevent illness
Center of delivery system	Hospital	Primary-care providers/ ambulatory settings	Physician offices, retail clinics
Focus of care	Episodic acute and chronic care	Population health, primary care	Preventive care
Driver of care decisions	Specialists	Primary-care providers/ patients	Patients with support of physicians, physician assistants, nurse practitioners
Metric of system success	Number of admissions	Number of enrollees	Patient outcomes
Performance optimization	Optimize individual provider performance	Optimize system-wide performance	Optimize patient outcomes
Utilization controls	Externally controlled	Internally controlled	Value based care
Quality measures	Defined as inputs to system	Defined as patient satisfaction	Value of care provided through measurement of patient outcomes
Physician role	Autonomous and independent	Member of care team; user of system-wide guidelines of care	Guiding physician assistants, providing oversight to nurse practitioners
Patient role	Passive receiver of care	Active partner in care	Primary driver of care

The introduction of new HCISs changes the workplace. Research has shown that in most cases the real value from an investment in information systems comes only when underlying work processes are changed to take advantage of the new information technology (see Figs. 16.4 and 16.5). At times, these changes can be substantial. The implementation of a new system offers an opportunity to rethink and redefine existing work processes to take advantage of the new information-management capabilities, thereby reducing costs, increasing productivity, or improving service levels. For example, providing electronic access to information that was previously accessible only on paper can shorten the overall time required to complete a multistep activity by enabling conversion of serial processes (completed by multiple workers using the same record sequentially) to concurrent processes (completed by the workers accessing an electronic record simultaneously). More fundamental business transformation is also possible with new technologies; for example, direct entry of medication orders by physicians, linked with a decision-support system, allows immediate checking for proper dosing and potential drug interactions, and the ability to recommend less expensive drug substitutes (Vogel 2003).

Few health care organizations today have the time or resources to develop entirely new information systems or redesign processes on their own; therefore, most opt to purchase commercial software products and to use consultants to assist them in the implementation of industry "best practices". Although these commercial systems allow some degree of custom tailoring, they also reflect an underlying model of work processes that may have evolved through development in other health care organizations with different underlying

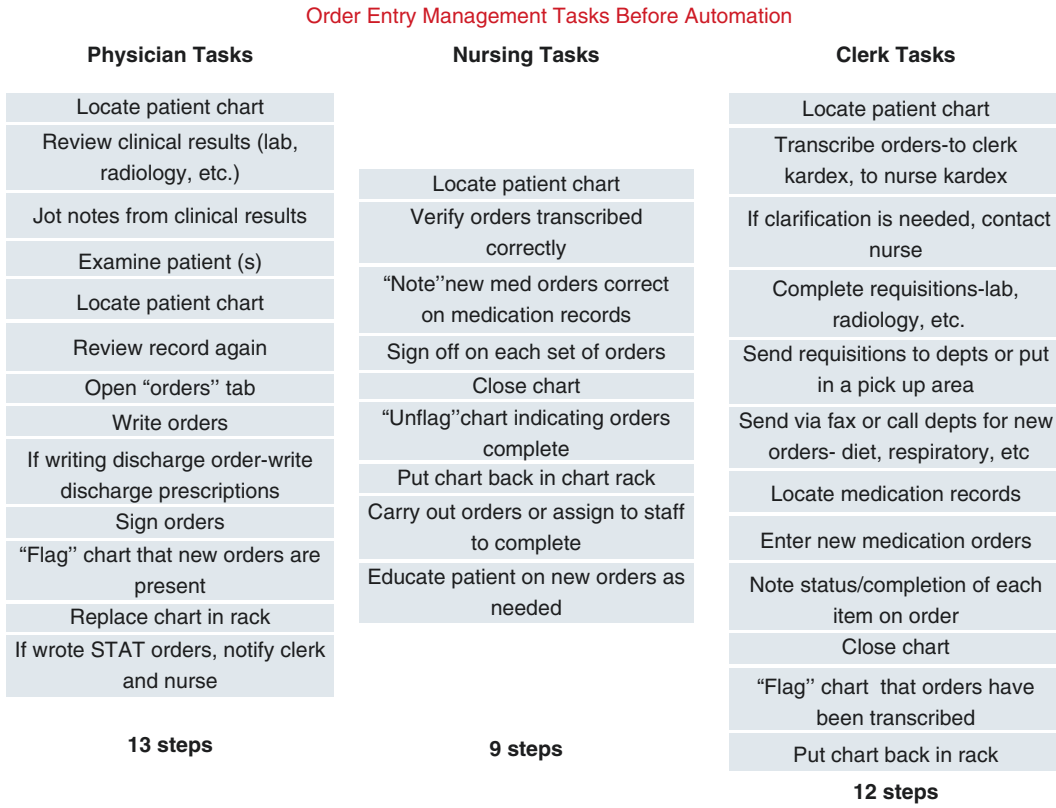


Fig. 16.4 The process of managing the manual creation of orders requesting services on behalf of patients in a hospital involves numerous tasks performed not only by the ordering physician, but by nursing and clerical staff

operational policies and procedures. To be successful, HCOs typically must adapt their own work processes to those embodied in the systems they are installing (For example, some commercial systems require care providers to discontinue and then reenter all orders when a patient is admitted to the hospital after being monitored in the emergency department). Furthermore, once the systems are installed and workflows have been adapted to them, they become part of the organization's culture—and any subsequent change to the new system may be arduous because of these workflow considerations. Decision-makers should take great care when selecting and configuring a new system to support and enhance desired work processes. Such organizational workflow adaptation represents a significant challenge to the HCO and its systems planners. Too often organizations are unable to realize the full potential return on their HCIS investments when they attempt to change the system to accommodate historical work flows, even before the new system is installed. Such management practices can significantly reduce much of the potential gains from the HCO's IT investment.

The **Health Information Technology for Economic and Clinical Health Act of 2009, (HITECH)** signed into law as Title XIII of **American Recovery and Reinvestment Act of 2009** (**ARRA**) economic stimulus bill, provided almost $30 Billion as an incentive for hospitals and physician practices to acquire and implement Electronic Medical Records (**EMR**s). As a result, during subsequent years, over 90% of hospitals and physician practices implemented EMRs. Consolidation among the vendor community, particularly for inpatient products, occurred during the same period with companies like Epic, Cerner, Meditech and Allscripts dominating the market and at the same time enhancing their product suites to incorporate most of the

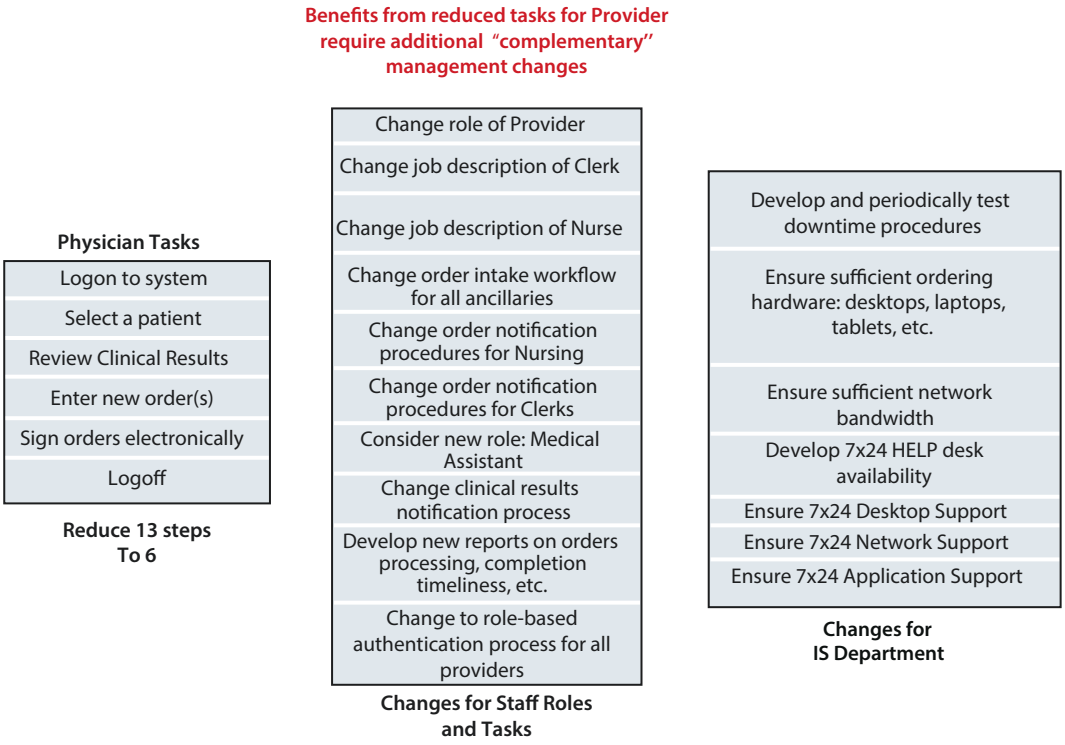


Fig. 16.5 The implementation of an electronic physician order entry system reduces the number of tasks that a physician needs to perform to enter an order, but such a system will only be successful if other "complementary" changes are made to both the workflow of staff and the responsibilities of the IS Department

HCIS functionality needed by hospitals. This has led to the demise of the "best of breed" strategy we mentioned earlier, as hospitals focused more on single vendor solutions. The market for ambulatory systems, on the other hand, has remained highly fragmented, with as many as 700 vendors competing for market share.

To meet the continually evolving financial and quality documentation requirements of today's health care environment, HCOs must continually evolve as well—and the analogy between changing an HCO and turning an aircraft carrier seems apt. Although an HCO's business plans and information-systems strategies may be reasonable and necessary, changing ingrained organizational behavior can be much more complex than changing the underlying information systems. Technology capabilities often exceed an HCO's ability to use them effectively and efficiently. Successful process integration requires not only successful deployment of the technology but also sustained commitment of resources to use that technology well; dedicated leadership with the willingness to make difficult, sometimes unpopular decisions; education; and possibly new performance incentives to overcome cultural inertia and politics. Government incentives to stimulate HCOs toward the **Meaningful Use** of information technology, which emerged from the 2010 Health care Reform legislation are a recent example of attempts to bring process integration and data integration together.

16.2.7 Security and Confidentiality Requirements

HCISs are one of the most frequent targets of hackers. They seek to gain access to health care data, which contains more personalized data (and hence is considered more valuable) than what is typically collected in other industries. Breeches of health care data stored

in both provider organizations and insurance companies have increased significantly over the past several years—often exceeding breeches of financial organization data found in organizations such as banks. In addition, the task of breaking into and stealing health care data has been superseded by ransomware--hackers blocking access to the data and demanding payment (often in bitcoin, which tends to be less traceable than cash payments). Using ransomware is much easier for hackers since data does not need to be moved and then offered for sale; payments come directly from the organization that was hacked in order to obtain the code needed to unblock access.

The protection of health information from unwanted or inappropriate use is governed not only by the trust of patients in their health providers but also by law. In accordance with the **Health Insurance Portability and Accountability Act (HIPAA) of 1996** (► Chap. 12), the Secretary of Health and Human Services recommended that "Congress enact national standards that provide fundamental privacy rights for patients and define responsibilities for those who serve them." This law and subsequent federal regulations now mandate standardized data transactions for sending data to payer organizations, the development and adherence to formal policies for securing and maintaining access to patient data, and under privacy provisions, prohibit disclosure of patient-identifiable information by most providers and health plans, except as authorized by the patient or explicitly permitted by legislation. Subsequent updates to the HIPAA regulations have strengthened considerably the requirements for security and privacy protections and have also given patients the right to pursue actions against both organizations and individuals when they feel that their personal information has been compromised. HIPAA also provides consumers with significant rights to be informed about how and by whom their health information will be used, and to inspect and sometimes amend their health information. Stiff criminal penalties including fines and possible imprisonment are associated with noncompliance or the knowing misuse of patient-identifiable information.

Computer systems can be designed to provide security, but only people can promote the trust necessary to protect the confidentiality of patients' clinical information. In fact, most breeches and inappropriate disclosures stem from human actions rather than from computer system failures. To achieve the goal of delivering coordinated and cost-effective care, clinicians need to access information on specific patients from many different locations. However, it is difficult to predict in advance which clinicians will need access to which patient data and from which locations. And unfortunately, there is an inverse relationship between ease of access and security robustness. Therefore, an HCIS must strike a balance between restricting information access, enabling health care workers to do their jobs, and ensuring the accountability of the users of patient information.

With federal requirements to provide patients with access to their data through portals, the risk of inappropriate access only increases. In addition, more and more devices (e.g., ventilators, monitoring machines, IV pumps) are connected to HCO's internal communications networks, making the **Internet of Things (IoT)**-- a reality in health care. To build trust with its patients and meet HIPAA requirements, an HCO should adopt a three-pronged approach to securing information. First, the HCO needs to designate a security officer (and typically a privacy officer as well) and develop uniform security and confidentiality policies, including specification of sanctions, and to enforce these policies rigorously. Second, the HCO needs to train employees so they understand the appropriate uses of patient-identifiable information and the consequences of violations. Third, the HCO must use electronic tools such as intrusion detection, access controls and audit trails not only to discourage misuse of information, but also to inform employees and patients that people who access confidential information without proper authorization or a "need to know", can be tracked and will be held accountable.

16.2.8 The Impact of Health Care Information Systems

On average, health care workers in administrative departments spend about three-fourths of their time handling information; workers in nursing units spent about one-fourth of their time on these tasks. The fact is that information management in health care organizations, even with significant computerization, is a costly activity in terms of both time and money. The collection, storage, retrieval, analysis, and dissemination of the clinical and administrative information necessary to support the organization's daily operations, to meet external and internal requirements for documentation and reporting, and to support short-term and strategic planning remain important and time-consuming aspects of the jobs of health-care workers.

Today, the justifications for implementing HCISs include cost reduction, productivity enhancements, and quality and service improvement, as well as strategic considerations related to competitive advantage, patient expectations, and regulatory compliance:

- *Cost reduction*. Much of the historical impetus for implementing HCISs was their potential to reduce the costs of information management in hospitals and other facilities largely by reducing the number of employees. HCOs continue to make tactical investments in information systems to streamline administrative processes and departmental workflows. Primary benefits that may offset some information-systems costs include reductions in labor requirements, reduced waste (e.g., dated surgical supplies that are ordered but unused or food trays that are delivered to the wrong destination and therefore are wasted), and more efficient management of supplies and other inventories. Large savings can be gained through efficient scheduling of expensive resources such as surgical suites and imaging equipment. In addition, HCISs can help to eliminate inadvertent ordering of duplicate tests and procedures. Once significant patient data are available online, information systems can reduce the costs of storing, retrieving, and transporting charts in the medical-records department.
- *Productivity enhancements*. A second area of benefit from an HCIS comes in the form of improved productivity of clinicians and other staff. With continuing (and at times increasing) constraints on reimbursements, HCOs are continually faced with the challenge of doing more with less. Providing information systems support to staff can in many cases enable them to manage a larger variety of tasks and data than would otherwise be possible using strictly manual processes. Interestingly, in some cases hospital investments in an HCIS support the productivity improvement of staff that are not employed by the hospital, namely the physicians, and can even extend to payers by lowering their costs. One of the major challenges with introducing a new HCIS is that the productivity of users may actually decrease in the initial months of the implementation. With complex clinical applications, learning new ways of working can lead to high levels of user dissatisfaction in addition to lowered productivity.
- *Quality and service improvement*. As HCISs have broadened in scope to encompass support for clinical processes, the ability to improve the quality of care has become an additional benefit. Qualitative benefits of HCISs include improved accuracy and completeness of documentation, reductions in the time clinicians spend documenting (and associated increases in time spent with patients), fewer drug errors and quicker response to adverse events, and improved provider-to-provider communication. Through telemedicine and remote linkages (see ▶ Chap. 20), HCOs can expand their geographical reach and improve delivery of specialist care to rural and outlying areas. Once patient data are converted from a purely transaction format to a format better suited for analytic work, the use of **clinical decision-support systems** in conjunction with a clinically focused HCIS can produce impressive benefits, namely improving the quality of care

while reducing costs (▶ Chap. 24) (Bates and Gawande 2003; James and Savitz 2011; Goldzweig, et al. 2009; Himmelstein and Woolhandler 2010; McCullough et al. 2010; Castaneda et al. 2015).

- *Competitive advantage*. Information technologies must be deployed appropriately and effectively; however, with respect to HCISs, the question is no longer whether to invest, but rather how much and what to buy. Although some organizations still attempt to cost justify all information-systems investments, many HCOs have recognized that HCISs are "**enabling technologies**" which means that the value comes not from the system itself but from what it "enables" the organization to do differently and better (Vogel 2003). If workflow and processes are not changed to take advantage of the technology, the value of the investment will largely go unrealized. And it is not just the ratio of financial benefits to costs that is important; access to clinical information is necessary not only to carry out patient management, but also to attract and retain the loyalty of physicians who care for (and thus control much of the HCO's access to) the patients. The long-term benefits of clinical systems include the ability to influence clinical practices by reducing large unnecessary variations in medical practices, to improve patient outcomes, and to reduce costs—although these costs might be more broadly economic and societal than related to specific reductions for the hospital itself (Leatherman et al. 2003; James and Savitz 2011; Gottlieb et al. 2015). Physicians ultimately control the great majority of the resource-utilization decisions in health care through their choices in prescribing drugs, ordering diagnostic tests, and referring patients for specialty care. Thus, providing physicians with access to information on "best practices" based on the latest available clinical evidence, as well as giving them other clinical and financial data to make appropriate decisions, is an essential HCIS capability.
- *Regulatory compliance*. Health care is among the most heavily regulated industries in our economy. State and federal regulatory agencies perform a variety of oversight activities, and these require increasingly sophisticated and responsive HCISs to provide the necessary reports. For example, the Food and Drug Administration now mandates the use of barcodes on all drugs. Similarly, HIPAA rules specify how personal health information must be managed as well as the required content and format for certain electronic data transactions for those HCOs that exchange data electronically. OSHA, the Department of Labor, the Environmental Protection Agency, the Nuclear Regulatory Commission, and a host of other agencies all have an interest in seeing that the health care provided by HCOs is consistent with standards of safety and fairness.
- *Consumerism and patient expectations*. As medical breakthroughs promise better outcomes for patients, the patients as consumers are becoming more involved. Federal mandates have driven the creation of **Patient Portals**, essentially dedicated web sites for patients to access and in some cases contribute to their own medical records. While there are limitations to patient involvement (e.g., to date the number of patients accessing and contributing through portals has not met expectations, and older, sicker patients often have limited abilities to access and/or understand their own data)—it is likely that trends toward more patient involvement will continue.

16.2.9 Managing Information Systems in a Changing Health Care Environment

Despite the importance of integrated information systems, implementation of HCISs has proved to be a daunting task, often requiring a multiyear capital investment of hundreds of millions of dollars and forcing fundamental changes in the types and ways that health care professionals perform their jobs. To achieve the potential benefits, HCOs must plan carefully and invest wisely. The grand challenge

for an HCO is to implement an HCIS that is sufficiently flexible and adaptable to meet the changing needs of the organization. Given the rapidly changing environment and the multiyear effort involved, people must be careful to avoid implementing a system that is obsolete functionally or technologically before it becomes operational. Success in implementing an HCIS entails consistent and courageous handling of numerous technical, organizational, and political challenges.

16.2.10 Changing Technologies

As we discussed in ▶ Chap. 6, dramatic changes in computing and networking technologies are continuing to occur. These advances are important in that they allow quicker and easier information access, less expensive computational power and data storage, greater flexibility, and other performance advantages. A major challenge for many HCOs is how to decide whether to support a best of breed strategy, with its requirement either to upgrade individual systems and interfaces to newer products or to migrate from their patchwork of legacy systems to a more integrated systems environment. Such migration requires integration and selective replacement of diverse systems that are often implemented with closed or nonstandard technologies and medical vocabularies. Unfortunately, the trade-off between migrating from best of breed to more integrated systems is that vendors offering more integrated approaches seldom match the functionality of historical best of breed environments, although with each new software version the gap lessens. As a result, best of breed strategies are becoming less of an option since commercial vendors are broadening and deepening the scope of their application suites to minimize the challenges of building and managing interfaces and to protect their market share. In a sense, it is the information content of the systems and the ability to implement them that is much more important than the underlying technology—as long as the data are accessible, the choice of specific technology is less critical.

16.2.11 Changing Culture

In the current health care environment, physicians are confronted with significant obstacles to the practice of medicine as they have historically performed it. Physicians' long history of entrepreneurial practice is changing as more and more physicians work directly as employees of HCOs and fewer and fewer own their own practices. As a result, physicians face significant adjustments as they are confronted by pressures to practice in accordance with institutional standards aimed at reducing variation in care, and to focus on the costs of care regardless of whether those costs are borne by HCOs or by third party payers. They are expected to assume responsibility not simply for healing the sick, but for the wellness of people who come to them not as patients but as members of health plans and health maintenance organizations. In addition, they must often work as members of collaborative patient-care teams. The average patient length of stay in a hospital is decreasing; at the same time, the complexity of the care provided both during and after discharge is increasing. The time allotted for an individual patient visit in an ambulatory setting is decreasing as individual clinicians face economic incentives to increase the number of patients for whom they care each day. Some HCOs, aided by federal funding incentives, are now instituting **pay-for performance** incentives to reward desired work practices. At the same time, it is well known that the amount of knowledge about disease diagnosis and treatment increases significantly each year, with whole new areas of medicine being added from major breakthroughs in areas such as genomic and imaging research. To cope with the increasing workload, greater complexity of care, extraordinary amounts of new medical knowledge, new skills requirements, and the wider availability of medical knowledge to consumers through the Internet, both clinicians and health executives must become more effective information managers, and the supporting HCISs must meet their ever more complex workflow and information requirements. As the health care culture and the roles

of clinicians and health executives continue to change, HCOs must constantly reevaluate the capabilities of information technology to ensure that the implemented systems continue to match user requirements and expectations.

16.2.12 Changing Processes

Developing a new vision of how health care will be delivered and managed, designing processes and implementing supporting information systems are all critical to the success of evolving HCOs. Changes in workflow processes affect the jobs that people do, the skills required to do those jobs, and the fundamental ways in which they relate to one another. For example, models of care management that cross organizational or specialty boundaries encourage interdisciplinary care teams to work in harmony to promote health as well as treat illness. Although information systems are not the foremost consideration for people who are redesigning processes, a poor information-systems implementation can institutionalize bad processes.

HCOs periodically undertake various process redesign initiatives (following models such as **Six Sigma, Kaizen** or **LEAN**), and these initiatives can lead to fundamental transformations of the enterprise. Indeed, work process redesign is essential if information systems are to become truly valuable "enablers" in HCOs. Too often, however, the lack of a clear understanding of existing organizational dynamics leads to a misalignment of incentives—a significant barrier to change—or to the assumption that simply installing a new computer system will be sufficient to generate value. Moreover, HCOs, like many organizations, are collections of individuals who often have natural fears about and resistance to change. Even under the best of circumstances, there are limits to the amount of change that any organization can absorb. The magnitude of work required to plan and manage organizational change is often underestimated or ignored. The handling of people and process issues has emerged as one of the most critical success factors for HCOs as they implement new work methods and new and upgraded information systems.

16.2.13 Changing Sources of Data

Historically the sources of data for electronic systems in health care were relatively limited. Transactions resulting from activities like lab tests, radiological studies, medications prescribed and given, and specific clinical activities such as inpatient encounters with physicians or outpatient visits were recorded and charges for these transactions were then posted to a patient's account. Once the inpatient stay or the ambulatory visit was completed, the charges were collected into a bill which was then sent to the patient's third-party payer (e.g., Medicare, Medicaid or a private insurance company) or to the patient directly. The amount of actual data processed from these transactions is relatively small. Even a complex series of clinical encounters between a patient and a physician can be captured in as little as 4–500 kb of data, or the equivalent of about 200 pages of text.

In recent years, however, there has been an explosion of data, both in terms of *volume* and of *complexity*. Data sources that create images (e.g., radiological studies, pathology slides) can generate data that easily grows. While an individual projection image can range from 8 to 32 MB in size, and a digital mammography image can range from 8 to 50 MB, CT and MRI scans generate relatively small individual images but complete studies can include literally thousands of images and grow to GBs in size. As imaging technology continues to develop with both more detailed individual images and greater number of images per study, the amount of data being collected and stored can be expected to grow exponentially.

In addition to the growth in data from these more traditional modalities, new data sources are being added to data already being collected and stored. As a result, HCISs must continue to evolve to incorporate access to this data. Since the completion of the initial Human Genome Sequencing project in 2003, for example, data from genomic studies has become an increasingly important source for clinicians. Sequencing machines today can produce a million times more data than what was collected in 2004, and more sequences

can be run in an hour than were produced in the previous decade. Not only is this a huge amount of data, it is in a format quite different from historical clinical data—hence the increase in *complexity* as well as *volume*.

We are also seeing an explosion of data from another source: wearables. These are devices that individuals wear on their arm or wrist, collecting data on various aspects of a person's physical health. It has been predicted that over the next few years, there could be as many as 6–700 million wearable devices worldwide.

One example of a device measuring health data is the continuous glucose monitor (CGM). CGMs can capture close to 300 data points each day or approximately 110,000 data points annually. With close to 30 million people diagnosed with diabetes in the US, even if only half of those use a CGM, there could be close to 3.3E12 data points collected annually.

New sources of data are being created almost constantly. As electronic sensors and connectivity are increasingly embedded in devices, many of them designed to monitor and manage the health of individuals, the amount of data collected will continue to grow. With the Internet as the backbone, there is almost no limit to what can be measured and connected electronically resulting in a true **Internet of Things (IoT)**.

The challenge of an HCIS is to enable clinicians to access these new sources of data, to understand their importance and relevance, and to then use them to enhance their diagnostic and therapeutic capabilities, leading hopefully to better outcomes for their patients.

16.2.14 Management and Governance

Figure 16.6 illustrates the information-technology environment of an HCO composed of two hospitals, an owned physician practice, affiliated nursing homes and hospice, and several for-profit service organizations. Even this relatively simple environment presents significant challenges for the management and governance of information systems. For example, to what extent will the information management function be controlled centrally versus decentralized to the individual operating units and departments? How should limited resources be allocated between new investment in strategic projects (such as office-based data access for physicians) and the often critical operational needs of individual entities (e.g., replacement of an obsolete laboratory information system)? Academic medical centers with distinct research and educational needs raise additional issues for managing information across operationally independent and politically powerful constituencies.

Trade-offs between functional and integration requirements, and associated contention between users and information-systems departments, will tend to diminish over time with the development and widespread adoption of technology standards and common clinical-data models and vocabulary. On the other hand, an organization's information-systems "wants" and "needs" will always outstrip its ability to deliver these services. Political battles will persist, as HCOs and their component operating units wrestle with the age-old issues of how to distribute scarce resources among competing, similarly worthy projects.

A formal HCIS governance structure with representation from all major constituents provides a critical forum for direction setting, prioritization, and resource allocation across an HCO. Leadership by respected clinical peers has proved a critical success factor for clinical systems planning, implementation, and acceptance. In addition, the creation of an Information Systems Advisory or Steering Committee composed of the leaders of the various constituencies within the HCO, can be a valuable exercise if the process engages the organization's clinical, financial, and administrative leadership and users and results in their gaining not only a clear understanding of the highest-priority information technology investment requirements but also provides a sense of accountability and ownership over the HCISs and their various functions (Vogel 2006). This supports one of the principles of information technology governance: *how* an institution makes IT investment deci-

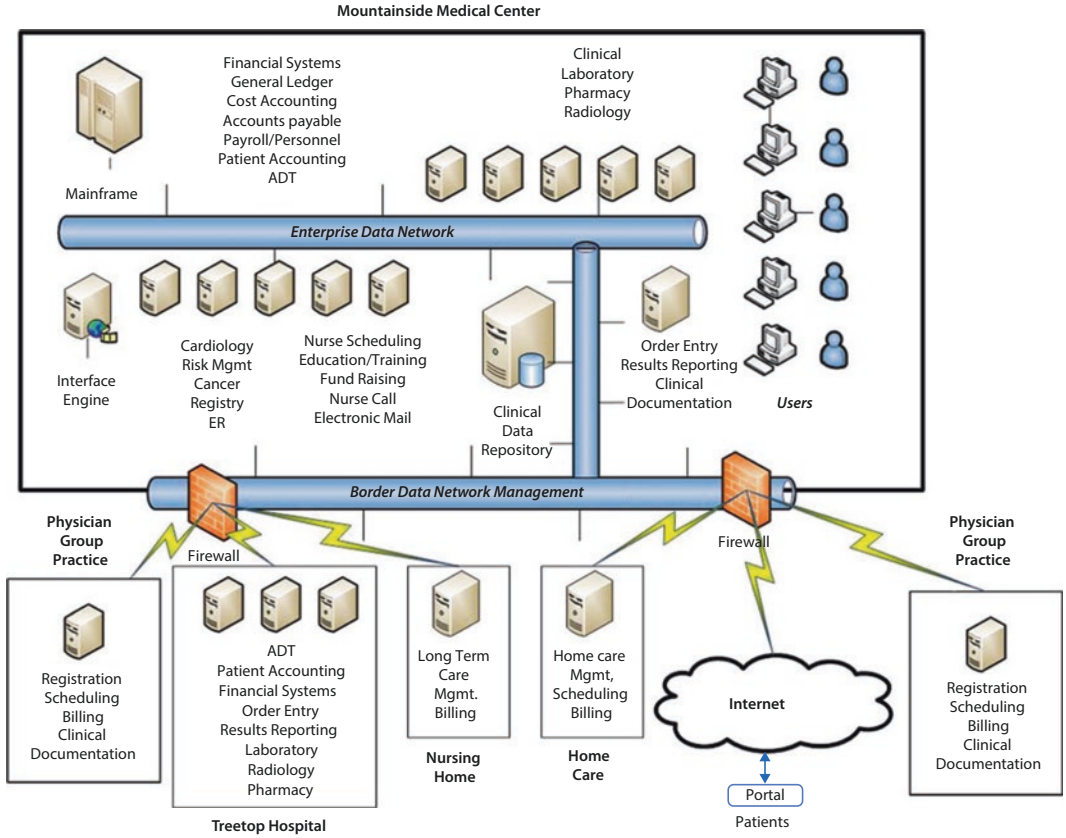


Fig. 16.6 An example of an information systems environment for a small integrated delivery network (IDN). Even this relatively simply IDN has a complex mix of information systems that pose integration and information management challenges for the organization

sions is often more important than *what* specific decisions are made (Weill and Ross 2004; Haddad et al. 2018). Because of the dynamic nature of both health care business strategies and the supporting technologies, many HCOs have seen the timeframes of their strategic information-management thinking shrink from 5 years to three, and then be changed yet again through annual updates.

16.3 Functions and Components of a Health Care Information System

Carefully designed computer-based information systems can increase the effectiveness and productivity of health professionals, improve the quality and reduce the costs of health services, and improve levels of service and of patient satisfaction. As described in ▶ Sect. 16.1, the HCISs support a variety of functions, ranging from the delivery and management of patient care to the administration of the HCO. From a functional perspective, HCISs typically consist of components that support five distinct purposes: (1) patient management and billing, (2) ancillary services, (3) care delivery and clinical documentation, (4) clinical decision support, (5) institutional financial and resource management (Fig. 16.7).

16.3.1 Patient Management and Billing

Systems that support patient management functions perform the basic HCO operations related to patients, such as registration, scheduling, admission, discharge, transfer among

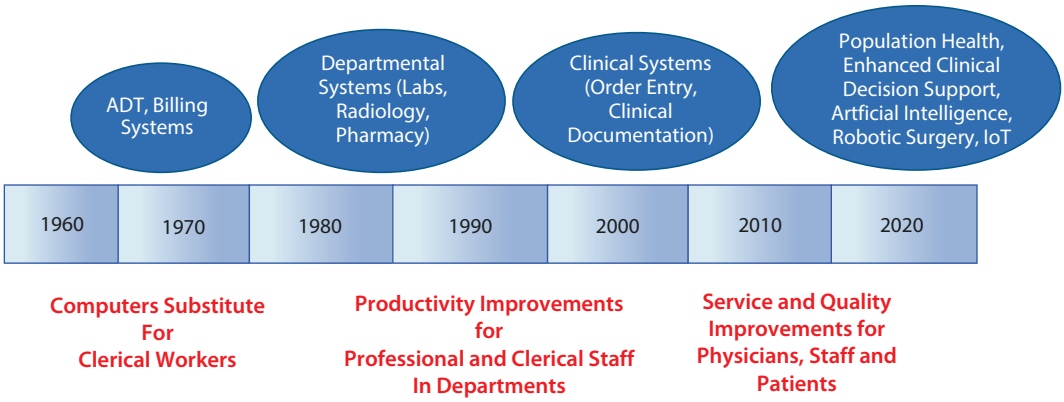


Fig. 16.7 The evolution of computing systems in hospitals has followed a path that parallels the evolution of computing systems in general. From mainframes to minicomputers to desktops, and more recently mobile devices, the purpose and function of systems in hospitals has followed a path from financial systems to departmental systems to systems designed specifically to enhance the productivity and raise the quality of health care services

locations, and billing. Historically within HCOs, maintenance of the hospital census and a patient billing system were the first tasks to be automated—largely because a patient's location determined not only the daily room/bed charges (since an ICU bed was more expensive than a regular medical/surgical bed) but where medications were to be delivered, and where clinical results were to be posted. Today, virtually all hospitals and ambulatory centers and many physician offices use a computer-based **master patient index** (**MPI**) to store patient-identification information that is acquired during the patient-registration process, and link to simple encounter-level information such as dates and locations where services were provided. The MPI can also be integrated within the registration module of an ambulatory care or physician-practice system or even elevated to an **enterprise master patient index** (**EMPI**) across several facilities. Within the hospital setting, the census is maintained by the **admission–discharge–transfer** (**ADT**) module, which is updated whenever a patient is admitted to the hospital, discharged from the hospital, or transferred from one bed to another.

Registration and patient census data serve as a reference base for the financial programs that perform billing functions. When an HCIS is extended to other patient-care settings—e.g., to the laboratory, pharmacy, and other ancillary departments—patient-management systems provide a common reference base for the basic patient demographic data needed by these systems. Without access to the centralized database of patient financial, demographic, registration and location data, these subsystems would have to maintain duplicate patient records. In addition, the transmission of registration data can trigger other activities, such as notification of hospital housekeeping when a bed becomes available after a patient is discharged. The billing function in these systems serves as a collection point for all the chargeable patient activity that occurs in a facility, including room/bed charges, ancillary service charges, and supplies used during a patient's stay.

Scheduling in a health care organization is complicated because patient load and resource utilization can vary by day, week, or season or even through the course of a single day simply due to chance, emergencies that arise, or to patterns of patient and physician behavior. Effective resource management requires that the appropriate resources be on hand to meet such fluctuations in demand. At the same time, resources should not remain unnecessarily idle since that would result in their inefficient use. The most sophisticated scheduling systems have been developed for the operating rooms and radiology departments, where scheduling challenges include matching the patient not only with the providers but also with special equipment and support staff

such as technicians. **Patient-tracking applications** monitor patient movement in multistep processes; for example, they can monitor, estimate, and manage patient wait times in the emergency department.

Within a multi-facility HCO, the basic tasks of patient management are compounded by the need to manage patient care across multiple settings, some of which may be supported by independent information systems. Is the Patricia C. Brown who was admitted last month to Mountainside Hospital the same Patsy Brown who is registering for her appointment at the Seaview Clinic? Integrated delivery networks ensure unique patient identification either through conversion to common registration systems or, more frequently, through implementation of an EMPI that links patient identifiers and data from multiple registration systems.

16.3.2 Ancillary Services

Ancillary departmental systems support the information needs of individual clinical departments within an HCO. From a systems perspective, those areas most commonly automated are the laboratory, pharmacy, radiology, blood-bank, operating rooms, and medical-records departments, but can also include specialized systems to support cardiology (for EKGs), respiratory therapy and social work. Such systems serve a dual purpose within an HCO. First, ancillary systems perform many dedicated tasks required for specific departmental operations. Such tasks include generating specimen-collection lists and capturing results from automated laboratory instruments in the clinical laboratory, printing medication labels and managing inventory in the pharmacy, and scheduling examinations and supporting the transcription of image interpretations in the radiology department. In addition, information technology coupled with robotics can have a dramatic impact on the operation of an HCO's ancillary departments, particularly in pharmacies (to sort and fill medication carts) and in clinical laboratories (where in some cases the only remaining manual task is the collection of the specimen and its transport to the laboratory's robotic system). Second, the ancillary systems contribute major data components to online patient records, including laboratory-test results and pathology reports, medication profiles, digital images (see ▶ Chap. 22), records of blood orders and usage, and various transcribed reports including history and physical examinations, operating room and radiology reports. HCOs that consolidate ancillary functions outside hospitals to gain economies of scale—for example, creating outpatient diagnostic imaging centers and reference laboratories—increase the complexity of integrated patient management, financial, and billing processes.

16.3.3 Care Delivery and Clinical Documentation

Electronic medical record **(EMR)** systems that support care delivery and clinical documentation are discussed at length in ▶ Chap. 14. Although comprehensive EMRs are the ultimate goal of most HCOs, many organizations today are still building more basic clinical-management capabilities. Automated **order entry** and **results reporting** are two important functions provided by the clinical components of an HCIS. Health professionals can use the HCIS to communicate with ancillary departments electronically, eliminating the easily misplaced paper slips or the transcription errors often associated with translating handwritten notes into typed requisitions, thus minimizing delays in conveying orders. The information then is available online, where it is easily accessible by any authorized health professional that needs to review a patient's medication profile or previous laboratory-test results. Ancillary departmental data represent an important subset of a patient's clinical record. A comprehensive clinical record, however, also includes various data that clinicians have collected by questioning and observing the patient, including the history and physical report, progress notes and problem lists. In the hospital, an HCIS can help health personnel perform an initial assessment when a patient is admitted to a unit, maintain patient-specific care plans, chart vital signs, maintain medication-administration records, record

diagnostic and therapeutic information, document patient and family teaching, and plan for discharge (also see ► Chap. 17). Many organizations have developed diagnosis-specific **clinical pathways** that identify clinical goals, interventions, and expected outcomes by time period. Using clinical pathways, case managers or care providers can document actual versus expected outcomes and are alerted to intervene when a significant unexpected event occurs. More hospitals are now implementing systems to support what are called **closed loop medication management systems** in which every task from the initial order for medication to its administration to the patient is recorded in an HCIS—one outcome of increased attention to patient safety issues.

With the shift toward delivering more care in outpatient settings, clinical systems have become more common in ambulatory clinics and physician practices. Numerous vendors have introduced software compatible with **smart phones**, **tablets**, and other mobile devices designed specifically for physicians in ambulatory settings, so that they can access appropriate information even as they move from one exam room to another. Such systems allow clinicians to record problems and diagnoses, symptoms and physical examinations, medical and social history, review of systems, functional status, active and past prescriptions, provide access to therapeutic and medication guidelines, etc. The most successful systems are integrated with a practice management system, providing additional support for physician workflow and typical clinic functions, for example, by documenting telephone follow-up calls or printing prescriptions. In addition, specialized clinical information systems have been developed to meet the specific requirements of intensive-care units (see ► Chap. 21), long-term care facilities, home-health organizations, and specialized departments such as cardiology and oncology.

16.3.4 Clinical Decision Support

Clinical decision-support systems (► Chap. 24) directly assist clinical personnel in data interpretation and decision-making. Once the basic clinical components of an HCIS are well developed, clinical decision-support systems can use the information stored there to monitor patients and issue alerts, to make diagnostic suggestions, to provide limited therapeutic guidance, and to provide information on medication costs. These capabilities are particularly useful when they are integrated with other information-management functions. For example, a useful adjunct to **computer-based physician order-entry** (**CPOE**) is a decision-support program that alerts physicians to patient food or drug allergies; helps physicians to calculate patient-specific drug-dosing regimens; performs advanced order logic, such as recommending an order for prophylactic antibiotics before certain surgical procedures; automatically discontinues drugs when appropriate or prompts the physician to reorder them; suggests more cost-effective drugs with the same therapeutic effect; or activates and displays applicable clinical-practice guidelines (see ► Chap. 24). **Clinical-event monitors** integrated with results-reporting applications can alert clinicians to abnormal results and drug interactions by electronic mail, text message or page. In the outpatient setting, these event monitors may produce reminders to provide preventive services such as screening mammograms and routine immunizations. The same event monitors might trigger access to the HCO's approved formulary, displaying information that includes costs, indications, contraindications, approved clinical guidelines, and relevant online medical literature (Kaushal et al. 2003). Unfortunately, an overabundance of alerts has also caused care givers to experience alarm fatigue resulting in them ignoring potentially critical warnings.

16.3.5 Financial and Resource Management

Financial and administrative systems assist with the traditional business functions of an HCO, including management of the payroll, human resources, general ledger, accounts payable, and materials purchasing and inventory. Most of these data-processing tasks are well structured and have been historically labor intensive and repetitious—ideal opportunities

for substitution with computers. Furthermore, with the exception of patient-billing functions, the basic financial tasks of an HCO do not differ substantially from those of organizations in other industries. Not surprisingly, financial and administrative applications have typically been among the first systems to be standardized and centralized in IDNs.

Conceptually, the tasks of creating a patient bill and tracking payments are straightforward, and financial transactions such as claims submission and electronic funds transfer have been standardized to allow **electronic data interchange** (**EDI**) among providers and payers. In operation, however, patient accounting requirements are complicated by the myriad and oft-changing reimbursement requirements of government and third-party payers. These requirements vary substantially by payer, by insurance plan, by type of facility where service was provided, and often by state. As the burden of financial risk for care has shifted from third party payers to providers (through **per diem** or **diagnosis-based reimbursements**), these systems have become even more critical to the operation of a successful HCO. As another example, managed care contracts add even more complexity, necessitating processes and information systems to check a patient's health-plan enrollment and eligibility for services, to manage referrals and preauthorization for care, to price claims based on negotiated contracts, and to create documentation required to substantiate the services provided.

As HCOs increasingly go "at risk" for delivery of health services by negotiating **per diem**, **diagnosis-based**, **bundled** and **capitated payments**, their incentives need to focus not only on reducing the cost per unit service but also on maintaining the health of members while using health resources effectively and efficiently. Similarly, the HCO's scope of accountability broadens from a relatively small population of sick patients to a much larger population of plan members (such as might be found in **ACO**s), most of whom are still well.

Provider-profiling systems support utilization management by tracking each provider's resource utilization (costs of drugs prescribed, diagnostic tests and procedures ordered, and so on) compared with severity-adjusted outcomes of that provider's patients such as their rate of hospital readmission and mortality by diagnosis. Such systems are also being used by government bodies and consumer advocate organizations as they publicize their findings, often through the Internet. **Contract-management systems** have capabilities for estimating the costs and payments associated with potential managed care contracts and comparing actual with expected payments based on the terms of the contracts. More advanced managed-care information systems handle **patient triage** and medical management functions, helping the HCOs to direct patients to appropriate health services and to proactively manage the care of chronically ill and high-risk patients. Health plans, and IDNs that incorporate a health plan, also must support payer and insurance functions such as claims administration, premium billing, marketing, and member services.

As HCOs continue to seek ways to reduce their expenses, they are introducing both more automation and more integration into "back office" systems. Enterprise Resource Planning (**ERP**) systems integrate human resources functions with payroll, accounts receivable and payment systems and overall supply management. ERP systems can be effectively integrated with other components of the HCIS, e.g., nurse staffing systems and surgical scheduling systems, in an attempt to use information to optimize operations and resource utilization. Implementing these systems can bring challenges similar to clinical systems, albeit with different types of employees. Long term employees often develop their own styles of managing information based on historical patterns and preferences, much of which must change when an ERP system is implemented.

16.4 Forces That Will Shape the Future of Health Care Information Systems Management

As we have discussed throughout this chapter, the changing landscape of the health-care industry and the strategic and operational

requirements of HCOs and IDNs have accelerated the acquisition and implementation of HCISs. The acquisition and implementation of **Electronic Medical Records** (**EMRs**) have been a particular focus, especially with the availability of federal stimulus funding through the provisions of the **Health Information Technology for Economic and Clinical Health** (**HITECH**) Act under the **American Recovery and Reinvestment Act of 2009** (**ARRA**). Although there are many obstacles to implementation and acceptance of smoothly functioning, fully integrated HCISs, few people today would debate the critical role that information technology plays in an HCO's success or in an IDN's efforts at clinical and operational management.

We have emphasized the dynamic nature of today's health care environment and the associated implications for HCISs. A host of new requirements loom that will challenge today's available solutions. We anticipate additional expectations and requirements associated with the changing organizational landscape, technological advances, and broader societal changes.

16.4.1 Changing Organizational Landscape

Although the concepts underlying HCOs and IDNs are no longer new, the underlying organizational forms and business strategies of these complex organizations continue to evolve. The success of individual HCOs varies widely. Some, serving target patient populations such as those with heart disease or cancer or age-defined groups such as children, have been relatively more successful financially than those attempting to serve patients across a wide range of illnesses or those attempting to combine diverse missions of clinical care, teaching and research. IDNs, on the other hand, have by and large failed to achieve the operational improvements and cost reductions they were designed to deliver. It is possible that entirely new forms of HCOs and IDNs will emerge in the coming years. Key to understanding the magnitude of the information systems challenge for IDNs in particular is recognizing the extraordinary pace of change—IDNs reorganize, merge, uncouple, acquire, sell off, and strategically align services and organizational units in a matter of weeks. While information technology is itself changing with accelerating frequency, today's state-of-the-art systems (computer systems and people processes) typically require months or years to build and refine.

The continuing pressures of the market place, including the downward trend in reimbursements and the evolving efforts by the Federal government to change the payment structure from **Procedure-based Payments** to **Value-Based Payments** has led to merger and acquisition activity across the health care industry. Although the outright merger or acquisition of one HCO by another HCO is still relatively infrequent, HCO's and IDNs have increasingly targeted physician practices for acquisition in order both to control the flow of patients to inpatient beds and to increase the standardization of care protocols across physician groups. In addition, in situations in which an outright merger or acquisition seems likely to be unsuccessful, affiliations between HCOs are an increasingly common strategy to ensure better coordination across historically competing boundaries.

All too frequently, business deals are cut with insufficient regard to the cost and time required to create the supporting information infrastructure. For IDNs even in the best of circumstances, the cultural and organizational challenges of linking diverse users and care-delivery settings will tax their ability to change their information systems environments quickly enough. These issues will increase in acuity as operational budgets continue to shrink—today's HCOs and IDNs are spending significant portions of their capital budgets on information-systems investments. In turn, these new investments translate into increased annual operating costs (costs of regular system upgrades, maintenance, user support, and staffing). Still most health care organizations devote at most 3–4% of their total revenues to their information systems operating budgets; in other information-intensive industries (e.g., financial services, air transportation), the percentage of operating budgets devoted to information technology investment can be three to four times higher (Weil 2001).

16.4.2 Changes Within the HCIS Organization

Information technology in the health care industry became a major focus for investment starting in the late 1980s and early 1990s, as developments in both hardware and software technology led both to increased affordability and enhanced functionality. We noted earlier the evolution from mainframes to minicomputers to networked PCs, as well as the transformation of application suites from enterprise-wide financial and billing systems to more departmentally-focused information technology investment decisions in provider organizations. This evolution led to an ever-more complex computing environment encompassing a diversity of hardware and software and networking capabilities to enable everything to work together. Organizational capabilities to manage these changes evolved as well, from a small group of technicians managing a mainframe environment to a growing number not only of technicians, but groups of dedicated application support staff, networking engineers and specialists in desktop support. Information Services Departments began to replace "Data Processing Departments" and their reach and importance to the organization. Leaders were no longer department directors but "Chief Information Officers" **(CIOs),** charged with responsibilities as diverse as keeping the systems running on a daily basis, overseeing the seemingly endless upgrades to both hardware and software, aligning information technology investments with business strategy, and keeping the personal data of patients secure from unauthorized access. As "CIOs" they were expected to lead the procurement of new hardware and software in conjunction with other "C-Suite" executives or physician leaders of the various departments seeking to introduce computers into their work environments. Most often these were systems supporting administrative and financial processes or ancillary departments like the laboratory or the radiology or pharmacy departments.

However, with the 1991 publication of The Institute of Medicine's *The Computer-Based Patient Record: An Essential Technology for Health Care* (Institute of Medicine 1991) an entire new era of information technology began to emerge with the goal of capturing all clinical activity electronically, giving rise to Electronic Medical Records. Followed a little more than a decade later by an executive order from then President George Bush with the goal of creating an electronic medical record within a decade, and then the financial incentives in the 2009 HITECH legislation to induce hospitals and physician offices to implement EMRs, the complexity of the health care computing environment increased dramatically. In addition, with the new emphasis on clinical computing, physicians as the primary users of clinical systems became much more invested not only in what functionality was being purchased, but who was leading the purchasing process.

During the 1990s and into the 2000s, an entire generation—including clinicians--became more sophisticated users of computer technology. As clinical systems came to be many organizations' largest information technology investment, clinicians sought to exercise more leadership in system selection and implementation. The position of Chief Medical Information (or Informatics) Officer (**CMIO**) was created, at times reporting to the Chief Information Officer (CIO), but at other times reporting outside of the historical information technology chain of command. Additional recognition for computer-savvy clinicians came with the creation in 2011 of a formal board certification in Clinical Informatics. In addition, with the increasing importance of clinical computing capabilities, provider organizations began looking to recognize formally the physician informatics leaders by appointing them to the CIO position. More traditional CIOs with backgrounds in business or computing became in some organizations relegated to operational roles—keeping the data center and the network running on the 24-hour basis that was essential in a health care environment. Some organizations created separate roles for Chief Technology Officers (**CTOs**) or Chief Innovation Officers that had little or no operational responsibilities. Additionally, as HCO's began to recognize the impact of social media

on their marketing efforts, the role of Chief Digital Officer also started to emerge.

These organizational changes can be expected to continue as responses to not only more complex clinical and computing environments, but also to the overwhelming importance of clinical activities overall to the success of any provider organization. The roles and responsibilities of these various information CxO positions can be expected to continue to evolve as both the overall organizational landscape and the capabilities of computer systems evolve as well.

16.4.3 Technological Changes Affecting Health Care Organizations

Future changes in technology are hard to predict. For example, although we have heard for over two decades that seamless voice-to-text systems are 5 years away from practical use, with the introduction of controlled vocabularies in areas such as radiology and pathology, we are beginning to see commercial products that can "understand" dictated speech and represent it as text that can then be structured for further analysis. Both the emergence of increasingly powerful processor and memory chips, and the decreasing cost of storage media will continue to be a factor in future health-systems design—although the tsunami of data coming from imaging modalities and from genomic medicine sequencing and analysis will continue to be a significant challenge (see ▶ Chaps. 2, 22, and 26). The ever expanding availability of Internet access, the increasing integration of voice, video, and data, and platforms which permit the integration of these various technologies as well as the availability of ever smaller platforms like tablets and smart phones, will challenge HCOs and IDNs—and HCISs--to have communications capacity not only within their traditional domain but also to an extended enterprise that may include patients' homes, schools, and workplaces. The design of modern software based on the replicability of code, coding standards such as **XML**, **C++, C#, Python,** and **Java** and frameworks such as **Microsoft's .NET** should eventually yield more flexible information technology systems.

One of the most significant technological challenges facing HCOs and IDNs today occurs because, while much of the health care delivered today continues to be within the four walls of a physician's office or a hospital, new venues such as retail clinics are growing in number and geographical distribution. Further, as the population ages, patients may seek care from a multitude of sources, including primary and specialty practices, multiple hospital visits (and even visits to multiple hospitals) and may increasingly be monitored in their homes. Health care information technologies (and clinical systems in particular) have focused historically on what happens within a physician's office or within a hospital, and not across physicians' office nor between the physicians' office and the hospital nor in the home of the patient. As technology developments increasingly permit care to be provided outside physician offices and hospitals, HCISs will be challenged to incorporate if not the data, then at least the access to these new sources of data.

In general, EMR products on the market today started with a single purpose: to automate the workflow of clinicians within a particular organizational setting. Among other features, EMRs focus on making data from previous encounters or activities easier to access and assuring that orders for tests and x-rays have the correct information, or that the next shift knows what went on previously. Despite visible successes and failures for all manner of products, EMRs in general can facilitate the automation of a complex workflow—of automating intra-organizational clinical processes as well as those that cross organizational boundaries.

Architectures that focus on what happens *within* organizational boundaries do not easily facilitate access to data *across* organizational boundaries. This is the challenge of **interoperability**. Recognizing that patients often receive care in a variety of organizational settings—hospitals, physicians offices, rehabilitation facilities, pharmacies, retail clinics, their homes, etc.—the challenge is to extend the internal workflow beyond the boundaries of individual organizations so

that data is available across a continuum of care. Interoperability then is not so much about what happens *within* an organization (although there can be challenges here as well), but what happens *across* organizational boundaries.

An *intra*-organizational architecture focuses on facilitating real time communications among providers, on optimizing the process of collecting data at the point of care, and on ensuring that clinical tasks are carried out in an appropriate sequence. An *inter*-organizational architecture needs to minimize the duplicate collection of data in different care settings, to facilitate quick searches of relevant data from a variety of (often external) sources, and to rank data in terms of relevance to a particular clinical question. Transitioning from *intra*- to *inter-organizational* data sharing is a significant technological challenge. While Health Information Exchanges (**HIE**s) and Health Record Banks (**HRB**s) are at the forefront of this transition (see ▶ Chap. 15), over time we can expect that the architectures of clinical systems that currently focus on what happens *within* an organization will need to transition to facilitate what happens *across* organizations.

Security and confidentiality concerns will likely increase as the emergence of a networked society profoundly changes our thinking about the nature of health care delivery. It is no longer only physicians or their orders that generate clinical data; increasingly patients are generating their own data either through wearables or through devices that measure their health in their own homes. Health services are still primarily delivered locally—we seldom leave our local communities to receive health care except under the most dire circumstances. In the future, providers and even patients will have access to health care experts that are dispersed over state, national, and even international boundaries. Distributed health care capabilities will need to support the implementation of collaborative models that could include virtual house calls and routine **remote monitoring** via telemedicine linkages (see ▶ Chap. 20).

16.4.4 Societal Change

At the beginning of the twenty-first century, clinicians find themselves spending less time with each patient and more time with administrative and regulatory—and often data entry--tasks. This decrease in clinician–patient contact has contributed to declining patient and provider satisfaction with electronic care-delivery systems. At the same time, empowered health consumers interested in self-help and unconventional approaches have access to more health information than ever before. These factors are changing the interplay among physicians, care teams, patients, and external (regulatory and financial) forces. The changing model of care, coupled with changing economic incentives to deliver measurable high-quality care at lower cost, places a greater focus on wellness and preventative and lifelong care. Although we might agree that aligning economic incentives with wellness is a good thing, this alignment also implies a shift in responsibility from care givers to patients.

Like the health care environment, the technological context of our lives is also changing. The Internet has already dramatically changed our approaches to information access and system design. Concurrent with the development of new standards of information display and exchange is a push led by the entertainment industry (and others) to deliver broadband multimedia into our homes. Such connectivity has the potential to change care models more than any other factor we can imagine by bringing fast, interactive, and multimedia capabilities to the household level. Finally, vast amounts of information can now be stored efficiently remotely, e.g., in the **cloud**, and on movable media such as **memory sticks**, which brings more flexibility as well as more risk, as such devices are both more convenient and more susceptible to being lost or misplaced. With the increase in the availability of consumer-oriented health information, including, for example, video segments that show the appearance and sounds of normal and abnormal conditions or demonstrate common procedures for home care

and health maintenance, we can expect even more changes in the traditional doctor/patient relationship.

With societal factors such as the focus on the outcomes of care pushing our HCOs and IDNs to change, cost constraints continuing to loom large, and the likely availability of extensive computing and communications capacity in the home, in the work place, and in the schools, HCOs and health providers are increasingly challenged to rethink the basic operating assumptions about how to deliver care. The traditional approach has been facility and physician centric—patients usually come to the hospital or to the physician's office at a time convenient for the hospital or the physician. The HCO and IDN of the twenty-first century may have to be truly "patient centric", operating within a health care delivery system without walls, where routine health management is conducted in nontraditional settings, such as homes and workplaces with increasing volumes and complexity of the data required to provide care.

Suggested Readings

Christensen, C., Grossman, J., & Hwang, J. (2009). The innovator's prescription. New York: McGraw-Hill This book builds on the author's previous work on disruptive innovation with specific applications to the health care industry. Christensen uses terms such as "precision medicine" to describe the advent of more personalized approaches to medical diagnosis and treatment and builds on his analysis of disruptive business models in other industries to analyze both the underlying problems and challenges of our health care delivery system.

Institute of Medicine, The Computer-Based Patient Record: An Essential Technology for Health Care, (1991) Washington, DC: The National Academies Press. https://doi.org/10.17226/5306. This book was one of first major attempts to argue for the use of computer technology to improve patient outcomes.

Lee, T., & Mongan, J. (2009). Chaos and organization in health care. Cambridge, MA: The MIT Press The authors describe the current health care situation as one simply of "chaos". Among the solutions they propose are increasing the use of electronic medical records and information technology in general for sharing knowledge.

Ong, K. (2011). Medical informatics: An executive primer (2nd ed.). Chicago: Health care and Management Information Systems Society An excellent overview of the challenges facing information technology applications in hospitals, physicians' offices, and in the homes of patients. Also includes a discussion of recent federal legislation intended to stimulate the use of electronic medical records and the challenges of measuring how to determine whether such investments are in fact "meaningfully used".

Porter, M., & Teisberg, E. (2006). Redefining health care: Creating value-based competition on results. Cambridge, MA: Harvard Business School Press The authors begin with a very straightforward assumption, which is that "the way to transform health care is to realign competition with value for patients" (p. 4), and proceed with an exhaustive discussion of the historical failures at reforming the health care system, the challenges inherent in physician-provider organization relationships, and how the only likely solution set to the current high cost of health care is to focus our efforts on what brings value to the patients.

Vogel, L. (2018). Who knew? Inside the complexity of American health care. New York: Taylor-Francis Identifies the major factors that combine to make health care the most complex industry in the American economy.

Questions for Discussion

1. Briefly explain the differences among an HCO's operational, planning, communications, and documentary requirements for information. Give two examples in each category. Choose one of these categories and discuss similarities and differences in the environments of an integrated delivery network, a community-based ambulatory-care clinic, and a specialty-care physician's office. Describe the implied differences in these units' information requirements.
2. Describe three situations in which the separation of clinical and administrative

information could lead to inadequate patient care, loss of revenue, or inappropriate administrative decisions. Identify and discuss the challenges and limitations of two methods for improving data integration.

3. Describe three situations in which lack of integration of information systems with clinicians' workflow can lead to inadequate patient care, reduced physician productivity, or poor patient satisfaction with an HCO's services. Identify and discuss the challenges and limitations of two methods for improving process integration.
4. Describe the trade-off between functionality and integration. Discuss three strategies currently used by HCOs to minimize this tradeoff.
5. Assume that you are the chief information officer of multi-facility HCO. You have just been charged with planning a new clinical HCIS to support a large tertiary care medical center, two smaller community hospitals, a nursing home, and a 40-physician group practice. Each organization currently operates its own set of integrated and standalone technologies and applications. What technical and organizational factors must you consider? What are the three largest challenges you will face over the next 24 months?
6. How do you think the implementation of clinical HCISs will affect the quality of relationships between patients and providers? Discuss at least three potential positive and three potential negative effects. What steps would you take to maximize the positive value of these systems?

References

Bates, D. W., & Gawande, A. A. (2003). Improving safety with information technology. *The New England Journal of Medicine, 348*(25), 2526–2534.

Blackstone, E. A., & Fuhr, J. P., Jr. (2003). Failed hospital mergers. *Journal of Health Law, 36*(2), 301–324.

Castaneda, C., Nalley, K., Mannion, C., Bhattacharyya, P., Blake, P., Pecora, A., Goy, A., & Suh, K. S. (2015). Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. *Journal of Clinical Bioinformatics, 5*(1), 1.

Goldzweig, C. L., Towfigh, A., Maglione, M., & Shekelle, P. G. (2009). Costs and benefits of health information technology: New trends from the literature. *Health Affairs (Millwood), 28*(2), w282–w293.

Gottlieb, L. M., Tirozzi, K. J., Manchanda, R., Burns, A. R., & Sandel, M. T. (2015). Moving electronic medical records upstream: Incorporating social determinants of health. *American Journal of Preventive Medicine, 48*(2), 215–218.

Haddad, P., McConchie, S., Schaffer, J. L., & Wickramasinghe, N. (2018). IS/IT governance in health care: An integrative model. In N. Wickramasinghe & J. Schaffer (Eds.), *Theories to inform superior health informatics research and practice. Healthcare delivery in the information age* (pp. 37–54). Cham, Switzerland: Springer.

Himmelstein, D. U., & Woolhandler, S. (2010). Obama's reform: No cure for what ails us. *BMJ, 340*, c1778.

James, B. C., & Savitz, L. A. (2011). How intermountain trimmed health care costs through robust quality improvement efforts. *Health Affairs (Millwood), 30*(6), 1185–1191.

Kastor, J. (2001). *Mergers of teaching hospitals in Boston, New York, and Northern California*. Ann Arbor: University of Michigan Press.

Kaushal, R., Shojania, K. G., & Bates, D. W. (2003). Effects of computerized physician order entry and clinical decision support systems on medication safety: A systematic review. *Archives of Internal Medicine, 163*(12), 1409–1416.

Leatherman, S., Berwick, D., Iles, D., et al. (2003). The business case for quality: Case studies and an analysis. *Health Affairs, 22*(2), 17–30.

McCullough, J. S., Casey, M., Moscovice, I., & Prasad, S. (2010). The effect of health information technology on quality in U.S. Hospitals. *Health Affairs (Millwood), 29*(4), 647–654.

Shortell, S. M., Gillies, R. R., Anderson, D. A., et al. (2000). *Remaking healthcare in America: The evolution of organized delivery* (2nd ed.). San Francisco: Jossey-Bass.

Vogel, L. H. (2003). Finding value from information technology investments: Exploring the elusive ROI in healthcare. *Journal for Health Information Management, 17*(4), 20–28.

Vogel, L. H. (2006, September 1) Everyone gets to play. CIO.

Weil, T. P. (2001). *Health networks: Can they be the solution?* Ann Arbor: University of Michigan Press.

Weill, P., & Ross, J. W. (2004). *IT governance: How top performers manage IT decision rights for superior performance*. Boston: Harvard Business School Press.

Patient-Centered Care Systems

Suzanne Bakken, Patricia C. Dykes, Sarah Collins Rossetti, and Judy G. Ozbolt

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_17

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What is patient-centered care? How does it differ from traditional, clinician-centric care?
- What are the information management challenges in patient-centered care?
- What are the roles of electronic health records and other informatics applications in supporting patient-centered care?
- What forces and developments have led to the emergence of patient-centered care systems?
- What collaborative processes are required to design patient-centered care systems and the electronic health records to support such care?
- How is current informatics research advancing progress toward collaborative, interdisciplinary, patient-centered care?

17.1 Information Management in Patient-Centered Care

Patient care is the focus of many clinical disciplines—medicine, nursing, pharmacy, nutrition, therapies such as respiratory, physical, and occupational, and others. Although the work of the various disciplines sometimes overlaps, each has its own primary focus, emphasis, and methods of care delivery. Each discipline's work is complex in itself, and collaboration among disciplines, an essential component of patient-centered care, adds another level of complexity. In all disciplines, the quality of clinical decisions depends in part on the quality of information available to the decision-maker. The systems that manage information for patient-centered care are therefore critical tools. Their fitness for the job varies, and the systems enhance or detract from patient-centered care accordingly. This chapter describes information management issues in patient-centered care, the emergence of patient-centered care systems in relation to these issues, the interdisciplinary collaboration required to develop patient-centered care systems, and current research. In so doing, it will demonstrate the necessity of a patient-centered perspective in the design of electronic health records (EHRs) and other patient-care systems.

As described later in this chapter, reports of the National Academies, Federal Government mandates, and a variety of social forces have called for transformation in the organization, delivery, financing, and quality of health care. The demand is for evidence-based, cost-effective, **patient-centered care**. Informatics is seen as essential to the provision, monitoring, and improvement of such care.

17.1.1 From Patient Care to Patient-Centered Care

Patient-centered care is a collaborative, interdisciplinary process focused on the care recipient in the context of the family, significant others, and community. A distinguishing feature of patient-centered care is the patient's active collaboration in shared decision-making, as contrasted to traditional clinician-centered care where the clinician holds the preponderance of power and authority. Patient-centered care empowers patients to actively participate in care by presenting treatment options that are consistent with patient values and preferences and in a format or context that is understandable and actionable (Krist and Woolf 2011; Payton et al. 2011). Typically, patient care includes the services of physicians, nurses, and members of other health disciplines according to patient needs: physical, occupational, and respiratory therapists; nutritionists; psychologists; social workers; and many others. Each of these disciplines brings specialized perspectives and expertise. Specific cognitive processes and therapeutic techniques vary by discipline, but all disciplines share certain commonalities in the provision of care.

In its simplest terms, the process of patient-centered care begins with collecting data and assessing the patient's current status and expressed concerns in comparison to cri-

teria or expectations of normality. Through cognitive processes specific to the discipline, diagnostic labels are applied, therapeutic goals are identified with timelines for evaluation, and therapeutic interventions are selected and implemented. The patient participates, as he or she is able, in determining therapeutic goals and selecting personally acceptable interventions from the options and their potential consequences as described by the clinician. At specified intervals, the patient is reassessed, the effectiveness of care is evaluated, and therapeutic goals and interventions are continued or adjusted as needed. If the reassessment shows that the patient no longer needs care, services are terminated. This process was illustrated for nursing in 1975 (Goodwin & Edwards 1975) and was updated and made more general in 1984 (Ozbolt et al. 1984). The flowchart reproduced in ◘ Fig. 17.1 could apply equally well to other patient-care disciplines.

Although this linear flowchart helps to explain some aspects of the process of care, it is, like the solar-system model of the atom, a gross simplification. Frequently, for example, in the process of collecting data for an initial patient assessment, the nurse may recognize (diagnose) that the patient is anxious about her health condition. Simultaneously with continuing the data collection, the nurse sets a therapeutic goal that the patient's anxiety will be reduced to a level that increases the patient's comfort and ability to participate in care. The nurse selects and implements therapeutic actions of modulating the tone of voice, limiting environmental stimuli, maintaining eye contact, using gentle touch, asking about the patient's concerns, and providing information. All the while, the nurse observes the effects on the patient's anxiety and adjusts his behavior accordingly. Thus, the complete care process can occur in a microcosm while one step of the care process—data collection—is underway. This simultaneous, nonlinear quality of patient care poses challenges to informatics in the support of patient care and the capture of clinical data.

Each caregiver's simultaneous attention to multiple aspects of the patient is not the only complicating factor. Just as atoms become molecules by sharing electrons, the care provided by each discipline becomes part of a complex molecule of interdisciplinary, patient-centered care. Caregivers and developers of informatics applications to support care must recognize that true patient-centered care is as different from the separate contributions of the various disciplines as an organic molecule is from the elements that go into it. The contributions of the various disciplines are not merely additive; as a therapeutic force acting upon and with the patient, the work of each discipline is transformed by its interaction with the patient and the other disciplines in the larger unity of patient-centered care.

17.1.2 Patient-Centered Care in Action

A 75-year-old woman with osteoarthritis, high blood pressure, and urinary incontinence is receiving care from a physician, a home-care nurse, a nutritionist, a physical therapist, and an occupational therapist. From a clinician-centered, additive perspective, each discipline could be said to perform the following functions:

1. Physician: diagnose diseases, prescribe appropriate medications, authorize other care services
2. Nurse: assess patient's understanding of her condition and treatment and her self-care abilities and practices; assess patient's concerns, values, and preferences regarding the management of her health; teach and counsel as needed; help patient to perform exercises at home; identify and help patient overcome barriers to self-care and participation in her recovery plan; report findings to physician and other members of care team
3. Nutritionist: assess patient's nutritional status and eating patterns; prescribe and teach appropriate diet to control blood pressure and build physical strength
4. Physical therapist: prescribe and teach appropriate exercises to improve strength and flexibility and to enhance cardiovascular health, within limitations of arthritis

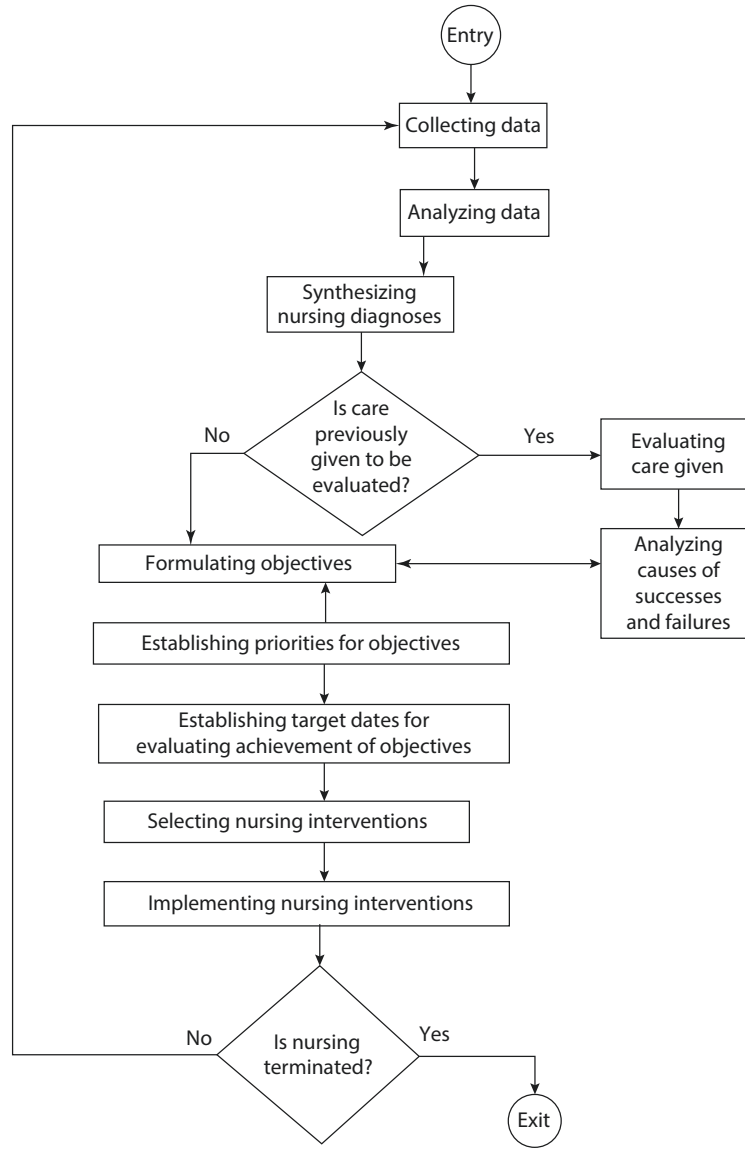


Fig. 17.1 The provision of nursing care is an iterative process that consists of steps to collect and analyze data, to plan and implement interventions, and to evaluate the results of interventions (Source: Adapted with permission from Ozbolt et al. (1985). © Springer Nature)

5. Occupational therapist: assess abilities and limitations for performing activities of daily living; prescribe exercises to improve strength and flexibility of hands and arms; teach adaptive techniques and provide assistive devices as needed

In a collaborative, interdisciplinary, patient-centered practice, the nurse discovers that the patient is not taking walks each day as prescribed because her urinary incontinence is exacerbated by the diuretic prescribed to treat hypertension, and the patient is embarrassed to go out. The nurse reports this to the physician and the other clinicians so that they can understand why the patient is not carrying out the prescribed regime. The physician then changes the strategy for treating hypertension while initiating treatment for urinary incontinence. The nurse helps the patient to understand the interaction of the various treatment regimes, provides practical advice and assistance in dealing with incontinence, and helps the patient to find personally acceptable ways to follow the prescribed treatments. The nutritionist works with the patient on the timing

of meals and fluid intake so that the patient can exercise and sleep with less risk of urinary incontinence. The physical and occupational therapists adjust their recommendations to accommodate the patient's personal needs and preferences while moving toward the therapeutic goals. Finally, the patient, rather than being assailed with the sometimes conflicting demands of multiple clinicians, is supported by an ensemble of services that meets shared therapeutic goals in ways consistent with her preferences and values.

This kind of patient-centered collaboration requires exquisite communication and feedback. The potential for information systems to support or sabotage patient-centered care is obvious.

17.1.3 Coordination of Patient-Centered Care

When patients receive services from multiple clinicians, patient-centeredness requires coordinating those services. Coordination includes seeing that patients receive all the services they need in logical sequence without scheduling conflicts and ensuring that each clinician communicates as needed with the others. Sometimes, a **case manager** or **care coordinator** is designated to do this coordination. In other situations, a physician or a nurse assumes the role by default. Sometimes, coordination is left to chance, and both the processes and the outcomes of care are put at risk. In recognition of this, the Institute of Medicine designated coordination of care as 1 of 14 priorities for national action to transform health care quality (Adams & Corrigan 2003). The **Health Information Technology for Economic and Clinical Health Act** (HITECH Programs 2009[1]) calls for patients to have a **medical home**, a primary care practice that will maintain a comprehensive problem list to make fully informed decisions in coordinating their care. In addition, Accountable Care Organizations seek to integrate providers and services to generate value for defined populations. Substantial federal investment in health information technology (HIT) through the HITECH and 21st Century Cures Acts dramatically increased the adoption and use of EHRs (Adler-Milstein & Jha 2017). While HIT spending increased markedly, care coordination technologies have not been a focus. HIT tools dedicated to coordination processes could improve care of complex patients through increased access to data, facilitated communication, timely shared decisions, and greater engagement of patients and their families as partners in their care plans. Well-designed information systems with patient facing-technologies (e.g., **personal health records** and **patient portals**) enable care coordination as they ensure that patients and providers have immediate access to accurate health information at home and across care settings (Ahern et al. 2011; Collins, Bavuso et al. 2017; Collins, Klinkenberg-Ramirez et al. 2017; Collins, Rozemblum et al. 2017).

17.1.4 Patient-Centered Care Across Multiple Patients

Delivering and managing interdisciplinary patient-centered care for an individual is challenging enough, but patient care has yet another level of complexity. Each clinician is responsible for the care of multiple patients. In planning and executing the work of patient-centered care, each professional must consider the competing demands of all the patients for whom she is responsible, as well as the exigencies of all the other professionals involved in each patient's care. Thus, the nurse on a post-operative unit must plan for scheduled treatments for each of her patients to occur near the optimal time for that patient. She must take into account that several patients may require treatments at nearly the same time and that some of them may be receiving other services, such as imaging or physician's visits, at the time when it might be most convenient for the nurse to administer the treatment. When unexpected needs arise, as they often do—an

1 ► http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__hitech_programs/1487 (Accessed: 4/26/13).

emergency, an unscheduled patient, observations that could signal an incipient complication—the nurse must set priorities, organize, and delegate to be sure that at least the critical needs are met. Similarly, the physician must balance the needs of various patients who may be widely dispersed throughout an institution. Decision-support systems have the potential to provide important assistance for both the care of individual patients and the organization of the clinician's workload.

17.1.5 Integrating Indirect-Care Activities

Finally, clinicians not only deliver services to patients, with all the planning, documenting, collaborating, referring, and consulting attendant on direct care; they are also responsible for **indirect-care** activities, such as teaching and supervising students, attending staff meetings, participating in continuing education, and serving on committees. Each clinician's plan of work must allow for both the direct-care and the indirect-care activities. Because the clinicians work in concert, these plans must be coordinated.

In summary, patient care is an extremely complex undertaking with multiple levels. To achieve patient-centered care, each clinician's contributions to the care of every patient must take into account not only that patient's values, preferences, and concerns, but also the ensemble of contributions of all clinicians involved in the patient's care and the interactions among them, and this entire suite of care must be coordinated to optimize effectiveness and efficiency. These very complex considerations are multiplied by the number of patients for whom each clinician is responsible. Patient care is further complicated by the indirect-care activities that caregivers must intersperse among the direct-care responsibilities and coordinate with other caregivers. The resulting cognitive workload frequently overwhelms human capacity. Systems that effectively assist clinicians to manage, process, and communicate the data, information, and knowledge essential to patient-centered care are critical to the quality and safety of that care.

17.1.6 Information to Support Patient-Centered Care

As complex as patient care is, the essential information for direct, patient-centered care is defined in the answers to the following questions:

- What are the patient's needs, concerns, preferences, and values?
- Who is involved in the care of the patient?
- What information does each clinician require to make decisions in his or her professional domain?
- From where, when, and in what form does the information come?
- What information does each clinician generate? Where, when, and in what form is it needed?

The framework described by Zielstorff, Hudgings, and Grobe (1993) provides a useful heuristic for understanding the varied types of information required to answer each of these questions. As listed in Table 17.1, this framework delineates three information categories: (1) patient-specific data about a particular patient acquired from a variety of data sources; (2) agency-specific data relevant to the specific organization under whose auspices the health care is provided; and (3) domain information and knowledge specific to the health care disciplines.

The framework further identifies four types of information processes that information systems may apply to each of the three information categories. *Data acquisition* entails the methods by which data become available to the information system. It may include data entry by the care provider, patient, or family, or acquisition from a medical device or from another computer-based system. *Data storage* includes the methods, programs, and structures used to organize data for subsequent use. Standardized coding and classification

Table 17.1 Framework for design characteristics of a patient-care information system with examples of patient-specific data, agency-specific data, and domain information and knowledge for patient care

Types of data	System processes			
	Acquiring	**Storing**	**Transforming**	**Presenting**
Domain-specific	Downloading relevant scientific or clinical literature or practice guidelines	Maintaining information in electronic journals or files, searchable by key words	Linking related literature or published findings; updating guidelines based on research	Displaying relevant literature or guidelines in response to queries
Agency-specific	Scanning, downloading, or keying in agency policies and procedures; keying in personnel, financial, and administrative records	Maintaining information in electronic directories, files, and databases	Editing and updating information; linking related information in response to queries; analyzing information	Displaying on request continuously current policies and procedures; sharing relevant policies and procedures in response to queries; generating management reports
Patient-specific	Point-of-care entry of data about patient assessment, diagnoses, treatments planned and delivered, therapeutic goals, and patient outcomes	Moving patient data into a current electronic record or an aggregate data repository	Combining relevant data on a single patient into a cue for action in a decision-support system; performing statistical analyses on data from many patients	Displaying reminders, alerts, probable diagnoses, or suggested treatments; displaying vital signs graphically; displaying statistical results

Source: Framework adapted with permission from *Next Generation Nursing Information Systems*, 1993, American Nurses Association, Washington, DC. Reused with permission.

systems useful in representing patient-centered care concepts are discussed in greater detail in ▶ Chap. 7. *Data transformation* (or *data processing*) comprises the methods by which stored data or information are acted on according to the needs of the end-user—for example, calculation of a pressure ulcer risk-assessment score at admission or calculation of critically ill patients' **acute physiology and chronic health evaluation (APACHE III)** scores. Figure 17.2 illustrates the transformation (abstraction, summarization, aggregation) of patient-specific data for multiple uses. *Presentation* encompasses the forms in which information is delivered to the end-user after processing.

Transformed patient-specific data can be presented in a variety of ways. Numeric data may be best presented in chart or graph form to allow the user to examine trends, whereas the compilation of potential diagnoses generated from patient-assessment data is better presented in an alphanumeric-list. Different types of agency-specific data lend themselves to a variety of presentation formats. Common among all, however, is the need for presentation at the point of patient care. For example, the integration of up-to-the-minute patient-specific data with agency-specific guidelines or parameters can produce alerts, reminders, or other types of notifications for immediate action. See ▶ Chap. 21, on patient-monitoring systems, for an overview of this topic. Presentation of domain information and knowledge related to patient care is most frequently accomplished through interaction with databases and **knowledge bases**, such as Medline or Micromedex. Commercial applications such as UpToDate™ are popular among clinicians because they provide easy

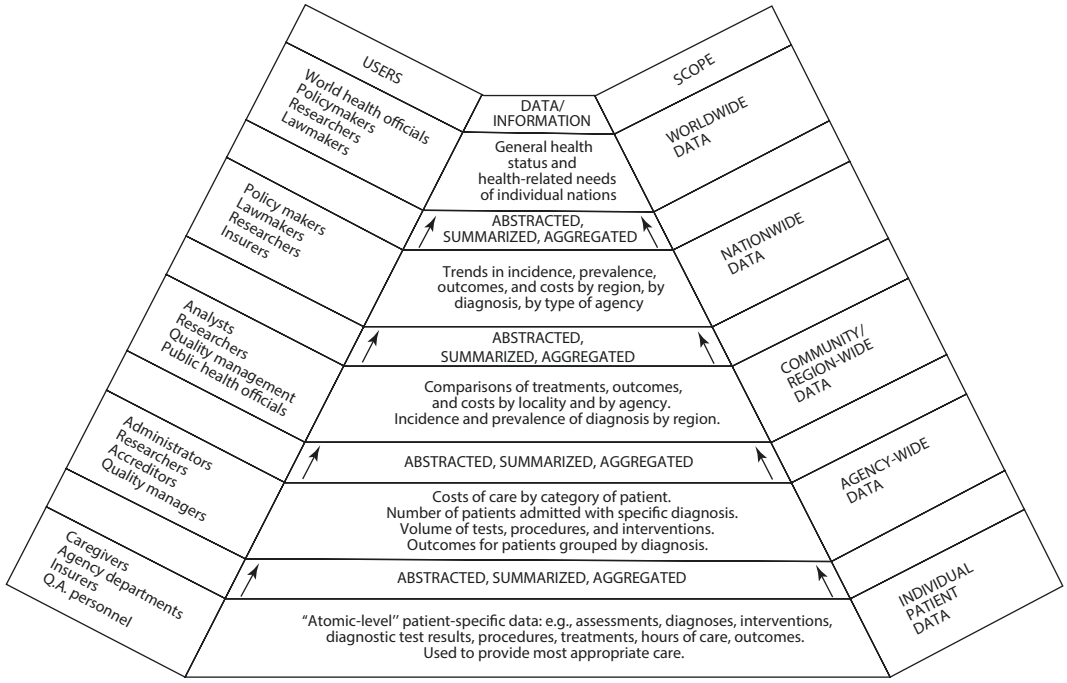


Fig. 17.2 Examples of uses for atomic-level patient data collected once but used many times (Source: Reprinted with permission from Zielstorff et al. (1993). © 1993 American Nurses Publishing, American Nurses Foundation/American Nurses Association, Washington, DC. Reproduced with permission.)

access to knowledge resources at the point of care. The **Infobutton**, developed at NewYork-Presbyterian Hospital, is an HL7 standard for context-aware knowledge retrieval (Del Fiol et al. 2012). Incorporated into EHRs, Infobuttons, along with associated management tools, can integrate data about the patient and the clinical context to provide immediate, point-of-care access to relevant knowledge resources (Cimino et al. 2013).

To support patient-centered care, information systems must be geared to the needs of all the clinicians involved in care. The systems should acquire, store, process, and present each type of information (patient-, agency-, and domain-specific) where, when, and how the information is needed by each clinician in the context of his or her professional domain. Systems designed for patient-centered care have the potential to go beyond supporting the collaborative, interdisciplinary care of individual patients. Through appropriate use of patient-specific information (care requirements), agency-specific information (clinicians and their responsibilities and agency policies and procedures), and domain information (guidelines), such systems can greatly aid the coordination of interdisciplinary services for individual patients and the planning and scheduling of each caregiver's work activities. Patient acuity is taken into account in scheduling nursing personnel, but historically has most often been entered into a separate system rather than derived directly from care requirements as recorded in the EHR. Fully integrated, patient-centered systems—still an ideal today—would enhance our understanding of each patient's situation, needs, and values, improve decision-making, facilitate communications, aid coordination, and use clinical data to provide feedback for improving clinical processes.

Clearly, when other **clinical information systems** designed to support patient-centered care fulfill their potential, they will not merely replace oral and paper-based methods of recording and communicating. They will be an integral and essential part of the transformation of health care to apply evidence-based interventions in accordance with patient needs

and values. How far have we come toward the ideal? What must we do to continue our progress?

17.2 The Emergence of Patient-Centered Care Systems

Events in the first decade of the Twenty-first Century planted the seeds of transformative change in patient care and clinical informatics. Over 10 years, the shared ideal of health care began to move from the Twentieth Century "doctor knows best" model toward a new vision of health care based on interdisciplinary teams drawing on a variety of knowledge and information resources to collaborate with one another and with patients and families to resolve or alleviate health problems and to achieve health goals. Recursive and iterative developments grew from reports of the **National Research Council** and the **Institute of Medicine** (now known as the **National Academy of Medicine)**; from government policies and initiatives; from changes in organizational and financial structures for health care delivery; and from advances in the informatics methods and technologies that have become integral to the provision, management, reimbursement, and improvement of health care. Much remains to be done to nurture continuing development, but with care and patience we can begin to harvest the benefits of better health care and better health for individuals and populations.

17.2.1 Publications of the National Academy of Sciences

With its seminal publication, *To Err is Human*: *Building a Safer Health System* (Kohn et al. 2000), the Institute of Medicine startled the world by estimating that clinical errors were killing up to 98,000 hospitalized Americans each year. The report called for a national focus to advance knowledge about safety, reporting efforts to identify and learn from errors, higher standards and expectations for safety, and implementation of safe practices and systems within health care organizations.

The follow-on report, *Crossing the Quality Chasm*: *A New Health System for the 21st Century* (Committee on Quality of Health Care in America 2001), addressed the need for fundamental change in the health care delivery system. Shortcomings included inadequate focus on quality and clinical infrastructures not sufficiently developed to provide the full range of services needed by persons with chronic conditions. Significantly, the report placed the blame not on individual health care professionals, but on inadequate and broken systems of care.

Crossing the Quality Chasm outlined a call for action by government, payers, providers, and the public to embrace a statement of purpose for the health care system as a whole—to reduce illness and improve health and functioning—and to adopt a shared agenda to achieve health care that would be safe, effective, patient-centered, timely, efficient, and equitable. The report recommended the redesign of care processes to achieve continuity in care relationships; customization in accordance with patient needs and values; the sharing of knowledge, information, and decision-making with patients; evidence-based decision-making; safety as a system property; transparency of information to facilitate informed decision-making by patients and families; anticipation of patient needs; continuous decrease in waste; and cooperation among clinicians. The report gave considerable attention to informatics as an essential methodology to achieve these aims and called for a renewed national commitment to a national health information infrastructure, with the elimination of most hand-written clinical information by 2010.

In *Patient Safety*: *Achieving a New Standard for Care* (Committee on Data Standards for Patient Safety 2004), the Committee highlighted the fact that a national health information infrastructure—a foundation of systems, technology, applications, standards, and policies—is required for error prevention and capture of data that facilitate local

and global learning from adverse events, near misses, and hazards. The report emphasized the need for data interchange standards as an essential building block.

The **National Academies** followed these three reports with a number of others that explored in greater depth aspects of the problems and recommendations described within them and made further recommendations for public and private actions to improve health care and its costs and outcomes. The National Research Council (Stead & Lin 2009) published *Computational Technology for Effective Health Care*: *Immediate Steps and Strategic Directions*. This report noted that many health information technologies in the current marketplace lacked the functionality to achieve the goals of improving health care. The central finding was that computer scientists, experts in health and biomedical informatics, and clinicians would need to collaborate to create technologies that would provide cognitive support to clinicians, patients, and family members as they sought to understand, resolve, or alleviate health challenges. The report recommended that federal and state governments and clinicians join forces to require vendors to provide systems that offer such "meaningful" support.

In its **Learning Health System** series, the National Academy of Medicine's Leadership Consortium for a Value & Science-Driven Health System synthesized the insights of a broad array of experts and explored the need for transformational change in the fundamental elements of health and health care. Multiple volumes reflect the fundamental role of informatics in a learning health system: *Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good* (Institute of Medicine 2010), *Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care – Workshop Series Summary* (Institute of Medicine 2011), *Digital Data Improvement Priorities for Continuous Learning in Health and Health Care – Workshop Summary* (Institute of Medicine 2013), and *Optimizing Strategies for Clinical Decision Support: Summary of Meeting Series* (Tcheng et al. 2017). Several priorities for collaborative action from the last are of particular relevance to enable scalable patient-centered care systems: create a national **Clinical Decision Support** (**CDS**) repository, develop tools to assess CDS efficacy, publish performance evaluations, promote financing and measurement to accelerate CDS adoption, develop a multi-stakeholder CDS learning community to inform usability, and establish an investment program for CDS research.

17.2.2 Federal Government Initiatives

The Health Information Technology for Economic and Clinical Health (HITECH) Act provided an unprecedented federal investment in HIT through a series of initiatives aimed at ensuring that all Americans benefit from EHR-supported patient-centered care. Administered by the Office of the National Coordinator for Health Information Technology, the activities are designed:

- To support the health care workforce through Regional Extension Centers for technical assistance for implementation of EHRs and training initiatives to ensure meaningful use of EHRs
- To enable coordination and alignment within and among states (State Health Information Exchange Cooperative Agreement Program)
- To establish connectivity to the public health community in case of emergencies (Beacon Community Program)

In addition, two federal rules support meaningful use of EHRs. The Incentive Programs for EHRs rule from the Centers for Medicare & Medicaid Services (CMS) defines minimum requirements that hospitals and eligible professionals must meet through their use of certified technology to qualify for incentive payments. Criteria related to providing patients with an electronic copy of their own health information and ability to electronically

exchange key clinical information are particularly important to patient-centered care. The complementary Standards and Certification Criteria for Electronic Health Records rule defines the criteria for certification of the technology. Also relevant to patient-centered care are **NHIN Direct** and **NHIN CONNECT**, which support health information exchange to enable patient-centered care.

The Agency for Healthcare Research and Quality (AHRQ) has also invested in advancing patient-centered care through investments in HIT. A particular focus is promoting systems engineering approaches to improve patient safety through evaluation of clinical processes resulting in improved work and information flows. This is reflected in the AHRQ Patient Safety Learning Labs grant portfolio. For example, the Brigham and Women's Hospital project, Making Acute Care More Patient-centered, focused on developing tools and care processes to engage patient, family, and professional care team members in reliable identification, assessment, and reduction of patient safety threats in real time before they manifest into actual harm. Widespread adoption of EHR systems and web-based technologies makes it possible to reuse EHR data to create visual displays at the bedside to provide decision support to the care team, patients, and family. ◘ Figure 17.3 is an example of a bedside screensaver display used at Brigham and Women's Hospital. The **icons** are driven by patient-specific information in the EHR. The display provides information to patients about their safety plan and provides decision support at the bedside related to patient-specific needs to professional and paraprofessional care team members.

Given these major investments in promoting EHR adoption and use for patient-centered care and research, the vision of every American reaping the benefits of EHRs is moving closer to reality. However, this will continue to be heavily influenced by associated changes in health care financial and organizational structures.

17.2.3 Financial and Organizational Structures in Health Care

The historical evolution of information systems that support patient care, and eventually patient-centered care, is not solely a reflection of the available technologies (e.g. Web 2.0, cloud computing). Societal forces—including delivery-system structure, practice model, payer model, and quality focus—have influenced the design and implementation of patient-care systems (◘ Table 17.2).

▪ Delivery-System Structure

Authors have noted the significant influence of the organization and its people on the success or failure of informatics innovations (Massaro 1993; Campbell et al. 2006; Ash et al. 2007). Others have documented unintended consequences of implementation of HIT and called for applications of models of processes, such as Iterative Sociotechnical Analysis, that take into account health care organizations' workflows, social interactions, culture, etc. to further elucidate the relationship between organizations and technology (Harrison et al. 2007; Koppel et al. 2005; Ozbolt et al. 2012). As delivery systems shifted from the predominant single-institution structure of the 1970s to the **integrated delivery networks** of the 1990s to the complex linkages of the Twenty-First Century, the information needs changed, and the challenges of meeting those information needs increased in complexity. The **patient centered medical home** (PCMH)[2] (also known as primary care medical home, advanced primary care, and health care home) is a model of primary care that delivers care designed to be patient-centered, comprehensive, coordinated, accessible, and continuously improved through a systems-based approach to quality and safety (Patient Centered Medical Home Resource Center 2011).[3] AHRQ and others (Bates and

2 ► http://www.ncqa.org/ (Accessed: 7/1/19).
3 ► https://pcmh.ahrq.gov/ (Accessed 7.8.19).

Fig. 17.3 A personalized screen saver used at Brigham and Women's Hospital to display patient-specific safety information to patients, family, and care team. (© Brigham and Women's Hospital, Center for Patient Safety, Research and Practice. Reused with permission.)

Bitton 2010) have noted the seminal role of HIT (e.g., health information exchange, disease registries, alerts and reminders) to support tasks related to **National Committee on Quality Assurance** PCMH standards for enhancing access and continuity, identifying and managing patient populations, planning and managing care, providing self-care and community support, tracking and coordinating care, and measuring and improving performance (Patient Centered Medical Home 2011). **Accountable Care Organizations** (ACOs) focus on care of designated populations. A recent systematic review (Kaufman et al. 2019) reveals consistent associations between ACO implementation and outcomes across payer types and reduced inpatient use, reduced emergency department visits, and improved measures of preventive care and chronic disease management. See ▶ Chaps. 11, 15, and 18 for discussions of managing clinical information in consumer-provider partnerships in care, in the public health information infrastructure, and in integrated delivery systems.

17

Professional Practice Models

Professional practice models have also evolved for nurses and physicians. In the 1970s, team nursing was the typical practice model for the hospital, and the nursing care plan—a document for communicating the plan of care among nursing team members—was most frequently the initial computer-based application designed for use by nurses. The 1990s were characterized by a shift to interdisciplinary-care approaches necessitating computer-based applications such as critical paths to support case management of aggregates of patients, usually with a common medical diagnosis, across the **continuum of care**. The Twenty-First Century sees advanced practice nurses increasingly taking on functions previously provided by physicians while maintaining a nursing perspective on collaborative, interdisciplinary care. This trend is likely to accelerate given the recommendations for facilitating full scope of practice for nurses and advanced practice nurses (e.g., certified nurse midwives, nurse practitioners) in the 2010 Institute of Medicine

Table 17.2 Societal forces that have influenced the design and implementation of patient-centered systems

	1970s	1980s	1990s	2000s	2010s
Delivery-system structure	Single institution	Single organization	Integrated delivery systems		Patient centered medical home Virtual care
Professional-practice model	Team nursing	Primary nursing	Patient-focused care, multi-disciplinary care, case management	Patient-centered care	Expansion of nurse and advanced practice nurse roles to legal scope of practice
	Single or small group physician practice	Group models for physicians	Variety of constellations of physician group practice models		
Payer model	Fee for service	Fee for service	Capitation	CMS P4P hospital initiative	Affordable Care Act of 2010
		Prospective payment	Managed care		Accountable Care Organizations Risk-bearing, coordinated care models
		Diagnosis-related groups			
Quality focus	Professional Standards Review Organizations (PSROs)	Continuous quality improvement	Risk-adjusted outcomes	Patient safety	Value-driven health care
	Retrospective chart audit	Joint Commission on Accreditation of Health Care Organization (JCAHO)'s agenda for change	Benchmarking	Learning organizations	Patient/consumer-centered outcomes Promoting interoperability
			Practice guidelines		
			Critical paths/care maps	Consumer-driven	
			Health Employer Data and Information Set (HEDIS)		

(continued)

Table 17.2 (continued)

	1970s	1980s	1990s	2000s	2010s
General technology trends			World Wide Web (Web 1.0)	Web 2.0	Cloud computing Digitization of the health care system, ehealth/ mhealth
				Social media	
				"Smart" mobile devices	

report on *The Future of Nursing*: *Leading Change*, *Advancing Health* (Committee on the Robert Wood Johnson Foundation Initiative on the Future of Nursing at the Institute of Medicine 2010). These changes broaden and diversify the demands for decision support, feedback about clinical effectiveness, and quality improvement as a team effort.

Physician practice models have shifted from single physician or small group offices to complex constellations of provider organizations. The structure of the model (e.g., staff model **health-maintenance organization (HMO)**, captive-group model health-maintenance organization, or independent-practice association) determines the types of relationships among the physicians and the organizations. These include issues— such as location of medical records, control of practice patterns of the physicians, and data-reporting requirements—that have significant implications for the design and implementation of patient-care systems. In addition, the interdisciplinary and distributed care approaches of the 1990s and the 2000s have given impetus to system-design strategies, such as the creation of a single patient problem list, around which the patient-care record is organized, in place of a separate list for each provider group (e.g., nurses, physicians, respiratory therapists). While a shared single patient problem list remains a goal of patient-centered care, few implementations have achieved successful integration of the medical problem list with the problem list from other health professions, in particular nursing, due to the persistence of siloed EHR terminology management (Collins, Bavuso et al. 2017; Collins, Klinkenberg-Ramirez et al. 2017; Collins, Rozemblum et al. 2017).

In the 2010s the use of shared clinical **dashboards** accelerated the operationalization of known patient-centered best practices, such as interdisciplinary clinical rounding and safety checklists, to drive increased situational awareness and promote common ground in relation to patient goals and preferences and prevention of harm (Collins, Gazarian et al. 2014; Mlaver et al. 2017). These shared, interdisciplinary, team-based dashboards are typically configured to a particular practice and specialty area (e.g., cardiac ICU, emergency department) or a clinical quality improvement strategic initiative (e.g., optimizing length of stay, promoting effective and early discharge planning) (Collins, Hurley et al. 2014). In the ambulatory setting shared dashboards are also used and configured to drive population health initiatives.

Payer Models

Changes in payer models have been a significant driving force for information-system implementation in many organizations. With the shift from fee for service to prospective payment in the 1980s, and then toward capitation in the 1990s, information about costs and quality of care has become an essential commodity for rational decision-making in the increasingly competitive health care mar-

ketplace. Because private, third-party payers often adopt federal standards for reporting and regulation, health care providers and institutions have struggled in the early 2000s to keep up with the movement toward data and information system standards accelerated by the **Health Insurance Portability and Accountability Act (HIPAA)**[4] and the initiatives to develop a National Health Information Network. With the advent of pay for performance (P4P), CMS has eliminated reimbursement for preventable conditions (e.g., catheter-associated urinary tract infections) that occur during hospitalizations ("CMS P4P,"). In this decade, there is no doubt that the implementation of the highly controversial Affordable Care Act of 2010 and evolving ACOs will profoundly impact patient-centered care and the information systems needed to support it.

Quality Focus

Demands for information about quality of care have also influenced the design and implementation of patient-care systems. The quality-assurance techniques of the 1970s were primarily based on retrospective chart audit. In the 1980s, continuous quality improvement techniques became the modus operandi of most health care organizations. The quality-management techniques of the 1990s were much more focused on concurrently influencing the care delivered than on retrospectively evaluating its quality. In the Twenty-first Century, patient-centered systems-based approaches—such as practice guidelines, alerts, and reminders tailored on patient clinical data and, in some instances, genomic data (i.e., personalized medicine)—are an essential component of **quality management**. In addition, institutions must have the capacity to capture data for benchmarking purposes and to report process and outcomes data to regulatory and accreditation bodies, as well as to any voluntary reporting programs to which they belong. Increasingly, concurrent feedback about the effectiveness of care guides clinical decisions in real time, and "dashboards" are used to display indicators related to different dimensions of quality.

Data science approaches using clinical data with a quality focus greatly expanded in the 2010s (Bates et al. 2014; Hruby et al. 2016). While much of this work has focused only on the subset of data within the EHR that are consistently structured and coded, such as medical diagnoses, allergies, and medication orders, there are increasing efforts to process and derive value from additional data sources, such as nursing assessment data and narrative notes (Klann et al. 2018) (see ► Chap. 8). The use of high volume clinical **metadata**, or data about the clinical data, also provides value for quality purposes, such as developing healthcare process models of care delivery processes, for a richer understanding of patient centered workflows and evaluation of best practices and gaps in care than previously feasible (Hripcsak & Albers 2013; Collins & Vawdrey 2012; Collins et al. 2013).

Of note, while advances in natural language processing continue, many of the computable processes outlined above, such as practice guidelines, alerts, and reminders, still require structured data capture, which typically requires manual data entry by clinicians (Collins, Couture et al. 2018). Evaluation of the value and downstream uses of *manually* captured structured data is an important consideration when configuring clinical systems to decrease burden on the clinician and promote 'top of license' practice. Data science approaches for continuous knowledge development are seen as an essential aspect of a Learning Health System. Innovations related to clinical data capture, such as increased capabilities and implementations of device-integrated data and voice-recognition tools, are expected to increase the volume, veracity, variety, and velocity of clinical data for data science and quality purposes. In turn, automated data capture promises to decrease clinician data entry burden and provide greater opportunity for clinical data consumption, interpretation, and decision-making tailored to patient preferences by the care team as part of a patient-centered system.

4 ► http://hhs.gov/ocr/privacy/ (Accessed: 4/26/13).

17.2.4 Advances in Patient-Centered Care Systems

The design and implementation of patient-care systems, for the most part, occurred separately for hospital and ambulatory-care settings. Early patient-care systems in the hospital settings included the University of Missouri-Columbia System (Lindberg 1965), the Problem-Oriented Medical Information System (PROMIS) (Weed 1975), the TriService Medical Information System (TRIMIS) (Bickel 1979), the Health Evaluation Logical Processing (HELP) System (Kuperman et al. 1991), and the Decentralized Hospital Computer Program (DHCP) (Ivers et al. 1983). Among the earliest ambulatory-care systems were the Computer-Stored Ambulatory Record (COSTAR), the Regenstrief Medical Record System (McDonald 1976), and The Medical Record (TMR). For a comprehensive review, see Collen (1995).

According to Collen (1995), the most commonly used patient-care systems in hospitals of the 1980s were those that supported nursing care planning and documentation. Systems to support capture of physicians' orders, communications with the pharmacy, and reporting of laboratory results were also widely used. Some systems merged physician orders with the nursing care plan to provide a more comprehensive view of care to be given. This merging, such as allowing physicians and nurses to view information in the part of the record designated for each other's discipline, was a step toward integration of information. It was still, however, a long way from support for truly collaborative interdisciplinary practice.

Early ambulatory-care systems most often included paper-based patient encounter forms. Some used a computer-scannable mark-sense format while others required clerical personnel to type the data into the computer. Current desktop, laptop, or handheld systems use keyboard, mouse, touchpad, or pen-based entry of structured information, with free text kept to a minimum. Current systems also provide for retrieval of reports and past records. Some systems provide decision support or alerts to remind clinicians to provide needed care, such as immunizations or screening examinations, and to avoid contraindicated orders for medications or unnecessary laboratory analyses. The best provide good support for traditional medical care. Support for comprehensive, collaborative care that gives as much attention to health promotion as to treatment of disease presents a challenge not only to the developers of information systems but also to practitioners and health care administrators who must explicate the nature of this practice and the conditions under which agencies will provide it.

Patient-care information systems in use today represent a broad range in the evolution of the field. Versions of some of the earliest systems are still in use, although most organizations have migrated to commercially available EHR systems. Internally developed systems were generally designed to speed documentation and to increase legibility and availability of the records of patients currently receiving care. As payment models shifted towards population health and value-based care, numerous organizations and providers merged to form healthcare networks. After weighing the costs and benefits associated with further developing the internally developed EHR system and expanding it to newly acquired healthcare providers and sites within their network, many organizations made the decision that the costs and resource requirements for long-term investment in the legacy EHR system were not sustainable. For some organizations this was a difficult decision because the internally developed legacy EHR systems had demonstrated positive outcomes in terms of improved quality and safety. However, substantial resources were required to support continuous development and modification of those systems.

The 2009 enactment of the HITECH Act and its requirements for meaningful use of EHR systems caused many of the commercially available EHR systems to include (or to plan to include) the core functionality needed to achieve the conditions of certification and to meet the population health and reporting requirements of the HITECH Act. Many

legacy EHR systems had inadequate functionality to enable compliance with the new regulations, nor could they support emerging value-based care models. Migration to a single commercial enterprise EHR helped to merge and consolidate clinical data to a single instance (e.g., one patient, one record) regardless of where a patient was seen within the expanding healthcare networks.

More recently developed systems attempt with varying success to respond to the edict "collect once, use many times." Selected items of data from patient records are abstracted manually or electronically to aggregate databases where they can be analyzed for administrative reports, for quality improvement, for clinical or health-services research, and for required patient safety and public health reporting. Such functionality is a key aspect of the federal requirements for meaningful use and interoperability. See ▶ Chap. 19 for a full discussion of public health informatics.

Some recently developed systems offer some degree of coordination of the information and services of the various clinical disciplines into integrated records and **care plans**. Data collected by one caregiver can appear, possibly in a modified representation, in the "view" of the patient record designed for another discipline. When care-planning information has been entered by multiple caregivers, it can be viewed as the care plan to be executed by a discipline, by an individual, or by the interdisciplinary team. Some patient-care systems offer the option to organize care temporally into clinical pathways and to have variances from the anticipated activities, sequence, or timing reported automatically. Others offer a patient "view" so that individuals can view and contribute to their own records. For example, patients hospitalized for cardiac conditions can review selected aspects of their records and enter data about their symptoms such as pain ratings into CUPID (Computerized Unified Patient Interaction Device), an iPad-based application (Vawdrey et al. 2011), or about their goals of care (Dykes et al. 2017). Today, care plans are generally limited to one discipline, disease, or care setting. Even after a decade of HITECH funding, a truly integrative electronic care plan that is comprehensive, cohesive, dynamic, and oriented around actionable patient goals and preferences is still not a reality. The concept of **longitudinal care plans** (LCPs) to support care coordination is referenced in the ACA and HITECH Acts and the ensuing regulations of **Meaningful Use** Stages 1 and 2. This legislation contained financial incentives and penalties, describing implementing LCPs as necessary and dependent on technical interoperability (McDonald et al. 2012). CMS also uses plans of care as part of their rules for eligible providers and facilities that serve Medicare patients with chronic conditions, and for performance-based incentive models (Agency for Healthcare Research and Quality 2014). However, there remains no consensus or definition of an LCP's contents, nor are there best practices for collaborating, updating, and reconciling the care plan across settings and with the patient and family (Dykes et al. 2014).

Electronic documentation of clinician progress notes has lagged behind other functions in EHRs (Doolan et al. 2003). The process of entering notes may occur through dictation, selecting words and phrases from structured lists, use of templates, and typing free text. Amid concerns that salience may be lost in electronic notes (Siegler 2010), Johnson et al. (2008) advocated for a hybrid approach that combines semi-structured data entry and natural language processing within a standards-based and computer-processible document structure. Thus, ability for data re-use is preserved while maintaining clinician efficiency and expressivity. Some progress has been made with sharing EHR notes with patients. **OpenNotes** is a national initiative to share clinician notes with patients. Early research indicates that clinician note transparency supports patient-centered care, empowers informal caregivers, and engages less educated and diverse patients (Chimowitz et al. 2018).

The publication of the Institute of Medicine's reports *To Err is Human* (2000) and *Crossing the Quality Chasm* (2001) resulted in increasing demands from health care providers for information systems that reduce errors in patient care. Information system vendors

are responding by developing such systems themselves and by purchasing the rights to patient care systems developed in academic medical centers that have demonstrated reductions in errors and gains in quality of care and cost control. **Closed loop** medication systems use technologies such as bar codes and decision support to guard against errors throughout the process of prescribing, dispensing, administering, and recording and have been identified as a key intervention to improve medication safety. In a before-and-after evaluation of the closed loop electronic medication administration system at Brigham and Women's Hospital, investigators found a significant reduction in the rates of transcription errors, medication errors, and potential adverse events (Poon et al. 2010). In other contexts, decision support systems offer **clinical practice guidelines**, protocols, and order sets as a starting point for planning individualized patient care; providing alerts and reminders; using knowledge bases and patient data bases to assess orders for potential contraindications; and offering point-and-click access to knowledge summaries and full-text publications. See ▶ Chap. 26 for more information about these systems.

Many health care organizations have substantial investment in legacy systems and cannot simply switch to more modern technology. Finding ways to phase the transition from older systems to newer and more functional ones is a major challenge to health informatics. To make the transition from a patchwork of systems with self-contained functions to truly integrated systems with the capacity to meet emerging information needs is even more challenging (see ▶ Chap. 16). Approaches to making this transition are described in the Journal of the American Medical Informatics Association (Stead et al. 1996). More recently, some institutions have applied Web 2.0 approaches to create configurable user interfaces to legacy systems. For example, MedWISE integrates a set of features that supports custom displays, plotting of selected clinical data, visualization of temporal trends, and self-updating templates as mechanisms for facilitating cognition during the clinical decision making and documentation process (Senathirajah, Bakken, & Kaufman 2014; Senathirajah, Kaufman, & Bakken 2014).

If patient-centered care systems are to be effective in supporting better care, health care professionals must possess the informatics competencies to use the systems. Consequently, many are integrating informatics competencies into health science education (See ▶ Chap. 25). For example, the Quality and Safety Education for Nurses initiative has named and described necessary competencies and associated curriculum to support patient-centered care, including competencies related to quality, safety, team work, and collaboration (Cronenwett et al. 2009). Recently, informatics competencies for nursing leaders were validated (Yen et al. 2017).

To what degree do patient-care disciplines need to prepare their practitioners for roles as informatics specialists? To the degree that members of the discipline use information in ways unique to the discipline, the field needs members prepared to translate the needs of clinicians to those who develop, implement, and make decisions about information systems. If the information needs are different from those of other disciplines, some practitioners should be prepared as system developers.

The mere existence of information systems does not improve the quality of patient care. The adoption and use of advanced features (such as CDS) that are sensitive to both workflow and human factors are needed to improve the quality of care (Stead & Lin 2009; Zhou et al. 2009). Recent safety reports, public policy, and reimbursement incentives raise awareness of the need for patient-centered care systems. Because traditional requirements for EHRs were provider-centric, existing information systems rarely provide the comprehensive suite of advanced features needed to support patient-centered care. However, the ability of systems to support patient-centered care is essential for achieving the vision of health care reform. What are the requirements for patient-centric information systems? How do these requirements drive the design of systems that will support patient-centered care?

17.3 Designing Systems for Patient-Centered Care

In the second decade of the Twenty-first Century, informaticians and clinicians increasingly share a vision for systems that support patient-centered care practices such as interdisciplinary care planning, care coordination, quality reporting, and patient engagement. This evolution is fueled in part by meaningful use requirements that aim to engage patients and families in their health care and to improve care coordination and the overall quality of care provided. Traditional EHR functionality must be expanded to support new features, functions, and care practices including seamless communication, interdisciplinary collaboration, and patient access to information. To optimize human and organizational factors and the integration of systems and workflow, these features must be built into information systems as core requirements, rather than as an afterthought.

The *Principles to Guide Successful Use of Health Care Information Technology* described by the National Research Council (Stead & Lin 2009) provide a comprehensive framework for defining a set of core requirements that will support the design of systems for patient-centered care. This framework defines nine principles related to both evolutionary (i.e., iterative, long-term improvements) and radical (i.e., revolutionary, new-age improvements) changes occurring in the United States' health care system. The principles and associated system design prerequisites are included in ◘ Table 17.3.

17.4 Current Research Toward Patient-Centered Care Systems

Friedman (1995) proposed a typology of the science in medical informatics. His four categories build from fundamental conceptualization to evaluation as follows:

- Formulating models for acquisition, representation, processing, display, or transmission of biomedical information or knowledge
- Developing innovative computer-based systems, using these models, that deliver information or knowledge to health care providers
- Installing such systems and then making them work reliably in functioning health care environments
- Studying the effects of these systems on the reasoning and behavior of health care providers, as well as on the organization and delivery of health care

While the Friedman typology continues to be useful more than 20 years after its inception, we propose extending the second and third categories in ▶ Sects. 17.4.2 and 17.4.3 to expand the focus from clinical informatics as a provider-centric discipline to a discipline that enables and supports patient-centric care. Following are examples of recent research with implications for patient-centered care.

17.4.1 Formulation of Models

For several decades, standards development organizations (SDOs) and professional groups alike have focused on the formulation of models that describe the patient care process and the formal structures that support management and documentation of patient care. The efforts of SDOs are summarized in ▶ Chap. 8. Early SDO efforts focused primarily on representing health care concepts such as professional diagnoses (e.g., medical diagnoses, nursing diagnoses) and actions (e.g., procedures, education, referrals). These efforts were complemented by professional efforts such as those of the Nursing Terminology Summit (Ozbolt 2000). As a result of multi-national efforts, **SNOMED CT** became an international standard that provides a formal model for concepts that describe clinical conditions and the actions of the multidisciplinary health care team (International Health Terminology Standards Development Organization 2011). In addition, SNOMED CT subsets have been developed for specific domains such nursing problems (Matney et al. 2012). Toward the goal of patient-centered care, attention has also been paid to approaches for formal

Table 17.3 Principles to guide successful use of health care information technology

	Principle	System design prerequisites
Evolutionary change	1. Focus on improvements in care—technology is secondary	Gaps in patient-centered care are clearly defined and operationalized. Health care IT is employed to enable the process changes needed to close gaps in patient-centered care.
	2. Seek incremental gain from incremental effort	An organization's portfolio of health care IT projects has varying degrees of investment. Each project is linked to measurable process changes to provide ongoing visible success with closing gaps in patient-centered care.
	3. Record available data so they can be used for care, process improvement, and research	Health care IT systems support auto capture of data about people, processes, and outcomes at the point of care. Data are used in the short term to support incremental improvements in patient-centered care processes. An expandable data collection infrastructure is employed that is responsive to future needs that cannot be anticipated today.
	4. Design for human and organizational factors	Clear consideration is given to sociological, psychological, emotional, cultural, legal, economic, and organizational factors that serve as barriers and incentives to providing patient-centered care. Health care IT should eliminate the barriers and enable the incentives, making it easy to provide patient-centered care.
	5. Support the cognitive functions of all caregivers, including health professionals, patients, and their families	Health care IT systems include advanced clinical decision support for high-level decision-making that is sensitive to both workflow and human factors.
Radical change	6. Architect information and workflow systems to accommodate disruptive change	Health care IT systems are designed using standard interconnection protocols that support the patient-centered care processes and roles of today while accommodating rapidly changing requirements dictated by new knowledge, care venues, policy, and increasing patient engagement.
	7. Archive data for subsequent re-interpretation	Health care IT systems support archival of raw data to enable ongoing review and analysis in the context of advances in biomedical science and patient-centered care practices.
	8. Seek and develop technologies that identify and eliminate ineffective work processes	Health care IT system design is preceded by a thorough investigation of current and future state work processes of all stakeholders (including patients and their families). Health care IT systems support efficient workflows that leverage ubiquitous access to information and communication and are not constrained by existing care venues or provider-centric practice patterns.
	9. Seek and develop technologies that clarify the context of data	Health care IT systems facilitate patient-centered care by presenting information in context with patient values and preferences and in a format that is understandable and actionable.

representation of terms that patients use to describe their problems and actions (Doing-Harris and Zeng-Treitler 2011).

In more recent years, the focus has turned to the development of **information models** (e.g., clinical elements model) (Oniki et al. 2016) and formal document structures that support sharing of data across heterogeneous information systems and care coordination. In terms of formal document structures, **Logical Observation Identifiers Names and Codes (LOINC)** provides a formal naming convention for document titles and sections (Hyun et al. 2009; Rajamani et al. 2015) and documents are represented according to the HL7 Consolidated **Clinical Document Architecture (C-CDA)** standard including the Release 2 Care Plan and the **Continuity of Care Document (CCD)**. Matney and colleagues illustrated the application of the C-CDA to support the nursing process (Matney, Warren et al. 2016; Matney, Dolin et al. 2016). A CCD designed specifically for low socioeconomic status persons living with HIV/AIDS (PLWH) enrolled in a special needs plan was implemented for viewing by PLWH, their clinicians, and case managers to promote coordination and quality of care (Schnall, Cimino et al. 2011; Schnall, Gordon et al. 2011). More recently, a set of scalable, standards-based approaches has been developed to support interaction of external systems with the native functions of vendor-based EHRs. The **Fast Health Interoperability Resource (FHIR)**, an HL7 standard, has gained traction as a mechanism for information exchange using a well-defined and limited set of resources. Of particular relevance to patient-centered care, Lee and colleagues (2016) developed a FHIR profile for cross-system exchange of a full pedigree-based family health history for applications used by clinicians, patients, and researchers. Built upon FHIR, the **Substitutable Medical Applications and Reusable Technologies (SMART)** platform enables EHR systems to behave as 'iPhone-like platforms' through an application programming interface and a set of core services that support easy addition and deletion of third party apps, such that the core system is stable and the apps are substitutable (Mandel et al. 2016). **CDS Hooks** is designed to invoke external CDS services from within the EHR workflow based upon a triggering event.[5] Services may be in the form of (a) *information cards* – provide text for the user to read; (b) *suggestion cards* – provide a specific suggestion for which the EHR renders a button that the user can click to accept, with subsequent population of the change into the EHR user interface; and (c) *app link cards* – provide a link to an app.

17.4.2 Development of Innovative Systems

For the purposes of developing innovative patient-centered care systems, the second category of the Friedman Typology described in ▶ Sect. 19.4 is expanded to address the use of models that deliver information or knowledge to both health care providers *and* patients. Consumers regularly use information and communication technology to support decision making in all aspects of their lives. However, access to tools to support health care decision making is suboptimal (Krist and Woolf 2011). Krist et al. (2010) proposed five levels of functionality for patient-centered health information systems.

- Level 1: Collects patient information related to health status, behaviors, medications, symptoms, and diagnoses (e.g., electronic version of traditional paper records maintained by patients)
- Level 2: Integrates patient information with clinical information (e.g., personal health record tethered to an EHR)
- Level 3: Interprets information to provide context in an appropriate level of health literacy
- Level 4: Provides tailored recommendations based on patient information, clinical information, and evidence-based guidelines
- Level 5: Facilitates patient decision-making, ownership, and action

The levels of functionality needed to support patient-centered health information

5 ▶ https://cds-hooks.org/(Accessed 5/31/18).

systems relate directly to several of the *Principles to Guide Successful Use of Health Care Information Technology* described by the National Research Council (Stead & Lin 2009) and outlined in ▶ Sect. 17.3; specifically principles 5, 8 and 9 (see ◘ Table 17.3).

Partners Health Care System in Boston, MA, developed Patient Gateway, a secure patient portal serving over 65,000 patient users from primary and specialty care practices affiliated with the Dana Farber Cancer Institute, Brigham and Women's Hospital, and Massachusetts General Hospital. Patient Gateway is a tethered **personal health record** (see ▶ Chap. 11) that provides functionality in line with the five levels described by Krist and Woolf. For example, tools for management of chronic illness are used by patients and providers to promote adherence with evidence-based health maintenance guidelines and to improve collaboration on diabetes self-management plans (Grant et al. 2006; Wald et al. 2009). Research on patient response and satisfaction with the Patient Gateway suggests that patients appreciate the ability to communicate electronically with providers, they welcome greater access to their health information including test results, and they believe that Patient Gateway enables them to better prepare for visits (Grant et al. 2006; Schnipper et al. 2008; Wald et al. 2009). Evaluations of patient satisfaction with personal health records with similar levels of functionality at other sites, including Geisinger Health System (Hassol et al. 2004), Group Health Cooperative (Ralston et al. 2007), and Virginia Commonwealth University (Krist et al. 2010), are consistent with the results reported at Partners Health Care System.

While traditionally patient portals have been associated with ambulatory care, some health care systems are providing modules within their patient portals to inform and activate patients and family during an acute hospitalization and associated transitions (Grossman et al. 2018). Early research indicates that in addition to the features commonly found in ambulatory patient portals (e.g., medications, labs, educational content, scheduling features), the acute care modules provide patient access to the plan of care, daily schedules, and direct care team communication. Some inpatient portals support patient-generated content such as notes, patient-provider messaging, and patient feedback related to their care plan or discharge plan. Research is needed on the use of acute care portals to explore their impact on patients' abilities to successfully navigate information-rich acute care hospitalizations and to examine the effects of portal use on patient activation, engagement, and the overall quality of care (Grossman et al. 2018).

The involvement of users has been identified as fundamental to well-designed systems that are usable and useful in the context of busy patient care workflows (Rahimi et al. 2009). Some examples of development activities where user involvement is needed are content standardization, workflow modeling, and **usability testing**.

- *Content standardization*: Content standardization includes identifying EHR content needed to support documentation of care provided and identification of data needed for reuse (e.g., decision support, quality reporting, and research). Content that is shared across disciplines and patients is identified. Content is modeled using standards to ensure data reuse and interoperability (Principle 3, ◘ Table 17.3) (Chen et al. 2008; Dykes et al. 2010; Kim et al. 2011).
- *Workflow modeling*: Sound modeling of the clinical workflow that underlies an electronic system is essential to designing systems that are usable by care team members (Peute et al. 2009). Workflow models are based on observations of current state clinical workflows including interactions with patients, staff, equipment, and supplies. Understanding of workflow interactions, including current state inefficiencies, informs effective and efficient future state workflows, use-case development, and system prototypes (Rausch & Jackson 2007; Mlaver et al. 2017). Workflow modeling of patient-centered systems includes clear evaluation of ways to use technology to identify and eliminate ineffective work processes (Principle 8, ◘ Table 17.3). Design of new systems is an opportunity

to provide ubiquitous access to information and communication by all care team members including patients. Applying these principles in workflow modeling assures that future state workflows are not constrained by existing care venues or provider-centric practice patterns.

- *Usability testing*: A key lesson learned from Computerized Physician Order Entry implementations is that electronic systems with poor usability interfere with clinical workflow. The unintended consequences of poorly designed systems are well known, and some widely disseminated papers (Ash et al. 2004, 2009; Koppel et al. 2005) have called into question the safety of using such systems with patients. Examples of common usability problems include overly cluttered screen design, poor use of available screen space, and inconsistencies in design. Involving end-users in design and enforcing usability design standards when building clinical systems prevents implementing systems that are difficult to use and interfere with, rather than support, patient-centered care (Principles 4, 5 and 8, ◘ Table 17.3).

Innovative systems to support patient care often take advantage of information entered in one context for use in other contexts.

For example, the Brigham and Women's Hospital Patient Safety Learning Lab in Boston developed a provider checklist and a patient-centered toolkit that used information from the order entry, scheduling, flowsheets (nursing documentation), and other systems to auto-populate a suite of tools used by clinicians and patients to improve team communication and patient safety. The iterative, participatory development process led to tools that are used every day in the medical intensive care units and that demonstrated significant reductions in adverse events and improvement in patient and family satisfaction (Mlaver et al. 2017).

The principle of entering information once for multiple uses also drove development of the bedside displays for inpatients and the care team at Brigham and Women's Hospital (Duckworth et al. 2017). A patient safety plan dashboard was developed that captures disparate data from the EHR and presents a personalized display as the screensaver on the bedside computer workstation. The dashboard aligns all care team members, including patients and families, in the safety plan. The screensaver content includes icons that provide actionable alerts related to patient-specific safety concerns. These bedside displays combine data from many sources to support the integrated care of physicians, nurses, and family members.

At Partners Health Care System, system developers are working with clinical teams to identify system requirements, to iteratively develop, and to test patient-centric systems that integrate decision support into the clinical workflow. For example, Dykes et al. (2010) developed a fall prevention toolkit that reuses fall risk assessment data entered into the clinical documentation system by nurses and automatically generates a tailored set of tools that provide decision support to all care team members, including patients and their family members at the bedside (Dykes et al. 2009). The fall prevention toolkit logic was developed from focus groups of professional and paraprofessional caregivers (Dykes et al. 2009), and of patients and family members (Carroll et al. 2010). As nurses complete and file the routine fall risk assessment scale, the documentation system automatically generates a tailored bed poster that alerts all team members about each patient's fall risk status and patient-appropriate interventions to mitigate risk. In addition, a patient education handout is generated that identifies why each individual patient is at risk for falls and what the patient and family members can do while in the hospital to prevent a fall. The icons used in the Fall TIPS poster and patient education handout have been developed and validated using a participatory design process with clinicians and patients (Leung et al. 2017; Hurley et al. 2009). In a randomized control trial of over 10,000 patients, the toolkit was associated with a 25% reduction in falls (Dykes et al. 2010). The Fall TIPS toolkit reduced falls by leveraging HIT to complete the three-step fall prevention process: (1) conduct fall risk assessments, (2) develop tailored fall prevention plans with

evidence-based interventions, and (3) consistently implement the plan. We learned that Fall TIPS was most effective at reducing falls and related injuries when patients and family were engaged in all three steps of the fall prevention process (Dykes et al. 2017).

17.4.3 Implementation of Systems

Much has been written about HIT failures and associated costs and consequences (Bloxham 2008; Booth 2000; McManus & Wood-Harper 2007; Ornstein 2003; Rosencrance 2006). Higgins and associates (see Rotman et al. 1996) described the lessons learned from a failed implementation of a computer-based physician workstation that had been designed to facilitate and improve ordering of medications. Those lessons are not identical to, but are consistent with, the recommendations of Leiner and Haux (1996) in their protocol for systematic planning and execution of projects to develop and implement patient-care systems. While long term follow-up of a vendor EHR implementation with advanced CDS identified lower prescribing error rates, achieving prior levels of perceived prescribing efficiency took nearly 2 years (Abramson et al. 2013, 2016).

In response to evidence of unintended consequences and clinicians voicing concerns after system implementations, the **American Medical Informatics Association (AMIA)** EHR 2020 Task Force on the Status and Future Directions of EHRs published a report in 2015 that outlined 10 recommendations that span five areas. These recommendations are in response to current barriers to quality care delivery experienced at many organizations after EHR system implementations. The five areas addressed were: simplify and speed documentation, refocus regulation, increase transparency and streamline certification, foster innovation, and support person-centered care delivery. Specific recommendations discussed decreasing documentation burden, improving the designs of interfaces so that they support and build upon how people think (i.e., cognitive-support design), and promoting the integration of EHRs into the full social context of care, moving beyond acute care and clinic settings to include home health, specialist care, laboratory, pharmacy, population health, long-term care, and physical and behavioral therapies (Payne et al. 2015).

Several studies have quantified documentation burden. In one setting, resident physicians spent 85 minutes per day authoring and viewing notes (Hripcsak et al. 2011). In another setting, on average, nurses perform 631–662 manual flowsheet data entries per 12-hour shift (excluding device integrated data), averaging to one data point every 0.82–1.14 minutes in acute care (Collins, Couture et al. 2018). Further, EHR log file analyses indicate nurses spend 21.4–38.2 minutes per day authoring notes, on average (Collins, Couture et al. 2018), yet fewer than 20% of nursing notes were read by physicians, and only 38% were read by other nurses (Hripcsak et al. 2011). There is an overall lack of standardization and consistency of data definitions within EHRs, leading to a proliferation of data elements that contribute to EHR burden and inhibit interoperability and automated reporting (Zhou et al. 2016; Collins, Bavuso et al. 2017; Collins, Klinkenberg-Ramirez et al. 2017; Collins, Rozemblum et al. 2017). Methods to increase consistency of data definitions have been published, but are often minimally implemented due to project timelines and limited resources (Collins et al. 2015, Collins, Bavuso et al. 2017, Collins, Klinkenberg-Ramirez et al 2017, Collins, Rozemblum et al 2017). Efforts by national organizations to improve consistency of data definitions include the American Medical Association's Integrated Health Model Initiative[6] and collaborations between the Office of the National Coordinator for Health Information Technology (ONC) and CMS.[7]

As these experiences demonstrate, the implementation of patient-care systems is far more complex than the replacement of one technology with another. Such systems

6 AMA. Integrated Health Model Initiative, 2018. ► https://ama-ihmi.org (Accessed 9/25/18).

7 ONC/CMS Reducing Clinician Burden Meeting. February 22, 2018. ► https://www.healthit.gov/news/events/onccms-reducing-clinician-burden-meeting (Accessed 9/25/18).

transform work and organizational relationships. If the implementation is to succeed, attention must be given to these transformations and to the disruptions that they entail. Southon et al. (1997) provided an excellent case study of the role of organizational factors in the failed implementation of a patient-care system that had been successful in another site.

To realize the promise of informatics for health and clinical management, people who develop and promote the use of applications must anticipate, evaluate, and accommodate the full range of consequences. In early 2003, these issues came to the attention of the public when a large academic medical center decided to temporarily halt implementation of its CPOE system due to mixed acceptance by the physician staff (Chin 2003; Ornstein 2003). A case series study by Doolan et al. (2003) identified five key factors associated with successful implementation: (1) having organizational leadership, commitment, and vision; (2) improving clinical processes and patient care; (3) involving clinicians in the design and modification of the system; (4) maintaining or improving clinical productivity; and (5) building momentum and support amongst clinicians. A collaboration of ten AMIA working groups and the **International Medical Informatics Association** Working Group on Organizational and Social Issues cosponsored a workshop to review factors that lead to implementation failure. These include poor communication, complex workflows, and failure to engage end-users in clearly defining system requirements. Recognizing that the problems encountered in failed implementations tend to be more administrative than technical, they recommended the following set of managerial strategies to overcome implementation barriers (1) provide incentives for adoption and remove disincentives; (2) identify and mitigate social, IT, and leadership risks; (3) allow adequate resources and time for training before and after implementation, including ongoing support; and (4) learn from the past and from others about implementation successes and failures and about how failing situations were turned around (Kaplan and Harris-Salamone 2009).

In 2014, the SAFER (Safety Assurance Factors for Electronic Health Record Resilience) Guides were first published to facilitate proactive risk assessments of EHR safety and usability related policies, processes, procedures, and configurations at healthcare organizations (Ash et al. 2016). These guides are endorsed by the ONC and available at ▶ HealthIT.gov. The SAFER Guides include nine guides organized into three broad groups: foundational guides, infrastructure guides, and clinical process guides. Recent evaluations indicate that health organizations would benefit from broader implementation of these guides and principles of safety and usability (Ash et al. 2016; Sittig et al. 2018).

For the purposes of promoting successful implementation of patient-centered systems, the third category of the Friedman typology is expanded to provide access to information to all team members including patients and their families or caregivers outside of traditional health care settings as follows: *Installing such systems and then making them work reliably in functioning health care environments and other settings where information is needed to promote health and wellbeing*. The majority of self-management occurs outside traditional health care settings. As noted in ◘ Table 17.3, a prerequisite for patient-centered systems is that they support efficient workflows with ubiquitous access to information and communication and that the systems are not constrained by existing care venues or provider-centric practice patterns (Principle #8). Clinical workflows are highly complex and data-rich, requiring formal analysis and evaluation before and after system implementation. For example, a time-motion study found that, on average, nurses engage in 31 communications and 52 hands-on tasks per hour, and multi-task 18.63% of the time (Yen et al. 2016). Strategies to involve users in system design or selection and customization will support successful implementation of systems that meet user expectations (Burley et al. 2009; Rahimi et al. 2009; Saleem, Russ, Justice et al. 2009; Saleem, Russ, Sanderson et al. 2009). User involvement in defining future workflows contributes to a shared understanding about the impact of information systems on clinical tasks and workflows (Leu et al. 2008).

Careful attention to the *Principles to Guide Successful Use of Health Care Information Technology* during system design will support successful implementation. For example, the principles related to evolutionary health care changes keep the focus on designing and implementing usable systems that enable patient-centered care practices. Principles related to radical change focus on development of flexible, adaptable systems that are architected to accommodate disruptive change and iterative development based on end-user feedback.

17.4.4 Effects of Clinical Information Systems on the Potential for Patient-Centered Care

Electronic health records and CPOE systems are intended to support safe, evidence-based, patient-centered care by examining patient-specific information, agency-specific information, and domain-specific information in the clinical context and proposing appropriate courses of action or alerting clinicians to potential dangers. Many current systems, however, fail to follow design principles that take into account the real contingency-driven, non-linear, highly interrupted, collaborative, cognitive, and operational workflow of clinical practice (Ash et al. 2004). These flaws can lead to errors in entering and retrieving information, cognitive overload, fragmentation of the clinical overview of the patient's situation, lack of essential operational flexibility, and breakdown of communication. Physicians, in particular, have found themselves chagrined by changes in the power structure as they have devoted more time to entering information and orders while other members of the health team have gained greater access to information and the concomitant capacity to make certain decisions without consulting the physician. Clinicians across the range of professions have expressed concern about the decrease in face-to-face communication with its verbal and non-verbal richness, negotiation, and redundant safety checks as more and more clinical information is exchanged via the computer (Campbell et al. 2006; Ash et al. 2007). These and other unintended consequences of EHRs and CPOE systems are the subject of ongoing research. Detecting and finding ways to prevent or mitigate the adverse, unintended consequences of these systems will be critical for supporting patient-centered care.

A number of unintended consequences stem from the incompatibility of system design with the clinician's cognitive workflow. For example, systems that make it difficult to find and retrieve information can interfere with patient-centered care. In a hospital preparing to implement a commercial CPOE system, investigators compared the efficiency, usability, and safety of information retrieval using the vendor's system, the current paper form, and a prototype CPOE developed on principles of **User Centered Design**. They found the prototype system to be similar to the paper form and both to be significantly superior to the vendor's system in efficiency, usability, and safety (Chan et al. 2011).

Other unintended consequences arise from over-reliance on the information system because of limited understanding of its design and capacities. To date, many CDSS are somewhat limited in their ability to incorporate patient-specific data into their decision algorithms. A synthesis of 17 systematic reviews conducted with sound methodology found that CDSS often improved providers' performance, especially in medication orders and preventive care. The reviewers noted, "These outcomes may be explained by the fact that these types of CDSS require a minimum of patient data that are largely available before the advice is (to be) generated: at the time clinicians make the decisions" (Jaspers et al. 2011, p. 327).

On the other hand, many systems offer functionalities that support patient-centered care. An important component of patient-centered care is the application of evidence in a plan of care tailored to the patient's needs. To increase the use of evidence-based order sets, investigators at Sinai-Grace Hospital in Detroit, Michigan embedded into the general admission order set specific, evidence-based order sets for the most common primary and secondary diagnoses for patients admitted to

their medical service. The result was a fivefold increase in the use of evidence-based order sets in the 16-month period following implementation (Munasinghe et al. 2011).

As in the examples above, most systems to support the cognitive workload have been directed toward physicians. Patient-centered care requires a broader perspective. A study of the information needs of case managers for PLWH found that the most frequent needs were for patient education resources (33%), patient data (23%), and referral resources (22%) (Schnall, Cimino et al. 2011). The investigators recommended that targeted resources to meet these information needs be provided in EHRs and continuity of care records through mechanisms such as the **Infobutton Manager**.

Key to patient-centered care is communication among all the health care professionals on the patient's team. A study at one academic medical center (Hripcsak et al. 2011) reviewed EHRs of hospitalized patients, along with usage logs, to make inferences about time spent writing and viewing clinical notes and patterns of communication among team members. In this setting, the core team for each patient consisted of one or more attending physicians, residents, and nurses, with social workers, dieticians, and various therapists joining the team later. Results showed that clinical notes were more likely to be reviewed within the same professional group, with attending physicians and residents viewing notes from nurses or social workers less than a third of the time. The investigators proposed that it might be useful to develop ways for EHRs "to summarize information and make it readily available, perhaps with the ability of the author to highlight information that may be critical and that has a high priority for communication" (p. 116) (Hirsch et al. 2014). They also noted that their study was limited to communications within the EHR and did not take into account face-to-face or telephone communications that might have occurred, especially in urgent situations. They suggested further research involving direct observation of clinicians, time-motion analyses, and think-aloud methods to develop deeper knowledge of how clinicians communicate about patient care, especially across the professions.

The quality of documentation tools can have a profound effect on whether information, even if communicated face-to-face, is acted upon in clinical care (Collins et al. 2011). In a neurovascular intensive care unit, therapeutic goals for patients were stated during daily interdisciplinary rounds. In this setting, the interdisciplinary team treated the attending physician's note as a common patient-focused source of information. Although the attending physician's note contained 81% of the stated ventilator weaning goals, it included only 49% of the stated sedation weaning goals. Overall, nearly a quarter of stated goals were not documented in the note. If a goal was not documented, it was 60% less likely to have a related action documented. Nurses' documentation rarely mentioned the goals, even if actions recorded were consistent with the goals as stated during rounds. Notably, the nurses' structured documentation system did not support sedation-related goals, even though sedation weaning was a nursing responsibility in this setting. The authors noted that the frequent omission of sedation goals from the attending physician's note might be because this nursing function was not a billable goal or act. They also expressed concern that the omission from the EHR of evidence of important clinical judgments nurses make could impair patient safety, quality improvement, and development of nursing knowledge. Thus, in this example, although the interdisciplinary team was collaborating in setting and reaching therapeutic goals, deficiencies in their processes and in the nurses' documentation system limited their achievement of patient-centered care.

A study at Vanderbilt University Medical Center also demonstrated both strengths and shortcomings in the ability of a clinical information system to support patient-centered care. Attending physicians in the Trauma and Surgical ICUs established protocols for Intensive Insulin Therapy that were built into a CDSS to advise nurses on insulin doses based on a patient's blood glucose and insulin resistance trends. In 94.4% of studied instances, nurses administered the recommended dose. When nurses overrode the recommended dose, they overwhelmingly administered less insu-

lin than the recommended dose, leading to a higher incidence of hyperglycemia than when the recommended dose was administered. Nurses appeared more concerned about hypoglycemia than hyperglycemia and to consider the patient's blood glucose but not the insulin resistance trend. They also noted that their workflow was impeded by the need to record information about the blood glucose, insulin dose, and primary dextrose source in two places—the CPOE that included the CDSS and the separate nurse charting system. The investigators' recommendations included displaying information about insulin resistance trends on screen (provided that this did not produce clutter and confusion) and developing clinical information systems that do not require double documentation. Strengths of this example for supporting patient-centered care include the collaboration of physicians and nurses in maintaining blood glucose in the desired range based on patient data. Shortcomings include the failure to present nurses with information about the patient's insulin resistance trend to aid their decision-making and the requirement that they record the same data in two places, thereby reducing time for direct patient care (Campion Jr et al. 2011).

Patient-centered care systems not only support the cognitive work and communications of clinicians; they also take into account the resources patients and families use to manage their health concerns. Increasingly, patients and family members engaged in promoting their own health turn to social networking sites on the Internet. While research on the impact of social networking sites on patient-related outcomes is in its infancy, early research indicates that social networking sites are used by patients and family members to get informational and social support for day-to-day management of chronic illnesses such as diabetes and heart failure (Mogi et al. 2017; Partridge et al. 2018). Several studies indicate that the primary objectives for using social networking sites were to request information, to provide information to others with similar conditions, to express emotion about one's own condition, to provide emotional support to others, and to promote a specific product (Greene et al. 2011; Mogi et al. 2017). Many health care professionals recognize the potential of social networking sites for the peer support that patients gain from participation, but they have concerns about the accuracy of the content that is offered on these sites. Available research indicates great variability in the accuracy and effectiveness of the clinical content offered on **social networking sites** (Greene et al. 2011; Mogi et al. 2017). One study found that approximately one-third of the posts on health-related social networking sites could be classified as an advertisement of a non-FDA-approved remedy or cure (Greene et al. 2011). It was also noted that requests for personal information were not uncommon, potentially making participants vulnerable to solicitations from product manufacturers or vendors. Clinicians should be aware of the social networking sites related to their area of expertise. Discussing social networking options with patients, including the advantages and disadvantages and the potential benefits and problems, can help patients to be more judicious consumers of social networking sites.

In patient-centered care, personal health records (see ► Chap. 13) are often viewed as a means of communication between patients and providers and as a method of engaging patients in understanding and acting in the interests of their own health. A 2016 international study to understand perspectives on sharing data with patients categorized countries that are focused on encouraging patients to receive access to their clinical data, within the following stages of maturity: Established, Emerging, and Limited. Countries with "Established" levels of maturity for sharing patient data had been focused on this work since the early 2000s and included Israel, England, Canada, Australia, and the United States (Prey et al. 2016). Expanded efforts to share patient data led to "Emerging" status for Austria, Argentina, Brazil, the Netherlands, Portugal, South Korea, Switzerland, and Uruguay. Iran, Japan, and Kenya were described as focused on EHR implementation with "limited" patient engagement and the potential for increased focus on patient engagement in the future. In 2013, ONC announced the Blue Button initiative, promoting patients' legal rights to receive their personal health informa-

tion and recognizing sites that enable consumers to download their health records. A 2017 ONC Data Brief reported that 52 percent of individuals have been offered online access to their medical record in the United States and over half of those individuals viewed their record within the past year, the equivalent of 28 percent of individuals nationwide (Patel & Johnson 2018).

Most of the clinical data being shared through portals is structured, coded information such as problem lists, allergies, medication lists, appointments, and laboratory results. Sharing of narrative notes is not widespread, but, as noted previously, the OpenNotes project has demonstrated the successful sharing of notes with patients. To date, over 28 million patients have online access to notes, and the network of providers participating in this project continues to grow.[8]

Patient portals, which evolved as a mechanism to access personal health data and populate personal health record platforms, began to take hold in the 2000s. Several pilots during this time, such as the Military Health System's MiCare portal in 2008, resulted in critical learnings and informed optimization of portal functionality over the following decade. Lessons learned include the following: (1) transfer of data upon specific patient request is more efficient than automatic transfer; (2) patient representatives prefer instant access to all their data, while many providers prefer an embargo time, particularly for release of sensitive results data; (3) inefficient provider access to personal health records and siloed repositories with incomplete information might pose the danger of ill-informed clinical decisions; and (4) giving patients the power to determine what medical information to share with the provider could similarly lead to clinical decisions made in the absence of vital information, with resulting harm to patients. The MiCare pilot concluded that while there is broad agreement on desired functionalities for portals, challenging tensions remain between patients' desire for access to and control of health information and providers' needs for full information about the patient and for appropriate opportunities for ethical disclosure of information to patients. Over the following decade throughout the 2010s, core functionality for portals continued to expand and mature. Common features now include secure messaging, prescription renewals and refills, appointment requests online, online bill pay, referral requests, medication reminders, links to reference materials, use of low health literacy terms, and increasing access to several types of clinical data. Clinical data released in portals include routine and sensitive laboratory results, genetic test results, medications, encounter information, allergies, immunizations, radiology and other diagnostic reports, problem lists, discharge summaries, and clinical notes. In many organizations embargo periods for releasing results have lessened or been eliminated, although variability remains.

As portal functionalities expanded and were optimized, acute care patient portals (portal access tailored to the hospital setting) began to emerge. Although some clinicians feared that patient portal access to data during a hospitalization would burden them with excessive patient inquiries, their fears were no more realized than were those of clinicians wary of the OpenNotes project (Collins, Bavuso et al. 2017; Collins, Klinkenberg-Ramirez et al. 2017; Collins, Rozemblum et al. 2017). User-centered design with patients and families and other key clinician and administrative stakeholders specified that acute care portals could provide value by humanizing the patient-clinician connection, facilitating the maintenance and sharing of verbal communication, and promoting ubiquitous and equitable access to information. Key features specified for portals in the acute care setting are the provision of clinical data, messaging with clinicians, glossary of clinical/hospital terms, patient education resources, patient diary, patient notepad for reminders, resources that support family involvement, and tiered displays for information-dense clinical data. New clinical workflows are required to integrate portals within the acute care setting, and these are facilitated by active clinician engagement and demonstration of improved patient outcomes and satisfaction (Collins, Bavuso et

8 ► https://www.opennotes.org/ (Accessed 9/25/18).

al. 2017; Collins, Klinkenberg-Ramirez et al. 2017; Collins, Rozemblum et al. 2017). Patient portals are increasing, but not yet in routine use, in the acute and post-acute setting.

In 2018, 7 key focus areas were identified for sociotechnical and evaluation research related to patient engagement and portal use in the acute and post-acute care settings. These identified research areas were (1) standards for interoperability, functionality, and patient-driven use case models; (2) appropriate access and policies for privacy and security; (3) user-centered design; (4) implementations that integrate with workflows for sustainable adoption; (5) data and content management and visualizations with inclusion of novel data sources (e.g., multimedia tools, safety reporting platforms, **social determinants of health** [SDOH] data); (6) CDS for patients and care partners; and (7) systematic evaluation of process, balance, and outcomes measurement (Collins, Dykes et al. 2018).

Importantly, patients expect that patient portals across ambulatory and acute care settings are seamless and integrated for ease of access within the patient's control. Seamless data access within the patient's control is not yet broadly realized, but several technical and policy initiatives offer promising paths that were not possible previously. The federal 21st Century Cures Act (Cures Act) includes provisions to improve patients' access to their health data and simplification of the patient's ability to electronically share their information. Aligned with those provisions, there are ongoing interoperability efforts with a primary focus on connecting mobile health apps and devices to EHRs using open **application programming interfaces** (APIs) to allow individuals to collect, manage, and share their health information. ONC, aligned with the Cures Act, promotes policy choices that will give consumers, clinicians, and innovators more options for getting to and using health information with specific HIT certification criteria that call for the development of modern APIs that do not require "special effort" to access and use (Rucker 2018). ONC promotes the use of the FHIR standard for representing clinical data with APIs.

Electronic health records and other computer-based information resources can influence the provision of patient-centered care even when the patient and the provider are in the same room. A study of computer use during acute pediatric outpatient visits found that female physicians were more likely than males to be communicating with patients and families while using the computer (Fiks et al. 2011). A recent study in ambulatory care revealed that although providers reported improvements attributable to EHRs (e.g., communication between providers, review of results with patients, and review of follow-up to testing results with patients), they perceived a negative effect on patient-provider connection (Sandoval et al. 2016). An observational study involving 20 primary care physicians and 141 of their adult patients showed how the inclusion of the computer in the clinical consultation can help patients shift the balance of power and authority toward shared decision-making and patient-centered care (Pearce et al. 2011). This Australian study found that about one-third of the patients actively included the computer as a party to the consultation, drawing the physician's attention to it as a source of information or authority. They concluded, *"In the future, computers will have greater agency, not less, and patients will involve themselves in the three-way consultation in more creative ways—for example, through online communication, or through the plugging into computers of their own electronic records, creating a situation where they co-own the information in the computer …. By democratizing and commoditizing information flows and authority in the consultation, we may in fact create truly patient-centered medicine, with the patient directing the action"* (p. 142).

As these examples illustrate, the complexity of collaborative, interdisciplinary, patient-centered care poses serious challenges to the design of clinical information systems. Many systems fall short in supporting cognitive work, even from a clinician-centric perspective. Supporting communications among clinicians, between clinicians and patients, and among patient and family support groups presents myr-

iad technical and ethical problems. Still, researchers and clinicians increasingly share a vision of patient-centered care that drives them to push the frontiers and develop support for this emerging model of care.

17.5 Outlook for the Future

Social and political forces have begun to transform health care in the United States, and HIT is advancing to support the changes. The transformation is rapid, disruptive, and not always smooth, but mandates and incentives are aligning with social and economic imperatives to maintain progress.

To meet demands for patient-centered care, changes must occur in clinician practice patterns and processes, in the organization and management of health services, and in the education of health care professionals and the public. To support patient-centered care, clinicians, informatics professionals, and computer scientists must develop health information and communication technologies that support collaboration; cognitive processes and operational workflow; communication and shared decision making between and among clinicians, patients, and family members; and trustworthy tools for the management of personal and family health.

Transformational change is daunting, and resistance is inevitable. Still, the chances for success have never been better. The vision of health care articulated by the National Academies of Sciences is guiding policy, research, and practical action by government agencies, health care providers, and the public.

Suggested Readings

Ahern, D. K., Woods, S. S., Lightowler, M. C., Finley, S. W., & Houston, T. K. (2011). Promise of and potential for patient-facing technologies to enable meaningful use. American Journal of Preventive Medicine, 40(5 Suppl 2), S162–S172. This article describes specific technologies that patients can use in the interests of their health and that support patient-centered care.

Chimowitz, H., Gerard, M., Fossa, A., Bourgeois, F., & Bell, S. K. (2018). Empowering Informal Caregivers with Health Information: OpenNotes as a Safety Strategy. Joint Commission journal on quality and patient safety, 44(3), 130–136.

Collins, S., Hurley, A., Chang, F., Illa, A., Benoit, A., Laperle, S., & Dykes, P. (2014). Content and functional specifications for a standards-based multidisciplinary rounding tool to maintain continuity across acute and critical care. Journal of the American Medical Informatics Association: JAMIA, 21(3), 438–447. This article describes the content and functional specifications for a standards-based multidisciplinary rounding tool.

Collins, S., Dykes, P., Bates, D. W., Couture, B., Rozenblum, R., Prey, J., et al. (2018). An informatics research agenda to support patient and family empowerment and engagement in care and recovery during and after hospitalization. Journal of the American Medical Informatics Association: JAMIA, 25(2):206–209. Based on a national workshop, this paper presents an informatics research agenda to support patient and family empowerment and engagement in care.

Dykes, P. C., Carroll, D. L., Hurley, A., Lipsitz, S., Benoit, A., Chang, F., et al. (2010). Fall prevention in acute care hospitals: A randomized trial. Journal of the American Medical Association: JAMIA, 304(17), 1912–1918. This article describes development and testing of an electronic fall prevention toolkit integrating decision support into clinical workflow at the bedside and intended for use by all team members including patients and family.

Grossman, L. V., Choi, S. W., Collins, S., Dykes, P. C., O'Leary, K. J., Rizer, M., Strong, P., Yen, P. Y., & Vawdrey, D. K. (2018). Implementation of acute care patient portals: recommendations on utility and use from six early adopters. Journal of the American Medical Informatics Association: JAMIA, 25(4), 370–379.

IOM (Institute of Medicine). (2013). Digital data improvement priorities for continuous learning in health and health care: Workshop summary. Washington, DC: The National Academies Press. This workshop summary describes digital data improvement priorities

for continuous learning in health and health care.

Matney, S. A., Warren, J. J., Evans, J. L., Kim, T. Y., Coenen, A., & Auld, V. A. (2012). Development of the nursing problem list subset of SNOMED CT®. Journal of Biomedical informatics, 45(4), 683–688.

Questions for Discussion

1. What is the utility of a linear model of patient care as the basis for a decision-support system? What are two primary limitations? Discuss two challenges that a nonlinear model poses for representing and supporting the care process in an information system.
2. Compare and contrast additive, clinician-centered versus coordinated, patient-centered models of interdisciplinary patient care. What are the advantages and disadvantages of each model as a mode of care delivery? What are the broad implications for design of information systems to support clinician-centered versus patient-centered models of care?
3. Imagine a patient-care information system that assists in planning the care of each patient independently of all the other patients in a service center or patient-care unit. What are three advantages to the developer in choosing such an information architecture? What would be the likely result in the real world of practice? Does it make a difference whether the practice setting is hospital, ambulatory care, or home care? What would be the simplest information architecture that would be sufficiently complex to handle real-world demands? Explain.
4. Zielstorff et al. (1993) proposed that data routinely recorded during the process of patient care could be abstracted, aggregated, and analyzed for management reports, policy decisions, and knowledge development. What are three advantages of using patient care data in this way? What are three significant limitations?
5. Over the past decade, many of the patient-care information systems designed in the 1970s have been replaced by vendor-based systems. What role have public policy and payment models had in driving this change? How do the practice models, payer models, and quality focus of today differ from those of the past? What differences do these changes require in information systems? What are two advantages and two disadvantages of the older, internally developed legacy systems versus vendor-provided systems?
6. What challenges exist in modeling information for patient-centered care? What considerations are important in designing patient-facing health information and communication technologies?

References

Abramson, E. L., Malhotra, S., Osorio, S. N., Edwards, A., Cheriff, A., Cole, C., & Kaushal, R. (2013). A long-term follow-up evaluation of electronic health record prescribing safety. *Journal of the American Medical Informatics Association, 20*, e52–e58.

Abramson, E., Patel, V., Pfoh, E., & Kaushal, R. (2016). How physician perspectives on e-prescribing evolve over time. *Applied Clinical Informatics, 7*, 994–1006.

Adams, K., & Corrigan, J. (Eds.). (2003). *Priority areas for national action: Transforming health care quality. A report of the Institute of Medicine*. Washington, DC: National Academy Press.

Adler-Milstein, J., & Jha, A. K. (2017). HITECH Act drove large gains in hospital electronic health record adoption. *Health Affairs (Millwood), 36*(8), 1416–1422.

Agency for Healthcare Research and Quality (AHRQ). (2014). *Care coordination measures atlas update: Table 4. Meaningful use measures that assess care coordination**. Available from: https://www.ahrq.gov/professionals/prevention-chronic care/improve/coordination/atlas2014/chapter4tab4.html

Ahern, D. K., Woods, S. S., Lightowler, M. C., Finley, S. W., & Houston, T. K. (2011). Promise of and potential for patient-facing technologies to enable meaningful use. *American Journal of Preventative Medicine, 40*(5 Suppl 2), S162–S172.

Ash, J. S., Berg, M., & Coiera, E. (2004). Some unintended consequences of information technology in health care: The nature of patient care information system-related errors. *Journal of the American*

Medical Informatics Association: JAMIA, 11(2), 104–112.

Ash, J. S., Sittig, D. F., Poon, E. G., Guappone, K., Campbell, E., & Dykstra, R. H. (2007). The extent and importance of unintended consequences related to computerized provider order entry. *Journal of the American Medical Informatics Association: JAMIA, 14*(4), 415–423.

Ash, J. S., Sittig, D. F., Dykstra, R. H., Campbell, E., & Guappone, K. (2009). The unintended consequences of computerized provider order entry: Results of a mixed methods exploration. *International Journal of Medical Informatics, 78*(Supplement 1), S69–S76.

Ash, J., Singh, H., & Sittig, D. (2016). *Safer guides*. HealthIT.gov. Available at: https://www.healthit.gov/topic/safety/safer-guides

Bates, D. W., & Bitton, A. (2010). The future of health information technology in the patient-centered medical home. *Health Affairs (Millwood), 29*(4), 614–621.

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs (Millwood) [Internet], 33*(7), 1123–1131.

Bickel, R. G. (1979). The TRIMIS concept. In *Proceedings of the 1979 annual symposium on computer applications in medical care* (pp. 839–842). Washington, DC.

Bloxham, A. (2008, August 10). £13 Billion NHS computer system failures affecting patient care. *The Telegraph*. Retrieved April 27, 2020, from https://www.telegraph.co.uk/news/uknews/2535099/13bn-NHS-computer-system-failures-affecting-patient-care.html

Booth, R. (2000). *Project failures costly, TechRepublic/Gartner study finds*. Retrieved April 27, 2020, from http://www.techrepublic.com/article/it-project-failures-costly-techrepublicgartner-study-finds/1062043

Burley, L., Scheepers, H., & Owen, L. (2009). *User involvement in the design and appropriation of a mobile clinical information system: Reflections on organisational learning*. New York: Springer.

Campbell, E. M., Sittig, D. F., Ash, J. S., Guappone, K. P., & Dykstra, R. H. (2006). Types of unintended consequences related to computerized provider order entry. *Journal of the American Medical Informatics Association: JAMIA, 13*(5), 547–556.

Campion, T. R., Jr., May, A. K., Waitman, L. R., Ozdas, A., Lorenzi, N. M., & Gadd, C. S. (2011). Characteristics and effects of nurse dosing overrides on computer-based intensive insulin therapy protocol performance. *Journal of the American Medical Informatics Association: JAMIA, 18*(3), 251–258.

Carroll, D. L., Dykes, P. C., & Hurley, A. C. (2010). Patients' perspectives of falling while in an acute care hospital and suggestions for prevention. *Applied Nursing Research, 23*(4), 238–241.

Chan, J., Shojania, K. G., Easty, A. C., & Etchells, E. E. (2011). Does user-centred design affect the efficiency, usability and safety of CPOE order sets? *Journal of the American Medical Informatics Association: JAMIA, 18*(3), 276–281.

Chen, E. S., Zhou, L., Kashyap, V., Schaeffer, M., Dykes, P. C., & Goldberg, H. S. (2008). Early experiences in evolving an enterprise-wide information model for laboratory and clinical observations. In *AMIA annual symposium proceedings, 2008* (pp. 106–110).

Chin, T. (2003, February 17). *American Medical News*.

Cimino, J. J., Overby, C. L., Devine, E. B., Hulse, N. C., Jing, X., Maviglia, S. M., & Del Fiol, G. (2013). Practical choices for infobutton customization: Experience from four sites. In *Proceedings of the AMIA annual symposium* (pp. 236–245).

Collen, M. F. (1995). *A history of medical informatics in the United States, 1950 to 1990*. Bethesda: American Medical Informatics Association.

Collins, S. A., & Vawdrey, D. (2012). "Reading between the lines" of flowsheet data: Nurses' optional documentation associated with cardiac arrest outcomes. *Applied Nursing Research, 25*(4), 251–257.

Collins, S. A., Bakken, S., Vawdrey, D. K., Coiera, E., & Currie, L. M. (2011). Agreement between common goals discussed and documented in the ICU. *Journal of the American Medical Informatics Association: JAMIA, 18*(1), 45–50.

Collins, S., Cato, K., Albers, D., Scott, K., Stetson, P., Bakken, S., & Vawdrey, D. (2013). Relationship between nursing documentation and patients' mortality. *American Journal of Critical Care, 22*(4), 306–313.

Collins, S., Gazarian, P., Stade, D., McNally, K., Morrison, C., Ohashi, K., Lehmann, L., Dalal, A., Bates, D., & Dykes, P. (2014). Clinical workflow observations to identify opportunities for nurse, physicians and patients to share a patient-centered plan of care. In *Proceedings of the AMIA annual symposium* (pp. 414–423).

Collins, S., Hurley, A., Chang, F., Illa, A., Benoit, A., Laperle, S., & Dykes, P. (2014). Content and functional specifications for a standards-based multidisciplinary rounding tool to maintain continuity across acute and critical care. *Journal of the American Medical Informatics Association: JAMIA, 21*(3), 438–447.

Collins, S. A., Gesner, E., Morgan, S., Mar, P., Maviglia, S., Colburn, D., et al. (2015). A practical approach to governance and optimization of structured data elements. *Studies in Health Technology and Informatics, 216*, 7–11.

Collins, S. A., Rozenblum, R., Leung, W. Y., Morrison, C. R. C., Stade, D. L., McNally, K., et al. (2017). Acute care patient portals: A qualitative study of stakeholder perspectives on current practices. *Journal of the American Medical Informatics Association: JAMIA, 24*(e1), e9–e17.

Collins, S., Bavuso, K., Swenson, M., Suchecki, C., Mar, P., & Rocha, R. (2017). Evolution of an

implementation-ready interprofessional pain assessment reference model. In *AMIA annual symposium proceedings* (p. 10).

Collins, S., Klinkenberg-Ramirez, S., Tsivkin, K., Mar, P. L., Iskhakova, D., Nandigam, H., Samal, L., & Rocha, R. A. (2017). Next generation terminology infrastructure to support interprofessional care planning. *Journal of Biomedical Informatics, 75*, 22–34.

Collins, S., Couture, B., Kang, M. J., Dykes, P., Schnock, K., Knaplund, C., Chang, F., & Cato, K. (2018). Quantifying and visualizing nursing flowsheet documentation burden in acute and critical care. In *Proceedings of the AMIA annual symposium*.

Collins, S., Dykes, P., Bates, D. W., Couture, B., Rozenblum, R., Prey, J., et al. (2018). An informatics research agenda to support patient and family empowerment and engagement in care and recovery during and after hospitalization. *Journal of the American Medical Informatics Association: JAMIA, 25*(2), 206–209.

Committee on Data Standards for Patient Safety. (2004). *Patient safety: Achieving a new standard for care. A report of the Institute of Medicine*. Washington, DC: The National Academies Press.

Committee on Quality of Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century. A report of the Institute of Medicine*. Washington, DC: National Academy Press.

Committee on the Robert Wood Johnson Foundation Initiative on the Future of Nursing at the Institute of Medicine. (2010). *The future of nursing: Leading change, advancing health*. Washington, DC: Institute of Medicine of the National Academies, The National Academies Press.

Cronenwett, L., Sherwood, G., & Gelmon, S. B. (2009). Improving quality and safety education: The QSEN Learning Collaborative. *Nursing Outlook, 57*(6), 304–312.

Del Fiol, G., Huser, V., Strasberg, H. R., Maviglia, S. M., Curtis, C., & Cimino, J. J. (2012). Implementations of the HL7 Context-Aware Knowledge Retrieval ("Infobutton") Standard: Challenges, strengths, limitations, and uptake. *Journal of Biomedical Informatics, 45*, 726–735.

Doing-Harris, K. M., & Zeng-Treitler, Q. (2011). Computer-assisted update of a consumer health vocabulary through mining of social network data. *Journal of Medical Internet Research, 13*(2), e37.

Doolan, D. F., Bates, D. W., & James, B. C. (2003). The use of computers for clinical care: A case series of advanced U.S. Sites. *Journal of the American Medical Informatics Association: JAMIA, 10*(1), 94–107.

Duckworth, M., Leung, E., Fuller, T., Espares, J., Couture, B., Chang, F., Businger, A. C., Collins, S., Dalal, A., Fladger, A., Schnipper, J. L., Schnock, K. O., Bates, D. W., & Dykes, P. C. (2017). Nurse, patient, and care partner perceptions of a personalized safety plan screensaver. *Journal of Gerontology Nursing, 43*(4), 15–22.

Dykes, P. C., Carroll, D. L., Hurley, A. C., Benoit, A., & Middleton, B. (2009). Why do patients in acute care hospitals fall? Can falls be prevented? *The Journal of Nursing Administration, 39*(6), 299–304.

Dykes, P. C., Carroll, D. L., Hurley, A., Lipsitz, S., Benoit, A., Chang, F., et al. (2010). Fall prevention in acute care hospitals: A randomized trial. *Journal of the American Medical Association, 304*(17), 1912–1918.

Dykes, P. C., Samal, L., Donahue, M., Greenberg, J. O., Hurley, A. C., Hasan, O., O'Malley, T. A., Venkatesh, A. I., Volk, L. A., & Bates, D. W. (2014). A patient-centered longitudinal care plan: Vison versus reality. *Journal of the American Medical Informatics Association: JAMIA, 21*(6), 1082–1090.

Dykes, P. C., Duckworth, M., Cunningham, S., Dubois, S., Driscoll, M., Feliciano, Z., Ferrazzi, M., Fevrin, F. E., Lyons, S., Lindros, M. E., Monahan, A., Paley, M. M., Jean-Pierre, S., & Scanlan, M. (2017). Pilot testing fall TIPS (tailoring interventions for patient safety): A patient-centered fall prevention toolkit. *The Joint Commission Journal on Quality and Patient Safety, 43*(8), 403–413.

Fiks, A. G., Alessandrini, E. A., Forrest, C. B., Khan, S., Localio, A. R., Gerber, A., et al. (2011). Electronic medical record use in pediatric primary care. *Journal of the American Medical Informatics Association: JAMIA, 18*(1), 38–44.

Friedman, C. P. (1995). Where's the science in medical informatics? *Journal of the American Medical Informatics Association: JAMIA, 2*(1), 65–67.

Goodwin, J. O., & Edwards, B. S. (1975). Developing a computer program to assist the nursing process: Phase I – from systems analysis to an expandable program. *Nursing Research, 24*, 299–305.

Grant, R. W., Wald, J. S., Poon, E. G., Schnipper, J. L., Gandhi, T. K., Volk, L. A., et al. (2006). Design and implementation of a web-based patient portal linked to an ambulatory care electronic health record: Patient gateway for diabetes collaborative care. *Diabetes Technology & Therapeutics, 8*(5), 576–586.

Greene, J. A., Choudhry, N. K., Kilabuk, E., & Shrank, W. H. (2011). Online social networking by patients with diabetes: A qualitative evaluation of communication with Facebook. *Journal of General Internal Medicine, 26*(3), 287–292.

Harrison, M. I., Koppel, R., & Bar-Lev, S. (2007). Unintended consequences of information technologies in health care–an interactive sociotechnical analysis. *Journal of the American Medical Informatics Association: JAMIA, 14*(5), 542–549.

Hassol, A., Walker, J. M., Kidder, D., Rokita, K., Young, D., Pierdon, S., et al. (2004). Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. *Journal of the American Medical Informatics Association: JAMIA, 11*(6), 505–513.

Health Information Technology for Economic and Clinical Health (HITECH) Act. (2009). Title XIII of Division A and Title IV of Division B of the American Recovery and Reinvestment Act of 2009 (ARRA).

Hirsch, J. S., Tanenbaum, J. S., Lipsky Gorman, S., Liu, C., Schmitz, E., Hashorva, D., Ervits, A., Vawdrey, D., Sturm, M., & Elhadad, N. (2014, October 28). HARVEST, a longitudinal patient record summarizer. *Journal of the American Medical Informatics Association., 22*(2), 263–274.

Hripcsak, G., & Albers, D. J. (2013). Next-generation phenotyping of electronic health records. *Journal of the American Medical Informatics Association: JAMIA, 20*(1), 117–121.

Hripcsak, G., Vawdrey, D., Fred, M., & Bostwick, S. (2011). Use of electronic clinical documentation: Time spent and team interactions. *Journal of the American Medical Informatics Association: JAMIA, 18*(1), 112–117.

Hruby, G. W., Matsoukas, K., Cimino, J. J., & Weng, C. (2016). Facilitating biomedical researchers' interrogation of electronic health record data: Ideas from outside of biomedical informatics. *Journal of Biomedical Informatics, 60*, 376–384.

Hurley, A. C., Dykes, P. C., Carroll, D. L., Dykes, J. S., & Middleton, B. (2009). Fall TIPS: Validation of icons to communicate fall risk status and tailored interventions to prevent patient falls. *Studies in Health Technology and Informatics, 146*, 455–459.

Hyun, S., Shapiro, J. S., Melton, G., Schlegel, C., Stetson, P. D., Johnson, S. B., & Bakken, S. (2009). Iterative evaluation of the health level-logical observation identifiers names and codes clinical document ontology for representing clinical document names: A case report. *Journal of the American Medical Informatics Association: JAMIA, 16*(3), 395–399.

International Health Terminology Standards Development Organization. (2011). *SNOMED CT* (*Systematized Nomenclature of Medicine-Clinical Terms*). Retrieved August 31, 2011, from http://www.ihtsdo.org/

IOM (Institute of Medicine). (2010). *Clinical data as the basic staple of health learning: Creating and protecting a public good: Workshop summary*. Washington, DC: The National Academies Press.

IOM (Institute of Medicine). (2011). *Digital infrastructure for the learning health system: The foundation for continuous improvement in health and health care: Workshop series summary*. Washington, DC: The National Academies Press.

IOM (Institute of Medicine). (2013). *Digital data improvement priorities for continuous learning in health and health care: Workshop summary*. Washington, DC: The National Academies Press.

Ivers, M. T., Timson, G. F., von Blankensee, H., Whitfield, G., Keltz, P. D., & Pfeil, C. N. (1983). Large scale implementation of compatible hospital computer systems within the Veterans Administration. In *Proceedings of the sixth annual symposium on Computer Applications in Medical Care* (pp. 53–56).

Jaspers, M. W., Smeulers, M., Vermeulen, H., & Peute, L. W. (2011). Effects of clinical decision-support systems on practitioner performance and patient outcomes: A synthesis of high-quality systematic review findings. *Journal of the American Medical Informatics Association: JAMIA, 18*(3), 327–334.

Johnson, S. B., Bakken, S., Dine, D., Hyun, S., Mendonca, E., Morrison, F., et al. (2008). An electronic health record based on structured narrative. *Journal of the American Medical Informatics Association: JAMIA, 15*(1), 54–64.

Kaplan, B., & Harris-Salamone, K. D. (2009). Health IT success and failure: Recommendations from literature and an AMIA workshop. *Journal of the American Medical Informatics Association: JAMIA, 16*, 291–299.

Kaufman, B. G., Spivack, B. S., Stearns, S. C., Song, P. H., & O'Brien, E. C. (2019). Impact of accountable care organizations on utilization, care, and outcomes: A systematic review. *Medical Care Research Review, 76*(3), 255–290.

Kim, H., Dykes, P. C., Thomas, D., Winfield, L. A., & Rocha, R. A. (2011). A closer look at nursing documentation on paper forms: Preparation for computerizing a nursing documentation system. *Computational Biology in Medicine, 41*(4), 182–189.

Klann, J. G., Collins, S., Cato, K., Waitman, L. R., & Westra, B. (2018, March 12–15). Nursing documentation and the clinical research informatics pipeline. In *AMIA 2018 informatics summits*. San Francisco, CA.

Kohn, L. T., Corrigan, J., & Donaldson, M. S. (2000). *To err is human: Building a safer health system*. Washington, DC: The National Academy Press.

Koppel, R., Metlay, J. P., Cohen, A., Abaluck, B., Localio, A. R., Kimmel, S. E., & Strom, B. L. (2005). Role of computerized physician order entry systems in facilitating medication errors. *JAMA : The Journal of the American Medical Association, 293*(10), 1197–1203.

Krist, A. H., & Woolf, S. H. (2011). A vision for patient-centered health information systems. *JAMA: The Journal of the American Medical Association, 305*(3), 300–301.

Krist, A., Woolf, S., & Rothemich, S. (2010). *Improving care through AHRQ health IT tools*. Retrieved July 10, 2011, from http://www.ahrq.gov/about/annualconf10/krist_rosenthal/krist.HTM

Kuperman, G. J., Gardner, R. M., & Pryor, T. A. (1991). *HELP: A dynamic hospital information system*. New York: Springer.

Lee, J., Hulse, N. C., Wood, G. M., Oniki, T. A., & Huff, S. M. (2016). Profiling Fast Healthcare Interoperability Resources (FHIR) of family health history based on the clinical elements models. In *AMIA annual symposium proceedings* (pp. 753–762).

Leiner, F., & Haux, R. (1996). Systematic planning of clinical documentation. *Methods of Information in Medicine, 35*, 25–34.

Leu, M. G., Cheung, M., Webster, T. R., Curry, L., Bradley, E. H., Fifield, J., et al. (2008). Centers speak up: The clinical context for health information technology in the ambulatory care setting. *Journal of General Internal Medicine, 23*(4), 372–378.

Leung, W. Y., Adelman, J., Bates, D. W., Businger, A., Dykes, J. S., Ergai, A., Hurley, A., Katsulis, Z., Khorasani, S., Scanlan, M., Schenkel, L., Rai, A., & Dykes, P. C. (2017, February 22). Validating fall prevention icons to support patient-centered education. *Journal of Patient Safety*. [epub ahead of print].

Lindberg, D. A. B. (1965). *Collection, evaluation, and transmission of hospital laboratory data*. Presented at the seventh IBM medical symposium, Poughkeepsie, NY.

Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health record. *Journal of the American Medical Informatics Association: JAMIA, 23*, 899–908.

Massaro, T. A. (1993). Introducing physician order entry at a major academic medical center. *Academic Medicine, 68*, 20–30.

Matney, S. A., Warren, J. J., Evans, J. L., Kim, T. Y., Coenen, A., & Auld, V. A. (2016). Development of the nursing problem list subset SNOMED CT®. *Journal of Biomedical Informatics, 45*, 683–688.

Matney, S. A., Dolin, G., Buhl, L., & Sheide, A. (2016). Communicating nursing care using the health level seven consolidated clinical document architecture release 2 care plan. *Computers Informatics Nursing, 34*(3), 128–136.

McDonald, C. J. (1976). Protocol-based computer reminders, the quality of care and the nonperfectibility of man. *The New England Journal of Medicine, 295*(24), 1351–1355.

McDonald, K. M., Schultz, E., Chapman, T., et al. (2012). *Prospects for care coordination measurement using electronic data sources (Prepared by Stanford University under subcontract to Battelle on Contract No. 290-04-0020—AHRQ SQI-II)*. Rockville: Agency for Healthcare Research and Quality.

McManus, J., & Wood-Harper, T. (2007). Understanding the sources of information systems project failure. *Management Services, 51*, 38–43.

Mlaver, E., Schnipper, J. L., Boxer, R. B., Breuer, D. J., Gershanik, E. F., Dykes, P. C., Massaro, A. F., Benneyan, J., Bates, D. W., & Lehmann, L. S. (2017). User-centered collaborative design and development of an inpatient safety dashboard. *The Joint Commission Journal on Quality and Patient Safety, 43*(12), 676–685.

Mogi, Y., Abedin, T., Ahmed, S., et al. (2017). Social networking sites for peer-driven health communication: Diabetes-related communities in Google+. *Diabetology International, 8*, 323–327.

Munasinghe, R. L., Arsene, C., Abraham, T. K., Zidan, M., & Siddique, M. (2011). Improving the utilization of admission order sets in a computerized physician order entry system by integrating modular disease specific order subsets into a general medicine admission order set. *Journal of the American Medical Informatics Association: JAMIA, 18*(3), 322–326.

Oniki, T. A., Zhuo, N., Beebe, C. E., Liu, H., Coyle, J. F., Parker, C. G., Solbrig, H. R., Marchant, K., Kaggal, V. C., Chute, C. G., & Huff, S. M. (2016). Clinical element models in the SHARPn consortium. *Journal of the American Medical Informatics Association: JAMIA, 23*(2), 248–256.

Ornstein, C. (2003, January 22). Hospital heeds doctors, suspends use of software. *Los Angeles Times*.

Ozbolt, J. (2000). Terminology standards for nursing: Collaboration at the summit. *Journal of the American Medical Informatics Association: JAMIA, 7*, 517–522.

Ozbolt, J. G., Schultz, S., Swain, M. A. P., Abraham, I. L., & Stein, K. F. (1984). Developing expert systems for nursing practice. In G. S. Cohen (Ed.), *Proceedings of the eighth annual symposium on Computer Applications in Medical Care* (pp. 654–657). Silver Spring: IEEE Computer Society.

Ozbolt, J. G., et al. (1985). A proposed expert system for nursing practice. *Journal of Medical Systems, 9*, 57–68.

Ozbolt, J. Sands, D. Z., Dykes, P. Pratt, W., Rosas, A. G., & Walker, J. M. (2012). Building Better Consumer eHealth. A report commissioned by the Office of the National Coordinator for Health Information Technology, under contract to Westat, Inc. Available at http://www.Healthit.gov/sites/default/files/final_report_building_better_consumer_ehealth.pdf.

Partridge, S. R., Gallagher, P., Freeman, B., & Gallagher, R. (2018). Facebook groups for the management of chronic diseases. *Journal of Medical Internet Research, 20*(1), e21.

Patel, V., & Johnson, C. (2018, April). Individuals' use of online medical records and technology for health needs. ONC Data Brief, No 40. https://www.healthit.gov/sites/default/files/page/2018-03/HINTS-2017-Consumer-Data-Brief-3.21.18.pdf

Patient Centered Medical Home. (2011). Retrieved July 17, 2011, from http://www.ncqa.org/

Patient Centered Medical Homes Resource Center. (2011). Retrieved July 17, 2011, from http://www.pcmh.ahrq.gov/portal/server.pt/community/pcmh_home/1483

Payne, T. H., Corley, S., Cullen, T. A., Ghandi, T. K., Harrington, L., Kuperman, G. J., Mattison, J. E., McCallie, D. P., McDonald, C. J., Tang, P. C., Tierney, W. M., Weaver, C., Weir, C. R., & Zaroukian, M. H. (2015). Report of the AMIA EHR-2020 Task Force on the status and future

direction of EHRs. *Journal of the American Medical Informatics Association, 22*, 1102–1110.

Payton, F., Pare, G., LeRouge, C., & Madhu, R. (2011). Health care IT: Process, people, patients and interdisciplinary considerations. *Journal of the Association for Information Systems, 12*(2/3), 1–13.

Pearce, C., Arnold, M., Phillips, C., Trumble, S., & Dwan, K. (2011). The patient and the computer in the primary care consultation. *Journal of the American Medical Informatics Association: JAMIA, 18*(2), 138–142.

Peute, L. W., Aarts, J., Bakker, P. J., & Jaspers, M. W. (2009). Anatomy of a failure: A sociotechnical evaluation of a laboratory physician order entry system implementation. *International Journal of Medical Informatics, 79*(4), e58–e70.

Poon, E. G., Keohane, C. A., Yoon, C. S., Ditmore, M., Bane, A., Levtzion-Korach, O., et al. (2010). Effect of bar-code technology on the safety of medication administration. *The New England Journal of Medicine, 362*(18), 1698–1707.

Prey, J. E., Polubriaginof, F., Kuperman, G. J., Tiase, V., Collins, S. A., & Vawdrey, D. K. (2016). International perspectives on sharing clinical data with patients. *International Journal of Medical Informatics, 86*, 135–141.

Rahimi, B., Vimarlund, V., & Timpka, T. (2009). Health information system implementation: A qualitative meta-analysis. *Journal of Medical Systems, 33*(5), 359–368.

Rajamani, S., Chen, E. S., Akre, M. E., Wang, Y., & Melton, G. B. (2015). Assessing the adequacy of the HL7/LOINC Document Ontology Role axis. *Journal of the American Medical Informatics Association: JAMIA, 22*(3), 615–620.

Ralston, J. D., Carrell, D., Reid, R., Anderson, M., et al. (2007). Patient web services integrated with a shared medical record: Patient use and satisfaction. *Journal of the American Medical Informatics Association: JAMIA, 14*(6), 798.

Rausch, T., & Jackson, J. (2007). *Using clinical workflows to improve medical device/system development*. Paper presented at the joint workshop on high-confidence medical devices, software, and systems and medical device plug-and-play interoperability.

Rosencrance, L. (2006, November 13). Problems abound for Kaiser e-health records management system. *Computerworld, 40*(46), 1–3.

Rotman, B. L., Sullivan, A. N., McDonald, T. W., Brown, B. W., DeSmedt, P., Goodnature, D., Higgins, M. C., Suermondt, H. J., Young, C., & Owens, D. K. (1996). A randomized controlled trial of a computer-based physician workstation in an outpatient setting: Implementation barriers to outcome evaluation. *Journal of the American Medical Informatics Association: JAMIA, 3*(5), 340–348.

Rucker, D. (2018, June 19). Achieving the interoperability promise of 21st century cures. *Health Affairs Blog*. Available at: https://www.healthaffairs.org/do/10.1377/hblog20180618.138568/full/

Saleem, J. J., Russ, A. L., Justice, C. F., Hagg, H., Ebright, P. R., Woodbridge, P. A., & Doebbeling, B. N. (2009a). Exploring the persistence of paper with the electronic health record. *International Journal of Medical Informatics, 78*(9), 618–628.

Saleem, J., Russ, A., Sanderson, P., Johnson, T., Zhang, J., & Sittig, D. (2009b). Current challenges and opportunities for better integration of human factors research with development of clinical information systems. *Yearbook of Medical Informatics, 2009*, 48–58.

Sandoval, M. B., Palumbo, M. V., & Hart, V. (2016). Electronic health record's effects on the outpatient office visit and clinical education. *Journal of Innovative Health Informatics, 23*(4), 765–771.

Schnall, R., Cimino, J. J., Currie, L. M., & Bakken, S. (2011). Information needs of case managers caring for persons living with HIV. *Journal of the American Medical Informatics Association: JAMIA, 18*(3), 305–308.

Schnall, R., Gordon, P., Camhi, E., & Bakken, S. (2011). Perceptions of factors influencing use of an electronic record for case management of persons living with HIV. *AIDS Care, 23*(3), 357–365.

Schnipper, J. L., Linder, J. A., Palchuk, M. B., Einbinder, J. S., Li, Q., Postilnik, A., & Middleton, B. (2008). "Smart forms" in an electronic medical record: Documentation-based clinical decision support to improve disease management. *Journal of the American Medical Informatics Association: JAMIA, 15*(4), 513–523.

Senathirajah, Y., Bakken, S., & Kaufman, D. (2014). The clinician in the driver's seat: Part 1 – A drag/drop user-composable electronic health record platform. *Journal of Biomedical Informatics, 52*, 165–176.

Senathirajah, Y., Kaufman, D., & Bakken, S. (2014). The clinician in the driver's seat: Part 2 – Intelligent uses of space in a drag/drop user-composable electronic health record. *Journal of Biomedical Informatics, 52*, 177–188.

Siegler, E. L. (2010). The evolving medical record. *Annals of Internal Medicine, 153*(10), 671–677.

Sittig, D. F., Salimi, M., Aiyagari, R., Banas, C., Clay, B., Gibson, K. A., et al. (2018). Adherence to recommended electronic health record safety practices across eight health care organizations. *Journal of the American Medical Informatics Association: JAMIA, 25*(7), 913–918.

Southon, F. C., Sauer, C., & Grant, C. N. (1997). Information technology in complex health services: Organizational impediments to successful technology transfer and diffusion. *Journal of the American Medical Informatics Association: JAMIA, 4*(2), 112–124.

Stead, W. W., & Lin, H. S. (Eds.). (2009). *Computational technology for effective health care: Intermediate steps and strategic directions National Research Council of the National Academies*. Washington, DC: The National Academies Press.

Stead, W. W., Borden, R., Bourne, J., Giuse, D., Giuse, N., Harris, T. R., Miller, R. A., & Olsen, A. J. (1996). The Vanderbilt University fast track to IAIMS: From planning to implementation. *Journal of the American Medical Informatics Association: JAMIA, 3*, 308–317.

Tcheng, J. E., Bakken, S., Bates, D. W., Bonner, H., III, Ghandi, T. K., Josephs, M., Kawamoto, K., Lomoton, E. A., MacKay, E., Middleton, B., Teich, J. M., Weingarten, S., & Hamilton-Lopez, M. (Eds.). (2017). *Optimizing strategies for clinical decision Support: Summary of meeting series*. Washington, DC: National Academies Press.

Vawdrey, D., Wilcox, L., Collins, S., Bakken, S., Feiner, S., & Boyer, A. (2011). A tablet computer application for patients to participate in their hospital care. In *Proceedings of the 2011 AMIA annual fall symposium* (pp. 1428–1435).

Wald, J. S., Grant, R. W., Schnipper, J. L., Gandhi, T. K., Poon, E. G., Businger, A. C., et al. (2009). Survey analysis of patient experience using a practice-linked PHR for type 2 diabetes mellitus. In *Proceedings of the 2009 AMIA annual symposium* (pp. 678–682).

Weed, L. L. (1975). *Your health care and how you can manage it*. Burlington: PROMIS Laboratory, University of Vermont.

Yen, P. Y., Kelley, M., Lopetegui, M., Rosado, A. L., Migliore, E. M., Chipps, E. M., & Buck, J. (2016). Understanding and visualizing multitasking and task switching activities: A time motion study to capture nursing workflow. In *AMIA annual symposium proceedings, 2016* (pp. 1264–1273).

Yen, P., Phillips, A., Kennedy, M., & Collins, S. (2017). Nursing Informatics Competency Assessment for the Nurse Leader (NICA-NL): Instrument refinement, validation, and psychometric analysis. *Journal of Nursing Administration, 47*(5), 271–277.

Zhou, L., Soran, C. S., Jenter, C. A., Volk, L. A., Orav, E. J., Bates, D. W., et al. (2009). The relationship between electronic health record use and quality of care over time. *Journal of the American Medical Informatics Association: JAMIA, 16*(4), 457–464.

Zhou L, Collins S, Morgan S, Zafar N, Gesner E, Fehrenbach M, Rocha R. (2016, November 12–16). A decade of experience in creating and maintaining data elements for structured clinical documentation in EHRs. In *Proceeding of the American Medical Informatics Association annual fall symposium* (pp. 1293–1302). Chicago, IL. PMID: 28269927.

Zielstorff, R. D., Hudgings, C. L., Grobe, S. J., & The National Commission on Nursing Implementation Project Task Force on Nursing Information Systems. (1993). *Next-generation nursing information systems: Essential characteristics for professional practice*. Washington, DC: American Nurses Publishing, American Nurses Foundation/American Nurses Association.

Population and Public Health Informatics

Martin LaVenture, David A. Ross, Catherine Staes, and William A. Yasnoff

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_18

Learning Objectives

After reading this chapter you should know the answers to these questions:

- What are the three core functions of public health, and how do they influence informatics requirements to achieve public health goals?
- What is the difference and relationship between public health and **population health**?
- What are the differences between **public health informatics** and other informatics specialty areas?
- What are the ways populations can be defined and how does that impact public health informatics?
- What are the categories and specific characteristics of informatics systems that are typically deployed in public health?
- What are the variations in the types of public health information systems needed at a local, regional, state, or national level?
- What factors influence the use of immunization information systems (IIS) and how can this model apply to other areas of the health system?
- What are some of the characteristics and factors that allow a public health information system to work in other countries but not in the U.S.?

18.1 Chapter Overview

The science and practice of Biomedical Informatics supports public health in its efforts to *promote* the health of populations, *prevent* disease and unhealthy exposures and behaviors, and *protect* populations exposed to human-caused or natural disasters. To optimize **population health**, one must address factors beyond the genetic and biologic make-up of individuals, such as the environment, behaviors, socio-economic status, occupation, access to care, and other influencers of health status. Although much of the variation in health status can be attributed to the zip code of one's residence, (Dwyer-Lindgren et al. 2017) behaviors (e.g., smoking and physical activity) are root determinants for the most common causes of death (Schroeder 2007). Public health measures leading to improved access to safe water and sanitation, nutrition, immunizations, and preventive care (particularly for pregnant women and children) are responsible for 25 of the 30 years gained in life expectancy in the US during the 20th Century (Bunker et al. 1994). Thus, effective improvement of the health status of populations requires the effective application of informatics strategies beyond the clinical care setting.

In this chapter, we first briefly describe public health science, the differences between "population health" and "public health," and explain key differences between clinical and public health practice that influence the needs and requirements for informatics-related interventions. Next, we define "**population** and **public health**" **informatics** as the systematic application of informatics methods and tools to support public health goals and outcomes, regardless of the setting. Finally, we describe specific example systems and applications that illustrate key challenges and opportunities.

18.2 What Is Public Health?

Public health is a complex discipline focused on promoting and protecting the health of people and communities where they live, learn, work and play (**American Public Health Association** (**APHA**) 2018[1]). Public health practice is guided by social justice and the needs of all persons within a population, not simply those accessing healthcare delivery systems. While medical care focuses on the detection, treatment, and management of injury and disease, public health practice and research involves a broad array of disciplines and diverse activities with an overarching emphasis on primary prevention, intervening at the earliest possible place in the

1 American Public Health Association, What is public health; Retrieval 10/12/2018: ► https://www.apha.org/what-is-public-health

causal chain leading to disease or disability. Prevention activities span improved access to safe food, clean water, air, and sanitation, vaccines, safe roadways and workplaces, and so forth – all in an effort to improve the health of communities, however defined. Public health achievements have been associated with major gains in life expectancy (CDC 1999[2]), and investment in disease prevention can yield significant cost savings and a healthier and less costly life (Trust for America's Health 2020[3]). Despite these achievements, global public health is challenged by the increasing mobility of populations and ongoing threats to security and safe environments which can result in regional outbreaks becoming pandemics (e.g. COVID-19).

It is useful to conceptualize public health in terms of three core functions: assessment, policy development, and assurance (Institute of Medicine (IOM) 1988). *Assessment* involves monitoring and tracking the health status of populations including identifying and controlling disease outbreaks. By relating health status to a variety of demographic, geographic, environmental, and other factors, it is possible to develop and test hypotheses about the etiology, transmission, and risk factors that contribute to health problems in a population and to develop and implement control strategies that contribute to improvements in population health.

Policy development, the second core function of public health, uses the results of assessment activities and etiologic research in concert with local resources, values and culture (as reflected via citizen input) to recommend public policies and interventions that improve health status. For example, the relationship between fatalities in automobile crashes and ejection of passengers from vehicles led to recommendations, and eventually laws, mandating seat belt use which contributed to a subsequent decrease in morbidity and mortality from automobile crashes.

Advances in information technology and widespread use of the internet, including social media sites and on-line discussion forums, as well as use of mobile apps, provide new opportunities for public health policy development. Given that public health is primarily a governmental activity, it depends upon and is informed by the consent of those governed. Policy development in public health is (or should be) based on science, but it is also guided by the values, beliefs, and opinions of each society it serves. Public health officials who wish to promote certain healthy behaviors, or to promulgate regulations (e.g., concerning fluoridated water, e-cigarettes, bicycle helmets, social distancing and so forth) would do well to tap into the online marketplace of ideas—both to understand the opinions and beliefs of their citizenry, and to inform and influence citizens to engage in those healthy behaviors.

Assurance, the third core function of public health, refers to the duty of public health agencies to assure their constituents that services necessary to achieve agreed upon goals are available. The services in question (including medical care) may be provided directly by the public health agency or by encouraging or requiring (through regulation) other public or private entities to deliver the services. For example, in some communities, local public health agencies provide direct clinical care to underserved or at-risk populations. The health department in Multnomah County, Oregon follows this model and offers health care services in multiple primary care clinics, schools, community sites and in people's homes. In other communities (e.g., Tacoma-Pierce County, Washington), local public health agencies have sought to minimize or eliminate direct clinical care services, instead working with and relying on community partners to provide such care. While there is great variation across jurisdictions, the fundamental function is unchanged: to assure that all

2 Centers for Disease Prevention and Control. (1999). Ten great public health achievements – United States, 1900–1999. *MMWR*; 48(12);241–243. Retrieval 10/02/2018: ► https://www.cdc.gov/mmwr/preview/mmwrhtml/00056796.htm

3 Trust for America's Health. (2020). Prevention for a healthier America: investments in disease prevention yield significant savings, stronger communities. Retrieval 1/25/2020: ► http://www.tfah.org/wp-content/uploads/2020/04/TFAH2020PublicHealthFunding.pdf

members of the community have adequate access to needed services, especially preventive care services and testing and diagnostic services in the context of an outbreak such as COVID-19.

The assurance function is frequently associated with clinical care, but also refers to assurance of the conditions that allow people to be healthy and free from avoidable threats to health—which include access to clean water, a safe food supply, responsive and effective public safety entities, and so forth.

This "core functions" framework is useful for describing the fundamental, overarching responsibilities of public health. The three core functions are operationalized through a set of ten essential public health services (Table 18.1) (Department of Health and Human Services (DHHS) 1994[4]). Although there is great variation in capacity to implement the ten services, they represent types of activities that public health agencies use to achieve their mission to assure conditions in which people can be healthy.

Whether one views public health through the lens of the three core functions or the ten essential services, managing and using information is a fundamental activity for public health effectiveness. For example, assessment, and several of the essential public health services, rely heavily on **public health surveillance**, the ongoing collection, analysis, interpretation, and dissemination of data to guide public health actions. The data may concern health conditions (e.g., breast cancer, communicable diseases, obesity), threats to health (e.g., smoking prevalence, drug overdose), healthcare delivery and quality (e.g., immunization rates or reports of health system quality monitoring), healthcare capacity (e.g., availability of immunization or medications, emergency or intensive care services, or other critical needs for delivering required care for a population), or other events (e.g., births) to guide public health action.

Table 18.1 Ten essential services of public health (DHHS 1994)

1. Monitor the health status of individuals in the community to identify community health problems
2. Diagnose and investigate community health problems and community health hazards
3. Inform, educate, and empower the community with respect to health issues
4. Mobilize community partnerships in identifying and solving community health problems
5. Develop policies and plans that support individual and community efforts to improve health
6. Enforce laws and rules that protect the public health and ensure safety in accordance with those laws and rules
7. Link individuals who have a need for community and personal health services to appropriate community and private providers
8. Ensure a competent workforce for the provision of essential public health services
9. Research new insights and innovate solutions to community health problems
10. Evaluate the effectiveness, accessibility, and quality of personal and population-based health services in a community

Source. Department of Health and Human Services. (1994). Essential public health functions. *Public Health in America*. Retrieval 08/12/2018: ► http://www.health.gov/phfunctions/public.htm

Public health surveillance data are often used to define priorities for public health actions, either to guide a public health response or policy development. Surveillance data may serve short-term needs (e.g., to respond to an acute infectious disease outbreak or pandemic such as COVID-19) or longer-term needs (e.g., to determine leading causes of premature death, injury, or disability), and are increasingly more available for querying and visualization through state and federal public health web sites (e.g., data.gov). Surveillance data are used by epidemiologists and researchers and can impact public understanding of health threats. For example, data

4 Department of Health and Human Services. (1994). Essential public health functions. *Public Health in America*. Retrieval 08/12/2018: ► http://www.health.gov/phfunctions/public.htm

used to manage the COVID-19 pandemic or data used to visualize the increasing prevalence of obesity in the U.S. over time both contributed to the tremendous energy and public focus brought to bear on these problems. Similarly, mortality data has been critical for understanding the evolving drug overdose epidemic in the U.S. (Seth et al. 2018). As is often the case though, no single data system provides all information required to appropriately tailor the public health response, particularly at a local level. For example, in addition to mortality data, more timely and comprehensive nonfatal and fatal overdose data are needed; therefore, other systems (such as **biosurveillance**, **syndromic surveillance** systems or an unintentional drug overdose reporting system) can be used to identify overdoses and emerging threats in local communities or improve collection of toxicology data to identify specific drugs involved (Seth et al. 2018).

While the core functions of public health described in the IOM framework have not changed for many years, rapid advances in technology and sources of data are changing the practice of public health. For example, there are (a) new data sources and methods to assess and understand the prevalence of disease in communities, the impact of public health response actions (e.g. contact tracing or stay at home orders associated with COVID-19), and the health status and determinants of disease in populations, (b) improved analytical and visualization software (e.g., **geographic information systems** (**GIS**)), and (c) improved ability to integrate and/or share health data across systems (Overhage et al. 2008). Informatics is therefore a foundational science for public health practice.

18.2.1 Public Health Versus Population Health

The phrase "population health" is increasingly used by researchers, practitioners, and policymakers in health care, public health, and other fields (Stoto 2013). For the purpose of conceptualizing population versus public health informatics, a helpful working definition for population health was proposed by Kharrazi et al. (2017):

> *Population health comprises organized activities for assessing and improving the health and well-being of a defined population. Population health is practiced by both private and public organizations. The target "population" can be a specific geographic community or region, or it may represent some other "denominator," such as enrollees of a health plan, persons residing in a provider's catchment area, or an aggregation of individuals with special needs. The difference between population health and public health is subtle, and there is not always a full consensus on these definitions. That said, public health services are typically provided by government agencies and include the "core" public health functions of health assessment, assurance, and policy setting. In the United States, most actions of public health agencies represent population health, but a considerable proportion, if not the majority, of population health services are provided by private organizations.* (Kharrazi et al. 2017).

Public health is typically focused on populations defined by a specific geographic community or region (Fig. 18.1). It may also leverage healthcare systems to implement strategies that meet public health goals such as the reporting and management of infectious diseases or administration of vaccine.

While "population health" means different things to different groups, it is always based on the underlying assumption that multiple common factors impact the health and well-being of specific populations, and that focused interventions early in the causal chain of disease will prevent morbidity and mortality and may also save resources.

In the context of health care reform, a population perspective has led to increased efforts to incorporate social and other determinants of health into medical care practice. This may include documenting this information in the **electronic health record** (**EHR**) in order to improve clinical decision making, and to better understand the health status of the community served. This has also led healthcare

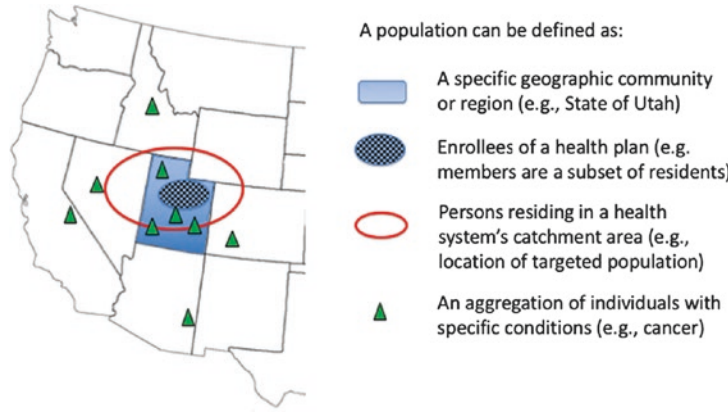


Fig. 18.1 The definition of a "population" varies according to need. (Courtesy of Catherine Staes, PhD, MPH, RN)

organizations to implement innovative strategies to improve health and reduce costs, such as providing housing, air conditioners, and/or transportation.

In the context of public health, a population perspective has always existed. However, now there are new opportunities to leverage data sources not traditionally used by public health (e.g., social media) and to use the information in EHRs to meet public health goals. For example, an EHR-based **clinical decision support system** (**CDSS**) and quality monitoring can be used to promote and monitor public health goals (e.g., improved cancer screening and immunization coverage, or early detection of health threats such as lead in water or an individual exposed to an infectious disease) among persons accessing clinical care.

18.3 Public Health Informatics

Public health informatics was first defined as the "systematic application of information science, computer science, and technology to public health practice, research, and learning" (Friede et al. 1995; Yasnoff et al. 2000). It is distinguished by its focus on populations (versus the individual), its orientation to prevention (rather than diagnosis and treatment), and its governmental context because it nearly always involves government agencies. It is a complex domain that is the focus of another entire textbook in this series (Magnuson and Dixon 2020).

The **Centers for Disease Control and Prevention** (**CDC**) has characterized public health informatics as developing and deploying methods for achieving a public health goal faster, better, or at lower cost by leveraging computer science, information science, or technology (Savel and Foldy, 2012). Public health systems are implemented across a wide range of settings (large and small, urban and rural) with variable infrastructure and capabilities, and for a workforce with a wide range of informatics experience and skills and access to technical resources and support. Given this complex context, we define public health informatics as:

> *The systematic application of informatics methods and tools to support public health goals and outcomes, regardless of the setting.*

The differences between public health informatics and other informatics specialty areas parallel the contrast between public health and medical care itself. Public health focuses on the health of the community as opposed to that of the individual patient. In the medical care system, individuals with specific diseases or conditions are the primary concern. In public health, the information and unit of analysis often relates to the community and may, for example, include sharing of information (such as disclosure of the disease status of an individual) to prevent further

spread of illness or isolating individuals to protect others. In addition, information about environmental and other factors (e.g., air and water quality, animal health, etc.) is also part of the public health domain. Finally, the focus on prevention and assessing health status across a population, rather than responding to diagnosis and treatment of individuals, necessitates the use of standards for health information exchange and large-scale analysis of data across multiple health systems.

18.4 Examples of Public Health Informatics Challenges and Opportunities

This section provides a high-level overview of the scope and function of information systems that support public health practice to illustrate the value (and challenges) of informatics methods and tools for optimizing public health outcomes. Public health practice and **epidemiology** have always been data-intensive endeavors; however, new opportunities to apply informatics-based strategies have arisen with the advances in computer technologies, increased use of EHRs and social media, and new techniques for mining data, delivering decision support, processing natural language, and system architectures that use standards to support interoperability. When considering how to apply informatics to address a public health need, it is important to understand the current "business" of public health and the history of systems that have successfully achieved their goals.

We first present an overview of the array of heterogeneous systems and applications used in the U.S. across the approximately 3000 local and 56 state and territorial health departments, and at the federal public health level, and describe informatics opportunities and challenges. Second, we focus on a robust, nationwide public health system built on informatics principles: immunization registries. Third, we describe a global public health informatics challenge and illustrate how successful informatics solutions can be applied in the context of the community/country where they are implemented.

18.4.1 Overview of Public Health Information Systems in the U.S.

Context

The fundamental science of public health is **epidemiology**, which is "*the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems*" (Last 2001). As a consequence, public health information systems may collect, use and report data at the level of ***an event***, ***a person***, or ***a population***, and may address a broad spectrum of topics along the causal chain of factors that impact health status. Public health efforts focus early in the causal chain to affect outcomes (Stiefel and Nolan 2012). For example, public health information systems may monitor exposures and risk factors, health events (e.g., vaccinations), persons with injuries or infectious or chronic disease, and other topics relevant for (a) understanding public health threats, and (b) designing and evaluating interventions. These information systems must support prevention and control efforts targeted at individuals and populations and, typically, also allow aggregate analysis to describe populations. For example, during the COVID-19 pandemic, public health authorities monitored laboratory tests and reported positivity rates and total tests performed, persons infected and hospitalized, and indicators of healthcare system capacity to respond to medical needs in a community. In contrast, nearly all medical information systems focus exclusively on supporting the processes of care for individuals. For example, EHRs and clinical laboratory systems are optimized so clinicians can quickly identify lab results for a specific individual, whereas public health practitioners want information about the patterns of antibiotic resistance over time and across multiple clinics relevant for the population in their jurisdiction. Public health

information systems should be optimized to collect relevant data, visualize and recognize epidemiology-relevant patterns, and identify emerging threats.

The "business" of public health involves a heterogeneous range of activities performed by a broad range of settings, including private but mostly government-based organizations. In the U.S., public health is organized across different levels of government, with each level having its own unique role. In a very general sense, information at the local level is required to take action in response to individual events (thus requiring personal identifiers) and to implement policies that impact populations, while information at the state level shifts towards monitoring events and supporting public health programs, and finally, information at the federal level focuses on national-level monitoring, policy development, funding of public health activities, and regulatory compliance (thus not requiring personal identifiers). In contrast, most other countries around the world have public health systems that are more centralized, which reduces barriers to information sharing and population-level analytics.

■ Systems at different levels of government in the U.S.

Local (City/County/Tribal) Public Health Practice

In the U.S., there are approximately 3000 city and county local health departments dedicated to promoting and protecting the health of people and animals in their community, and additional tribal jurisdictions each with their own public health organizations. At the local level, the work of public health often involves direct interaction with individuals or businesses in the community. Local health departments (LHDs) collect data to serve the needs of individuals, e.g., to support client follow up in LHD clinics. It is typical for local health department staff (e.g., a public health nurse, epidemiologist, or environmental health specialist) to investigate and gather identifiable person-specific information. This information is needed to ensure that infected persons are properly treated and that exposed individuals receive prophylaxis or follow-up, or that control measures are properly implemented particularly in settings where transmission may occur (e.g., in day care, long term care or food service settings). Unlike clinical systems, the systems used by LHD staff must support contact tracing, identification of exposed individuals, and management of isolation, follow-up testing and other control measures. During outbreaks of highly contagious infections such as measles or COVID-19, these systems are critical for supporting LHD activities. In addition, LHDs need to summarize data across their jurisdiction. Each LHD operates under the legal and policy framework of their respective state or territorial health department.

Information systems at the local public health level vary tremendously, depending on the size of the jurisdiction, funding levels, and the activities the local agency is required to perform. For example, the scope of practice for a rural health department with four staff members is very different from the multi-thousand employee New York City Health Department. In view of this variation, it is not surprising that information systems also range from simple spreadsheets to complex electronic record systems. The effective application of informatics principles to functions and data is currently often limited, but the ecosystem is advancing. For example, it is becoming more common for state-based surveillance systems to be web-based and support local information needs as well. For example, the systems to manage persons with sexually transmitted infections, tuberculosis, or COVID-19 may be state-wide systems that allow access for local public health staff to perform case management and investigations.

State Public Health Practice

In the U.S., there are 56 state and territorial health departments charged to carry out the responsibility of the laws and policies of their respective jurisdictions. Each state and territory has a health officer who usually reports to the governor and is tasked with leading the health agenda. The business of public health is data intensive and the systems currently used have evolved one by one over time based on needs and funding availability.

State health department information systems are often dedicated to a particular disease or condition such as infections disease, cancer, or injury. In Minnesota, for example, there are at least 21 such information systems that maintain individual level information and exchange information with hospitals, clinics and other health settings in the community (Table 18.2). The systems presented in the table are common but not representative of all states. For example, some states have integrated communicable disease surveillance systems, while others continue to operate separate systems for HIV/AIDS or TB surveillance, depending on state laws and funding. In addition, depending on the relationship between the state and its local (city/county/tribal) public health agencies, the systems may be the same, integrated, or separate. Finally, public health agencies have other systems as well (e.g., for **electronic health information exchange** (**HIE**)), so this list is not comprehensive, but rather illustrates the diversity of systems and the kinds of data that can be made available for national aggregation.

The systems listed in Table 18.2 are categorized as monitoring, workflow management, or case/care management systems, based on their primary function and key features. Monitoring systems typically rely on clinical or laboratory data as their source of information and attempt to aggregate data across health systems. They are often dependent on providers to identify the records that need to be shared and on evolving health data and interoperability standards to efficiently and accurately share the information used to generate a population-level view. The events may differ (e.g., a birth, a 'case' of measles or COVID-19, a birth defect, any blood lead level result, a health claim, etc.) but the event detection, information summarization, and reporting processes are more similar than different. In contrast, the workflow management systems are internally-focused on scheduling and collecting information, but may require interactions with external systems to process requests or reports. While such systems are common in many industries, the health and personal data managed in public health impose additional security and privacy requirements. Finally, the case and care management systems are similar in functionality to outpatient EHRs where persons (either effected by a condition or exposed to a health threat) must receive diagnostic testing, preventive services, treatment, and management to ensure ongoing monitoring and/or treatment to prevent further spread in the community.

Before planning to develop a new public health information system, one should first determine which of these three fundamental system categories is required as a prerequisite to deciding whether modifying an existing system or building/buying an entirely new one can most effectively address the need. Establishing the system type also enables the development team to more effectively seek out and learn from the experience of those who have previously built or managed similar systems.

The systems are often designed with special features for population-level analysis and context, with multiple variables indexed, to support sophisticated statistical and **Geographic Information System** (**GIS**) support capabilities. The system may be optimized for retrieval from very large (multi-million) record databases, and to quickly cross-tabulate data to study seasonal and secular trends and look for patterns by person, place, and time.

National Public Health Practice

There are numerous federal-level public health agencies that directly and indirectly support and fund public health activities at the local and state level, provide direct clinical services (e.g., through the Indian Health Service or other federal agencies), provide specific services in response to critical events (e.g., in response to bioterrorism threats or events that occur offshore), perform regulatory oversight, or aggregate information from across the U.S. to provide nationwide information and guidance for policy development. The agencies supporting the public health mission range from the CDC, Food and Drug Administration (FDA), Environmental Protection Agency (EPA), Department of Agriculture, Consumer Product Safety Commission, and Occupational Safety and Health Administration, among others. These

Table 18.2 Sample of information systems used by the Minnesota State Department of Health, classified by primary function and key features

Sample of systems	Sample key features
Monitoring systems for surveillance of knowledge, attitudes, or health events that may impact health	
Vital Statistics System tracking births and deaths[a]	Data collection may occur (a) in response to a triggered event[a], or (b) using a pre-planned sampling strategy[b] After a triggering event occurs, a set of data is transferred to a public health agency or an organization acting on the behalf of public health Triggered data: Often originates in clinical systems May be individual events or summarized reports May be reported at intervals ranging from near real-time to hourly, daily, monthly, etc. Often, systems must be able to associate the data with other potentially-related events.
Immunization Information System[a]	
Cancer Surveillance System[a]	
Birth Defects Information System[a]	
Sudden Unexplained Infant Death/Sudden Death in Youth Program[a]	
Traumatic Brain and Spinal Cord Injury System[a]	
Communicable disease surveillance systems[a]: Infectious Disease Surveillance Sexually Transmitted Infections (STD/SDI) Surveillance AIDS/HIV Surveillance	
Drug Overdose and Substance Abuse Surveillance System[a]	
Blood Lead Information System[a]	
All-payer claims database[a]	
Injury surveillance[a]	
Antibiotic Resistance and Stewardship[a]	
Animal health surveillance[a]	
Syndromic surveillance[a]	
Breast/Cervical Cancer Screening quality monitoring[b]	
Environmental monitoring of air & water[b]	
Surveys[b]: Behavioral Risk Factor surveillance system (BRFSS) National Health and Nutrition examination survey (NHANES)	
Process Control systems to manage workflow	
Public Health Laboratory Information System	Internal systems with 'industry-standard' functionality May interact with external systems to receive requests and report out information
Women, Infants and Children (WIC)	
Food-service inspection system	
Vector control operations management system	
Vaccine distribution system	
Medical Cannabis Information System	

Table 18.2 (continued)

Sample of systems	Sample key features
Clinical care and case management systems	
Public Health Clinics for Targeted Services: Sexually transmitted disease clinic Immunization clinic Public Health Contact Management Systems Sexually transmitted disease contacts Tuberculous contacts COVID-19 contacts	Person-based electronic health records to manage care of persons with specific needs Manage screening and follow-up of identified populations to ensure appropriate care, either directly provided or referred
Children with Special Health Needs System	
Tuberculosis Control System	
Newborn Hearing Screening/Early Hearing Detection and Intervention	
Refugee/Immigrant Health Information System	

[a]Data collection occurs in response to a triggered event
[b]Data collection occurs using a pre-planned sampling strategy

agencies may or may not have a regulatory role, but they illustrate the diversity and complexity of public health at the federal level. They all require some level of cross-state aggregation of information and typically do not gather person-specific data.

In the U.S., public health data is processed via distributed information systems with minimal aggregation at the federal level. In fact, it is only reportable conditions such as infectious diseases, births, and deaths that are uniformly and relatively completely reported on a national basis by the CDC. The U.S. lacks a unifying public health information infrastructure. Thus, the U.S. relies on state and local health departments to develop and use independent information systems to support their public health needs. This presents significant challenges, as described in the next section. In contrast, France, Great Britain, Denmark, Norway and Sweden have comprehensive systems in selected areas, such as occupational injuries, infectious diseases, and cancer. No country, however, has complete reporting for every problem, but many are often able to deliver timely answers to important public health questions by having information on the entire population.

Informatics Challenges and Opportunities

U.S. Public Health Information Infrastructure

The inadequate national infrastructure to detect and respond to public health threats and support the management of local, state and national prevention and control programs are amplified by the COVID-19 pandemic with its devastating morbidity and financial impact. The current infrastructure in the U.S. has a patchwork of diverse information systems due, in part, to siloed and short-term funding and the U.S. Constitution which is silent on the responsibility for protecting the public's health, leaving state law to govern public health actions. Federal leadership can provide overarching structure and resources, consolidate information, and present the rationale for a unified approach. However, each state functions independently to deal with the multiplicity of response, management and recovery decisions presented by health threats.

The grand challenge for public health is to create and maintain a unified public health information infrastructure based on informatics principles that supports multi-jurisdictional (local, state, federal) and cross-jurisdictional needs, and seamlessly interacts

with other relevant information systems such as those containing clinical, laboratory or environmental data (see ► Chap. 15).

High level informatics characteristics and considerations for such a public health information infrastructure include the ability to:

- Support both routine and emergent public health functions using the same infrastructure, avoiding the development of systems that compete for resources in the context of an outbreak.
- Rapidly scale existing surveillance systems to make local, statewide, and national data actionable in a timely manner to meet demands at each level in a crisis.
- Provide data and information in sufficient granularity to meet the needs of local, state and national public health needs.
- Rapidly add new information system capability (e.g., contact tracing for COVID-19 control) in the context of an outbreak.
- Routinely onboard, validate, and use new electronic data sources
- Ensure high quality, timely, complete and accurate data.
- Incubate innovations in technology and foster workforce capacity
- Operate with leadership and governance that is independent from jurisdictional boundaries and with authority to define requirements and engage public and private partners.
- Educate leaders and the public regarding the purpose of the infrastructure and the essential role of science to inform public policy.

Emerging Innovations and Opportunities

Advances in informatics methods and tools, CDSS and health interoperability standards, and the increasing availability of clinical and other novel sources of data (e.g., individual genomic sequences, patient-generated health data from wearable devices), are all impacting the way business is done at public health agencies, leading to new opportunities. Public health data collection systems are a potential gold mine for applying novel analytic or visualization tools, and advances in health data standards are improving opportunities to exchange data and knowledge, allowing for 2-way communication between public health and clinical systems. Closing the information loop with providers by enabling data aggregated by public health to be used for clinical decision making is a long-time goal that is becoming more achievable over time.

The development of sophisticated knowledge management and decision support methods represents a growing opportunity to more effectively use public health resources and EHR data. For example, there are an increasing number of public health guidelines that, if structured and encoded as standardized digital algorithms, can be widely distributed to provide CDS at the point of care through EHRs (e.g., laboratory testing criteria for coronavirus). Similarly, there are advances in harmonizing and standardizing case definitions, reporting logic, and patient summaries that are relevant for identifying conditions of interest to be reported from an EHR to a public health agency (i.e., to support electronic case reporting or death certificate records reporting). These knowledge resources are a necessary component for automating event detection and the generation of electronic case reports but are not sufficient. Ongoing efforts are focused on improving data and exchange standards, testing and evaluating various implementation strategies, and increasing collaboration between the multiple stakeholders involved (e.g., the vendor, clinical, and public health communities). Eventually, successful implementation requires 2-way communication with providers, which is particularly relevant for case reporting where it is important for providers to be promptly alerted about unusual disease or risk factors in their local area, such as disease clusters, environmental hazards, and antibiotic resistance patterns.

18.4.2 Immunization Information Systems: A Public Health Informatics Exemplar

Immunization Information Systems (IIS), also known as immunization registries, are confidential, computerized, population-based systems that collect and consolidate vaccination

data from vaccine providers and offer tools for designing and sustaining effective immunization strategies at the provider and program levels (CDC 2013). In the U.S., an IIS is operating in almost every state and can share data with other IIS using national standards. IIS are a critical resource for rapidly identifying and reporting gaps in immunization coverage, such as the precipitous decline in child vaccinations observed as the COVID-19 pandemic emerged in spring 2020 and well-child visits transitioned to telemedicine video conferences. Administration of immunizations require an in-person visit. Using IIS data from Michigan, public health authorities were able to report that "vaccination coverage declined in all milestone age cohorts, except for birth-dose hepatitis B coverage, which is typically administered in the hospital setting. Among children aged 5 months, up-to-date status for all recommended vaccines declined from approximately two thirds of children during 2016–2019 (66.6%, 67.4%, 67.3%, 67.9%, respectively) to fewer than half (49.7%) in May 2020" (Bramer et al. 2020).

The success of IIS in the U.S. is a compelling story that illustrates the principles of public health informatics and describes the multi-decade efforts that were needed to implement large-scale population-level IIS in each state and link them into a reliable national network. The story highlights how informatics capabilities in leadership, collaboration, and collective problem solving have been critical to overall success. In addition to their orientation to prevention, IIS can function properly only through continuing interaction with the health care system; in fact, they were designed to optimize connections for use in the clinical setting. Although IIS are among the largest and most complex public health information systems, the successful implementations and extensive interoperability in 49 states show conclusively that it has been possible over time to overcome the challenging informatics problems they present.

The major functions of IIS include (a) the ability to accept immunization records electronically using a variety of file formats and messaging standards, with online secure access to patient immunization records 24/7, (b) providing vaccine forecasting/decision support based on a patient's consolidated and **de-duplicated** immunization history, (c) supporting vaccine inventory management and vaccine ordering, producing official immunization records for school and other institutional enrollment, (d) generating immunization coverage reports for an individual provider, clinical practice or jurisdiction, and (e) supporting the national Vaccine Adverse Event Reporting System (VAERS).

History, Context and Success of IIS

Childhood immunizations are among the most successful public health interventions, resulting in the near elimination of nine vaccine preventable diseases that historically extracted a major human toll in terms of both morbidity and mortality (IOM 2000). The need for IIS stems from the challenge of assuring complete immunization protection for the approximately 10,388 children born each day in the U.S. in the context of three complicating factors: the scattering of immunization records among multiple providers given that a complete history is essential to providing an accurate forecast of vaccines needed at a visit; new vaccines, vaccine combinations. and antigen formulations regularly made available leading to a schedule that is increasingly complex as the number of vaccines has increased to over 25 doses recommended by age 6; and the conundrum that the very success of mass immunization has reduced the incidence of disease, lulling parents and providers into a sense of complacency where potential vaccine side effects often get more attention than the diseases.

The IIS history from the 1990s is remarkable for a number of key success factors including the committed leadership and shared vision. It is a story of a long slow discovery of challenges and the successful collaboration of many stakeholders to overcome those challenges over several decades. Understanding the lessons learned is foundational to successful implementation of similar large-scale public health information systems in the future.

During 1989–91, a large-scale measles epidemic occurred in the U.S. that highlighted the dangers of inadequate vaccination coverage among preschool aged children and resulted in approximately 165 deaths and 60,000 cases of disease. In response, a scientific community was established in the early 1990s to focus on testing, building and sharing knowledge to advance the impact of IIS. The complex and challenging issues confronted included policy, organizational concerns, funding, and a rapidly changing technical landscape. The collaborative nature of the scientific learning community led to open sharing of successes and failures which led to consistent IIS advancement.

Leading the way towards IIS, the Robert Wood Johnson Foundation (RWJF) was a key funder of the pioneering All Kids Count (AKC) program established in the 1990s. Under this program, an IIS learning community was created and led by the Taskforce for Global Health based in Atlanta, Georgia. The AKC program provided critical national leadership early in the IIS movement by distributing and managing competitive grants to over a dozen cities, counties and states that funded IIS implementations. The AKC, CDC and others were also key disseminators of the rapidly growing body of IIS knowledge, including the sponsorship of multiple meetings of the IIS learning community. As private funding ended, the **American Immunization Registry Association (AIRA)** was founded and became a key national focal point for the IIS community for coordination, cooperation, collaboration and policy development and engagement with the stakeholders, vendors and others in the private sector and expanded the community of interest nationwide. These efforts were funded by CDC as well as states and local jurisdictions.

An example of successful collaboration came from the need for timely exchange of a high volume of immunization information accurately and consistently with health providers. This led to the first public health HL7 version 2 messaging standards and guides for immunization information (see ▶ Chap. 7), beginning in 1995.

As knowledge was gained from early implementation successes and failures, the IIS community also developed Essential Infrastructure Functional Standards, codifying years of experience in refining system requirements (Table 18.3), (CDC 2018[5]). Version 4.0 of the standards identifies eight critical functions needed for IIS implementation that are supported by 44 detailed standard requirements, which has led to greater IIS functional uniformity nationwide.

Table 18.3 Essential IIS Infrastructure Functional Standards v4.0 2017 (CDC 2018)

1.0 The IIS contains complete and timely demographic and immunization data for children, adolescents, and adults residing or immunized within its jurisdiction.
2.0 The IIS identifies, prevents, and resolves duplicated and fragmented patient records using an automated process.
3.0 The IIS identifies, prevents, and resolves duplicate vaccination events using an automated process.
4.0 The IIS implements written and approved confidentiality policies that protect the privacy of individuals whose data are contained in the system.
5.0 The IIS implements comprehensive account management policies consistent with industry security standards.
6.0 The IIS is physically and digitally secured in accordance with industry standards for protected health information, security, encryption, uptime, and disaster recovery.
7.0 The IIS supports IIS users who access and use the IIS functions and submit or access IIS data.
8.0 The IIS exchanges data with health information systems in accordance with current interoperability standards endorsed by CDC for message content, format, and transport.

Source. Centers for Disease Control and Prevention. (2018). IIS functional standards, v4.0 Atlanta, GA. Retrieval 08/31/2018: ▶ https://www.cdc.gov/vaccines/programs/iis/functional-standards/func-stds-v4-0.html

5 Centers for Disease Control and Prevention. (2018). IIS functional standards, v4.0 Atlanta, GA. Retrieval 08/31/2018: ▶ https://www.cdc.gov/vaccines/programs/iis/functional-standards/func-stds-v4-0.html

Key Informatics Issues in Immunization Information Systems

The implementation, upgrading and management of IIS present challenging informatics issues in at least six areas: (1) Stakeholder collaboration and interdisciplinary communications; (2) Legislative and policy issues including privacy; (3) Funding, sustainability and governance; (4) Data quality and monitoring; (5) System design and interoperability; and (6) Limited prior experience with similar types of systems. While the specific manifestations of these issues are unique to IIS, these six areas represent the typical domains that must be addressed and overcome in most public health informatics projects. Over nearly three decades, many factors have contributed to the success of an individual IIS and also to the network of state IISs operating as a national system. Table 18.4 shows key informatics factors that are important contributors to IIS success.

Table 18.4 Key Informatics factors contributing to the success of IIS Nationwide

Establish a shared vision and goals and ensure effective leadership that includes:
Understanding the complexity of establishing a population-based information system and that it operates within a multidimensional health information ecosystem.
Planning for change and reassess how to accomplish goals.
Developing an effective sustainability/business plan from the beginning.
Developing a robust communications plan and comprehensive communications strategy.
Focusing on immunization information as the primary asset; recognize that technology is a means to that end.
Implement with workforce, stakeholders, and end users in mind and include:
Supporting a learning community to create and share knowledge and address common problems collaboratively.
Involving stakeholders, particularly end users, from the beginning.
Designating key roles for informaticians.
Training all staff in informatics principles.
Create well-designed and effectively used information systems and include:
Defining the system requirements to support users' needs.
Developing and implementing according to standards (and create standards if none are available).
Using the registry information as early in the implementation process as possible (even if it's not perfect).
Establishing and using metrics to evaluate progress and quantify impact.

Stakeholder Collaboration and Interdisciplinary Communications

The organizational and collaborative issues involved in operating and upgrading an IIS are challenging because of the large number and wide variety of users, most of whom are outside the IIS organization. Each of the user groups has distinct needs, including clinicians (to ensure age-appropriate vaccination), clinic managers (for vaccine management and ordering), schools (to ensure student adherence to state school immunization laws), health plans (to measure immunization coverage among beneficiaries and perhaps by provider), local health departments (to assess immunization coverage in their jurisdiction and identify children who have fallen behind and require outreach), and CDC (for accountability for federally-funded vaccines and IIS funding). Ensuring these diverse needs are understood, balanced and effectively met can be daunting on the typically slim governmental budgets on which IIS have been developed and must operate.

Interdisciplinary communication is a key challenge in any biomedical informatics project—it is certainly not specific to public health informatics. To be useful, a public health information system must accurately represent and enable the complex concepts and processes that underlie the specific business functions required. Information systems represent a highly abstract and complex set of data, processes, and interactions. This complexity needs to be discussed, specified, and

understood in detail by a variety of personnel with little or no expertise in the terminology and concepts of information technology. Therefore, successful IIS implementation and enhancements require clear communication among public health specialists, immunization specialists, providers, IT specialists, and related disciplines, an effort complicated by gaps in a shared vocabulary and differences in the usage of common terms from the various domains.

To deal with the communications challenges, particularly between IT and public health specialists, it is essential to identify and engage a public health informatician who has familiarity with both information technology and public health, can perform system analyses, and understands business processes. The primary role of this informatician is to be able to effectively translate concepts between domains that use vocabularies that are unfamiliar to others (e.g., between IT specialists and clinicians). Also, the informatician should have a deep understanding of the information processing context of both the current and proposed systems. It is also important for individuals from all the user communities related to the project to have representation in the decision-making processes. A clear understanding and set of working definitions for common terms and terminology is essential for effective communication.

Legislative and Policy Issues Including Privacy

Legislative and policy issues are important aspects of the informatics challenges of IIS. Federal laws including the **Health Insurance Portability and Accountability Act (HIPAA)** are important but State laws typically govern who has access to IIS data for what purposes, so system design must accommodate multiple levels of role-based access to functionality. A major issue is whether patient/parent consent is required before submission of immunization data to IIS, and, if so, how that consent is communicated and managed in the system. IIS must also be able to record that the patient has declined to receive vaccines for religious or other reasons as defined in state law and use that information to suppress vaccine forecasting/decision support messages and reminder-recall notices. Some jurisdictions have enacted regulations requiring providers to submit immunization data to IIS. Such a regulatory approach to ensuring information completeness is less burdensome as automated electronic file submissions have largely replaced manual data entry. Negotiating and implementing policies for interstate access and data exchange with providers and other IIS is another example of a key issue that can be problematic.

Funding, Sustainability, and Governance

Funding and sustainability are continuing challenges for all IIS. Naturally, an important tool for securing funding is development of a persuasive business case that shows the anticipated costs, benefits and value for the IIS investment. A substantial body of evidence now shows benefits, effectiveness and costs of IIS (Guide to Community Preventive Services 2015). However, many of the currently operational IIS had to develop and effectively justify their value before such information was readily available.

Specific benefits associated with IIS include preventing duplicate immunizations, eliminating the necessity to review vaccination records for school and day care admission, efficiencies in provider offices from the immediate availability of complete immunization history information, patient-specific vaccine schedule recommendations, and managing records during disasters (Boom et al. 2007). The careful review of the evidence on effectiveness, costs and benefits of specific immunization IIS functions may also be helpful in prioritizing system enhancement requirements.

Governance issues are also critical to success of implementation and enhancements to IIS. All the key stakeholders need to be represented in the ongoing, open and transparent decision-making processes, guided by a mutually acceptable governance mechanism. IIS require established rules for identifying needed enhancements, prioritizing them across the often-disparate needs of diverse user groups, and effectively managing and communicating the changes as they are developed and implemented. Governance can also

be used for establishing metrics for progress, such as number of provider sites enrolled and trained, setting other priorities, and for ongoing review of confidentiality policies.

Data Quality and Monitoring

Ensuring high quality data (including completeness, accuracy and timeliness) is vital to the success of an IIS. Provider use of an IIS increases with the confidence that they can trust that an IIS query will find the patient needing an immunization, the immunization history is complete and accurately consolidated, and the forecast of the immunizations that are due is reliable. High quality data is essential to support these core functionalities and thereby provide benefit to all the stakeholders. A variety of methods are used to maintain high quality data, including quality checks at the time of data acquisition and maintaining robust data feedback programs so all stakeholders can contribute to data quality and integrity.

Monitoring levels of vaccine use also relies on high quality data. Examples include calculations of immunization rates across jurisdictions such as Medicaid, **HEDIS**, and supporting the National Immunization Survey (NIS). Quality data is also needed for outbreak support to provide rapid information on vulnerable populations and ongoing assessment of the response. In addition, local clinical community and small area analysis can help identify groups in need of community outreach for assistance. Special studies on intervention trends in immunizations and effectiveness studies are other key uses of the data that require high quality data.

Examples of challenges for data quality include unmerged records on individuals, or wrongly combined records; unreported mobility where address, and other location information is not recorded or wrongly recorded; errors in data arriving from the providers or the birth records; duplicate records and inadequately de-duplicated records on individuals and on vaccines reported from several sources; and partial reporting of vaccines administered. Each of these data quality problems needs strong mitigation and monitoring strategies.

High data quality is the cornerstone of successfully reaching immunization program-related goals. IIS Functional Standards related to data quality are woven into the CDC AIRA IIS Essential Infrastructure Functional Standards (Table 18.3) and are reflected in multiple goals in that document (CDC 2018). This underscores the importance of planning for and ensuring data quality in all aspects of access and use of IIS data and functionality.

CDC and state IIS Directors have established a detailed monitoring and measurement system that uses about 100 measures for tracking progress to annually assess adherence to desired capabilities and standards. For example, ▶ Fig. 18.2 shows the percentage point differences between NIS and IIS for

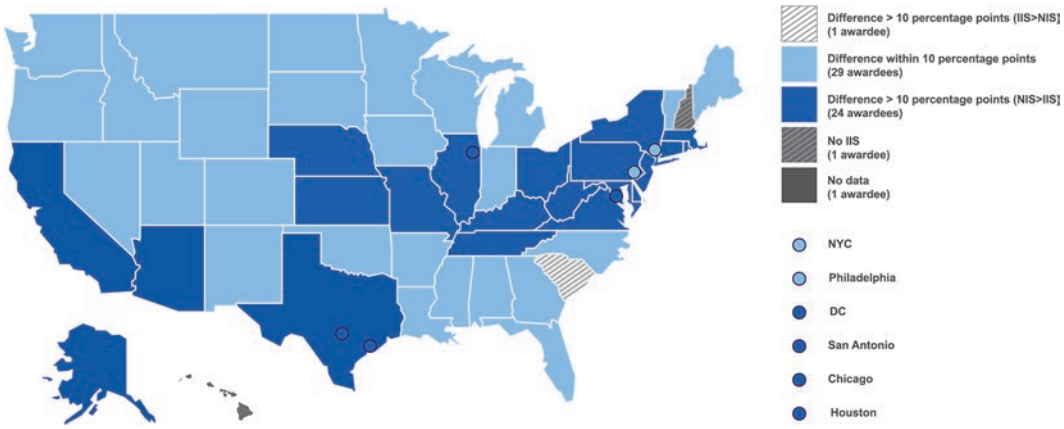


Fig. 18.2 Percentage point differences between National Immunization Survey (NIS) and IIS for combined 7-vaccine series* completion — IIS Annual Report, United States, 2017* (Source: CDC 2018)

combined 7-vaccine series completion. This type of monitoring and analysis demonstrates that more universal use of IIS information has the potential to replace existing, expensive surveys and enable more timely data to support users and community needs. In addition, AIRA has facilitated the development of numerous best practice guidelines to help ensure ongoing quality improvement, efficiency and increased standardization.

System Design and Information Architecture

System design and information architecture are important factors in the success of IIS. Difficult design issues include data acquisition methods, database organization, identification and matching of individuals, generating immunization recommendations, access to data, protocols for electronic exchange and interoperability, and reports related to clinical practice and community rates of immunization. Acquiring immunization data has been a challenging system design issue and an area of considerable IIS change as EHR use has become more common, new adolescent and adult immunizations are added to IIS, and information is submitted from a broader range of settings like pharmacies. Within the context of busy pediatric and primary care practices (where the majority of immunizations are given), the data acquisition strategy must by necessity be extremely efficient and result in minimal additional work for participating providers. Use of EHR systems can effectively support this strategy. Although most physician practices are using EHRs, only a minority have enabled bidirectional exchange with IIS.

Database design must support the desired IIS functions and allow efficient implementation of these capabilities. The design must consider operational needs for data access for an individual record and calculating individual forecasts of needed immunizations, and the requirements for population-based immunization assessment, management of vaccine inventory, and generating recall and reminder notices. One example of a particularly important database design decision for IIS is whether to represent immunization information by vaccine or by antigen. Vaccine-based representations map each available preparation, including those with multiple antigens, into its own specific data element. Antigen-based representations translate multi-component vaccines into their individual antigens prior to storage. In some cases, it may be desirable to represent the immunization information both ways. Specific consideration of required response times for specific queries must also be factored into these key design decisions.

Identification and matching of individuals within IIS is another critical issue. Because it is very common for an individual to receive immunizations from multiple providers, any system must be able to match information from multiple sources to assemble a complete unduplicated record of immunizations and retain the sources of that information. In the absence of a national unique patient identifier, most IIS assign an arbitrary number to each individual and use a matching algorithm, which utilizes multiple items of demographic information to assess the probability that two records are really from the same person and can detect duplicate reports of an immunization. Development of such algorithms and optimization of their parameters has been the subject of extensive investigation in the context of IIS, particularly with respect to de-duplication (PHII 2006[6]).

Another critical design issue is generating vaccine recommendations from an individual's prior immunization history, based on guidance from the CDC's Advisory Committee on Immunization Practices (ACIP). As more vaccines have become available, both individually and in various combinations, the immunization schedule has become increasingly complex, especially if any delays occur in receiving doses, an individual has a contraindication, or local issues require special consideration. The language used in the written guidelines can be

6 Public Health Informatics Institute. (2006). The Unique Records Portfolio. Decatur, GA: Public Health Informatics Institute. Retrieval 08/29/18: ► https://phii.org/resources/view/4380/unique-records-portfolio-guide-resolving-duplicate-records-health-information

ambiguous with respect to definitions, e.g., for ages and intervals, making implementation of CDSS problematic. Considering that the recommendations are updated relatively frequently, maintaining software that produces accurate immunization recommendations is a continuing challenge. Accordingly, the implementation, testing, and maintenance of decision support systems to produce vaccine recommendations has also been the subject of extensive study (Yasnoff 2014).

Finally, easy access to the information in IIS is essential. While independent web-based interfaces are common, the ideal is to provide a seamless query launched within the context of the provider's EHR workflow that returns IIS information and forecast to be incorporated into the EHR. Similarly, the design should support efficient access to summary reports on immunization rates for a clinic or community, reports on children who are behind schedule and support delivery of electronic reminder or recall notices to support prevention. Consumers' direct access to their own immunization record is desirable; however, there are design considerations regarding security, allowable data views and editing rights, so this is currently the subject of considerable experimentation and testing.

18.4.3 Global Health Perspective and Opportunities

Global health has been defined as "an area for study, research, and practice that places a priority on improving health and achieving equity in health for all people worldwide." (Koplan et al. 2016) Improving health globally requires accurate and timely data that can be effectively applied to education, research, innovation and service. Thus, global health work provides a vision of a future of informatics closely integrated with public health practice.

■ Current state

Today, we live in a globally interconnected world where no person is more than 36 hours away from any other person (Friedman 2007). Almost every corner of the world has access to the Internet and in most countries, even the poorest, many segments of those societies have cellular and even smart phone access to the Internet. This rapid diffusion of these information technologies has sparked increased informatics activities to promote health progress.

For example, the U.S. Global AIDS Initiative, which became known as the President's Emergency Plan for AIDS Relief, or **PEPFAR,** which began in 2003, has relied heavily on informatics infrastructure. Today, PEPFAR is working in more than 120 countries and has resulted in massive investment in health system infrastructure: laboratories, primary care services, medical supplies, drug supply chains, and information systems. These same health systems have benefitted further from funding from the Bill and Melinda Gates Foundation (BMGF), which has sparked marked progress across vaccine preventable diseases, neglected tropical diseases and many other conditions that impact the poorest throughout the world. These social entrepreneurs also bring strong accountability to global health – they expect their funds to produce a specific health outcome. Tracking measured outcomes requires informatics.

Recognizing these trends, in 2013 the U.S. government funded the global health security agenda (GSHA), which promotes a range of health development activities across the lower income tier of countries. The agenda is designed to improve disease surveillance, laboratory and diagnostic capacity, workforce capacity, and basic health services. The GSHA has set the stage for widespread recognition that informatics is an essential tool towards health progress in these countries.

It is clear from these efforts that informatics will drive health improvement in low and middle income countries. Consequently, most countries now acknowledge, as does the World Health Organization, that their health efforts must be data informed and data driven.

Example of Informatics Innovation Leading to Health Impact: Global Trachoma Mapping

Trachoma is a bacterial disease that spreads through contact with the eyes or nose of an infected individual, shared towels or clothing, and vectors such as flies. Repeated infections cause eyelashes to turn inward, scratching and scarring the cornea, leading to blindness. About 1.9 million people in 42 countries are currently blind or visually impaired as a result. Treating the infection along with access to clean water and sanitation leads to elimination throughout entire populations.

In 2009, the International Trachoma Initiative (ITI), a program of the Task Force for Global Health (TFGH) and their partners recognized the need for better information about the prevalence of trachoma to inform decisions about where to implement specific intervention strategies such as eye surgery, antibiotic treatment, facial hygiene, and environmental sanitation (ITI 2012). In 2012, in partnership with ITI, the UK's Department for International Affairs announced the three-year Global Trachoma Mapping Project (GTMP) to complete *The Global Atlas of Trachoma* to map global trachoma prevalence.[7] Field teams used mobile devices to record findings of trachoma infection. Data management software converted the household-level findings to prevalence maps at the village and district levels. An electronic data capture and management tool called LINKS was adapted for GTMP. The interface was designed to be basic enough for field staff to understand and secure enough to prevent accidental data loss or breaches.

By 2016, more than 2.6 million people had been surveyed for trachoma to create prevalence maps of 29 countries, the largest infectious disease map ever created. The availability of high-quality data on trachoma had widespread impact. The *Atlas* helped inform the scale-up of the global trachoma elimination program. With the availability of prevalence data, ITI was able to make better decisions about where to allocate antibiotic supplies and coordinate production. That same year, ITI reached scale when Pfizer donated 120 million antibiotic treatments to 32 countries, bringing the cumulative program total to 627 million treatments (Taskforce for Global Health 2018).

Emerging Opportunities and Directions

The African Union formed Africa CDC in 2017. Rather than being disease focused, it is built around the key disciplines that protect and promote the health of populations, i.e., informatics, surveillance, drug and medical supply chain, diagnostics/laboratory, and policy. As a result, a key goal is to educate a new cadre of informaticians throughout the African continent to assure that timely, accurate, and relevant data inform action across the spectrum of health needs.

Ubiquitous and reliable high-speed bandwidth also now readily enables worldwide collaborations. The notion of global to local shared learning is now a reality, with communities of practice forming that involve members from all continents. Learning is now multidirectional.

In the global health community, the lack of legacy institutions combined with very limited resources has created strong incentives for innovation. When combined with the insistence on specific, real-time quantitative results and outcomes from funding sources, the extensive use of informatics has been an inevitable consequence to ensure efficient and effective progress in health improvement.

18.5 Public and Population Health Informatics Conclusion

Public health informatics is the systematic application of informatics methods and tools to support public health goals and outcomes, regardless of the setting. Effective public health information systems and communi-

7 Taskforce for Global Health. (2018). The global trachoma mapping project: determining prevalence to help eliminate trachoma by 2020. Retrieval 10/20/18: ► https://www.taskforce.org/case-study/global-trachoma-mapping-project-determining-prevalence-help-eliminate-trachoma-2020

cation between clinical and other systems can help to assure prevention actions, timely monitoring of disease patterns, and rapid responses to epidemics, thereby saving lives and money.

Public health information and the development of **health information infrastructure** (HII) (see ▶ Chap. 15) are closely related. Public health informatics supports the population assessment, assurance and policy development roles of public health. In contrast, HIIs focus on medical care to individuals while also connecting providers and patients within a population. Ideally, these two areas work together supporting both community health assessment and individual care. In the past, public health and health care have not traditionally interacted as closely as they should. Both domains focus on the health of communities—public health does this directly, while the medical care system does it one patient at a time. However, it is now clear that medical care must also focus on the community to integrate the effective delivery of services across all care settings for all individuals (IOM 2011; Sittig and Singh 2020). An effective HII could allow many public health information needs currently met through independently operated and maintained systems to be more efficiently addressed via periodic HII queries (e.g., to assess relationships between various diseases, conditions, treatments, and possible risk factors) or through automatic real-time reporting of relevant information from the HII to public health (e.g., to support surveillance and control of COVID-19).

Successful public health and population health informatics requires an informatics-savvy organization that has a clear vision, strategy, and governance for information management and use; a workforce skilled in using information and information technologies; and well-designed and effectively used information systems (Baker et al. 2016). The information imperative is urgently driven by the increasing digitization of data coming into health departments from an increasing number of sources, the need for timely information to inform increasingly complex public health decisions, and the growing costs of operating aging public health information systems (Brand et al. 2018a). Information innovation to address growing needs requires an agency-wide informatics-savvy organizational approach (LaVenture et al. 2014c, 2014d; 2017c, 2017d) and an appreciation that there are key stages of innovation to ensure the successful integration of research in the public health practice space (Xu et al. 2011).

Public health systems frequently involve non-health organizations such as law enforcement and parks and recreation departments. Thus, public health informaticians must adopt methodologies that bridge professional and organizational divides, such as the Public Health Informatics Institute's Collaborative Requirements Methodology (PHII 2011).

Despite the focus of many current public health informatics activities on population-based extensions of the medical care system (leading to the orientation of this chapter), applications beyond this scope are possible, desirable, and many innovative strategies and applications are under development or in use. Indeed, the phenomenal contributions to health made by the hygienic movement of the nineteenth and early twentieth centuries suggest the power of large-scale environmental, legislative, and social changes to promote human health (Rosen 1993). The effective application of informatics to populations through public health is a key challenge of the twenty-first century. It is a challenge we must accept, understand, and overcome if we want to create an efficient and effective health care system as well as truly healthy communities for all.

Suggested Readings

American Immunization Registry Association (AIRA). (2018). Fundamentals of IIS. Retrieval Oct 24, 2018 http://repository.immregistries.org/resource/fundamentals-of-iis/from/major-iis-topics/best-practices-2/. A summary of fundamental understanding of

three essential IIS topics. Staff, Data Quality, Interoperability and HL7 basics..

Brand, B., LaVenture, M., Lipshutz, J., Stevens, W. F., & Baker, E. L. (2018). The information imperative for public health: A call to action to become informatics-savvy. *Journal of Public Health Management and Practice:* JPHMP, 24 (6), 586–589. A summary of reasons for the urgent need for informatics and some strategies for moving organizations forward.

Centers for Disease Control and Prevention (CDC). (2013). Progress in immunization information systems – United States, 2011. MMWR Morb Mortal Wkly Rep. *62*(3), 48–51.

Centers for Disease Prevention and Control. (1999). Ten great public health achievements – United States, 1900–1999. MMWR; 48(12);241–243.

Centers for Disease Control and Prevention. (2018). IIS functional standards, v4.0 Atlanta, GA: Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention. (2012). CDC's Vision for Public Health Surveillance in the 21st Century. MMWR 2012;61(Suppl; July 27, 2012).

Department of Health and Human Services. (1994). Essential public health functions. Public Health in America. Washington DC: Department of Health and Human Services.

Hinman, A. R., & Ross, D. A. (2010). Immunization registries can be building blocks for national health information systems. Health Affairs, 29(4), 676–682. Describes the components of IIS and how they are also key elements in the national health infrastructure.

Infectious Diseases Society of America (ISDS). Stories of Surveillance in Action. Retrieval October, 25th, 2018 https://www.surveillancerepository.org/search/success-stories and https://www.healthsurveillance.org/default.aspx Contains many practical examples on how public health surveillance is used to support clinical care and community health.

International Trachoma Initiative (ITI). (2012). Global Scientific Meeting on Trachomatous Trichiasis (TT). DeCatur, GA: The Task Force for Global Health.

Koo, D., O'Carroll, P. W., & LaVenture, M. (2001). Public health 101 for informaticians. Journal of the American Medical Informatics Association, 8(6), 585–597. An accessible document that introduces public health thinking and approach.

LaVenture, M., Brand, B., & Baker, E. L. (2014a). Developing an informatics-savvy health department: I: Vision and core strategies. *Journal of puBlic Health Management and Practice:* JPHMP, 20 (6), 667–669.

LaVenture, M., Brand, B., & Baker, E. L. (2014b). Developing an informatics-savvy health department: II: Operations and tactics. *Journal of Public Health Management and Practice:* JPHMP, 21(1), 96–99.

LaVenture, M., Brand, B., & Baker, E. L. (2017a). Developing an informatics-savvy health department: from discrete projects to a coordinating program. Part I: Assessment and governance. *Journal of Public Health Management and Practice:* JPHMP, 23(6), 325–327.

LaVenture, M., Brand, B., Ross, D. A., & Baker, E. L. (2017b). Developing an informatics-savvy health department: from discrete projects to a coordinating program. Part II: Creating a skilled workforce. *Journal of Public Health Management and Practice:* JPHMP, 23(6), 638–640.

Magnuson, J. A., & Dixon, B. E. (Eds.). (2020). Public health informatics and informatics systems. New York: Springer. A comprehensive textbook.

Overhage, M., Grannis, S., & McDonald, C. J. (2008). A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. American Journal of Public Health, 98(2), 344–350.This study documents the improved quality and timeliness of electronic lab reporting of diseases to public health.

Public Health Informatics Institute. (2006). The Unique Records Portfolio. Decatur, GA: Public Health Informatics Institute.

Taskforce for Global Health. (2018). The global trachoma mapping project: determining prevalence to help eliminate trachoma by 2020. Decatur, GA: Taskforce for Global Health.

Trust for America's Health. (2020). Impact of Chronic Underfunding on America's Public Health System: Trends, Risks, and Recommendations. Trust for America's Health.

US Department of Health and Human Services. (2018). Healthy people 2030 framework. Washington, DC: US Department of Health and Human Services. Retrieved at https://www.healthypeople.gov/2020/About-Healthy-People/Development-Healthy-People-2030/Framework (Accessed 23 Oct 2018). This is an essential guide to the national public health goals.

Yasnoff, W. A., O'Carroll, P. W., Koo, D., Linkins, R. W., & Kilbourne, E. M. (2000). Public health informatics: Improving and transforming public health in the information age. *Journal of Public Health Management and Practice,* 6(6), 67–75. A concise yet comprehensive introduction to the field.

Questions for Discussion

1. How might the trend of widespread adoption of electronic health records and interest in population health affect public health informatics?
2. Compare and contrast the types of data needed and functions required in an information system for clinical versus public health information systems. Explain it from non-technical and technical perspectives.
3. How can the successful model of immunization registries be used in other domains of public health (be specific about those domains)? How might it fail in others? Why?
4. A significant and increasing percentage of the U.S. GDP is spent on medical care. How could population and public health informatics help to use those monies more efficiently? Or lower the figure absolutely?
5. If public health informatics (PHI) involves the application of information technology in any manner that improves or promotes human health, does this necessarily involve a human "user" that interacts with the PHI application? For example, could the information technology underlying anti-lock braking systems be considered a public health informatics application? Provide other examples.
6. How might cloud computing (shared configurable computing resources including networks, servers, storage, applications, and services), and mobile technology transform public health informatics?

References

American Public Health Association (APHA) (2018). What is Public Health. Washington DC: American Public Health Association.

Baker, E. L., Brand, W., Davidson, A., LaVenture, M., Singletary, V., & Smith, P. (2016). Building the business case for public health information systems. *Journal of Public Health Management and Practice, 22*(6), 603–606.

Boom, J. A., Dragsbaek, A. C., & Nelson, C. S. (2007). The success of an immunization information system in the wake of Hurricane Katrina. *Pediatrics, 119*(6), 1213–1217.

Bramer, C. A., Kimmins, L. M., Swanson, R., et al. (2020). Decline in child vaccination coverage during the COVID-19 pandemic — Michigan care improvement registry, May 2016–May 2020. *MMWR. Morbidity and Mortality Weekly Report, 69*, 630–631.

Brand, B., LaVenture, M., & Baker, E. L. (2018a). Developing an informatics-savvy health department: From discrete projects to a coordinating program. Part III: Ensuring well-designed and effectively-used information systems. *Journal of Public Health Management and Practice, 24*(6), 181–184.

Brand, B., LaVenture, M., Lipshutz, J., Stevens, W. F., & Baker, E. L. (2018b). The information imperative for public health: A call to action to become informatics-savvy. *Journal of Public Health Management and Practice, 24*(6), 586–589.

Bunker, J. P., Frazier, H. S., & Mosteller, F. (1994). Improving health: Measuring effects of medical care. *Milbank Quarterly, 72*, 225–258.

Department of Health and Human Services. (1994). Essential Public Health Services: Public Health in America. Washington DC: Department of Health and Human Services.

Dwyer-Lindgren, L., Stubbs, R. W., Bertozzi-Villa, A., Morozoff, C., Callender, C., Finegold, S. B., et al. (2017). Variation in life expectancy and mortality by

cause among neighbourhoods in King County, WA, USA, 1990–2014: A census tract-level analysis for the Global Burden of Disease Study 2015. *The Lancet Public Health, 2*(9), e400–e410.

Friede, A., McDonald, M. C., & Blum, H. (1995). Public health informatics: How information-age technology can strengthen public health. *Annual Review of Public Health, 16*, 239–252.

Friedman, T. (2007). *The world is flat 3.0: A brief history of the twenty-first century*. New York: Picador/ Farrar, Straus and Giroux.

Guide to Community Preventive Services. (2015). Community Preventive Services Task Force Recommendation for Use of Immunization Information Systems to Increase Vaccination Rates. *Journal of Public Health Management and Practice*: May/June 2015 - Volume 21 - Issue 3 - p 249–252.

Institute of Medicine. (1988). *The future of public health*. Washington, DC: National Academy Press.

Institute of Medicine. (2000). *Calling the shots: Immunization finance polices and practices*. Washington, DC: National Academy Press.

Institute of Medicine. (2011). *Engineering a learning healthcare system: A look at the future*. Washington, DC: National Academies Press.

Kharrazi, H., Lasser, E. C., Yasnoff, W. A., et al. (2017). A proposed national research and development agenda for population health informatics: Summary recommendations from a national expert workshop. *Journal of the American Medical Informatics Association, 24*(1), 2–12.

Koplan JP, Bond TC, Merson MH, Reddy KS, Rodriguez MH, Sewankambo NK, et al. Towards a common definition of global health. *Lancet*. 2016;373:1993–5.

Last, J. M. (Ed.). (2001). *Dictionary of epidemiology* (4th ed., p. 61). New York: Oxford University Press.

LaVenture, M., Brand, B., & Baker, E. L. (2014c). Developing an informatics-savvy health department: I: Vision and core strategies. *Journal of Public Health Management and Practice, 20*(6), 667–669.

LaVenture, M., Brand, B., & Baker, E. L. (2014d). Developing an informatics-savvy health department: II: Operations and tactics. *Journal of Public Health Management and Practice, 21*(1), 96–99.

LaVenture, M., Brand, B., & Baker, E. L. (2017c). Developing an informatics-savvy health department: From discrete projects to a coordinating program. Part I: Assessment and governance. *Journal of Public Health Management and Practice, 23*(6), 325–327.

LaVenture, M., Brand, B., Ross, D. A., & Baker, E. L. (2017d). Developing an informatics-savvy health department: From discrete projects to a coordinating program. Part II: Creating a skilled workforce. *Journal of Public Health Management and Practice, 23*(6), 638–640.

Magnuson, J. A., & Dixon, B. E. (Eds.). (2020). *Public health informatics and informatics systems*. New York: Springer.

Overhage, J. M., Grannis, S., & McDonald, C. J. (2008). A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. *American Journal of Public Health, 98*(2), 344–350.

Public Health Informatics Institute. (2011). *Collaborative requirements development methodology*. Public Health Informatics Institute.

Rosen, G. (1993). *History of public health*. Baltimore: Johns Hopkins University Press.

Savel, T. G., & Foldy, S. (2012). The role of public health informatics in enhancing public health surveillance. MMWR. *Morbidity and Mortality Weekly Report, Supplement, 61*(3), 20–24.

Schroeder, S. A. (2007). We can do better – Improving the health of the American people. *New England Journal of Medicine, 357*, 1221–1228.

Seth, P., Scholl, L., Rudd, R. A., & Bacon, S. (2018). Overdose deaths involving opioids, cocaine, and psychostimulants United States, 2015–2016. *MMWR, 67*, 349–358.

Sittig, D. F., & Singh, H. (2020). COVID-19 and the need for a national health information technology infrastructure. *JAMA, 323*(23), 2373–2374. https://doi.org/10.1001/jama.2020.7239.

Stiefel, M., & Nolan, K. A. (2012). *Guide to measuring the triple aim: Population health, experience of care, and per capita cost. IHI innovation series white paper*. Cambridge: Institute for Healthcare Improvement.

Stoto, M. (2013). *Population Health in the Affordable Care Act Era*. Washington DC: Academy Health.

Xu, W., Pettey, W., Livnat, Y., Gesteland, P., Rajeev, D., Reid, J., Samore, M., Evans, R. S., Rolfs, R. T., & Staes, C. (2011). Strengthening partnerships along the informatics innovation stages and spaces: Research and practice collaboration in Utah. *Online Journal of Public Health Informatics, 3*(3).

Yasnoff, W. A., O'Carroll, P. W., Koo, D., Linkins, R. W., & Kilbourne, E. M. (2000). Public health informatics: Improving and transforming public health in the information age. *Journal of Public Health Management and Practice, 6*(6), 67–75.

Yasnoff, W. A., & Miller, P. L. (2014). Decision support and expert systems in public health. In J. A. Magnuson & P. C. Fu (Eds.), *Public health informatics and information systems* (2nd ed., pp. 449–467). New York: Springer.

mHealth and Applications

Eun Kyoung Choe, Predrag Klasnja, and Wanda Pratt

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_19

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What is mobile health (mHealth), and how has it evolved over time?
- What are the current and potential values and benefits of mHealth for (a) patients and caregivers, (b) the general public, (c) clinicians, and (d) researchers?
- What are the key features of mHealth in supporting personal health management?
- What ethical considerations and issues do mHealth technologies raise regarding health disparities?

19.1 Introduction

The field of **mobile health (mHealth)** focuses on the uses of mobile technologies, such as mobile phones and **wearables**, to support both the delivery of healthcare services as well as individuals' efforts to manage their health in everyday lives. mHealth applications are highly varied, ranging from clinical tools for remote patient monitoring and shared decision making to patient-centered tools intended to help individuals better manage their health in daily life, such as increasing physical activity or controlling one's blood glucose levels. Although applications of mHealth overlap with other areas of informatics research and practice—such as personal health informatics (▶ Chap. 11) and telehealth (▶ Chap. 20), in the context of mHealth, these applications have a distinct flavor largely due to the omnipresence of mobile technology and the unique features that this omnipresence enables. In this chapter, we trace the history of mHealth, highlight key features of these technologies that make them uniquely suited for delivery of health interventions, and provide an overview of health services and health promotion for which they are used. We conclude with a brief discussion of ethical issues raised by the rapid growth of mHealth.

19.1.1 The Omnipresence of Mobile

The rapid advancements and widespread adoption of mobile technologies have altered how healthcare providers practice medicine, how patients access healthcare and manage their health and well-being, and how researchers design and evaluate health interventions. Called mHealth, both industry and research sectors have been increasingly developing devices and applications that leverage mobile and wireless communication to deliver healthcare services or to support people managing their own health and well-being.

Mobile technologies are omnipresent. Since 1984 when the first handheld mobile phone called DynaTAC (see ◘ Fig. 19.1) was introduced (Murphy 2013), mobile computing and communication devices have been significantly advanced and widely adopted. As of 2018, 95% of U.S. adults own a mobile phone of some kind, 77% own a **smartphone**, and 53% own a **tablet** computer.[1] Mobile phone ownership is widespread worldwide: 59% of people own a smartphone and a further 31% own other types of mobile devices such as flip phones.[2] mHealth devices include **smartwatches** and other wearables, technologies well-suited for monitoring activity and physiological data and for the delivery of lifestyle change interventions. As of 2018, the

1 Mobile fact sheet. 2018. Pew Research Center. Retrieval June 3, 2019: ▶ https://www.pewinternet.org/fact-sheet/mobile/

2 Poushter J, Bishop C, Chwe H. June 2018. Social media use continues to rise in developing countries but plateaus across developed ones: digital divides remain, both within and across countries. Pew Research Center. Retrieval June 3, 2019: ▶ https://www.pewresearch.org/global/2018/06/19/social-media-use-continues-to-rise-in-developing-countries-but-plateaus-across-developed-ones/

Fig. 19.1 A DynaTAC 8000X, a first commercially available mobile phone from 1984. By Redrum0486 - ► http://en.wikipedia.org/wiki/File:DynaTAC8000X.jpg,CCBY-SA3.0,► https://commons.wikimedia.org/w/index.php?curid=6421950

global market for wearable devices in the healthcare sector reached over 2 billion U.S. dollars, which is a 600% increase over the market size in 2016.[3]

Mobile devices have permeated every aspect of people's lives. In a 2011 study, Dey and colleagues (Dey et al. 2011) found that people keep their mobile device in close proximity (within the same room or closer) almost 90% of the time. With respect to usage time, American adults spend about 5 hours/day on mobile devices.[4] The extensive use of mobile devices generates massive data from mobile sensors and usage behaviors, from which we can infer people's behavioral and psychological patterns, such as sleep, activity, mood (LiKamWa et al. 2013), and even psychological traits (Lee et al. 2014). The inference drawn from the data collected from mobile devices is meaningful and reliable as long as people keep using the device. Therefore, many mHealth technologies are designed to promote high engagement and long weartime, so that people can "stick" with the technology. However, as smartphone overuse and addiction have become societal problems (Kwon et al. 2013), supporting healthy engagement with mHealth technology warrants careful consideration and future research.

19.1.2 Evolution of mHealth Technologies

mHealth technologies have evolved in parallel with the advancements in mobile communication technologies and portable computing devices. Although we cannot provide a detailed account of the history of mHealth technologies in this chapter, we hereby aim to provide key milestones and devices in the evolution of mHealth technologies to establish a broad context for later discussion. Specifically, we cover **personal digital assistants (PDAs)**, cellular phones and smartphones, and wearables, each of which sparked important innovations in mHealth applications. See Table 19.1 for the summary of representative mHealth devices.

19.1.2.1 PDAs and Cellular Phones

Developed in 1993, the Apple Newton was marketed as the first PDA device (see Fig. 19.2a). Early applications for Newton

3 Statista. 2019. Projected size of the global market for wearable devices in the healthcare sector from 2015 to 2021 (in million U.S. dollars) [Data file]. Retrieval June 3, 2019: ► https://www.statista.com/statistics/607982/healthcare-wearable-device-revenue-worldwide-projection/

4 Khalaf S, Kesiraju L. March 2, 2017. U.S. consumers time-spent on mobile crosses 5 hours a day. Flurry Analytics Blog. Retrieval June 3, 2019: ► https://flurrymobile.tumblr.com/post/157921590345/us-consumers-time-spent-on-mobile-crosses-5

Table 19.1 Representative mHealth devices and their key technical specs and supported applications

Device Category	Example Device (Release year)	Key Technical Specs	Supported Applications
Personal Digital Assistant (PDA)	Palm Pilot Personal (1997)	Palm OS 2.0, 16 MHz CPU, 512 KB Memory, 160 × 160 pixel monochrome display, touchscreen LCD, speaker, serial hotsync port, desktop cradle	Note taking (patient progress notes), calculation, to-do list, calendar, reference materials (ePocrates, 5-Minute Clinical Consult)
Cellular Phone	Nokia 5500 Sport (2006)	Symbian OS 9.1, 235 MHz CPU, 8 MB Memory, 1.7″ (208 × 208 pixel) display, camera, microphone, speaker, GSM network, Bluetooth connectivity, accelerometer,	Sports (stopwatch, steps, calories burn) tracking, multimedia messaging, email, text to speech
Smartphone	Samsung Galaxy S10 (2019)	Android OS 9.0, Octa-core CPU, 8GB Memory, 128GB Storage, 6.1″ Quad HD+ (3040 × 1440 pixel) display, triple camera, GSM+CDMA/4G LTE network, Wi-Fi, Bluetooth connectivity, capacitive multi-touch, stereo speakers, microphone, accelerometer, barometer, gyro sensor, geomagnetic sensor, RGB light sensor, proximity sensor	Multiple health and fitness applications via Google Play Store
Tablet	Apple iPad 7th Generation (2019)	10.2″ Retina display, LED-backlit with Multi-Touch and IPS technology, 2160 × 1620 pixel display, compatible with Apple Pencil and Keyboard, Cellular data network, Wi-Fi, Bluetooth connectivity, cameras, two speaker audio, fringerprint identity sensor, gyro sensor, accelerometer, ambient light sensor, barometer	Multiple health and fitness applications via App Store
Smartwatch	Fitbit Versa 2 (2019)	1.34″ OLED display, Wi-Fi, Bluetooth connectivity, microphone, Accelerometer, optical heart rate monitor, altimeter, vibration motor, SpO2 sensor, NFC, ambient light sensor	Floors climbed, activity (e.g., walk, run, swim) tracking, sleep tracking with light, deep, and REM, personalized reminders, guided breathing, 24/7 heart rate tracking, resting heart rate, calories burn

were designed to assist everyday tasks such as note-taking, calculation, and to-do list management. The Newton was one of the first devices to feature pen-based handwriting recognition, which eventually became popular in other handheld devices. Due to the technical limitations, however, the Newton was not widely adopted and was eventually discontinued in 1998. The Pilot 1000, introduced in 1996 by Palm Computing, was the first widely accepted PDA; it offered a 160 × 160 pixels monochrome touch-sensitive display, a separate area for pen text input (using a special script called Graffiti), and sufficient memory to store reference materials. Over the next decade, the Pilot 1000 was followed by a large

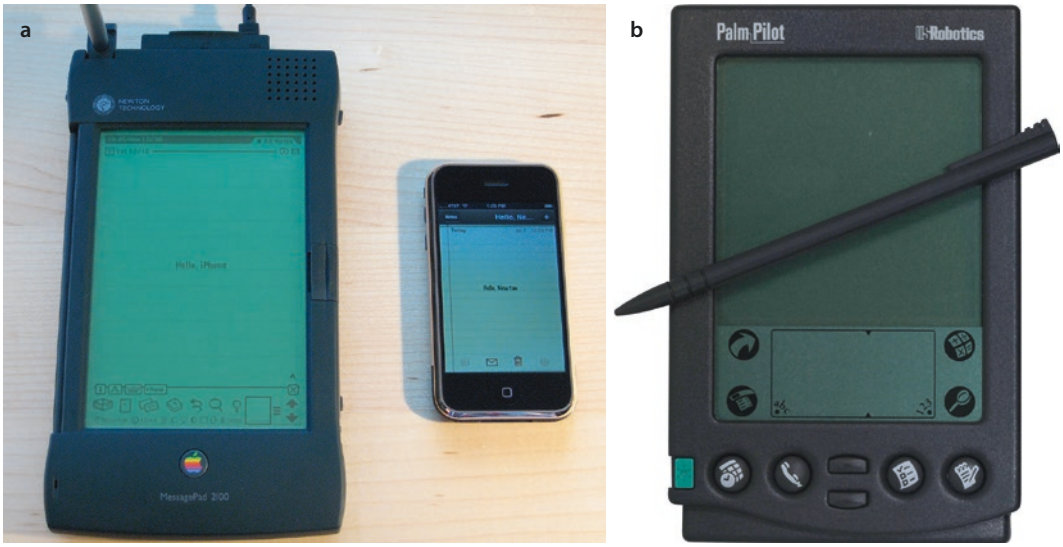


Fig. 19.2 **a** The Apple Newton MessagePad 2100, running Newton OS, alongside the original iPhone running iOS. By Blake Patterson from Alexandria, VA, USA - Newton and iPhone: ARM and ARM, CC BY 2.0, ► https://commons.wikimedia.org/w/index.php?curid=7039806. **b** Figure PalmPilot with stylus. By Rama & Musée Bolo - Own work, CC BY-SA 2.0 fr, ► https://commons.wikimedia.org/w/index.php?curid=36959631

number of more advanced Palm devices (e.g., the PalmPilot Personal, Fig. 19.2b), as well as Windows Mobile devices (e.g., the Pocket PC), which incorporated color displays, higher resolution screens, and eventually, cellular connectivity. Many medical apps and mHealth research prototypes were designed for these devices.

The new modes of data entry using pen-based gestures, paging navigation, and menu palettes were shown to be effective in capturing structured data in healthcare settings. These interface characteristics were effective in overcoming some of the limitations of paper-based data capture, such as high capture burden and low compliance (Lane et al. 2006). As a means to ease the data capture burden, PDA applications were designed for doctors to write patient progress notes (Poon et al. 1996), for patients to capture symptoms (Stratton et al. 1998), and for researchers to collect sensitive health data in resource-poor field sites (e.g., Jaspan et al. 2007).

PDAs also served as medical reference tools for medical students and doctors (e.g., with applications such as ePocrates Rx, 5-Minute Clinical Consult, MedMath, MedCalc). In a 2006 review paper on PDAs and medical education, the authors reported that 60–70% of medical students and residents used PDAs for educational purposes or patient care (e.g., patient tracking and documentation) (Kho et al. 2006).

For researchers, PDAs were a good prototyping platform to test mHealth applications. Intille, Kukla, Farzanfar, and Bakr (2003) used a standard PDA with a barcode scanner plug-in to create an application that helps people compare food items at the time of making food purchasing decisions. The application delivered tailored, motivational information with an aim to help people make "just-in-time" incremental changes to their diet (Intille et al. 2003). Siek et al. (2006) leveraged PDAs' voice recording feature, as well as a barcode scanner, to create a food monitoring tool for patients with chronic kidney disease.

Traditionally, PDAs did not support phone services. However, cellular phones, which became popular around the same time as the early PDAs, did by connecting to a wireless communications network through radio waves or satellite transmissions. Early cellular phones were equipped with voice and text communication capabilities, allowing

researchers to design health interventions that deliver health messages (e.g., a text messaging smoking cessation program) (Free et al. 2011). As with PDAs, cellular phones became more powerful over time, adding more advanced features, such as multimedia messaging (e.g., sending and receiving images), text to speech, and motion sensors. These features were promptly incorporated into health applications. For example, the Nokia 5500 Sport was the first mobile phone that had a built-in 3D accelerometer, and it came with an application that automatically detected running and walking. The phone also included a diary for planning and tracking workouts, as well as enabling users to add workouts that the phone did not detect automatically.

19.1.2.2 Smartphones and Tablets

IBM and BellSouth's Simon is largely considered the first "smartphone" because it featured a phone with a touchscreen, email and many other capabilities that became available to the public in 1994.[5] Simon did not last long, but it presaged other smartphones, such as the BlackBerry and Windows Mobile.[6] Smartphones gained popularity in 2007 when Apple announced the first generation of the iPhone, which featured a large capacitive touchscreen, multi-touch interactions (e.g., pinching for zooming), and thin slate-like form factor. In the following year, Apple opened the App Store, a marketplace to distribute applications, which became popularly known as "apps," for the iPhone (and subsequently other iOS devices). Soon after the iPhone was launched, in 2008, Google released a new mobile OS platform, Android. The choice of the OS determines which phone to choose and which apps the user can run. Google's Android runs on a wide range of devices manufactured by a variety of companies, whereas Apple's iOS runs on Apple's iPhone devices only.

Due to the broad user base and convenient marketplaces (platforms) to distribute apps, the iPhone and Android smartphones have led to a rapid growth in digital health in the past few years. As of 2017, more than 325,000 health and fitness apps were available in the major mobile app stores, with 3.6 billion downloads.[7] As platforms evolve, so do the mHealth apps. For example, Epocrates, a medical reference app, has changed over time from its earlier version for Palm OS devices to its recent versions for iOS/Android devices.

Although 73% of mHealth apps in 2015 were designed for supporting general wellness—for example, apps for tracking exercise and diet, and for managing stress, the mHealth market is shifting toward the support for managing specific health conditions. Such applications constituted 40% of mHealth apps in 2017.[8] Some of the key categories of mHealth apps include: symptom checkers, healthcare professional finders, managing clinical and financial records, health-condition education and management, self-monitoring, remote patient monitoring, rehabilitation programs, and prescription filling and compliance (or adherence). Based on a review of commercial mHealth apps, the most prevalent conditions that mHealth apps are targeting are diabetes, depression, migraine, asthma, low vision, and hearing loss (Martínez-Pérez et al. 2013). In addition, apps for women's health and pregnancy, and medication reminders are now common.

5 Aamoth, D. August 18, 2014. First Smartphone Turns 20: Fun Facts About Simon TIME. Retrieval July 7, 2019: ► https://time.com/3137005/first-smartphone-ibm-simon/

6 Pothitos, A. October 31, 2016. The History of the Smartphone Mobile Industry Review. Retrieval July 7, 2019: ► http://www.mobileindustryreview.com/2016/10/the-history-of-the-smartphone.html

7 Pohl M. 2017. 325,000 mobile health apps available in 2017 – Android now the leading mHealth platform. Research2Guidance. Retrieval June 3, 2019: ► https://research2guidance.com/325000-mobile-health-apps-available-in-2017/

8 The growing value of digital health: Evidence and impact on human health and the healthcare system. November 7, 2017. IQVIA Institute. Retrieval June 3, 2019: ► https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health

Fig. 19.3 A person wearing a smartwatch By Crew - ► https://pixabay.com/en/smartwatch-gadget-technology-smart-828786/,CC0, ► https://commons.wikimedia.org/w/index.php?curid=46644979

19.1.2.3 Wearable Devices

Wearable devices include activity trackers (e.g., wrist-worn devices, rings, chest bands, belt-type, and earpieces), smartwatches (see Fig. 19.3), and smart clothing (e.g., shirts, bras, socks, and pants). They track a variety of activities, such as walks, runs, workouts, smoking, and eating, as well as physiological processes, such as sleep, heart rate, breathing, and sun exposure. Activity trackers and smartwatches in particular have become increasingly popular (Choe et al. 2014; Fritz et al. 2014; Lupton 2014; Rooksby et al. 2014). These devices employ low-burden self-monitoring by leveraging wearable and mobile sensing to capture data; they commonly provide behavioral feedback with coaching, goal-setting, and reminders. Such devices are available from a variety of commercial companies (Apple, Fitbit, Garmin, Fossil, Polar, etc.) and come in a number of form factors, from thin bands that resemble a bracelet (e.g., Fitbit Alta) to devices that aim to look like traditional watches (e.g., Garmin Fenix, and watches from Skagen).

Although the adoption rate for activity trackers and smart watches in the US is 17.6%, they have not been taken up equally by all age groups, with older generation showing a significantly lower adoption rate (4.6% for age 65 and up) than younger generations (30.8% among 25 to 34-year-olds).[9] In addition, the abandonment of wearable device is high; about one third of people who own a wearable device abandon the device after six months.[10] Researchers have identified a variety of reasons why people abandon these devices. For instance, common reasons for abandoning wearable devices include difficulty in deciding what to do with the data and disappointment with the level of information the devices provide (Lazar et al. 2015). Given that a key way that these devices try to help people maintain and change behavior is by making the monitored activities more salient, the abandonment of wearable device means that the benefits of wearable devices might fade away soon after the abandonment (Klasnja et al. 2011). To help enhance the device weartime and sustain the benefits of self-monitoring, these devices should be better integrated in people's everyday life, fostering user engagement.

Contrary to other computing equipment, wearable devices could serve as a fashion item. In this regard, companies attend to the device's form factor, providing diverse customization options. Customizability can enhance a person's sense of identity, which in turn is associated with more favorable attitude, higher exercise intention, and a greater sense of attachment towards the track (Kang et al. 2017).

Although many personalized options are available on the hardware side, designing effective and personalized behavioral feedback on the software side warrants future research. There are growing bodies of litera-

9 US wearable user penetration, by age, 2017 (% of population in each group). December 19, 2016. eMarketer. Retrieval June 4, 2019: ► https://www.emarketer.com/Chart/US-Wearable-User-Penetration-by-Age-2017-of-population-each-group/202360

10 Endeavour Partners. April 21, 2017. Inside wearables: how the science of human behavior change offers the secret to long-term engagement. Medium. Retrieval June 13, 2019: ► https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3

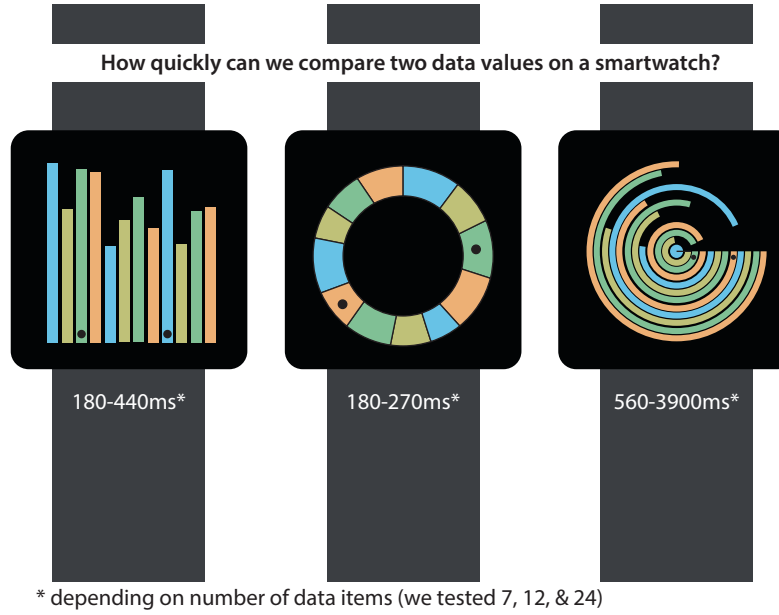


Fig. 19.4 Glanceable visualization on smartwatch. Example images of the stimuli used in the two perception studies in Blascheck, Besançon, Bezerianos, Lee, & Isenberg (2019)

ture in the ubiquitous computing (Amini et al. 2017; Gouveia et al. 2016) and visualization (Blascheck et al. 2019 (see Fig. 19.4); Brehmer et al. 2018) communities that have begun to examine ways to create effective feedback on a small screen. More cross-disciplinary collaboration on this front is needed to design and build effective intervention components for wearable devices (Lee et al. 2018).

Wearables have been popular subjects of research in healthcare domains due to their potential to serve as an intervention in a variety of contexts. See a systematic mapping study of how "Internet of Things," including wearables have been deployed and evaluated in the medical field (Sadoughi et al. 2020).

19.1.3 Current Platforms

The smartphone user base is sharply divided between Google's Android platform and Apple's iOS. As of June 2018, 54.1% percent of U.S. smartphone users were using a Google Android device, and 44.5% were using an Apple device.[11] The divide in the market share and the idiosyncratic characteristics of individual phones pose problems for developers and researchers. To create a native app (i.e., an app that has been developed for use on a particular platform), developers must redouble their efforts to create and maintain versions for each platform. Researchers, who often lack skills and resources to build native mobile apps for multiple mobile platforms, must pick either of the two platforms instead of supporting both, which can introduce a selection bias in recruitment.[12] To address such concerns, efforts have been devoted to developing frameworks and

11 Statista. 2019. Subscriber share held by smartphone operating systems in the United States from 2012 to 2019 [Data file]. Retrieval June 13, 2019: ▶ https://www.statista.com/statistics/266572/market-share-held-by-smartphone-platforms-in-the-united-states/

12 The U.S. Mobile App Report. 2014. Retrieval January 15, 2020: ▶ https://www.comscore.com/Insights/Infographics/iPhone-Users-Earn-Higher-Income-Engage-More-on-Apps-than-Android-Users?cs_edgescape_cc=US

platforms to handle device and operating system compatibility. For example, Apache Cordova[13] enables developers to build cross-platform mobile apps with HTML5, CSS, and JavaScript, with extensions that provide access to some hardware features such as camera, **GPS**, and Bluetooth. Titanium[14] is a mobile development environment that allows the creation of native apps across different mobile devices and operating systems, while reusing a large part of the codes across apps. Such cross-platform support enables developers of mHealth technologies to reach potential users of either Android or Apple devices.

19.1.4 Data Access & Data Standards

Accessing mHealth data could be valuable for many stakeholders, including self-trackers who want to learn insights about themselves (Choe et al. 2014), researchers who want to incorporate mHealth data in their research (e.g., Althoff et al. 2016; Jakicic et al. 2016; Fitabase[15]), and app developers who want to integrate multiple data sources in a new service (e.g., Exist.io,[16] Gyroscope,[17] and Instant[18]). For some patients, accessing their personal health data is a matter of life and death, as in the case of Type 1 diabetes patients, who have been struggling for a long time to have direct access to their **continuous glucose monitor (CGM)** data (Kaziunas et al. 2018). In clinical contexts, doctors can use the patients' data collected outside the hospital to diagnose patients accurately and monitor them closely (Chung et al. 2015; Kim et al. 2017; West et al. 2016; Zhu et al. 2016).

People can access their data collected from mHealth apps through downloading a file (e.g., CSV, FIT, GPX, KML, and TCX) from an app, website, or a health platform (e.g., Apple Health), and using application programming interfaces (APIs). However, it is typically difficult for lay people to download and repurpose the data, even if they are the ones who contributed to collecting them (Kim et al. 2019). Although recent regulations—such as the European General Data Protection Regulation (GDPR)[19]—have the potential to enhance personal data accessibility, the industry lags behind the reforms.

Data standards are defined as "documented agreements on representations, formats, and definitions of common data" (Fenton et al. 2013).[20] They support interoperability among heterogeneous systems, and are critical in leveraging patient-generated data in the current mHealth ecosystem. Among a few existing health platforms in the market (e.g., Apple Health, Google Fit, and Samsung Health), Apple Health and its frameworks (HealthKit and ResearchKit) are widely being adopted, and have become the standard interface to fitness and medical devices.[21] Apple's HealthKit supports data integration from iOS apps by allowing third party apps to transfer data to and from the HealthKit. However, data loss could happen during the transfer, and HealthKit is not compatible with Android apps, failing to accommodate more than half of the smartphone users. As in the case of the Open mHealth ini-

13 Apache Cordova. 2015. The Apache Software Foundation. Retrieval June 13, 2019: ► https://cordova.apache.org/

14 Titanium Mobile Development Environment. 2017. Axway. Retrieval June 13, 2019: ► https://www.appcelerator.com/Titanium/

15 Fitabase. 2019. Retrieval June 11, 2019: ► https://www.fitabase.com/

16 Exist. 2019. Hello Code. Retrieval June 11, 2019: ► https://exist.io/

17 Gyroscope. 2019. Gyroscope Innovations, Inc. Retrieval June 11, 2019: ► https://gyrosco.pe/

18 Instant. 2015. Retrieval June 11, 2019: ► https://instantapp.today

19 European Union General Data Protection Regulation. Retrieval September 9, 2019: ► https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en

20 Data Standards. 2019. Public Health Data Standards Consortium. Retrieval June 13, 2019: ► http://www.phdsc.org/standards/health-information/d_standards.asp

21 Apple is going after the healthcare industry, starting with personal health data. January 8, 2019. CB Insights. Retrieval June 13, 2019: ► https://www.cbinsights.com/research/apple-healthcare-strategy-apps/

tiative[22] and Human API,[23] some efforts have been made in creating standard for personal health data schema. However, the uptake of the Open mHealth initiative and Human API has been slow, and no incentives encourage companies to follow the standardization just yet. Health platform companies and health app developers need to make a concerted effort to establish standards for personal data schemas to support interoperability and to prevent data loss.

On the clinical informatics side, app developers have been building 3rd party clinical apps—some of which are mHealth apps—that can access patients data directly from EHR by leveraging **SMART on FHIR,**[24] an open, standards-based platform for medical apps. Through SMART on FHIR, healthcare organizations can plug third-party medical apps into their EHR that use those standard data types, which poses immense opportunities for clinicians and patients to use EHR data for diverse purposes (e.g., patient engagement, disease management, and research) on mobile devices.

19.2 Key Features of mHealth Technologies

A great deal of utility of mobile technology for health applications comes from the ways in which these tools can collect data about individuals' health status, behavior, and context. Unique to mobile technology is its ability to collect data *passively* via various types of sensors embedded in mobile phones and wearable devices. With the number of mobile sensors and the quality of the inference increasing at a rapid rate, mobile health tools are able to detect a broad range of behaviors and states continuously and in the background, with minimal user interaction, enabling collection of rich, longitudinal data that can be used for assessment and treatment. When self-report is needed, mobile tools offer ways to collect self-report data in context and with low burden, greatly increasing the ecological validity of user-provided information.

19.2.1 Sensing to Collect Data

A key feature of many mHealth applications is that they make use of information generated via sensors contained in mobile phones or worn on the body. Modern mobile phones contain a large number of sensors, including GPS, accelerometers, gyroscopes, cameras, and microphones. In recent years, the use of on-body sensors that continuously monitor individuals' activities and states has exploded. Examples include wrist-worn fitness trackers as well as a large variety of specialty sensors such as those that detect galvanic skin response, oxygen saturation, blood glucose, heart rate variability, core body temperature, and blood pressure.

The data from such on-body sensors are typically transmitted to a mobile phone, where, along with data collected on the phone itself, it is used to provide feedback to users or is made available to third-party applications. Sensor data are used for three basic purposes: (1) for assessing physiological processes, such as resting heart rate or blood glucose; (2) for inferring individual's activities, such as physical activity, eating, and sleep; and (3) for inferring context, such as location and social environment.

19.2.1.1 Assessing Physiological Processes

An important use of sensing in mHealth technologies is to assess physiological processes and states that are important for supporting health management. Such sensing is usually done through devices that users wear on their bodies (e.g., bands worn on the wrist, or instrumented adhesive patches worn on the torso) or even through their phone. Regardless of the sensor form factor, physiological data obtained through mobile sensing are typically

22 Open mHealth. 2015. The Tides Center. Retrieval June 13, 2019: ► http://www.openmhealth.org/

23 Human API. 2019. Retrieval June 13, 2019: ► https://www.humanapi.co/

24 SMART Health IT. Retrieval January 15, 2020: ► https://smarthealthit.org/

transmitted to a mobile phone via a low-power radio (e.g., Bluetooth), where it can be used to drive intervention delivery through an mHealth app or uploaded to remote servers for monitoring by the healthcare team.

The most common type of sensors used to monitor physiological processes are wrist-worn wearables. The exact set of sensors—and, thus, what they are able to detect—varies by device, but the main physiological sensor found in such devices is the optical heart-rate sensor, which uses green and orange light emitting diodes (LEDs) and a photodetector to detect the pulse waveform (Alexander et al. 1989). Recently, heart rate sensors have begun to add pulse oximetry functionality as well, by incorporating a red LED, and, as of 2018, these more advanced heart rate sensors are beginning to show up even in devices in the $150 range, such as Fitbit Charge 3 and Garmin Vivosmart 4, which represent a large segment of the wearables market.

The chief physiological metrics sensed by heart rate sensors are momentary heart rate and resting heart rate. Combining momentary heart rate data with physical activity data (sensed via an accelerometer), wrist-worn devices also try to estimate energy expenditure, although the quality of these calculations is variable (see Consolvo et al. 2014 for examples of design implications of this variability in inference). Certain wristbands—notably, recent devices made by Garmin—attempt to characterize users' stress levels, which can be estimated from the inter-beat interval data obtained from optical heart rate sensors (Hovsepian et al. 2015). Finally, wrist-based devices with pulse oximeters are able to detect blood oxygen saturation, although the current generation of devices do not provide continuous monitoring of this metric.

Although wristbands are currently the most common form factor for commercial wearable sensors, the range of physiological data that can be sensed unobtrusively at the wrist is limited. As such, several other form factors exist, including smart rings (e.g., Moodmedtric and Motiv), instrumented clothing (e.g., Hexoskin, OMsignal, and Skiin), and adhesive tags that can be attached to clothing (Spire). Such devices are able to detect other physiological metrics beyond heart rate—including respiration, body temperature, and electrodermal activity—enabling, among other things, more robust assessment of stress (e.g., Moodmetric and Spire).

In addition to consumer-oriented commercial devices, physiological sensing using wearable sensors is a thriving research area. The current crop of commercial devices described above are based on decades of sensing research, with the research on optical pulse oximetry dating back to the 1980s (see Alexander et al. 1989 for a brief review), and stress detection drawing on ten years of work on multimodal sensing (Ertin et al. 2011; Hovsepian et al. 2015). More recently, two key topics in mobile physiological sensing research have been cuff-less, continuous measurement of blood pressure and noninvasive measurement of blood glucose.

Over the last few years, there has been a quickly growing body of research aimed at developing unobtrusive methods for blood pressure (BP) measurement. Much of this work attempts to repurpose optical heart rate sensors already present in the wrist-based activity trackers to estimate blood pressure from photoplethysmography (PPG) measurements. Recent work (Zhang and Feng 2017; Patil et al. 2017) has shown that machine learning techniques can be used to estimate both systolic and diastolic BP from PPG signals with around 90% accuracy. Similarly, Carek et al. (2017) have used the accelerometer and optical heart rate sensor in a smart watch to estimate BP from the pulse transit time, with similar accuracy. Although these accuracy rates are still too low for widespread clinical applications, this line of work is quickly developing, and accurate continuous BP monitoring is likely not too distant.

The other key target for physiological sensing in the context of mHealth has been blood glucose measurement. Traditionally, blood glucose could only be measured from blood samples, typically obtained from a finger prick or, in the case of continuous glucose monitors, from a needle semi-permanently inserted into the subcutaneous tissue. In recent years, both the finger prick-based glucose monitors and CMGs have started to incorpo-

rate Bluetooth connectivity, enabling glucose readings to be automatically uploaded to and logged in a smartphone app, greatly facilitating keeping of accurate logs and provision of data-driven self-management coaching. For instance, BlueStar (► www.welldoc.com), an FDA-approved app for diabetes self-management, uses patients' glucose readings uploaded from a connected glucometer to provide just-in-time advice for specific actions patients can take to keep their glucose levels in the prescribed range. An early randomized clinical trial of BlueStar showed a 2% HbA1c improvement in the BlueStar arm compared to the .68% improvement in the control condition (Quinn et al. 2008). As with BP measurement, recent research has focused on making blood glucose measurement more convenient and less invasive. Researchers have explored multiple approaches to detecting glucose levels non-invasively, including optical, by shining light into the skin to detect changes in glucose concentration in the blood, chemical, to detect glucose levels in the saliva, and electrochemical, to detect glucose via a smart contact lens (see Eadie and Steele 2017 for a review). A number of systems that are using these approaches are currently undergoing clinical trials.

Finally, many research projects aim to use sensors found in mobile phones to enable detection of physiological processes that have traditionally required expensive, medical-grade equipment. Among other metrics, projects in this category used cell phone sensors to detect lung function via the phone microphone (Larson et al. 2012), hemoglobin concentration in the blood using the phone camera (Wang et al. 2016), intraocular pressure from the video captured on a smartphone (Mariakakis et al. 2016), and blood alcohol level from interaction patterns with app user interfaces (Mariakakis et al. 2018). The purpose of projects such as those listed above is to make expensive medical tests that require specialized equipment cheaper and more accessible to decrease health disparities and increase access to care.

19.2.1.2 Inferring Activities

A key role of sensing in the context of mobile health is detection of individuals' activities and states. The most common application of sensing in this domain has been detection of physical activity. Accelerometers and gyroscopes are found not only in wearable fitness trackers but also in smartphones, as well in a growing segment of "hybrid" watches—standard quartz or mechanical timepieces which contain a small number of sensors and ability to receive phone notifications. Accelerometers in all these devices are continuously assessing users' movements to determine how active they are. The chief metric derived from this data is the number of steps that an individual has taken, but many devices and mobile apps also combine accelerometer data with heart rate data to estimate the number of active minutes, usually calculated to correspond to minutes of moderate to vigorous physical activity (MVPA), a metric that is commonly used in physical activity guidelines (e.g., Piercy et al. 2018). Additionally, more advanced wearables are increasingly attempting to detect exactly which physical activity a user is performing (e.g., biking, swimming, elliptical, etc.) to help users keep more accurate exercise logs and improve calorie expenditure estimation. As with many other areas of mobile health, detection of specific physical activities was pioneered in the research setting (Choudhury et al. 2008), but its accuracy has only recently become sufficiently high to make it feasible for broad inclusion in commercial devices.

Another common activity detected via mobile devices is sleep. Most wrist-based activity trackers (from Fitbit, Garmin, etc.) track sleep using accelerometry, and many of them try to categorize different stages of sleep as well, providing users with summaries of the amounts of time spent in deep, light, and REM sleep. A recent systematic review of reliability and validity of consumer activity trackers found that these devices typically overestimate total sleep time and sleep efficiency when compared to polysomnography but underestimate wake-after-sleep onset (Everson et al. 2015). Given that these devices

estimate sleep duration using wrist-worn accelerometers, part of the problem appears to be that evening activities that involve little movement (reading, watching a movie, etc.) can be confused for sleep, leading to the overestimation of total sleep time.

An important focus of recent work has been automated detection of eating. Diet tracking has been consistently shown to be one of the most effective self-management strategies for weight loss (Michie et al. 2009; Webb et al. 2010), and it is of great importance to a range of epidemiological research. Yet, consistent manual diet tracking is notoriously difficult to achieve over extended periods of time. For this reason, a great deal of recent research has focused on trying to automate diet tracking in various ways. mHealth researchers have taken a number of approaches to address this problem. One approach has been to simplify tracking by enabling users to take pictures of their meals using the phone's camera. To extract information about what was eaten, the photo-based approach relies either on computer vision (e.g., Kitamura et al. 2010), or on crowdsourcing, where the food images are analyzed through a sequence of tasks assigned to workers on **Amazon Mechanical Turk** (Noronha et al. 2011). The photo-based approach still requires active user engagement, but the user burden is reduced compared to manual logging, albeit at a higher financial cost (for crowdsourcing) or at the cost of lower data accuracy (for computer vision-based approaches).

As an alternative to photo-logging of food intake, recent research has attempted to facilitate diet tracking by automating detection of eating episodes. To do so, researchers have used a number of sensing approaches, including using microphones to detect the sound of chewing and swallowing (Alshurafa et al. 2014; Makeyev et al. 2012), using wrist-based accelerometers to detect hand movements indicative of eating or drinking (Amft et al. 2005; Kyritsys et al. 2017), and using a small neck-worn camera to detect when food is brought to the mouth (Sun et al. 2014; Chen et al. 2013). Unlike photo-logging, which aims to produce the same kind of data as traditional manual diet tracking (a log of the foods eaten with corresponding micro and macronutrient information), these approaches focus on detecting the *activity* of eating, rather than detecting *what* is eaten. As such, detection of eating currently does not result in accurate nutrient information, although it can provide some information about the types of food eaten (fresh fruit and vegetables are crunchier and require more chewing than, say, a hamburger), as well as the amount (by virtue of detecting the total number of eating episodes and their duration). Even with this limitation, automatic eating detection can be very useful for a range of mobile health applications, from supporting independent living of the elderly by verifying they are eating regularly, to interventions for weight management that attempt to decrease snacking and mindless eating.

mHealth systems have used sensor data to detect a range of other activities, including medication adherence via instrumented pill bottles (Hayes et al. 2006; Abbey et al. 2012), falls (Dai et al. 2010; Fang et al. 2012) smoking (Ali et al. 2012; Parate et al. 2014), and activities of daily living, such as washing and cooking. The latter category is particularly relevant for supporting independent living, but detecting such activities with a high level of precision often involves combining mobile devices with instrumenting the environment, such as adding RFID tags to common objects like pots (Buettner et al. 2009), or using cameras (Liu et al. 2014).

19.2.1.3 Inferring Context

In addition to detecting physiological processes and user activities, mobile devices are commonly used to detect user's context to tailor intervention delivery to the user's current situation. The key contextual variable used by mHealth apps is location. Both iOS and Android contain robust location capabilities that leverage GPS, Wi-Fi, and cell tower information to determine user's location in a battery-efficient way and enable third-party applications to obtain location information even while the application is in the background (i.e., the phone screen is off or the user is interacting with another application). These capabilities enable mHealth apps to identify nearby resources (e.g., healthy restaurants or

emergency rooms), and to provide intervention content that is appropriate to the user's current context. For instance, HeartSteps, an mHealth physical activity intervention, sends user's activity suggestions that are tailored to their location, weather, time of day, and day of the week (Klasnja et al. 2018).

Similarly, MyBehavior (Rabbi et al. 2015), another physical activity intervention, tracks users' daily movements and then, based on the current location, provides recommendations for walking routes a user can take to increase their steps, balancing the user's step goal with the suggested activity's feasibility (e.g., how much time it would take). In the addiction space, location is often used to determine high-risk situations (e.g., proximity to a bar) to provide just-in-time support, such as coping strategies (e.g., see Gustafson et al. 2014). Such **just-in-time adaptive interventions (JITAIs)**, enabled by knowledge of the individual's current context and activity, hold great promise for achieving goals of precision behavioral health by providing support only when it's most likely to be effective and when individuals are receptive to it (Nahum-Shani et al. 2015; Spruijt-Metz and Nilsen 2014).

Other contextual variables that can be passively sensed by current smartphones include weather (by lookup via the user's location), the user's calendar (e.g., free/busy blocks), ambient noise levels, and proximity to other people. A number of research efforts (Kanhere 2011; Devarakonda et al. 2013) have examined incorporation of chemical sensors into the phone that would allow sensing of pollution levels. Although such pollution levels can currently be approximated by looking up pollution data based on the user's location, sensing on the phone could enable more timely and potentially more effective just-in-time support for individuals with asthma and other respiratory problems.

19.2.2 Collecting Self-Reported Data

Across physiological processes, user activities, and context, mobile devices are increasingly able to detect information about a person that can be used to characterize that person's health-related needs and provide timely support to address those needs. However, not everything can be sensed, and a range of important metrics needed to monitor patients' health (e.g., pain levels and medication side effects) and support self-management require self-report data. For these cases, mobile devices enable self-report that is precise, timely, and minimally burdensome. Collecting and leveraging in-situ self-report data using mobile devices has recently gained tremendous interest among researchers in multiple fields, such as ubiquitous computing, behavioral sciences, psychology, public health, and machine learning.

19.2.2.1 In-Situ Data Collection Methods

Research methods such as diary study and **ecological momentary assessment (EMA)** (Shiffman et al. 2008; Stone et al. 2007) help researchers understand details about a person's context, intentions, and actions that traditional research methods (e.g., interview, survey, and system log analysis) cannot reveal. Diary study and EMA share the same purpose of collecting data in people's natural environment to maximize ecological validity. In diary study, people capture self-report data—usually once a day—using either pen and paper, mobile diary apps, or online survey. Some diaries are structured (e.g., sleep diary for Cognitive Behavioral Therapy for Insomnia) whereas others are less structured (e.g., free-form note). Although diary study helps researchers collect in-situ data relatively easily, it is subject to low adherence (if participants forget to fill it out) and recall bias (if there is a large time interval between when an event happened and when the event was captured). More recently, researchers increasingly use EMA, which refers to frequent, brief collection of self-report data about the person's current situation and experiences. As a data collection method, EMA is intended to decrease recall biases and memory limitations inherent in retrospective self-report, while—by collecting self-report frequently (often 5 to 8 times a day)—enabling researchers to understand how behavior, psychological processes,

and individuals' experiences change over time and are influenced by time-varying factors such as the person's physical environment.

19.2.2.2 Mobile Research Platforms for In-Situ Data Collection

Mobile devices are particularly well suited for supporting diary and EMA studies in multiple ways. Researchers and commercial companies have developed a number of software platforms for collecting in-situ data that enable researchers to create data-collection schedules with no or minimal programming. These tools typically support to the collection of self-report data, passive logging of smartphone sensors (e.g., GPS, bluetooth, device status, and activity), or some combination of both (semi-automated tracking) (Choe et al. 2017).

One of the early projects of this kind, MyExperience (Froehlich et al. 2007) ran on Windows Mobile device, and it enabled researchers to construct sophisticated EMA surveys and schedules by writing a single configuration file in XML. MyExperience surveys could collect traditional questionnaire-style data (i.e., multiple-choice questions, text responses, etc.), as well as use the phone's camera and microphone to collect rich multimedia data. They also allowed complex survey schedules where different questionnaires were delivered to participants at different times and based on different triggering conditions.

Many similar systems have since been developed for different mobile platforms, and researchers can now choose among both commercial solutions where survey configuration is done by a commercial company that has developed a proprietary EMA platform (e.g., Life Data Corp and Ilumivu) and open-source platforms that researchers can configure and deploy themselves (e.g., Momento (Carter et al. 2007), MyExperience (Froehlich et al. 2007), AWARE (Ferreira et al. 2015), Jeeves (Rough and Quigley 2015), PACO (Evans 2016), Sensus (Xiong et al. 2016), Extrasensory App (Vaizman et al. 2018), OmniTrack (Kim et al. 2017), and TEMPEST (Batalas et al. 2018)). Among the latter group, the most mature current offering is PACO,[25] a cross-platform (Android and iOS) system built by the Google engineer Bob Evans and his colleagues. Like MyExperience, PACO supports construction of multiple surveys and schedules, a broad range of question types, and sophisticated questionnaire triggering. All configuration of PACO questionnaires can be done through a web interface without any programming, and its website provides study management functionality, including enrolling participants, monitoring their adherence, and downloading response data in standard formats for statistical analyses. To further extend people's data capture capability, OmniTrack enables people to create an Android native tracking app without programming (Kim et al. 2017). By combining various data fields and integrating external data services (e.g., Fitbit) into a single tracking app, people can create a customized tracking app and configure the app to be used as a general diary app or an EMA tool. Researchers can design and deploy an OmniTrack app using OmniTrack for Research (see ◘ Fig. 19.5), which handles app deployment, updates, and participant monitoring.[26] The ability to construct data capture instruments and conduct studies from the web with no programming greatly increases opportunities for researchers outside of technical disciplines to integrate rich in-situ data into their research projects.

Many of the EMA platforms support not just traditional time-based prompting where questionnaires are triggered based on time, but also event-based prompting, where questionnaires are triggered when the phone detects that a particular kind of event has occurred. To do so, EMA systems use sensors in the phone and connected devices to monitor users' state and behavior, and then trigger a questionnaire when a particular set of conditions are met. The sensors and types of

25 PACO: The Personal Analytics Companion. Retrieval June 13, 2019: ► https://pacoapp.com/

26 Kim YH, Lee B, Choe EK, Seo J. 2019. OmniTrack for Research. GitHub. Retrieval June 13, 2019: ► https://omnitrack.github.io/research/

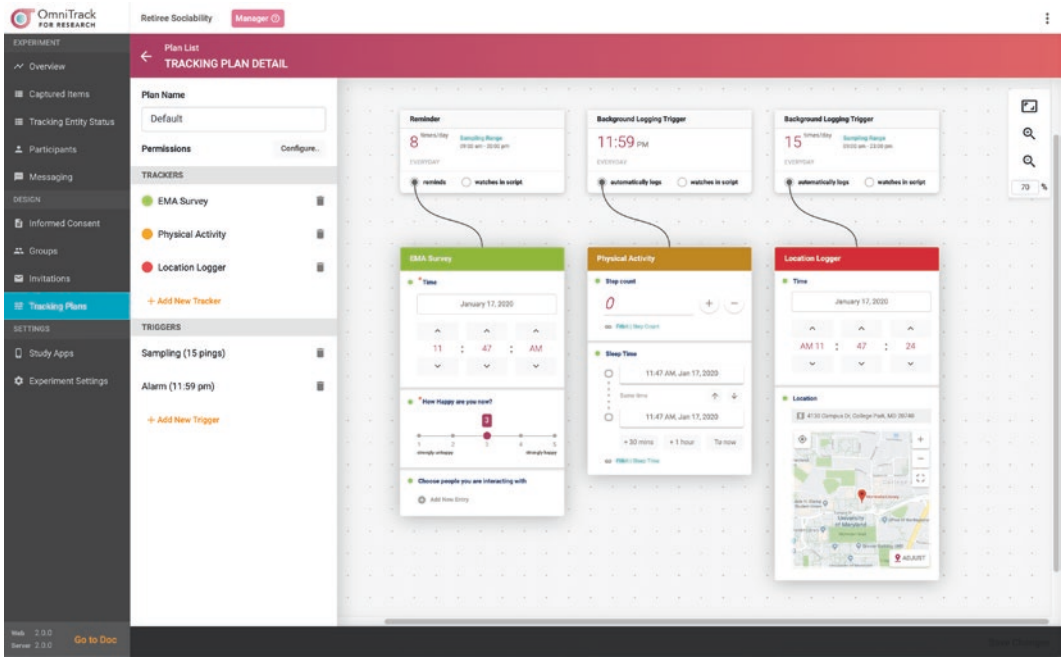


Fig. 19.5 OmniTrack for Research. Researchers can use the graphical user interfaces to create mobile data collection apps equipped with both manual and automated data collection functionalities

events that can be used for triggering vary from platform to platform, but it is common to be able to define events based on location (e.g., a questionnaire can trigger when the person gets home or stops moving), phone activity (a questionnaire can trigger after the person finishes a phone call), or data from accelerometers (a questionnaire can trigger when the person finishes a physical activity). Event-based prompting allows researchers to collect data at precise times when an event of interest (e.g., physical activity) occurs, maximizing their ability to get timely information about the person's experience and surrounding events and minimizing recall biases and the risk of forgetting.

Finally, some EMA platforms enable not just collection of EMA data but also the use of the provided answers to trigger interventions. Such ecological momentary interventions (EMIs; Heron and Smyth 2010) are particularly useful in the substance use arena (e.g., Gustafson et al. 2014; Dennis et al. 2015), where a person's answers to EMA questions can be used to calculate risk of relapse, which, if it reaches a predefined threshold, can trigger coping interventions or facilitate contact with a recovery coach. The ability of EMA systems to trigger a request for self-report at times of high risk (e.g., when a person recovering from alcoholism is near a bar) and immediately respond to provided answers enables the kind of tailored, just-in-time support that was impossible prior to the development of modern mobile technology.

19.2.2.3 Incorporating Self-Reporting Into User Interactions

In addition to diary and EMA studies, mobile technology can facilitate collection of self-report by incorporating self-reporting into interactions that users already perform on their mobile devices. Particularly interesting attempts to do this involved appropriating the unlock gesture on Android phones to enable a user to provide a single piece of self-report data as part of the process of unlocking the phone. The first project that took route, Slide to X (Truong et al. 2014), found that individuals unlocked their phones between 20 and 100 times per day on average. To take advantage of this frequent interaction, Truong et al. built three applications that turned the

phone's standard unlocking interface into a data-collection tool. One of them, Slide To QuantifySelf, enabled users to answer a single question (e.g., "how happy are you right now?") on a Likert scale in place of using performing a standard unlock gesture. Users could specify multiple questions with which to be prompted, as well as when during the day each question should be asked (e.g., asking about whether they ate breakfast mid-morning), providing a low-burden way to collect a rich self-report dataset that can help an individual better manage her health. A recent project called LogIn (Zhang et al. 2016) expanded on this idea by developing gesture-based unlock interactions to track pleasure and accomplishment, sleepiness, and mood. Unlike Slide to QuantifySelf, which used a single Likert scale item to record self-report, LogIt used more sophisticated gesture-based interactions, such as measuring mood on Russell's affect grid (Russell et al. 1989).

Self-report can be incorporated into applications as well. Here too the underlying idea is to decrease the burden of providing self-report by tying it to an action the user is already performing. In HeartSteps (Klasnja et al. 2018), an mHealth physical activity intervention, users can receive prompts to go for a brief walk or move to disrupt prolonged sitting. Like many mHealth interventions, these prompts are provided as push notifications to the user's phone. Other than dismissing them through a simple OK button, however, HeartSteps provides users with three different buttons to dismiss the notification, each intended to indicate how the user perceived that particular prompt. Users can press a thumbs-up icon to indicate that they liked the activity suggestion and that it came at a good time, a thumbs-down icon to indicate that the suggestion was not helpful or came at a bad time, and a button to turn off future prompts for a certain period of time, indicating that they will be busy or unavailable for the intervention. The data from a study of HeartSteps showed that participants were significantly more active after the prompts they marked thumbs-up than those to which they responded in either of the two other ways, suggesting that the self-report obtained in this manner is a useful indicator of individuals' receptivity to intervention.[27] In the new version of HeartSteps, these data are being used by a learning algorithm that personalizes intervention provision for each HeartSteps user. Adding a single question to an application dashboard, or triggering a question by observing the use of other applications (e.g., when a user quits a social media app) are other ways when interactions already taking place on the phone can be leveraged to collect self-report data without the need for additional disruptive prompting of the user.

19.2.2.4 Easing Data Entry

Finally, mobile devices offer a number of ways to ease data entry during self-report. As we mentioned, most EMA platforms enable users to take pictures and use the phone's microphone to speak their answers to the questions. Such multimedia responses enable collection of rich qualitative data that would be too burdensome or even impossible to collect by requiring users to type their responses into a text form. Multimedia capture enables additional types of data analyses that cannot be done on traditional questionnaire data. For instance, PlateMate (Noronha et al. 2011) lets users track what they are eating by taking pictures of meals. The system then uploads these images to Mechanical Turk, where they are processed through a number of steps that extract detailed nutritional information from the images, enabling logging of calories and micro-nutrients with much lower self-report burden than is involved in traditional database-based nutritional logging. Similarly, machine learning algorithms can be used to process audio recordings of data individuals enter by speaking into the phone not only to extract content of those recordings, but also to detect user's mood or features such as latency or pitch that can be indicative of changes in the individual's mental health, such as an onset of a manic episode (e.g., Gideon et al. 2016).

27 Predrag Klasnja, personal communication, 8/31/2019.

Self-report can be made more convenient by placing self-reporting interfaces into easily accessible locations and minimizing how much user interaction is required. Choe et al. (2015) enabled low-burden self-monitoring of sleep by placing a self-tracking widget on the phone's lock screen, making it visible every time a person reached for her phone, and by recording sleep quality via a single tap. Similarly, Fitabase Engage, a new platform from Fitabase, enables simple questions to show up on individuals' Fitbit smartwatches, and the questions to be answered with a single tap. The speed of this interaction, and not needing to reach for the phone, greatly decreases the perceived burden on answering questions delivered via Fitabase Engage.

19.2.3 Providing Interventions in Individuals' Daily Lives

By facilitating passive data collection via sensors embedded in the phone and wearable devices, and by providing ways to collect brief self-report data at the right time and in the right context, mobile technology has greatly enhanced collection of data that are needed to understand health behaviors, to monitor patients' health, and to provide interventions. We now turn to how this information can be used by mHealth systems to provide interventions at times when those interventions are most needed and when individuals are most receptive to them.

Given the immense number of mHealth apps, their content can be categorized in a number of different ways (e.g., see Klasnja & Pratt, 2012 for one classification). For simplicity, many mHealth apps can be seen as falling into one or more of the following five categories:

1. Reminders. One of the simplest functions of mHealth interventions is to provide reminders. Reminders exist for many different health behaviors, including attending scheduled clinic visits, taking medications, and applying sunscreen on sunny days. Sometimes reminders are a stand-alone intervention (e.g., a text messaging intervention to increase adherence to clinic visits (Koshy et al. 2008; Leong et al. 2006)), while other times they are embedded into more complex interventions, such as those for chronic disease management.
2. Support for behavior change. A large number of mHealth interventions are intended to help individuals make health-promoting changes in their behavior. mHealth interventions exist for improving medication adherence, increasing wellness behaviors like physical activity and healthy diet, helping with cessation of addictive behaviors like smoking and substance use, preventing relapse, and adhering to health management practices like monitoring of glucose or blood pressure. Some of the applications for the management of mental health (e.g., depression and bipolar disorder) fall into this category as well, as they often focus on helping individuals enact therapeutic practices like behavioral activation, or increase the regularity of sleep, social contact, and other behaviors that support mental well-being.
3. Discovery of patterns. A related category is mHealth interventions that support individuals in discovering patterns in their behavioral or physiological responses. For instance, TummyTrials (Karkar et al. 2017) is a recent mHealth intervention intended to help individuals detect food triggers that aggravate irritable bowel syndrome. Other interventions, like Health Mashups, (Bentley et al. 2013), try to help individuals understand what factors influence their physical activity, sleep, and other wellness behaviors. Although interventions in this category do not have explicit behavior-change features, such as goal-setting or planning, the **self-experimentation** they support is usually in service of making health-promoting changes in one's behavior, making this category a close relative of the behavior-change interventions.
4. Detection or prediction of critical health events. An important use of sensors and self-report data in the mHealth context is to enable interventions that detect and/or predict critical health events that may

require prompt medical attention. mHealth solutions exist to detect a broad range of acute health events, including falls, alarming levels of chemotherapy toxicity, imminent risk of a heart attack, onset of a manic episode, and decompensation in chronic heart failure. Depending on the severity of the detected event, such interventions either provide guidance to the individual on how to manage the event or automatically contact the emergency services to get the individual medical help as quickly as possible.

5. Communication with the healthcare system. A growing number of mHealth interventions are intended to facilitate communication with the health system through the support of remote patient monitoring, secure messaging, prescription refills, accessing labs and imaging, and scheduling appointments. Unlike the categories reviewed above, mHealth apps in this category are almost always provided by and tied to a particular health system, pharmacy, or a clinic. To make such mHealth interventions work, clinical workflows often have to be restructured to accommodate the use of mHealth tools by patients and the data that are generated as part of that use.

Although the above categories are not intended to be exhaustive, they do cover many of the common types of mHealth applications. Thus, they demonstrate the range of the intervention work that has been done in the field of mHealth.

19.2.4 Providing Just-in-Time Adaptive Interventions

From the earliest days of mHealth, a major promise of mobile technology, given its constant proximity to the person, has been its ability to provide support for health management when that support is most needed (Intille 2004; Patrick et al. 2008; Nilsen et al. 2012): at times when a person is at risk (e.g., of a substance use relapse) and needs help with coping, when there is an opportunity to engage in health-promoting behaviors, or when the sensors detect that there is a critical change in the person's health (e.g., high likelihood of a heart attack) and she needs to receive immediate medical attention. The development of such just-in-time interventions is made possible by continuous, low-burden data collection, always-on Internet connectivity, powerful phone-based information processing, and the ability to prompt the person via a push notification or a text message to deliver an intervention. A great deal of recent work on mHealth interventions has aimed to realize this promise of smart, timely intervention delivery.

Although the promise of timely, in-context information has been acknowledged for many years, it is only recently that the technical, algorithmic, and methodological developments have enabled the development of interventions that realize this goal. In recent literature, such mHealth interventions have been referred to as just-in-time adaptive interventions (JITAIs; Nahum-Shani et al. 2015; Nahum-Shani et al. 2016; Spruijt-Metz and Nilsen 2014). JITAIs refer to mHealth systems that use *decision rules*—if-then rules or algorithms that specify when, where, and how interventions are delivered to individuals—to attempt to provide the right type of support at the right times and in the right contexts. JITAIs use sensors and low-burden self-report to continuously monitor individuals' state, behavior, and the environment, and when they detect that an individual is in a state of high risk or has an opportunity to engage in a health-promoting behavior, they make a decision about whether to intervene. How these decisions are made varies. Simple JITAIs use deterministic if-then rules that determine what the system should do when a situation of risk or opportunity is detected. For instance, a JITAI might send a person recovering from alcohol use disorder a push notification with a coping strategy every time that person is within a certain distance of a bar (Gustafson et al. 2014). Or the decision rules may be stochastic, where intervention is not provided every time a situation of risk or opportunity is encountered but only with certain probability—usually to reduce user burden (e.g., Klasnja et al. 2018). Finally, the system

may incorporate algorithms that evolve the decision rules over time to maximize their effectiveness for each individual (we will return to this point shortly).

What form the decision rules take depends on the nature of a JITAI's intervention components, as well as the situations it is trying to target. For situations that occur frequently—such as getting stressed or high likelihood of lighting up a cigarette—interventions have to be spaced out to manage user burden and reduce the risk of system abandonment. Stochastic rules are well suited for meeting these criteria, as they can reduce the overall number of interventions while creating a sense of unpredictability, which may increase the effectiveness of the interventions by reducing habituation. On the other hand, for risk states that occur rarely—or for which the consequences of not intervening are severe (the risk of suicide being an extreme example)—simple deterministic decision rules may be both appropriate and adequate.

As we mentioned above, decision rules can also evolve over time. Although standard deterministic and stochastic decision rules do not change over the course of system use and are typically the same for all users of a system, mHealth systems with evolving rules aim to personalize deliver of interventions for each system user. The idea is that the system learns patterns in user behavior and intervention response over time, and it can then adjust how it provides interventions to maximize their effectiveness for each individual user and to minimize user burden.

Two approaches to intervention personalization are currently being investigated: reinforcement learning and control systems engineering. JITAIs based on reinforcement learning (RL) use algorithms to continually adjust the probability of intervening based on observations of the outcome of previously delivered interventions (Sutton and Barto 1998). Observations of successful outcomes lead the algorithm to increase the probability of intervening in the similar context in the future; unsuccessful outcomes decrease that probability. RL algorithms can focus either on short-term outcomes—the most immediate outcomes of individual intervention provisions (e.g., the number of steps a person walks on a day when the system sent a motivational message in the morning)—or long-term outcomes that prioritize how the system performs over time (e.g., the user's average daily step count over the course of a month). Much of the foundational research in RL has been done in areas like robotics where both the definition of success and the relevant state variables are relatively unambiguous. As human behavior is inherently more messy and the system's knowledge of the person's state and environment is far more noisy, it's an open research question whether mHealth systems would work better by employing algorithms that focus on clearer but shorter-term outcomes, or if the more sophisticated algorithms that focus on the long-term can be made to work in this setting.

The other approach to personalized JITAIs draws on control systems engineering (Hekler et al. 2018; Phatak et al. 2018). Control systems engineering focuses on the development of systems that are capable of controlling complex processes, such as the flight of an airplane or blood glucose metabolism. At the heart of this approach is the development of mathematical models—called **dynamical systems models**—that encode what is known about the influences on the process or behavior that needs to be controlled. Those models are then used by control systems such as the plane autopilot or the artificial pancreas to make decisions about how the system should intervene to maximize the likelihood that the process will behave in a desired way (the flight will get to the destination city, the blood glucose will be in the healthy range, etc.). In the context of mHealth, the approach is being used to form dynamical models of health behaviors (e.g., an individual's daily steps), and the models are then used by an mHealth system's "controller" to decide when and how to intervene to move the behavior in the desired direction (e.g., to increase the daily step count). What is notable about this approach is that the controller is constantly updating its model based on continual observation of the person's behavior and response to interventions, so that subsequent intervention decisions are always based on the updated model, one that

progressively better describes the idiosyncrasies of each user's behavior. Although control systems engineering has been used successfully to manage complex processes for decades, its application to mobile health is still in its infancy (see Hekler et al. 2018 for the state of the field).

19.2.5 Supporting Self-Experimentation

We see a growing interest in supporting patients and laypeople to design and conduct self-experimentation in personalized health. Although traditional clinical studies (e.g., epidemiological surveys, longitudinal studies, and randomized controlled trial) provide relevant knowledge at the population level, they do not provide the necessary knowledge for any given individual. On the other hand, in self-experimentation (or n-of-1 trials), an individual serves as their own control, allowing them to systematically explore a specific hypothesis (e.g., Does caffeine impact my sleep?) of their interest. To support laypeople in designing and conducting scientifically rigorous self-experimentations, researchers designed self-experimentation platforms leveraging mobile devices' continuous monitoring and capture capabilities. PACO and IFTTT[28] can be used as a general-purpose self-experimentation platform due to their random notification and tracking features (Evans 2016). Domain-specific self-experimentation platforms can provide customized support. For example, TummyTrials (Karkar et al. 2017) helps people identify food triggers, and SleepCoacher (Daskalova et al. 2016) helps identify connections between potential sleep disruptors and sleep quality. Despite some underlying limitations of self-experimentations—such as carryover effects and blinding—self-experimentation augmented with mHealth technologies has great potential to leverage personal knowledge and advance personalized medicine (Lillie et al. 2011).

28 IFTTT. Retrieval June 13, 2019: ► https://ifttt.com/

19.3 Broad Considerations and Challenges

All health technologies have associated pragmatic challenges and ethical issues regarding privacy and security, impact on people's work, the potential to increase health disparities, and regulatory issues. In this section, we focus on aspects of those issues that are unique to or particularly problematic for mHealth applications.

19.3.1 Privacy and Security

For mHealth technologies, the key privacy and security concerns center around the personal data that can be obtained from a device that is always or nearly always with you. Information that could be collected from one's mobile device includes a great deal that people often want to keep private, such as their location, time spent with the device on, other apps used (including the duration and frequency of use), websites visited, search terms used, etc. Often those data are stored "in the cloud," making it vulnerable to hackers, and allowing the applications to sell or use the data for other purposes. An alternative to such cloud-based systems would be to store all the information only on one's mobile device, so that no other system has access to it. However, then the user is vulnerable to the loss of such important data if the mobile device is lost or damaged. Although many applications claim to use or sell only aggregate, anonymized data, the amount and kind of data collected from one's mobile device make it particularly vulnerable to reidentification. In addition, people tend to agree to terms of service without reading them, and thus, likely have little knowledge about what data are available to others.

An additional challenge for mHealth applications comes from its ubiquitous nature—others could easily observe notifications or reminders on someone's mobile screen because the screen is often visible to many others throughout the day. Much harm can come from the disclosure of such private information, whether it is from observing the

information on the screen or from gathering stored data. For example, location logs could be used to disclose businesses or locations a person has visited when and or for how long, or to predict where a person will be when, which could endanger someone who needs to conceal their whereabouts for safety reasons. Although most mobile operating systems now attempt to make such location tracking optional, many mHealth apps need such data to function fully. For example, fitness trackers need location data to accurately record users' activity. Thus, users often must choose between using an mHealth application and keeping their data private.

19.3.2 Changes in Clinician Work

The proliferation of these mHealth apps brings the potential for substantive changes in clinician work as well as for clinician-patient interactions. Because many people are now using mHealth apps to generate huge amounts of detailed health data, people could expect clinicians to use those data to gain new insights into a person's health and positively influence their care. However, the volume and variety of data as well as applications that generated the data make it challenging for clinicians to incorporate the new data into their workflow. A literature review of empirical studies of self-tracking tools identified many clinician work-related concerns about information quality and the lack of standards for representing or viewing that data (West et al. 2016, 2017). Yet, recent research points to the value of human-centered design approaches to addressing these concerns. For example, one study of DataMD showed that such an approach could help clinicians develop a new workflow that would allow them incorporate this kind of data and improve their counseling skills and support more in-depth conversations (Kim et al. 2017).

Yet, irrespective of good design, the real-time sensing nature of the data creates other workflow concerns, particularly ethical and legal expectations that clinicians respond to the sensed data promptly. For example, mental health providers could be expected to intervene immediately when a mobile sensor indicates a high likelihood that their client could harm themselves or others. Although such responsiveness could improve outcomes, these expectations could lead to further problems with clinician burnout and challenges with patient autonomy and confidentiality. Physicians are already experiencing a greater increase in burnout and reduction in satisfaction with work-life balance than peer adults in the U.S. (Shanafelt et al. 2015). However, a scoping review of physician well-being in the mHealth context showed that these technologies are playing important roles in the improvement of physicians' well-being too (Chen et al. 2018). Such advances are additionally changing both the nature and amount of patient-clinician interactions. We have much to learn about how clinicians' use of these new mHealth technologies and their response to patients' use of the technologies will affect clinicians' work.

19.3.3 Changes in Patient Work

This increasing prevalence of mHealth apps are changing the work and personal lives of the people who use these tools to assist in their everyday health and well-being. With constant sensing of our health comes the potential for unhealthy disruptions to daily life. Even simple apps that count steps for physical fitness typically interrupt people when they have reached their step goal or remind people to get up and move during the day. Although these rewards and reminders can help people meet their health goals, such disruptions can come at inopportune times or places and negatively impact people's overall well-being. Many apps allow people to disable those disruptive features, but then they risk missing important information or reminders that could be key to its successful use. Frequent tracking and constant reminders of one's health can also lead to detrimental obsessions. For example, one study examined the effect of tracking technology on college students and found that those students who used fitness trackers had higher levels of eating concerns and symptoms of eating disorders (Simpson

and Mazzeo 2017). Other studies have shown that such tracking can make healthy activities that people used to enjoy feel more like work and lead to a decrease in those activities and subjective well-being (Etkin 2016). Furthermore, people could become overly reliant on the objectively sensed but not necessarily accurate data and discount their own or others' subjective experiences.[29]

Nonetheless, such widespread availability of mobile health apps also brings unprecedented power to everyday people in their ability to collect new forms of data about their own health and to interpret that data independently from clinicians. For example, HemaApp allows people with anemia or pulmonary illnesses to detect and monitor total hemoglobin in their blood using a smartphone's camera and flash, rather than requiring a visit to a clinician's office for a blood draw and follow up visit about their results (Wang et al. 2016). Many other apps, such as BiliScreen for jaundice detection and pancreatic cancer screening (Mariakakis et al. 2017) and SpiroSmart for assessing lung function (Larson et al. 2012), allow people to detect health problems or monitor existing problems that previously required a clinician visit.

19.3.4 Health Disparities

Despite the increasing uptake of mobile phones by people—regardless of race, age, or socioeconomic status—and the high prevalence of smartphone usage (81% of all U.S. adults),[30] concerns still remain about whether increased reliance upon such mobile technologies will exacerbate existing health disparities. A recent systematic review examined the research literature to investigate mHealth interventions for vulnerable populations (Stowell et al. 2018). They identified familiarity with the technology, use of engaging multimedia content, frequent delivery of content, and personalization as facilitators in successful interventions. However, costs and concerns about confidentiality and privacy (particularly for those who hadn't completed their immigration paperwork) were substantial barriers. Although a slight majority of the reviewed studies showed significant improvements in the evaluated health measures, their meta-analysis failed to show that the mHealth interventions successfully impacted health outcomes in vulnerable populations. Researchers are beginning to work together to identify opportunities for sociotechnical interventions to reduce health disparities (Siek et al. 2019), but much work remains.

19.3.5 Regulatory Issues

The staggering number of mHealth apps that are available in the marketplace makes it challenging for clinicians, patients, and the general public to discern which apps are effective and safe to adopt. Many worry about the proliferation of these mHealth apps, particularly the wide variation in quality, potential misleading or unsubstantiated claims, and the vulnerability of disclosure of personal health information. The US Food and Drug Administration (FDA) is the regulatory body that could provide such safety and effectiveness oversight of mHealth apps, but its role and influence has been changing. In 2013, 2015, and 2019, the FDA revised its guidance about what it will regulate in the mHealth space and is now focusing only on apps that pose a great risk if they do not work as intended.[31] The regulations pertain to apps that diagnose, treat, or prevent a health condition. To help developers determine what laws and

29 Siek, K. Why fitness trackers may not give you all the 'credit' you hoped for. January 15, 2020. The Conversation. ► https://theconversation.com/why-fitness-trackers-may-not-give-you-all-the-credit-you-hoped-for-128585 Retrieved January 15, 2020.

30 Pew Research Center. June 12, 2019. Mobile Fact Sheet. 1–6. ► http://www.pewinternet.org/factsheet/mobile/ Retrieval January 15, 2020.

31 FDA. Device Software Functions Including Mobile Medical Applications. ► https://www.fda.gov/medical-devices/digital-health/device-software-functions-including-mobile-medical-applications Retrieval January 15, 2020.

regulations apply to their mHealth apps, the FDA created an interactive tool that poses questions and summarizes the applicability of the laws based on the answers.[32] In the private sector, Xcertia serves as an mHealth app collaborative effort of the American Medical Association (AMA), the American Heart Association (AHA), DHX Group and the Healthcare Information and Management Systems Society (HIMSS) "to foster safe, effective, and reputable health technologies."[33] They have developed a set of guidelines for assessing operability, privacy, security, content, and usability. Clearly, the safety and regulation of mHealth apps will remain a key issue for the future.

19.4 Future Directions

The rapidly changing nature of technology makes writing a book chapter on any aspect of it challenging, but anticipating future directions is even more daunting. All aspects of this book face that challenge, but the field of mHealth and the applications that characterize it are especially dynamic. Exact statistics are hard to find and rapidly out of date, but nonetheless paint a picture of mHealth's growing influence. A 2017 report from IQIA reports that over 318,000 mHealth apps were available worldwide and that more than 200 health apps are added each day.[34] According to App Annie's State of Mobile 2019 Report, the global download of mHealth apps exceeded 400 million in 2018 with the growth coming from many different countries.[35] Grand View Research predicts that the global market for mHealth apps will reach 236 billion in US dollars by 2026 with fitness as the largest type of mHealth app.[36]

Some see the future of mHealth through the eyes of science fiction. In particular, many have envisioned Star Trek's® tricorder-like technology of a small, hand-held device that could quickly diagnose a variety of medical conditions. Qualcomm incentivized this vision with its $10 million XPRIZE competition to develop a mHealth device that could accurately diagnose 13 medical conditions, capture 5 real-time health vital signs, and provide a compelling consumer experience, without input from a healthcare professional or facility.[37] Although no one was able to meet all their criteria, Final Frontier Medical Devices (now Basil Leaf Technologies[38]) received the top prize of $2.5 million with their DxtER device that employed non-invasive sensors to collect vital signs, body chemistry, and biological functions.[39] Many others have alternative, grand views of the future of mHealth apps. One point is clear: mHealth applications will continue to influence all aspects of healthcare—from wellness and prevention through

32 Developing a mobile health app? Find out which federal laws you need to follow. ▶ https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-apps-interactive-tool Retrieval January 15, 2020

33 Xcertia: mHealth App Guidelines. ▶ https://xcertia.org/ Retrieved January 15, 2020.

34 The Growing Value of Digital Health: Evidence and Impact on Human Health and the Healthcare System. Nov 07, 2017. ▶ https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health Retrieved January 15, 2020.

35 Sydow, L. Medical Apps Transform How Patients Receive Medical Care. April 16, 2019. ▶ https://www.appannie.com/en/insights/market-data/medical-apps-transform-patient-care/ Retrieved January 15, 2020.

36 Grand View Research. mHealth Apps Market Size, Share & Trends Analysis Report By Type (Fitness, Lifestyle Management, Nutrition & Diet, Women's Health, Medication Adherence, Healthcare Providers/Payers), And Segment Forecasts, 2019–2026. June 2019 ▶ https://www.grandviewresearch.com/industry-analysis/mhealth-app-market Retrieved January 15, 2020

37 Qualcomm Tricorder XPRIZE: Empowering Personal Healthcare. ▶ https://www.xprize.org/prizes/tricorder Retrieved January 15, 2020.

38 Basil Leaf Technologies. ▶ http://www.basilleaftech.com/ Retrieved January 15, 2020.

39 Family-led team takes top prize in qualcomm tricorder xprize competition for consumer medical device inspired by Star Trek® April 13, 2017. ▶ https://www.xprize.org/prizes/tricorder/articles/family-led-team-takes-top-prize-in-qualcomm-tricor

patient-provider interaction and even surgery—and that influence will likely follow an unpredictable but pivotal path.

Suggested Readings

Klasnja, P., & Pratt, W. (2012). Healthcare in the pocket: Mapping the space of mobile-phone health interventions. Journal of Biomedical Informatics, 45(1), 184–198.

Choe, E. K., Lee, N. B., Lee, B., Pratt, W., & Kientz, J. A. (2014). Understanding quantified-selfers' practices in collecting and exploring personal data. In Proceedings of the 32nd annual ACM conference on human factors in computing systems (pp. 1143–1152). New York: ACM.

Kim, Y. H., Jeon, J. H., Lee, B., Choe, E. K., & Seo, J. (2017). OmniTrack: a flexible self-tracking approach leveraging semi-automated tracking. Proceedings of the ACM conference on interactive, mobile, wearable and ubiquitous technologies. 1(3):1:1–1:28.

Stowell, E., Lyson, M. C., Saksono, H., Wurth, R. C., Jimison, H., Pavel, M., & Parker, A. G. (2018). Designing and evaluating mHealth interventions for vulnerable populations: A systematic review. In Proceedings of the 2018 CHI conference on human factors in computing systems. Paper (Vol. 15, pp. 1–17).

Questions for Discussion

- What challenges does the rapid evolution of mobile technologies and platforms (OS) pose to the people who develop mHealth interventions?
- What are some of the ways in which mHealth technologies are designed for health professionals? In designing such tools, what are some key design considerations?
- How can lay individuals access their own data collected from mHealth technologies? In what cases such data access might be useful for individuals?
- In addition to creating new ways to deliver health interventions, mHealth has been seen as having the potential to greatly advance health research, including fields such as epidemiology. How can mHealth tools advance our understanding of factors that shape individuals' health?
- What are just-in-time adaptive interventions and how do mHealth tools enable this type of health intervention?
- Most patient-centered mHealth tools are discretionary use technologies, in the sense that individuals can choose whether, how much, and for how long to use these devices and applications. Yet, for these tools to have a hope of being effective, they have to be used. Given what learned in this chapter, what aspects of mHealth tools can facilitate and hinder engagement? How can mHealth designers make these tools more engaging so individuals can benefit from them?

References

Abbey, B., Alipour, A., Gilmour, L., Camp, C., Hofer, C., Lederer, R., et al. (2012). A remotely programmable smart pillbox for enhancing medication adherence. In *2012 25th IEEE international symposium on computer-based medical systems (CBMS)* (pp. 1–4). IEEE.

Alexander, C. M., Teller, L. E., & Gross, J. B. (1989). Principles of pulse oximetry: Theoretical and practical considerations. *Anesthesia & Analgesia, 68*(3), 368–376.

Ali, A. A., Hossain, S. M., Hovsepian, K., Rahman, M. M., Plarre, K., & Kumar, S. (2012). mPuff: Automated detection of cigarette smoking puffs from respiration measurements. In *Proceedings of the 11th international conference on information processing in sensor networks* (pp. 269–280). ACM.

Alshurafa, N., Kalantarian, H., Pourhomayoun, M., Sarin, S., Liu, J. J., & Sarrafzadeh, M. (2014). Non-invasive monitoring of eating behavior using spectrogram analysis in a wearable necklace. In *2014 IEEE healthcare innovation conference (HIC)* (pp. 71–74). IEEE.

Althoff, T., White, R. W., & Horvitz, E. (2016). Influence of Pokémon Go on physical activity: Study and implications. *Journal of Medical Internet Research., 18*(12), 1–23.

Amft, O., Junker, H., & Troster, G. (2005). Detection of eating and drinking arm gestures using inertial body-worn sensors. In *Ninth IEEE international symposium on wearable computers (ISWC'05)* (pp. 160–163). IEEE.

Amini, F., Hasan, K., Bunt, A., & Irani, P. (2017). Data representations for in-situ exploration of health and fitness data. In *Proceedings of the 11th EAI international conference on pervasive computing technologies for healthcare* (pp. 163–172). New York: ACM.

Batalas, N., Khan, V. J., & Markopoulos, P. (2018). Using TEMPEST: End-user programming of web-based ecological momentary assessment protocols. *Proceedings of the ACM on Human-Computer Interaction. 2(EICS), 3*, 1–3:24.

Bentley, F., Tollmar, K., Stephenson, P., Levy, L., Jones, B., Robertson, S., et al. (2013). Health mashups. *ACM Transactions on Computer-Human Interaction, 20*(5), 1–27.

Blascheck, T., Besançon, L., Bezerianos, A., Lee, B., & Isenberg, P. (2019). Glanceable visualization: Studies of data comparison performance on smartwatches. *IEEE Transactions on Visualization and Computer Graphics, 25*(1), 630–640.

Brehmer, M., Lee, B., Isenberg, P., & Choe, E. K. (2018). Visualizing ranges over time on mobile phones: A task-based crowdsourced evaluation. *IEEE Transactions on Visualization and Computer Graphics, 25*(1), 619–629.

Buettner, M., Prasad, R., Philipose, M., & Wetherall, D. (2009). Recognizing daily activities with RFID-based sensors. In *Proceedings of the 11th international conference on ubiquitous computing* (pp. 51–60). New York: ACM.

Carter, S., Mankoff, J., & Heer, J. (2007). Momento: support for situated ubicomp experimentation. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 125–134). New York: ACM.

Carek, A. M., Conant, J., Joshi, A., Kang, H., & Inan, O. T. (2017). SeismoWatch: wearable cuffless blood pressure monitoring using pulse transit time. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 40.

Chen, H. C., Jia, W., Yue, Y., Li, Z., Sun, Y. N., Fernstrom, J. D., & Sun, M. (2013). Model-based measurement of food portion size for image-based dietary assessment using 3D/2D registration. *Measurement Science and Technology, 24*, 105701.

Chen, S., Ju, X., Guo, X., & Tian, W. S. (2018). *Factors influencing physician's well-being management within the mobile health context: A scoping review*. Proceedings of healthcare informatics & health information technology (Sighealth) AMCIS 2018.

Choe, E. K., Lee, N. B., Lee, B., Pratt, W., & Kientz, J. A. (2014). Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the 32nd annual ACM conference on human factors in computing systems* (pp. 1143–1152). New York: ACM.

Choe, E. K., Lee, B., Kay, M., Pratt, W., & Kientz, J. A. (2015). SleepTight: Low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (pp. 121–132). New York: ACM.

Choe, E. K., Abdullah, S., Rabbi, M., Thomaz, E., Epstein, D. A., Cordeiro, F., et al. (2017). Semi-automated tracking: A balanced approach for self-monitoring applications. *IEEE Pervasive Computing, 16*(1), 74–84.

Choudhury, T., Borriello, G., Consolvo, S., Haehnel, D., Harrison, B., Hemigway, B., et al. (2008). The Mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing, 7*(2), 32–41.

Chung, C. F., Cook, J., Bales, E., Zia, J., & Munson, S. A. (2015). More than telemonitoring: Health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. *Journal of Medical Internet Research, 17*(8), 1–25.

Consolvo, S., Klasnja, P., McDonald, D. W., & Landay, J. A. (2014). Designing for healthy lifestyles: Design considerations for mobile technologies to encourage consumer health & wellness. *Foundations and Trends in Human-Computer Interaction, 6*(3–4), 167–315.

Dai, J., Bai, X., Yang, Z., Shen, Z., & Xuan, D. (2010). Mobile phone-based pervasive fall detection. *Personal and Ubiquitous Computing, 14*(7), 633–643.

Daskalova, N., Metaxa-Kakavouli, D., Tran, A., Nugent, N., Boergers, J., McGeary, J., & Huang, J. (2016). SleepCoacher: A personalized automated self-experimentation system for sleep recommendations. In *Proceedings of the 29th annual symposium on user interface software and technology* (pp. 347–358). New York: ACM.

Dennis, M. L., Scott, C. K., Funk, R. R., & Nicholson, L. (2015). A pilot study to examine the feasibility and potential effectiveness of using smartphones to provide recovery support for adolescents. *Substance Abuse, 36*(4), 486–492.

Devarakonda, S., Sevusu, P., Liu, H., Liu, R., Iftode, L., & Nath, B. (2013). Real-time air quality monitoring through mobile sensing in metropolitan areas. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing* (p. 15). ACM.

Dey, A. K., Wac, K., Ferreira, D., Tassini, K., Hong, J. H., & Ramos, J. (2011). Getting closer: An empirical investigation of the proximity of user to their smart phones. In *Proceedings of the 13th international conference on ubiquitous computing* (pp. 163–172). New York: ACM.

Eadie, M., & Steele, R. J. (2017). Non-invasive blood glucose monitoring and data analytics. In *Proceedings of the international conference on compute and data analysis* (pp. 138–142). ACM.

Ertin, E., Stohs, N., Kumar, S., Raij, A., Al'Absi, M., & Shah, S. (2011). AutoSense: Unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In *Proceedings of the 9th ACM conference on embedded networked sensor systems* (pp. 274–287). ACM.

Etkin, J. (2016). The hidden cost of personal quantification. *Journal of Consumer Research, 42*(6), 967–984.

Evans, B. (2016). PACO—Applying computational methods to scale qualitative methods. *Ethnographic Praxis in Industry Conference Proceedings, 2016*(1), 348–368.

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity, 12*(1), 159.

Fang, S. H., Liang, Y. C., & Chiu, K. M. (2012). Developing a mobile phone-based fall detection system on android platform. In *2012 computing, communications and applications conference* (pp. 143–146). IEEE.

Fenton, S., Giannangelo, K., Kallem, C., & Scichilone, R. (2013). Data standards, data quality, and interoperability (updated). *Journal of AHIMA, 84*(11), 64–69.

Ferreira, D., Kostakos, V., & Dey, A. K. (2015). AWARE: Mobile context instrumentation framework. *Frontiers in ICT, 2*(6), 1–6.

Free, C., Knight, R., Robertson, S., Whittaker, R., Edwards, P., Zhou, W., et al. (2011). Smoking cessation support delivered via mobile phone text messaging (txt-2stop): A single-blind, randomised trial. *The Lancet, 378*(9785), 49–55.

Fritz, T., Huang, E. M., Murphy, G. C., & Zimmermann, T. (2014). Persuasive technology in the real world: A study of long-term use of activity sensing devices for fitness. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 487–496). New York: ACM.

Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., & Landay, J. A. (2007). MyExperience: A system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th international conference on mobile systems, applications and services* (pp. 57–70). New York: ACM.

Gideon, J., Provost, E. M., & McInnis, M. (2016). Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. In *Proceedings of IEEE international conference on acoustics, speech, and signal processing (ICASSP)* (pp. 2359–2363). Piscataway: IEEE.

Gouveia, R., Pereira, F., Karapanos, E., Munson, S. A., & Hassenzahl, M. (2016). Exploring the design space of glanceable feedback for physical activity trackers. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing* (pp. 144–155). New York: ACM.

Gustafson, D. H., McTavish, F. M., Chih, M. Y., Atwood, A. K., Johnson, R. A., Boyle, M. G., et al. (2014). A smartphone application to support recovery from alcoholism: A randomized clinical trial. *JAMA Psychiatry, 71*(5), 566–572.

Hayes, T. L., Hunt, J. M., Adami, A., & Kaye, J. A. (2006). An electronic pillbox for continuous monitoring of medication adherence. In *2006 international conference of the IEEE engineering in medicine and biology society* (pp. 6400–6403). IEEE.

Hekler, E. B., Rivera, D. E., Martin, C. A., Phatak, S. S., Freigoun, M. T., Korinek, E., et al. (2018). Tutorial for using control systems engineering to optimize adaptive mobile health interventions. *Journal of Medical Internet Research, 20*(6), e214.

Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology, 15*(1), 1–39.

Hovsepian, K., al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M., & Kumar, S. (2015). cStress: Towards a gold standard for continuous stress assessment in the mobile environment. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (pp. 493–504). New York: ACM.

Intille, S. S., Kukla, C., Farzanfar, R., & Bakr, W. (2003). Just-in-time technology to encourage incremental, dietary behavior change. In *AMIA annual symposium proceedings* (p. 874). Bethesda, MD: American Medical Informatics Association.

Intille, S. S. (2004). Ubiquitous computing technology for just-in-time motivation of behavior change. *Studies in Health Technology and Informatics, 107*(2), 1434–1437.

Jakicic, J. M., Davis, K. K., Rogers, R. J., King, W. C., Marcus, M. D., Helsel, D., et al. (2016). Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: The IDEA randomized clinical trial. *Journal of the American Medical Association, 316*(11), 1161–1171.

Jaspan, H. B., Flisher, A. J., Myer, L., Mathews, C., Seebregts, C., Berwick, J. R., et al. (2007). Brief report: Methods for collecting sexual behaviour information from South African adolescents—a comparison of paper versus personal digital assistant questionnaires. *Journal of Adolescence, 30*(2), 353–359.

Kaziunas, E., Lindtner, S., Ackerman, M. S., & Lee, J. M. (2018). Lived data: Tinkering with bodies, code, and care work. *Human-Computer Interaction, 33*(1), 49–92.

Kang, J., Binda, J., Agarwal, P., Saconi, B., & Choe, E. K. (2017). Fostering user engagement: Improving sense of identity through cosmetic customization in wearable trackers. In *Proceedings of the 11th EAI international conference on pervasive computing technologies for healthcare* (pp. 11–20). New York: ACM.

Kanhere, S. S. (2011). Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces. In *2011 IEEE 12th international conference on mobile data management* (Vol. 2, pp. 3–6). IEEE.

Karkar, R., Schroeder, J., Epstein, D. A., Pina, L. R., Scofield, J., Fogarty, J., et al. (2017). Tummytrials: A feasibility study of using self-experimentation to detect individualized food triggers. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 6850–6863). New York: ACM.

Kho, A., Henderson, L. E., Dressler, D. D., & Kripalani, S. (2006). Use of handheld computers in medical education. *Journal of General Internal Medicine, 21*(5), 531–537.

Kim, Y., Heo, E., Lee, H., Ji, S., Choi, J., Kim, J. W., et al. (2017). Prescribing 10,000 steps like aspirin. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 5787–5799). New York: ACM.

Kim, Y., Lee, B., & Choe, E. K. (2019). Investigating data accessibility of personal health apps. *Journal of the American Medical Informatics Association, 26*(5), 412–419.

Kim, Y. H., Jeon, J. H., Lee, B., Choe, E. K., & Seo, J. (2017). OmniTrack: a flexible self-tracking approach leveraging semi-automated tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. *1*(3):1:1–1:28.

Kitamura, K., de Silva, C., Yamasaki, T., & Aizawa, K. (2010). Image processing based approach to food balance analysis for personal food logging. In *Multimedia and expo (ICME), 2010 IEEE international conference on* (pp. 625–630).

Klasnja, P., Consolvo, S., & Pratt, W. (2011). How to evaluate technologies for health behavior change in HCI research. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3063–3072). New York: ACM.

Klasnja, P., & Pratt, W. (2012). Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics, 45*(1), 184–198.

Klasnja, P., Smith, S., Seewald, N. J., Lee, A., Hall, K., Luers, B., et al. (2018). Efficacy of contextually tailored suggestions for physical activity: A micro-randomized optimization trial of HeartSteps. *Annals of Behavioral Medicine, 53*(6), 573–582.

Koshy, E., Car, J., & Majeed, A. (2008). Effectiveness of mobile-phone short message service (SMS) reminders for ophthalmology outpatient appointments: Observational study. *BMC Ophthalmology, 8*(9), 1–6.

Kwon, M., Lee, J. Y., Won, W. Y., Park, J. W., Min, J. A., Hahn, C., et al. (2013). Development and validation of a smartphone addiction scale (SAS). *PLoS One, 8*(2), e56936:1–e56936:7.

Kyritsis, K., Diou, C., & Delopoulos, A. (2017). Food intake detection from inertial sensors using lstm networks. In *International conference on image analysis and processing* (pp. 411–418). Cham: Springer.

Lane, S. J., Heddle, N. M., Arnold, E., & Walker, I. (2006). A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. *BMC Medical Informatics and Decision Making, 6*(23), 1–10.

Larson, E. C., Goel, M., Boriello, G., Heltshe, S., Rosenfeld, M., & Patel, S. N. (2012). SpiroSmart: Using a microphone to measure lung function on a mobile phone. In *Proceedings of the 2012 ACM conference on ubiquitous computing* (pp. 280–289). New York: ACM.

Lazar, A., Koehler, C., Tanenbaum, J., & Nguyen, D. H. (2015). Why we use and abandon smart devices. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (pp. 635–646). New York: ACM.

Lee, B., Brehmer, M., Isenberg, P., Choe, E. K., Langer, R., & Dachselt, R. (2018). Data visualization on mobile devices. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (pp. 1–8). New York: ACM.

Lee, Y. K., Chang, C. T., Lin, Y., & Cheng, Z. H. (2014). The dark side of smartphone usage: Psychological traits, compulsive behavior and technostress. *Computers in Human Behavior, 31*, 373–383.

Leong, K. C., Chen, W. S., Leong, K. W., Mastura, I., Mimi, O., Sheikh, M. A., et al. (2006). The use of text messaging to improve attendance in primary care: A randomized controlled trial. *Family Practice, 23*(6), 699–705.

LiKamWa, R., Liu, Y., Lane, N. D., & Zhong, L. (2013). Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceedings of the 11th annual international conference on mobile systems, applications, and services* (pp. 389–402). New York: ACM.

Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., & Schork, N. J. (2011). The n-of-1 clinical trial: The ultimate strategy for individualizing medicine? *Personalized Medicine, 8*(2), 161–173.

Liu, H., Philipose, M., Pettersson, M., & Sun, M. T. (2014). Recognizing object manipulation activities using depth and visual cues. *Journal of Visual Communication and Image Representation, 25*(4), 719–726.

Lupton, D. (2014). Self-tracking cultures: Towards a sociology of personal informatics. In *Proceedings of the 26th Australian computer-human interaction conference on designing futures: The future of design* (pp. 77–86). New York: ACM.

Makeyev, O., Lopez-Meyer, P., Schuckers, S., Besio, W., & Sazonov, E. (2012). Automatic food intake detection based on swallowing sounds. *Biomedical Signal Processing and Control, 7*(6), 649–656.

Mariakakis, A., Wang, E., Patel, S., & Wen, J. C. (2016). A smartphone-based system for assessing intraocular pressure. In *2016 38th annual international conference of the IEEE engineering in medicine and biology society* (pp. 4353–4356). Piscataway: IEEE.

Mariakakis, A., Parsi, S., Patel, S. N., & Wobbrock, J. O. (2018). Drunk user interfaces: Determining blood alcohol level through everyday smartphone tasks. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (p. 234:1-234:13). New York: ACM.

Mariakakis, A., Banks, M. A., Phillipi, L., Yu, L., Taylor, J., & Patel, S. N. (2017). *Biliscreen: smartphone-based scleral jaundice monitoring for liver and pancreatic disorders*. Proceedings of the ACM conference on interactive, mobile, wearable and ubiquitous technologies. 1, 2, article 20. 26 pages. https://doi.org/10.1145/3090085

Martínez-Pérez, B., De La Torre-Díez, I., & López-Coronado, M. (2013). Mobile health applications for the most prevalent conditions by the World Health Organization: Review and analysis. *Journal of Medical Internet Research, 15*(6), 1–19.

Michie, S., Abraham, C., Whittington, C., McAteer, J., & Gupta, S. (2009). Effective techniques in healthy eating and physical activity interventions: A meta-regression. *Health Psychology, 28*(6), 690.

Murphy, T. (2013). 40 years after the first cell phone call: Who is inventing tomorrow's future? *IEEE Consumer Electronics Magazine, 2*(4), 44–46.

Nahum-Shani, I., Hekler, E. B., & Spruijt-Metz, D. (2015). Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health Psychology, 34*(S), 1209–1219.

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2016). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine, 52*(6), 446–462.

Nilsen, W., Kumar, S., Shar, A., Varoquiers, C., Wiley, T., Riley, W. T., Pavel, M., & Atienza, A. A. (2012). Advancing the science of mHealth. *Journal of Health Communication, 17*(Sup1), 5–10.

Noronha, J., Hysen, E., Zhang, H., & Gajos, K. Z. (2011). PlateMate: Crowdsourcing nutritional analysis from food photographs. In *Proceedings of the 24th annual ACM symposium on user Interface software and technology (UIST '11)* (pp. 1–12). New York: ACM.

Parate, A., Chiu, M. C., Chadowitz, C., Ganesan, D., & Kalogerakis, E. [2014]. Risq: Recognizing smoking gestures with inertial sensors on a wristband.

Patil, O. R., Gao, Y., Li, B., & Jin, Z. (2017). CamBP: A camera-based, non-contact blood pressure monitor. In *Proceedings of the 2017 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2017 ACM international symposium on wearable computers* (pp. 524–529). ACM.

Patrick, K., Griswold, W. G., Raab, F., & Intille, S. S. (2008). Health and the mobile phone. *American Journal of Preventive Medicine, 35*(2), 177–181.

Phatak, S. S., Freigoun, M. T., Martín, C. A., Rivera, D. E., Korinek, E. V., Adams, M. A., et al. (2018). Modeling individual differences: A case study of the application of system identification for personalizing a physical activity intervention. *Journal of Biomedical Informatics, 79*, 82–97.

Piercy, K. L., Troiano, R. P., Ballard, R. M., Carlson, S. A., Fulton, J. E., Galuska, D. A., et al. (2018). The physical activity guidelines for Americans. *JAMA, 320*(19), 2020–2028.

Poon, A. D., Fagan, L. M., & Shortliffe, E. H. (1996). The PEN-ivory project: Exploring user-interface design for the selection of items from large controlled vocabularies of medicine. *Journal of the American Medical Informatics Association, 3*(2), 168–183.

Quinn, C. C., Clough, S. S., Minor, J. M., Lender, D., Okafor, M. C., & Gruber-Baldini, A. (2008). WellDoc™ mobile diabetes management randomized controlled trial: Change in clinical and behavioral outcomes and patient and physician satisfaction. *Diabetes Technology & Therapeutics, 10*(3), 160–168.

Rabbi, M., Aung, M. H., Zhang, M., & Choudhury, T. (2015). MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 707–718).

Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. C. (2014). Personal tracking as lived informatics. In *Proceedings of the 32nd annual ACM conference on human factors in computing systems* (pp. 1163–1172). New York: ACM.

Rough, D., & Quigley, A. (2015). Jeeves—A visual programming environment for mobile experience sampling. In *2015 IEEE symposium on visual languages and human-centric computing (VL/HCC)* (pp. 121–129). Piscataway: IEEE.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*(3), 493–502.

Sadoughi, F., Behmanesh, A., & Sayfouri, N. (2020). Internet of things in medicine: A systematic mapping study. *Journal of Biomedical Informatics, 103*, 103383.

Siek, K. A., Connelly, K. H., Rogers, Y., Rohwer, P., Lambert, D., & Welch, J. L. (2006). When do we eat? An evaluation of food items input into an electronic food monitoring application. In *2006 pervasive health conference and workshops* (pp. 1–10). Piscataway: IEEE.

Siek, K., Veinot, T., & Mynatt, B. (2019). Research opportunities in sociotechnical interventions for health disparity reduction. Computing Community Consortium; 2019 Jun. Available at https://cra.org/ccc/wp-content/uploads/sites/2/2018/01/17602-CCC-Health-Disparities-ReportFinal.pdf

Shanafelt, T. D., Hasan, O., Dyrbye, L. N., Sinsky, C., Satele, D., Sloan, J., & West, C. P. (2015). Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. In *Mayo Clinic proceedings* (Vol. 90, No. 12, pp. 1600–1613). Elsevier.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32.

Simpson, C. C., & Mazzeo, S. E. (2017). Calorie counting and fitness tracking technology: Associations with eating disorder symptomatology. *Eating Behaviors, 26*, 89–92.

Spruijt-Metz, D., & Nilsen, W. (2014). Dynamic models of behavior for just-in-time adaptive interventions. *IEEE Pervasive Computing, 13*(3), 13–17.

Stone, A., Shiffman, S., Atienza, A., & Nebeling, L. (2007). *The science of real-time data capture: Self-reports in health research*. Oxford: Oxford University Press.

Stowell, E., Lyson, M. C., Saksono, H., Wurth, R. C., Jimison, H., Pavel, M., & Parker, A. G. (2018). Designing and evaluating mHealth interventions for vulnerable populations: A systematic review. In *Proceedings of the 2018 CHI conference on human factors in computing systems. Paper* (Vol. 15, pp. 1–17).

Stratton, R. J., Stubbs, R. J., Hughes, D., King, N., Blundell, J. E., & Elia, M. (1998). Comparison of the traditional paper visual analogue scale questionnaire with an apple Newton electronic appetite rating system (EARS) in free living subjects feeding ad libitum. *European Journal of Clinical Nutrition, 52*(10), 737–741.

Sun, M., Burke, L. E., Mao, Z. H., Chen, Y., Chen, H. C., Bai, Y., Li, Y., Li, C., & Jia, W. (2014). eButton: A wearable computer for health monitoring and personal assistance. In *Proceedings of the 51st annual design automation conference*. San Francisco: 1–5 June 2014.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge: MIT Press.

Truong, K. N., Shihipar, T., & Wigdor, D. J. (2014). Slide to x: Unlocking the potential of smartphone unlocking. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3635–3644). New York: ACM.

Vaizman, Y., Ellis, K., Lanckriet, G., & Weibel, N. (2018). Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 554:1–554:12). New York: ACM.

Wang, E. J., Li, W., Hawkins, D., Gernsheimer, T., Norby-Slycord, C., & Patel, S. N. (2016). HemaApp: Noninvasive blood screening of hemoglobin using smartphone cameras. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing* (pp. 593–604). New York: ACM.

Webb, T., Joseph, J., Yardley, L., & Michie, S. (2010). Using the internet to promote health behavior change: A systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *Journal of Medical Internet Research, 12*(1), e4.

West, P., Giordano, R., Van Kleek, M., & Shadbolt, N. (2016). The quantified patient in the doctor's office. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 3066–3078). New York: ACM.

West, P., Van Kleek, M., Giordano, R., Weal, M., & Shadbolt, N. (2017). Information quality challenges of patient-generated data in clinical practice. *Frontiers in Public Health, 5*, 284.

Xiong, H., Huang, Y., Barnes, L. E., & Gerber, M. S. (2016). Sensus: A cross-platform, general-purpose system for mobile crowdsensing in human-subject studies. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing* (pp. 415–426). New York: ACM.

Zhang, X., Pina, L. R., & Fogarty, J. (2016). Examining unlock journaling with diaries and reminders for in situ self-report in health and wellness. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5658–5664). New York: ACM.

Zhang, Y., & Feng, Z. (2017). A SVM method for continuous blood pressure estimation from a PPG signal. In *Proceedings of the 9th International Conference on Machine Learning and Computing* (pp. 128–132).

Zhu, H., Colgan, J., Reddy, M., & Choe, E. K. (2016). Sharing patient-generated data in clinical practices: An interview study. In *AMIA annual symposium proceedings* (pp. 1303–1312). Bethesda, MD: American Medical Informatics Association.

Telemedicine and Telehealth

Michael F. Chiang, Justin B. Starren, and George Demiris

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_20

Learning Objectives

After reading this chapter you should know the answers to these questions:

- What are the key informatics requirements for successful implementation of telehealth systems?
- What are some key benefits from and barriers to implementation of telehealth systems?
- What are the most promising emerging application domains for telehealth?

20.1 Introduction

Complexity and collaboration characterize health care in the early twenty-first century. Complexity arises from increasing sophistication in the understanding of health and disease, wherein etiological models must acknowledge both molecular processes and physical environments. Collaboration reflects not only inter-professional collaboration, but also a realization that successful attainment of optimal well-being and effective management of disease processes necessitate active engagement of clinicians, lay persons, family members, communities, and society as a whole. This chapter introduces the concepts of **telemedicine** and **telehealth**, and illustrates how advanced networks make possible the collaborations necessary to achieve the full benefits of our growing understanding of health promotion, disease management and rehabilitation. Consider the following situation:

Samuel is a 76-year-old man with coronary artery disease, poorly-controlled Type II diabetes, and high blood pressure. He lives alone in a rural area and does not drive. His daughter lives further away but visits occasionally. One of his neighbors visits regularly to check on him and assist with various errands. In the past, Samuel has been unable to keep medical appointments consistently because of difficulty arranging transportation. He had a recent acute hyperglycemic episode that required hospitalization. After 4 days he is medically stable and ready for discharge. He is able to measure his blood glucose and can safely administer the appropriate dose of insulin. The nurse notes that Samuel sometimes has trouble calibrating his insulin dose to the blood glucose reading.

20.1.1 Telemedicine and Telehealth to Reduce the Distance Between the Consumer and the Health Care System

Historically, health care has usually involved travel. Either the health care provider traveled to visit the patient, or more recently, the patient traveled to visit the provider. Patients with diabetes, like Samuel whom we will be discussing later and throughout the chapter, typically meet with their physician every 2–6 months to review data and plan therapy changes. Travel has costs, both directly, in terms of gasoline or transportation tickets, and indirectly, in terms of travel time, delayed treatment, and lost productivity. In fact, travel has accounted for a significant proportion of the total cost of health care (Starr 1982). Because of this, both patients and providers have been quick to recognize that rapid electronic communications have the potential to improve care by reducing the costs and delays associated with travel. This has involved both access to information resources, as well as communication among various participants, including patients, family members, primary care providers and specialists whether it is synchronous communication (where all stakeholders interact at the same time) or asynchronous (where information is exchanged with a time lag).

As is the case with informatics, the formal definitions of telemedicine and telehealth tend to be very broad. **Telemedicine** involves the use of modern information technology, especially two-way interactive audio/video communications, computers and telemetry to deliver health services to remote patients and to facilitate information exchange between primary care physicians and specialists at some distance from each other (Bashshur et al. 2009). **Telehealth** is a somewhat newer and broader term referring to remote health care that includes clinical and social services provided

using telemedicine, as well as interactions with automated systems or information resources. Because of its broader scope, we are using the term telehealth in this chapter.

As is the case with biomedical informatics, there are many different sub-domains within telehealth. For nearly every clinical domain, there is a "tele-X" or "X telehealth", where X is the clinical domain. Examples include: **Teleradiology** (see ▶ Sect. 20.3.4); **Teleophthalmology** (see ▶ Sect. 20.3.4); **Telepsychiatry** (see ▶ Sect. 20.3.5); and, **home telehealth** (see ▶ Sect. 20.3.5). Some sub-domains do not fit neatly into this naming paradigm. **Correctional Telehealth** (see ▶ Sect. 20.3.5) refers to the location of the patient in a prison. It is discussed separately because of the unique business model, and the fact that it represents an early and sustained success. **Remote Intensive Care** (see ▶ Sect. 20.3.5) is the term used to describe the use of telehealth technologies in an ICU setting. **Teleconsultation** is a general term describing the use of telehealth technologies to support discussions between clinicians, or between a clinician and a patient. The archetypal teleconsultation occurs when the patient and the generalist clinician are in a rural or remote location and a specialist is at a distant tertiary referral facility. **Telepresence** (see ▶ Sect. 20.3.6) refers to high-speed, multi-modality telehealth interactions, such as **Telesurgery**, that gives the feeling of "being there". In this chapter we will review how some of these sub-domains may play a role in supporting Samuel manage their health care needs more effectively.

It is clear from the definition above that there is considerable overlap between telehealth and biomedical informatics. In fact, one will frequently find papers on telehealth systems presented at biomedical informatics conferences and presentations on informatics at telehealth and telemedicine meetings. Some groups, especially in Europe, have adopted the rubric **health information and communication technology** (HICT). The major distinction is one of emphasis. Telehealth and telemedicine emphasize the notion of distance, especially the provision of care to remote or isolated patients and communities. In contrast, biomedical informatics emphasizes methods for handling the information moving between the participants, irrespective of the distance between patient and provider.

Consumer health informatics (CHI), also called **personal health informatics** (**PHI**), is a related domain that bridges the distance between patients and health care resources, and that typically emphasizes interactions with computer-based information such as websites or information resources. Collectively, CHI and telehealth deliver health care knowledge and expertise to where they are needed, and are ways to involve the patient as an active partner in care. Despite their similarities, CHI and telehealth come from very different historical foundations. Telehealth derived from traditional patient care, while CHI derived from the self-help movements of the 1970's. Largely owing to this historical separation, practitioners and researchers in the two fields tend to come from different backgrounds. For these reasons, we are presenting CHI and telehealth as two distinct, but closely related domains (see ▶ Chap. 11 for more information on personal health informatics).

20.2 Historical Perspectives

The use of communication technology to convey health-related information at a distance is nothing new. The earliest known example may be the use of so-called "leper bells" carried by individuals during Roman times. Sailing ships would fly a yellow flag to indicate a ship was under quarantine and awaiting clearance by a doctor, or a yellow and black "plague flag" to indicate that infected individuals were on board. By some accounts, when Alexander Graham Bell said "Mr. Watson. Come here. I need you" in 1876, it was because he had spilled acid on his hand and needed medical assistance. In 1879, only 3 years later, the first description of telephone use for clinical diagnosis appeared in a medical journal (Practice by Telephone 1879).

20.2.1 Early Experiences

One of the earliest and most long-lived telehealth projects is the Australian Royal Flying Doctor Service (RFDS), founded in 1928 and continuing to this day. In addition to providing air ambulance services, the RFDS provides telehealth consultations. These consultations first used Morse Code, and later voice, leveraging radio communications to the remote sheep stations in the Australian outback. Lay people played a significant role here, clearly communicating their concerns and clinical findings to the RFDS and carefully carrying out instructions while awaiting, if necessary, the arrival of the physician. The RFDS is most famous for its standardized medical supply chest, introduced in 1942. The chest contains diagnostic charts and medications, identified only by number. This allowed the consulting clinician to localize symptoms by number and then prescribe care, such as "take one number five and two number fours."

Modern telehealth can be traced to 1948 when the first transmission of a radiograph over a phone line was reported. Video-based telehealth can be traced to 1955 when the Nebraska Psychiatric Institute began experimenting with a closed-circuit video network on its campus. In 1964 this was extended to a remote state mental health facility to support education and **teleconsultation**. In 1967, Massachusetts General Hospital (MGH) was linked to Logan International Airport via a microwave audio-video link (Bird 1972; Murphy et al. 1973). In 1971 the National Library of Medicine began the Alaska Satellite Biomedical Demonstration project linking 26 remote Alaskan villages utilizing NASA satellites (Hudson and Parker 1973).

The period from the mid 1970's to the late 1980's was a time of much experimentation, but few fundamental changes in telehealth. A variety of pilot projects demonstrated the feasibility and utility of video-based telehealth. The military funded a number of research projects aimed at developing tools for providing telehealth care on the battlefield. The early 1990's saw several important advances. Military applications developed during the previous decades began to be deployed. Military **teleradiology** was first deployed in 1991 during Operation Desert Storm. Telehealth in military field hospitals was first deployed in 1993 in Bosnia. Several states, including Georgia, Kansas, North Carolina and Iowa implemented statewide telehealth networks. Some of these were pure video networks, based on broadcast television technology. Others were built using evolving Internet technology. During this same period, correctional telehealth (see ▶ Sect. 20.3.5) became much more common. For example, in 1992 East Carolina University contracted with the largest maximum-security prison in North Carolina to provide telehealth consultation.

Telehealth projects in the early 1990s continued to be plagued by two problems that had hampered telehealth since its inception: high cost and poor image quality. Both hardware and **high-bandwidth** connections were prohibitively expensive. A single telehealth station typically cost over $50,000 and connectivity could cost thousands of dollars per month. Most programs were dependent on external grant funding for survival. Even with this, image resolution was frequently poor and **motion artifacts** were severe.

The Internet revolution that began in the late 1990s drove fundamental change in telehealth. Advances in computing power both improved image quality and reduced hardware costs to the point that, by 2000, comparable systems cost less than a tenth of what they had a decade earlier. Improvements in **image compression** made it possible to transmit low-resolution, full-motion video over standard telephone lines, enabling the growth of telehome care. With the increasing popularity of the World Wide Web, high-bandwidth connections became both more available and less expensive. Many telehealth applications that had relied on expensive, dedicated, point-to-point connections were converted to utilize commodity Internet connections. The availability of affordable hardware and connectivity also made access to health-related electronic resources from the home, school

or work place possible and fueled the growth of consumer health information. In 2020, the COVID-19 pandemic highlighted the potential of telehealth in facilitating essential health care services and led to an expedited adoption of telehealth services across many health systems worldwide (Jain et al. 2020).

20.2.2 Recent Advances in Medical-Grade Broadband Technology

As telemedicine applications are being increasingly used in critical medical situations such as emergency care and remote surgery applications, quality of service (QOS) becomes extremely important. It is important to note that optimally provisioning a network for medical-grade QOS does not simply imply that the network will provide "quality" in the sense of reliability, consistency and bandwidth performance, although these characteristics are certainly important requirements. Any network, no matter the bandwidth available, can become congested – overwhelmed with the volume of traffic to the extent that sessions are interrupted and data lost. Bandwidth availability limitations are particularly prevalent in rural locations where high-capacity circuits may be unavailable or prohibitively expensive.

Newer network routing technologies such as **multiprotocol label switching** (MPLS) can, in addition to providing superior network throughput performance, permit explicit prioritization of clinical traffic while simultaneously providing access to lower priority administrative and other non-clinical traffic. The individual data packets of high priority traffic (e.g., telehealth or patient monitoring sessions) are "tagged" with a numerical priority flag. As the QOS-tagged packets traverse the network, each routing/switching device recognizes the priority tag and preferentially processes and forwards the packets. This explicit QOS combined with advanced security and privacy features within a broadband network has been characterized as "Medical Grade" broadband.

20.3 Bridging Distance with Informatics: Real-World Systems

There are many ways to categorize telehealth resources, including classifications based on participants, bandwidth, information transmitted, medical specialty, immediacy, health care condition, and financial reimbursement. The categorization in ◘ Table 20.1 is based loosely on bandwidth and overall complexity. This categorization was chosen because each category presents different challenges for informatics researchers and practitioners.

A second categorization of telehealth systems that overlaps the previous one is the separation into synchronous (or real-time) and asynchronous (or **store-and-forward** systems). Video conferencing is the archetypal synchronous telehealth application. Synchronous telehealth encounters are analogous to conventional office visits. Telephony, chat-groups, and **telepresence** (see ▶ Sect. 20.3.6) are also examples of synchronous telehealth. A major challenge in all synchronous telehealth is scheduling. All participants must be at the necessary equipment at the same time.

Store-and-forward, as the name implies, involves the preparation of a dataset at one site that is sent asynchronously to a remote recipient. Remote interpretation, especially teleradiology, is the archetypal example of store-and-forward telehealth. Images are obtained at one site and then sent, sometimes over very low bandwidth connections, to another site where the domain expert interprets them. Other examples of store-and-forward include access to Web sites, e-mail and text messaging. Some store-and-forward systems support the creation of multimedia "cases" that contain multiple clinical data types, including text, scanned images, wave forms and videos.

20.3.1 The Forgotten Telephone

Until recently, the telephone was a forgotten component in telehealth. The field of telemedicine and telehealth focused on video

Table 20.1 Categories of telehealth and consumer health informatics

Telehealth category	Bandwidth	Applications
Information resources	Low to moderate	Web-based information resources, patient access to electronic medical records
Messaging	Low	E-mail, chat groups, consumer health networks, personal clinical electronic communications (PCEC)
Telephone	Low	Scheduling, triage
Remote monitoring	Low to moderate	Remote monitoring of pacemakers, diabetes, asthma, hypertension, Congestive Health Failure (CHF).
Remote interpretation	Moderate	PACS, remote interpretation of radiographic studies and other images, such as dermatologic and retinal photographs.
Videoconferencing	Low to high	Wide range of applications, from telehome care to telementoring and telepsychiatry
Telepresence	High	Remote surgery, telerobotics

and largely ignored the audio-only telehealth. This is paradoxical given that up to 25% of all primary care encounters occur via the telephone. These include triage, case management, results review, consultation, medication adjustment and logistical issues, like scheduling. In part, this can be traced to the fact that telephone consultations are not reimbursed by most insurance carriers.

More recently, increased interest in cost control through case management has driven renewed interest in use of audio-only communication between patients and providers. Multiple articles have appeared on the value of telephone follow-up for chronic conditions (Downes et al. 2017; Jayakody et al. 2016). Several managed care companies have set up large telephone triage centers. The National Health Service in the UK is investing £123 million per year in NHS Direct, a nation-wide telephone information and triage system that handles 27,000 calls per day.

20.3.2 Electronic Messaging

Electronic text-based messaging has emerged as a popular mode of communication between patients and providers. It began with patients sending conventional e-mails to physicians. The popularity of this grew so rapidly that national guidelines were developed (Kane and Sands 1998). However, e-mail has a number of disadvantages for health-related messaging: delivery is not guaranteed; privacy and security are problematic; e-mail is transient (there was no automatic logging or audit trail); and the messages are completely unstructured.

To address these limitations, a variety of Web-based messaging solutions, called **personal clinical electronic communications**, have been developed (Sarkar and Starren 2002). Because the messages never leave the Web site, many of the problems associated with conventional e-mail are avoided. Web-based messaging is a standard feature of **patient portals** see ▶ Chap. 11) associated with many EHRs. The inclusion of messaging as a Meaningful Use requirement for Certified EHRs significantly increased the use of web-based messaging to provide telehealth (Fig. 20.1).

20.3.3 Remote Monitoring

Remote monitoring is a subset of telehealth focusing on the capture of clinically relevant data in the patients' homes or other locations outside of conventional hospitals, clinics or health care provider offices, and the subsequent transmission of the data to central locations for review. The conceptual model

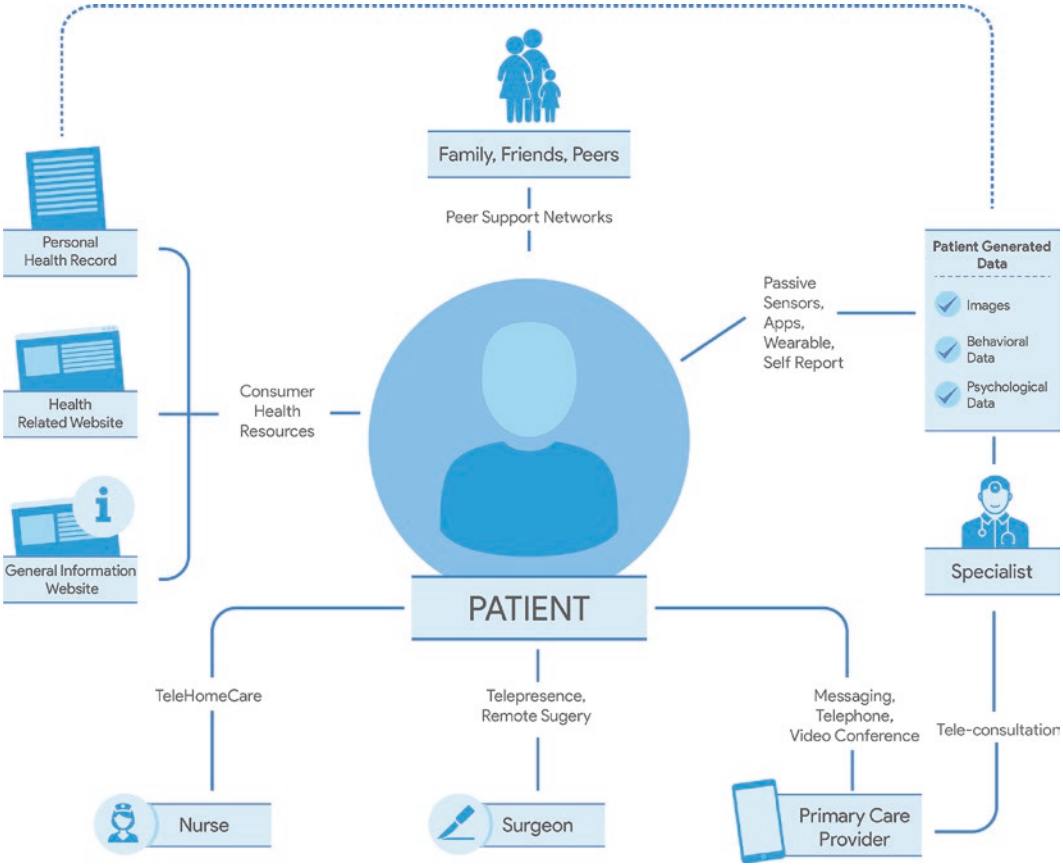


Fig. 20.1 Connections. The figure shows different ways that electronic communications can be used to link patients with various health resources. Only connections directly involving the patient are shown (e.g., use of the EHR by the clinician is not shown). Patient generated data are created via passive monitoring (such as home-based sensors) or active monitoring (such as wearables or other active monitoring devices) in the patient's home or other community settings. Other resources, such as remote surgery or imaging, would require the patient to go to a telehealth-equipped clinical facility

underlying nearly all remote monitoring is that clinically significant changes in patient condition occur between regularly scheduled visits and that these changes can be detected by measuring physiologic parameters.

The care model presumes that, if these changes are detected and treated sooner, the overall condition of the patient will be improved. An important distinction between remote monitoring and many conventional forms of telemedicine is that remote monitoring focuses on management, rather than on diagnosis. Typically, remote monitoring involves patients who have already been diagnosed with a chronic disease or condition. Remote monitoring is used to track parameters that guide management. Any measurable parameter is a candidate for remote monitoring. The collected data may include continuous data streams or, more commonly, discrete measurements.

Another important feature of most remote monitoring is that the measurement of the parameter and the transmission of the data are typically separate events. The measurement devices have a memory that can store multiple measurements. The patient will send the data to the caregiver in one of several ways. For many studies, the patient will log onto a server at the central site (either over the Web or by direct dial-up) and then type in the data. Alternately, the patient may connect the

measurement device to a personal computer or specialized modem and transfer the readings electronically.

More recently, a variety of monitoring devices have been developed that either connect directly to mobile telephones or transmit the data to the mobile phone using Bluetooth wireless. The mobile phone then transmits the data to a provider for review. A major advantage of direct electronic transfer is that it eliminates problems stemming from manual entry, including falsification, number preference and transcription errors. The role of mobile telephones in providing health services has grown so rapidly that the term **mobile health** (or "mHealth") has been coined. The term appeared first, one time, in 2004 in PubMed. See ▶ Chap. 19 for more on mHealth systems. Additionally, the emergence of interconnected sensors and devices referred to as **Internet of Things (IoT)** described later in the chapter, have the potential to contribute to remote monitoring systems.

Any condition that is evaluated by measuring a physiologic parameter is a candidate for remote monitoring. The parameter most measured in the remote setting is blood glucose for monitoring diabetes. A wide variety of research projects and commercial systems have been developed to monitor patients with diabetes. Patients with asthma can be monitored with peak-flow or full-loop **spirometers**. Patients with hypertension can be monitored with automated blood pressure cuffs. Patients with congestive heart failure (CHF) are monitored by measuring daily weights to detect fluid gain. Remote monitoring of pacemaker function has been available for a number of years and has recently been approved for reimbursement. Home coagulation meters have been developed that allows the monitoring of patients on chronic anticoagulation therapy. See a discussion on remote intensive care later in the chapter and also ▶ Chap. 21 for more on patient monitoring systems.

Several factors limit the widespread use of remote monitoring. First is the question of efficacy. While these systems have proven acceptable to patients and beneficial in small studies, few large-scale controlled trials have been done. Second is the basic question of who will review the data. Research studies have utilized specially trained nurses at centralized offices, but it is not clear that this will scale up. Third is money—for most conditions, remote monitoring is still not a reimbursed activity.

20.3.4 Remote Interpretation

Although Samuel was diagnosed with Type II diabetes over 20 years ago and realizes that visual loss can be a serious complication, he has only rarely received dilated eye exams for retinopathy screening. There is no eye doctor conveniently located near his home, and he feels that the appointments are always too long and that he has no problems such as blurred vision. However, his primary care doctor has recently implemented a new retinal screening machine in the office. During a routine medical examination, Samuel receives a retinal photograph from an office technician that is then interpreted by a remote ophthalmologist. Samuel is told that he has high-risk diabetic retinopathy that requires treatment to prevent visual loss. He is emergently referred to an ophthalmologist, who performs a successful laser procedure to treat the diabetic retinopathy.

Remote interpretation is a category of store-and-forward telehealth that involves the capture of images, or other data, at one site and their transmission to another site for interpretation. This may include radiographs (*teleradiology*), photographs (*teledermatology*, *teleophthalmology*, *telepathology*), wave forms such as ECGs (e.g. *telecardiology*), and text-based medical data.

The store-and-forward telehealth modalities have benefited most from the development of the **commodity Internet** and the increasing availability of affordable high bandwidth connections that it provides. The shared commodity Internet provides relatively high bandwidth, but the available bandwidth is continuously varying. This makes it much better suited for the transfer of text-based data and image files, rather than for streaming data or video connections. Although image files are often tens or hundreds of megabytes

in size, the files are typically transferred to the interpretation site and cached there for later interpretation. From a logistical perspective, multiple remote interpretations may be batched and performed together, thereby providing important workflow and convenience advantages over traditional medical examinations or real-time video telehealth paradigms.

Teleradiology

By far, teleradiology is the largest category of remote interpretation, and probably the largest category of telehealth. Teleradiology (along with telepathology) represents the most mature clinical domain in telehealth. With the deployment of **picture archiving and communications systems** (PACS) that capture, store, transmit and displays digital radiology images, the line between teleradiology and conventional radiology is blurring. In fact, routine medical care in radiology and pathology is increasingly being delivered primarily through "telehealth" strategies (Radiology image management is discussed in more detail in ▶ Chap. 22).

Many factors have contributed to the more rapid adoption of telehealth in domains such as radiology and pathology. One important factor is the relationship between these specialists and their patients. In both domains, the professional role is often limited to the interpretation of images, and the specialist rarely interacts directly with the patient. To patients, there is therefore little perceived difference between a radiologist in the next building and one in the next state.

An important factor driving the growth of teleradiology is that it is reimbursable by insurance payers. Because image interpretation does not involve direct patient contact, few payers make any distinction about where the interpretation occurred. Rapid dissemination of teleradiology systems has also been supported by widespread adoption of vendor-neutral image storage and transmission standards such as **Digital Imaging and Communication in Medicine** (DICOM; discussed in more detail in ▶ Chaps. 7 and 22). Finally, numerous evaluation studies have demonstrated that digital image interpretation by through teleradiology has comparable, or potentially even better, accuracy and efficiency compared to traditional film-based radiological examination (Franken et al. 1992; Mackinnon et al. 2008; Reiner et al. 2002).

Teleophthalmology

Another area of remote interpretation that is growing rapidly is teleophthalmology, particularly for retinal disease screening. As one example, diabetic retinopathy (retinal disease) is a leading cause of blindness that can be treated if detected early. However, it has been found that nearly 50% of diabetics are non-compliant with guidelines recommending annual screening eye examinations (Brechner et al. 1993). Systems have been developed that allow nurses or technicians in primary care offices to obtain high quality digital retinal photographs. These images are sent to regional centers for interpretation. If diabetic retinopathy is identified or suspected, the patient is referred for full ophthalmologic examination.

Large-scale operational systems have been implemented by the Veterans Health Administration and by other institutions, particularly in areas with limited accessibility to eye care specialists (Cavalleranno et al. 2005; Cuadros and Bresnick 2009). In fact, remote interpretation of retinal images by certified reading centers, when taken after dilation of the eyes using standard photographic protocols originally developed for clinical research trials, has been demonstrated to classify diabetic retinopathy more accurately than traditional dilated eye examination. This is likely because retinal abnormalities found on photographs may be reviewed in more detail than what is generally feasible during traditional eye examinations.

Another application of teleophthalmology is in retinopathy of prematurity (ROP), a leading cause of blindness in premature infants, whereby hospitalized infants are examined regularly to identify treatment-requiring disease. However, these examinations are logistically difficult and time consuming, and the number of ophthalmologists willing to perform them has decreased. As a result, systems have been developed in which trained nurses capture retinal photographs and transmit

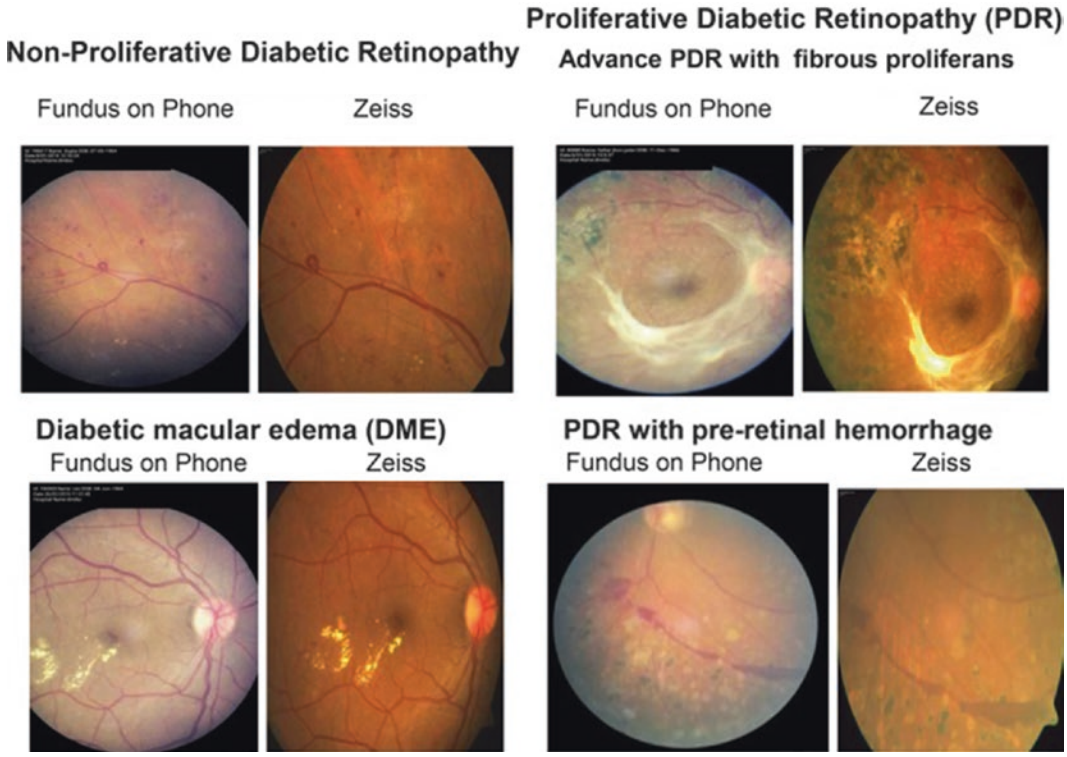


Fig. 20.2 Retinal images of diabetic retinopathy obtained via a fundus on phone (FOP) smartphone system compared to a professional Zeiss camera traditionally used by optometrists. (Source: Rajalakshmi et al. 2015) Creative Common Attribution

them to experts for remote interpretation (Richter et al. 2009). The proliferation of smartphones has introduced additional ways to promote teleopthamology, using a "fundus on phone" (FOP) camera to facilitate a smartphone-based cost-effective retinal imaging system (Fig. 20.2).

20.3.5 Video-Based Telehealth

To many people telehealth is videoconferencing. Whenever the words "telehealth" or "telemedicine" are mentioned, most people have a mental image of a patient talking to a doctor over some type of synchronous video connection. Indeed, most early telehealth research did focus on synchronous video connections. For many of the early studies, the goal was to provide access to specialists in remote or rural areas. Nearly all of the early systems utilized a hub-and-spoke topology where one hub, usually an academic medical center, was connected to many spokes, usually rural clinics.

Many of the early telehealth consults involved the patient and the primary care provider at one site conferring with a specialist at another site. Most of the state-wide telehealth networks operated on this model. This was so engrained in the telehealth culture, that the first legislation allowing Medicare reimbursement of telehealth consults required a "presenter" at the remote site.

This requirement for a "presenter" exacerbated the scheduling problem. Because synchronous video telehealth often uses specialized videoconferencing rooms, the televisits need to be scheduled at a specific time. Getting the patient and both clinicians (expert and presenter) at the right places at the right time has forced many telehealth programs to hire a full-time scheduler. The scheduling problem, combined with the advent of more user-friendly equipment, ultimately led

Medicare to drop the presenter requirement. Even so, scheduling is often the single biggest obstacle to greater use of synchronous video consultations.

A second obstacle has been the availability of relevant clinical information. Because of the inability to interface between various EHRs, it was not unusual for staff to print out results from the EHR at one site and then to fax those to the other site prior to a synchronous video consultation.

Unlike store-and-forward telehealth, synchronous video requires a stable data stream. Although video connection can use conventional phone lines (commonly referred to as **plain old telephone service**, or POTS) that provide 64 bits-per-second (64 kbs) transmission speed, diagnostic quality video typically requires at least 128 Kbs and more commonly 384 Kbs. In order to guarantee stable data rates, synchronous video in clinically critical situations still relies heavily on dedicated circuits, either **Integrated Service Digital Network** (ISDN) connections or leased lines. Within single organizations, or in consultative or educational settings, Internet Protocol (IP) based video conferencing has become the dominant modality. While POTS-based telehealth systems were common in 1990s and even early 2000s, the diffusion of high-speed Internet has led to a much wider adoption of IP based videoconferencing. The anticipated growth of **5G (fifth generation wireless systems**) facilitating far higher speeds and connections with massive capacity and low latency for consumer devices, is expected to accelerate the use of telehealth for a broad spectrum of applications and target populations.

Synchronous video telehealth has been used in almost every conceivable situation. In addition to traditional consultations, the systems have been used to transmit grand rounds and other educational presentations. Video cameras have been placed in operating rooms at hub sites to transmit images of surgeries for educational purposes. Video cameras have been placed in emergency departments and operating rooms at spoke sites to allow experts to "telementor" less experienced physicians in the remote location. Video cameras have also been placed in ambulances to provide remote triage.

More recently, the growing popularity of mobile devices is creating potential for new strategies involving real-time video communication between patients and health care providers. This is especially promising because mobile networks are low-cost and widely-available for consumers, and because they are increasingly accessible even in developing countries. However, health information exchange using mobile networks raises concerns about privacy, security, and compliance with **Health Insurance Portability and Accountability Act** (HIPAA). With appropriate encryption settings, wireless video communication using mobile device applications may be HIPAA-compliant (e.g. FaceTime; Apple Computer, Cupertino, CA). There are already various commercially available solutions that allow patients to download smartphone apps to access clinicians. Some of these apps use chatbot technology to screen symptoms before matching patients with clinicians who can communicate with text, images and videos and can e-prescribe to local pharmacies. In the future, these mobile technologies may provide additional opportunities for increased communication between patients and providers.

Prior to the adoption of IP-based videoconferencing, programs that begun with grant funding ended soon after the grant funding ended. Even after the advent of IP-based conferencing, many programs continued to struggle. This was in spite of the fact that Medicare had begun reimbursing for synchronous video under limited circumstances. The COVID-19 pandemic introduced short-term policy changes and led to an accelerated growth of telehealth as we discuss later (see ► Sect. 20.3.8).

Some rural health care providers, such as the Marshfield Clinic in Wisconsin, have integrated synchronous telehealth into their standard care model to provide routine specialist services to outlying location. Some categories of synchronous video telehealth have developed sustainable models: telepsychiatry, correctional telehealth; home telehealth, emergency telehealth, and remote intensive care.

Telepsychiatry

In many ways, psychiatry is the ideal clinical domain for synchronous video consultation. Diagnosis is based primarily on observing and talking to the patient. The interactive nature of the dialog means that store-and-forward video is rarely adequate. Physical examination is relatively unimportant, so that the lack of physical contact is not limiting. There are very few diagnostic studies or procedures, so that interfacing to other clinical systems is less important. In addition, state offices of mental health deliver a significant fraction of psychiatric services, minimizing reimbursement issues. This is illustrated by two projects. In 1995, the South Carolina Department of Mental Health established a telepsychiatry network to allow a single clinician to provide psychiatric services to deaf patients throughout the state (Afrin and Critchfield 1997). The system allowed clinicians, who had previously driven all over the state, to spend more time in patient care and less time traveling.

The system was so successful that it was expanded to multiple providers and roughly 20 sites. The second example comes from the New York State Psychiatric Institute (NYSPI), which is responsible for providing expert consultation to mental health facilities and prisons throughout the state. As in South Carolina, travel time was a significant factor in providing this service. To address the problem, the NYSPI created a videoconference network among the various state mental health centers. The system allows specialists at NYSPI in New York City to provide consultations in a timelier manner, improving care and increasing satisfaction at the remote sites.

Correctional Telehealth

Prisons tend to be located far from major metropolitan centers. Consequently, they are also located far from the specialists in major medical centers. Transporting prisoners to medical centers is an expensive proposition, typically requiring two officers and a vehicle. Depending on the prisoner and the distance, costs for a single transfer range from hundreds to thousands of dollars. Because of the high cost of transportation, correctional telehealth was economically viable even before the advent of newer low cost systems.

Correctional telehealth also improves patient satisfaction. A fact surprising to many is that inmates typically do not want to leave a correctional institution to seek medical care. Many perceive it as stigmatizing to navigate a medical facility in prison garb. In addition, the social structure of prisons is such that any prisoner who leaves for more than a day risks losing privileges and social standing. Correctional telehealth follows the conventional model of providing specialist consultation to supplement to on-site primary care physicians. This has become increasingly important with the rising prevalence of AIDS in the prison population.

Home Telehealth

After Samuel misses two scheduled visits, the Diabetes Educator calls see what the matter is. Samuel explains that it is a 1-h drive from his home to the diabetes center, that his daughter had trouble taking time off from work to drive him, and that he would have difficulty leaving his wife home alone because she has been ill recently. The Diabetes Educator notes that Samuel lives in a rural area and is eligible to receive educational services via telehealth. She signs Samuel up to receive a Home Telehealth Unit and schedules delivery. The unit is initially difficult for him to use because he is not familiar with computer systems. However, after this initial learning process, Samuel rarely misses a video education session. At one visit, Samuel complains that his daughter who lives further away, is always "on his case" about his injections. The nurse schedules the next video visit during an evening when Samuel's daughter can join the video-call. She also schedules Samuel to have a video visit with the dietician.

Somewhat paradoxically, one of the most active areas of telehealth growth is at the lowest end of the bandwidth spectrum—telehealth activities into patients' homes. In the late 1990s, many believed that home broadband access would soon become ubiquitous and a number of vendors abandoned POTS-based systems in favor of IP-based video

solutions. The broadband revolution was slower than expected, especially in rural and economically depressed areas most in need of home telehealth services. A few research projects paid to have broadband or ISDN installed in patients' homes. In response to this, the American Telemedicine Association released new guidelines for Home Telehealth in 2002 in which synchronous video was provided over POTS connections. However, more recently high speed Internet and wireless networks have significantly expanded coverage in the US and abroad leading to a growth of high speed Internet based video delivery products. In addition to video, home telehealth systems typically have data ports for connection of various peripheral devices, such as a digital stethoscope, glucose meter, blood pressure meter, or spirometer or allow for Bluetooth connection.

Home telehealth can be divided into two major categories. The first category, often called **telehome care**, is the telehealth equivalent of home nursing care. It involves frequent video visits between nurses and, often homebound, patients. With the advent of prospective payment for home nursing care, telehome care is viewed as a way for home care agencies to provide care at reduced costs and potentially lead to a reduction of rehospitalization for home care patients with complex care needs. As with home nursing care, telehome care tends to have a finite duration, often focused on recovery from a specific disease or incident. Several studies have shown that telehome care can be especially valuable in the management of patients recently discharged from the hospital and can significantly reduce readmission rates.

The second category of home telehealth centers on the **management of chronic diseases**. Compared with telehome care, this type of home telehealth frequently involves a longer duration of care and less frequent interactions. Video interactions tend to focus on patient education, more than on evaluation of acute conditions. An important distinction between telehome care and disease management telehealth is that interactions in the former are initiated and managed by the nurse. Measurements, such as blood pressure, are typically collected during the video visit and uploaded as part of the video connection. For disease management, the system also needs to support remote monitoring, patient-initiated data uploads and, possibly, Web-based access to educational or disease management resources.

One of the earliest, and the first large-scale project to examine the value of telemedicine systematically in the home setting, was the Informatics for Diabetes Education and Telemedicine (IDEATel) project (Starren et al. 2002). Started in 2000, the IDEATel project was an 8-year, $60 million demonstration project funded by the Center for Medicare and Medicaid Services (CMS) involving 1665 diabetic Medicare patients in urban and rural New York State. In this randomized clinical trial, half of the patients received Home Telemedicine Units (HTU), and half continued to receive standard care. In addition to video, the HTU allowed patients to interact in multiple ways with their online charts. When patients measured blood pressure or fingerstick glucose, the encrypted results were transmitted to the Columbia University Web-based Clinical Information System (WebCIS; Hripcsak et al. 1999) at New York Presbyterian Hospital (NYPH). Nurse case managers monitored patients by reviewing the generated data and potential alerts, and providing consultation to patients.

In 2012 Steveton et al. (2012) published findings from one of the largest home telehealth randomized clinical trials to date. The trial was conducted in the UK and involved 179 general practices and 3230 people with diabetes, chronic obstructive pulmonary disease or heart failure who were randomly assigned to either usual care or the telehalth group that also received a set top box connected to their television capturing symptom questions and educational messages and various peripheral devices such as pulse oximeters, glucometers and digital weight scale for capturing and transmitting vital signs. The study demonstrated that home telehealth was associated with lower mortality and emergency department (ED) admission rates. That same year findings from another clinical trial (Takahashi et al. 2012) in the US revealed

different trends. In this study, 205 participants were randomly assigned to a telemonitoring group (including video, peripheral devices for vital signs and symptom reporting) or to usual care. No significant differences in hospitalizations and ED visits were found between the two groups; mortality however was higher in the telemonitoring group. This study did focus on frail older adults (with an average age over 80 years) and followed a different design and analytic approach.

The advancement of sensor technologies has led to the concept of "smart homes", namely residential settings with embedded passive monitoring technologies to facilitate monitoring of residents with the goal to maximize their well-being and safety (Demiris and Hensel 2008). Passive monitoring tools utilize sensors to facilitate functional, safety or physiological monitoring or cognitive support or sensory aids, to monitor security or address social isolation. Examples include a bed sensor that detects restlessness at night or sleep interruptions, motion sensors that capture overall activity levels in the home, sedentary behaviors or bathroom visits, door sensors that measure time spent inside or number of visitors, gait sensors to assess gait characteristics and changes over time as well as fall risk (Liu et al. 2016; Reeder et al. 2013). Insight into behavioral health and activity levels along with more traditional home telehealth data sets such as vital signs and symptom reporting provide a more comprehensive assessment of one's well-being (including not only the physiological but also the physical, social, mental and cognitive aspects of wellness (Dawadi et al. 2016)) introducing a new era for home telehealth. The rise of the Internet of Things (IoT), namely the diffusion of networks of devices, appliances and sensors that are interconnected and enable different passive monitoring components to exchange data and be remotely controlled (Lhotska et al. 2018), allows not only for monitoring of behavioral data but also for providing tailored responses to these observations (for example, adjusting lighting if unstable gait is detected or allowing a clinician to remotely adjust environmental parameters). These concepts are still emerging and technical, clinical and ethical implications have not been fully examined but it is certain that the future of home telehealth will encompass new data sets and tools, and expanded roles and responsibilities for clinicians, patients and families.

Emergency Telemedicine

Samuel develops slurred speech and weakness on the right side of his body. His daughter, who happens to be with him at the time, calls 911. The ambulance crew notifies the emergency room that they are in route with a possible stroke victim. On arrival, the rural emergency department (ED) physician does a quick evaluation and connects via telemedicine with a stroke neurologist at an academic health center. The neurologist talks with the Samuel and his daughter, and participates in the examination with the ED physician. Following laboratory work and a CT negative for hemorrhage, the ED physician again consults with the neurologist who confirms the diagnosis of ischemic stroke and institutes thrombolytic therapy via pre-arranged protocol. Samuel is transferred to the intensive care unit for close monitoring of his diabetes, hypertension, and evolving stroke.

"Just in time" consultation in the emergency setting potentially represents one of the most beneficial uses of telehealth. Emergency telemedicine has been used in a variety of ways and has demonstrated significant benefits, including in such area as tele-trauma care, burn care, and critical care pediatric specialists consulting on critically ill or injured children (Heath et al. 2009; Ricci et al. 2003; Saffle et al. 2009). Telehealth in the emergency setting is likely to have the greatest benefit when time-limited critical decision making by a specialist physician regarding a specific intervention is necessary.

An important and increasingly frequently used application demonstrating this is in the evaluation and treatment of the stroke patient. Best practice management of ischemic stroke in appropriate patients now includes the use of thrombolytic therapy such as tissue plasminogen activator (tPA), which has been shown to have statistically significant clinical and financial benefits. Recommendations and drug labeling limit the use of intravenous tPA

to within 3 h of when the patient was last seen as well or had witnessed onset of symptoms.

This therapy, however, has significant complications, particularly in patients with hemorrhagic rather than ischemic events – requiring urgent specialty consultation, along with rapid expert interpretation of imaging and laboratory work. Many settings lack the specialty expertise to have on-site "stroke teams" to accomplish best practice. Telemedicine can bring specialty expertise to a remote location for emergency evaluation of the patient directly, while transmit images and laboratory work for immediate interpretation.

This model of care, first called "telestroke" care by Levine and Gorman, has been increasingly used throughout the country (Levine 1999). The efficacy of this model, compared to traditional telephone consultation, was evaluated by Meyer et al. (2008). These investigators found that telestroke care resulted in more accurate decision making than did telephone consultation. Based on a comprehensive review of evidence, the American Heart Association and American Stroke Association concluded that "evidence supporting the equivalence of telestroke to in-person care is accumulating (Wechsler et al. 2017)". In their report, they review models of telestroke and provide suggestions for standardizing and adopting quality measures (highlighting among others the responsibility for collecting quality data as a core component of the agreement between telestroke sites and a coordinating stroke center or distributed partner), and recommendations for licensing, credentialing, training and documentation.

Remote Intensive Care

Samuel was admitted to the intensive care unit (ICU) in his local hospital with the diagnosis of stroke, diabetes and hypertension. He is being treated with thrombolytic therapy. During the night, Samuel's blood pressure begins to rise significantly above the recommended level for patients under treatment with thrombolytic therapy. This is quickly recognized by a remote tele-ICU team that provides coverage for all of the ICU beds in Samuel's rural hospital. This remote intensive care team has complete access to Samuel's electronic health record and bedside monitors and they also have video and audio connectivity into the room. The remote critical care team is able to quickly connect to Samuel's room and do a neurological exam with the assistance of the on-site nursing team. They determine that the exam is unchanged from the emergency room. They are able to order appropriate medications, recommend more frequent neurological checks, and directly follow his blood pressure response.

Consultation models in the in-patient setting using telemedicine in a variety of specialties have been reported. Including intensive care where timely consults are often essential (Assimacopoulos et al. 2008; Marcin et al. 2004). Although, these consultation models in critical care have shown benefit, a comprehensive multi-modality model has becomeh more common. This is often referred to as tele-ICU, and is defined as care provided to critically ill patients with at least some of the managing physicians and nurses in a remote location.

Some of the initial work in this area, done by Rosenfield and Bresslow in the Sentara Health System, demonstrated improved mortality, reduced lengths of stay and decreased costs (Rosenfeld et al. 2000). Remote intensive care has grown significantly over time with an estimated 10% of all ICU beds in the U.S. covered under this model of care, in large part due to a shortage of critical care physicians Typically, a single "Command Center" can cover multiple intensive care units over a large geographic region creating significant efficiencies and economies of scale.

This model of care integrates several of the technologies discussed in this book and is primarily enabled using electronic health records, evidenced based decision support tools, connections to bedside monitoring systems and audio/video based telemedicine into patient rooms. Most commonly, critical care health professionals co-manage care from a Command Center led by board-certified critical care physicians. Protocols and treatments reviews for patient management are incorporated into the care process using data from the monitoring and alert systems that indicate when changes in care should take place. The goal is to assure adherence to best practice,

achieve shorter response times to alarms, abnormal laboratory values and more rapid initiation of life saving interventions (Lilly et al. 2011).

Published studies have shown mixed results in terms of the benefits of tele-ICU. Lilly et al. reported that in a single academic medical center, implementation of tele-ICU was associated with reduced mortality and LOS, as well as lower rates of preventable complications (Lilly et al. 2011). A recent study compared inter-hospital transfer rates in hospitals with a tele-ICU with transfer rates of facilities with no telemedicine program in the Veterans Health Administration system (examining 52 ICUs in 23 acute care facilities) and found that ICU telemedicine was associated with a decrease in inter-hospital ICU transfers (Fortis et al. 2018). Another successful demonstration of the concept of the tele-ICU is the eICU at Emory University. In 2012, Emory launched an innovative plan to develop a collaborative network supporting intensive care units remotely throughout Georgia and more recently even partnered with clinicians in Australia to ensure 24/7 monitoring by experts. In the eICU experts can monitor the patient and speak directly to a care provider at the patient's bedside in Atlanta, while also talking with the patient and family caregivers. The use of specialized cameras, video monitors, microphones and speakers installed in Emory's ICU rooms, at four of its hospitals and one non-Emory hospital connect providers throughout the state of Georgia and more recently also to care teams in Australia. The eICU was found to reduce length of patient stay, resulted in fewer readmissions, reduced costs while addressing the shortage of intensivists (Buchman et al. 2017). In a systematic review and meta-analysis of studies examining outcomes of tele-ICUs, the authors concluded that the tele-ICU may reduce the ICU and hospital mortality and shorten the ICU length of stay but have no significant effect in hospital length of stay (Chen et al. 2018). This analysis also highlighted that further examination of the cost-effectiveness of a tele-ICU is needed.

20.3.6 Telepresence

Telepresence involves systems that allow clinicians to not only view remote situations, but also to act on them. The archetypal telepresence application is telesurgery. The most basic surgical telepresence systems simply permit two-way audio-video communications, by which remote surgeons can observe, teach, and collaborate with local surgeons while they operate on patients.

More advanced surgical telepresence systems allow procedures to actually be performed remotely. Although largely still experimental, a trans-Atlantic gall bladder operation was demonstrated in 2001 (Kent 2001). The military has funded considerable research in this area in the hope that surgical capabilities could be extended to the battlefield. Telepresence requires high bandwidth, low **latency** connections. Optimal telesurgery requires not only teleoperation of robotic surgical instruments, but also accurate **force feedback** (or **haptic feedback**) that requires extremely low network latencies. Accurate millisecond force feedback has been historically limited to distances under 100 miles. The endoscopic gall bladder surgery mentioned above is an exception to this general principle because that specific procedure relied almost exclusively on visual information. It used a dedicated and custom configured 10 Mb/s fiberoptic network with a 155 ms latency.

Providing tactile feedback over large distances actually requires providing the surgeon with simulated feedback while awaiting transmission of the actual feedback data. Such simulation requires massive computing power and is an area of active research. Telesurgery also require extremely high-reliability connections. Loss of a connection is an annoyance during a consultation; it can be fatal during a surgical procedure.

Robotic surgery systems have been commercially available since the early 2000s. In these systems, surgical instruments and a camera are introduced into the patient through small incisions. The surgeon controls these instruments remotely, while he or she is

viewing a magnified three-dimensional camera image of the patient's anatomical structures. These systems are currently being used in some medical centers for small-incision surgery, typically performed by surgeons seated adjacent to their patients. The increasing availability and use of these robotic surgery systems creates possibilities for an increasing number of telesurgery applications.

To date, robotically-assisted surgery has been most common in fields such as cardiothoracic surgery, gynecology, and urology. Potential advantages of remote robotically-assisted surgery may include smaller incisions, improved anatomic visualization, and finer control of surgical instrumentation. Several clinical studies comparing robotically-assisted surgery with traditional surgery have suggested that the outcomes are similar (Ficarra et al. 2009). However, additional research is required to determine the optimal role of robot-assisted surgery and its applications to telesurgery.

A novel form of telepresence gives clinicians the ability not only to see, but also to walk around. Since the early 2000s, a commercially-available system has combined conventional video telehealth with a remotely controlled robot (◘ Fig. 20.3). It allows clinicians literally to make remote video rounds. A frequent problem with telehealth systems is having the equipment where it is needed. With this system, the telehealth equipment is able to take itself to wherever it is needed. Remote monitoring may also be performed by interfacing digital devices such as stethoscopes or imaging systems to the remote-controlled robot. These remote-controlled systems are most often used by physicians and nurses to examine patients in nursing homes or other long-term facilities, to improve health care access in rural areas, and eaperform postoperative examinations. Croghan et al. (2018) designed and tested a remotely controlled mobile audiovisual drone to access inpatients in surgical wards based on a lightweight device that is freely mobile and "emulates human interaction by swiveling and adjusting height to patients' eye-level." As technological advancements in robotics introduce new innovative models of telepresence, identification of relevant outcome measures and rigorous evaluation studies are needed to assess both the effectiveness and unintended consequences of such solutions.

◘ **Fig. 20.3** Telehealth robot. This is controlled by a remote clinician, and includes videoconferencing and remote monitoring capabilities. In this example, a specialist is connecting with a nurse during a patient transfer. Image courtesy of InTouch Health, reproduced with permission

20.3.7 Delivering Specialty Knowledge to a Network of Clinical Peers

Telemedicine is used not only to provide direct services to patients but also to facilitate continuing education and peer support for clinicians with the goal to ultimately improve health care outcomes. The Project ECHO (Extension for Community Healthcare Outcomes) was originally established by the University of New Mexico as a partnership of academic medicine, public health offices and community clinics to use videoconferencing in order to promote knowledge networks and connect clinicians in rural areas with specialists in order to study cases of patients with unique needs (Arora et al. 2007). The first program focused on Hepatitis C but has since been adopted in numerous settings in the United States and worldwide, for various chronic conditions and populations including among

others, care of children with autism (Mazurek et al. 2017), geriatric mental health (Fisher et al. 2017) and diabetes (Swigert et al. 2014). The wide adoption of ECHO aims to utilize telehealth to strengthen workforce capacity in underserved areas and address health disparities. As of 2018 Project ECHO operated more than 220 hubs for more than 100 conditions or diseases in 31 countries. In a recent study, a program informed by the original ECHO model called SCAN-ECHO (Specialty Care Access Network-Extension for Community Healthcare Outcomes) was introduced by the Veterans Health Administration (VHA) to improve care for patients with liver disease in rural and underserved areas where care to specialized care can be challenging. The study collected 5 years of clinical data from 62,237 veterans with liver disease in the region. Only 513 of these veterans had a primary care physician who participated in SCAN-ECHO where they could discuss their patient cases with specialists but they had a 54% higher survival rate compared to the rest of their cohort even when adjusted for other variables (Su et al. 2018).

20.3.8 The Emergence of Telehealth during a Global Pandemic

As telehealth bridges geographic distance, it enables continuity in delivery of services even at times when populations may not have access to travel or even be restricted by physical distancing and quarantine. The COVID-19 pandemic in 2020 highlighted this potential and led to rapidly accelerated growth of telehealth. Worldwide health systems quickly adopted telehealth solutions. In the US, insurers expanded coverage to include all telemedicine and telehealth visit types including home visits, and licensure requirements were relaxed (Centers for Medicare and Medicaid Services 2020). The US Department of Health and Human Services (2020) waived enforcement of HIPAA regulations to allow the use of video-conferencing for telemedicine visits including the use of widely available video-conferencing solutions. In one large health system (NYU Langone Health) which was at the COVID-19 outbreak epicenter at that time, telemedicine visits increased by 135% in urgent care and by 4345% increase in non-urgent care between March 2 and April 14, 2020 (Mann et al. 2020). In addition to using telehealth for the delivery of traditional health care services, this platform also played a role for decision making and self-triage in the context of the pandemic. One such example is a telehealth patient portal for self-triage and scheduling that was created at the University of California San Francisco to enable asymptomatic patients to report exposure history and for symptomatic patients to be triaged and paired with appropriate levels of care (Judson et al. 2020). The rapid expansion of telehealth in times of this pandemic highlights the significance of investing in infrastructure and training to better prepare health systems in times of public health emergencies.

20.4 Challenges and Future Directions

As telehealth evolves from research novelty to being a standard way that health care is delivered, many challenges must be overcome. Some of these challenges arise because the one patient, one doctor model no longer applies. Basic questions of identity and trust become paramount. At the same time, the shifting focus from treating illness to managing health and wellness requires that clinicians know not only the history of the individuals they treat but also information about the social and environmental context within which those individuals reside. In the diabetes example, knowledge of the family history of risk factors, diseases, and the appropriate diagnostic and interventional protocols, aid the clinical staff in providing timely and appropriate treatment.

20.4.1 Challenges to Using the Internet for Telehealth Applications

Because of the public, shared nature of the Internet, its resources are widely accessible by citizens and health care organizations.

This public nature also presents challenges to the security of data transmitted along the Internet. The openness of the Internet leaves the transmitted data vulnerable to interception and inappropriate access. In spite of significant improvements in the security of Web browsing several areas, including protection against viruses, authentication of individuals and the security of email, remain problematic.

Ensuring every citizen access to the Internet represents a second important challenge to the ability to use it for public health purposes. Access to the Internet presently requires computer equipment that may be out of reach for persons with marginal income levels. Majority-language literacy and the physical capability to type and read present additional requirements for effective use of the Internet. Preventing inequalities in access to health care resources delivered via the Internet will require that health care agencies work with other social service and educational groups to make available the technology necessary to capitalize on this electronic environment for health care. A 2019 Report by the U.S. Federal Communications Commission indicated that more than 20 million Americans lack advanced broadband Internet access (defined as download speeds of at least 25 megabits per second with upload speeds of 3 Mbps) highlighting that many rural settings depend on satellite Internet for access at a higher cost. The COVID-19 pandemic intensified the disparities that emerge from this digital divide and strengthened ongoing efforts to disseminate higher bandwidth to rural settings.

As health care becomes increasingly reliant on Internet-based telecommunications technology, the industry faces challenges in insuring the quality and integrity of many devices and network pathways. These challenges differ from previous medical device concerns, because the diversity and reliability of household equipment is under the control of the household, not the health care providers. There is an increased interdependency between the providers of health services, those who manage telecommunication infrastructure and the manufacturers of commercial electronics. Insuring effective use of telehealth for home and community-based care requires that clinical services be supported by appropriate technical resources. The challenge of the digital divide that highlights varying degrees of access for patients to infrastructure and tools necessary for telehealth, must be addressed when designing and implementing such systems. This consideration is necessary to ensure that telehealth systems, meant to bridge geographic distance and increase access, do not end up further exacerbating inequities and raising additional barriers to high quality care. Additionally, consumer education is necessary so that patients and families fully understand risks and benefits of using telehealth software and hardware integrated into the care they receive. Educational initiatives need to address a wide spectrum of consumers' literacy and health literacy but also data literacy, namely consumers' ability to process, extract meaning and communicate knowledge generated by data.

20.4.2 Licensure and Economics in Telehealth

Licensure is frequently cited as the single biggest problem facing telemedicine involving direct patient-provider interactions. This is because medical licensure in the United States is state-based, while telemedicine frequently crosses state or national boundaries. The debate revolves around the questions of whether the patient "travels" through the wire to the clinician, or the clinician "travels" through the wire to the patient. Several states have passed legislation regulating the manner in which clinicians may deliver care remotely or across state lines. Some states have enacted "full licensure models" that require practitioners to hold a full, unrestricted license in each state where a patient resides. Many of these laws have been enacted specifically to restrict the out-of-state practice of telemedicine. To limit Web-based prescribing and other types of asynchronous interactions, several states have enacted or are considering regulations that would require a face-to-face encounter before any electronically delivered care

would be allowed. In contrast, some states are adopting regulations to facilitate telehealth by exempting out-of-state physicians from in-state licensure requirements provided that electronic care is provided on an irregular or episodic basis. Still other models would include states agreeing to either a mutual exchange of privileges, or some type of "registration" system whereby clinicians from out of state would register their intent to practice via electronic medium.

At the same time, national organizations representing a variety of health care professions (including nurses, physicians and physical therapists) have proposed a variety of approaches to these issues. While the existing system is built around individual state licensure, groups that favor telemedicine have proposed various interstate or national licensure schemes. The Federated State Board of Medical Examiners has proposed that physicians holding a full, unrestricted license in any state should be able to obtain a limited telemedicine consultation license using a streamlined application process. The American Medical Association is fighting to maintain the current state-based licensure model while encouraging some reciprocity. The American Telemedicine Association supports the position that—since patients are "transported" via telemedicine to the clinician—the practitioner need only be licensed in his or her home state. The National Council of State Boards of Nursing has promoted an Interstate Nurse Licensure Compact (NLC) whereby licensed nurses in a given state are granted multi-state licensure privileges and are authorized to practice in any other state that has adopted the compact. By 2015, 25 states had adopted the NLC; in 2017 the eNLC went into effect as an enhanced compact that addressed the challenge of uniform criminal background checks. As of early 2020, 34 states had enacted the eNLC.

The second factor limiting the growth of telehealth is reimbursement. Prior to the mid-1990s there was virtually no reimbursement for telehealth outside of teleradiology. For many years Medicare routinely reimbursed for synchronous video only for rural patients. In January 2015, the Centers for Medicare and Medicaid Services (CMS) created a new chronic care management (CCM) code that provides for non-face-to-face consultation which introduces options for reimbursement for asynchronous remote monitoring. Furthermore, starting in 2018 CMS allowed providers to get reimbursed separately for time spent on the collection and interpretation of health-related data that were generated remotely. As technologies advance and play a pervasive role in our health care system, we anticipate the incremental changes in telehealth legislation to accelerate. In 2017, 210 telehealth related bills were active across thirty states and as new technological capabilities are introduced, we anticipate further legislative efforts. Over 25 states have laws mandating private insurers to reimburse for telehealth services (with at least another 10 additional states have pending or proposed laws to do so). Numerous insurers provide reimbursement for electronic messaging and online consultations. The Medicare Telehealth Parity Act of 2015 has led to advancements in reimbursement for teleradiology and teledermatology including payments for store-and-forward telehealth; however, restrictions still apply to types of technologies used, services provided and populations covered. Few groups have even considered reimbursement for telehealth services that do not involve patient-provider interaction. An expert system could provide triage services; tailored on-line educational material, or customized dosage calculations. Such systems are expensive to build and maintain, but only services provided directly by humans are currently reimbursed by insurance.

Historically, patients have been perceived as reluctant to pay directly for telehealth services, especially when face-to-face visits were covered by insurance. This trend is changing as consumers are more familiar with the use of various technologies to bridge geographic distance and embrace various innovative tools to bridge geographic distance. In a 2017 study (Chang et al. 2017) estimating US households' willingness to pay for telehealth, the representative household was willing to pay $4.39 per month for telehealth. This valuation increased for household with higher opportu-

nity costs and even more for households living more than 20 miles away from their nearest medical facility (to $6.22 per month).

Finally, home telehealth monitoring may reduce the health care costs associated with unreimbursed hospital readmissions. For example, some insurance payers do not reimburse for hospital readmissions that occur within 30 days of discharge, and there are anecdotal reports of health systems paying for 32 days of home monitoring post-discharge. Determining whether, and how much, to pay for telehealth services will likely be a topic of debate for years to come. Starting in 2020 Accountable Care Organizations (ACOs) with Medicare fee-for-service beneficiaries will have the option to expand telehealth services to include the home as an eligible originating site (without being subject to the current Medicare geographic requirements for the telehealth originating site). Public health developments affect the legal landscape as the recent COVID-19 pandemic highlighted where several of the regulatory and licensing barriers were temporarily lifted. The regulatory landscape will evolve over time as both technology and medical knowledge advance and societal needs change.

20.4.3 Logistical Requirements for Implementation of Telehealth Systems

Telehealth systems must be carefully evaluated before implementation for routine use in individual disease situations, to ensure that they have sufficient diagnostic accuracy and reproducibility for clinical application. Appropriate training and credentialing standards must be developed for personnel who capture clinical data and images from patients locally, as well as for physicians and nurses who perform remote interpretation and consultation. Clear rules and responsibilities must be developed for remote patient management, including the appropriate response for situations in which data are felt to be of insufficient quality for telehealth. Guidelines for medicolegal liability must be established. Software that displays clinical information required for remote management, and that integrates into existing workflow patterns and maximizes efficiency through good usability principles will be required. More specifically in the context of usability, when patients are asked to utilize telehealth equipment, several factors should be taken into consideration such as patients' previous experience with and comfort in using technology, potential functional or cognitive limitations, and the availability of family members or informal caregivers who may be able to assist. Consider our example of Samuel who lives alone and has some visual limitations. The decision to use software and/or hardware for home-based monitoring should address Samuel's living arrangements and residential infrastructure, as well as his ability and willingness to operate the telehealth system (and the potential role his neighbor may play who is involved in Samuel's care). Methods for providing added value from technology toward telehealth diagnostic systems through strategies such as links to consumer health resources or computer-based diagnosis may be explored (Koreen et al. 2007). Finally, studies have suggested that patient satisfaction with telehealth systems is high (Lee et al. 2010). However, the practitioner-patient relationship is fundamental to health care delivery, and mechanisms must be developed that this bond is not lost from telehealth.

20.4.4 Telehealth in Low Resource Environments

In many parts of the developing world, the density of both health care providers and of technology is quite low. Thus, the demand for telehealth is high, but the ability to deliver it is challenged. Many of these regions have largely skipped traditional land-line telephony and moved directly to cellular infrastructure (Foster 2010). This, combined with advances in low-cost laptop computers that do not depend on stable power-grids, has allowed the development of a wide variety of telehealth and tele-education applications. The majority

of these are based on an asynchronous model. Transport media range from standard broadband in the urban areas, to satellite connections, to cellular data, to **SMS messaging**. The largest group of applications focuses on the provision of remote consultations for difficult cases using computer-based systems, while general health education and remote data collection have been the primary applications using cellular telephony applications. However, the development of smart mobile telephones with high-resolution cameras is rapidly blurring this distinction. Successful implementation of telehealth in various low resource settings is demonstrated in a series of projects captured by an e-book by Wootton and Bonnardot (2015).

20.4.5 Future Directions

Telehealth validation studies across a range of clinical domains have demonstrated good diagnostic accuracy, reliability, and patient satisfaction. Based on these results, numerous real-world telehealth programs have been implemented throughout the world. In the long term, successful large-scale expansion of these programs will require addressing the above challenges.

Beyond these practical factors, traditional medical care uses a workflow model based on synchronous interactions between clinicians and individual patients. The workflow model is also a sequential one in that the clinician may deal with multiple clinical problems or data trends but only within the context of treating a single patient at a time. Medical records, both paper and electronic, as well as billing and administrative systems all rely on this sequential paradigm, in which the fundamental unit is the "visit." Advances in telehealth are disrupting this paradigm. Devices have been developed that allow remote electronic monitoring of diabetes, hypertension, asthma, congestive heart failure (CHF), and chronic anticoagulation. As a result, clinicians may become inundated by large volumes of electronic results. This may mean that clinicians will no longer function in an assembly-line fashion, but will become more like dispatchers or air-traffic controllers, electronically monitoring many processes simultaneously. Clinicians will no longer ask simply, "How is Mrs. X today?" They will also ask the computer "Among my 2,000 patients, which ones need my attention today?" Neither clinicians, nor EHRs, are prepared for this change.

Many service industries such as the travel and transportation sectors have more recently experienced dramatic changes due to the concept of "shared economy" that promotes a shift from strictly regulated frameworks for transactions to decentralized approaches where community networks promote identifying and optimizing resources based on needs identified by the community members. This paradigm shift has started to also permeate the health care field introducing an expanded perspective of telehealth whereby consumers use an app to arrange for on-demand home visits or "virtual visits" enabled by videoconferencing. One example is the Pager app that helps consumers find a doctor who will contact and visit within a guaranteed two-hour window. Similarly, various apps like Mend or HealApp arrange for on-demand video-consultations or visits. Other examples of sharing economy apps utilize crowd sourcing to support diagnosis processes. These trends introduce opportunities and challenges as we consider the extent to which regulatory and quality safeguards may be necessary to maximize benefits and reduce unintended consequences. Furthermore, the use of wearables and smart home technologies expands the traditional models of telehealth so that in the near future comprehensive telehealth systems may integrate physiological, behavioral, social, cognitive, environmental and genomic data sources to deliver "precision medicine" in a continuum of care. Artificial intelligence and predictive analytics will play a key role in this expansion of the telehealth paradigm.

Perhaps the greatest long-term effect of the information and communication revolution will be the breaking down of role, geographic, and social barriers. Medicine is already benefiting from this effect. Traditional "doctors and nurses" are collaborating with public health professionals, and anyone with computer access can potentially communicate with patients or experts around the world. The

challenge will be to facilitate productive collaborations among patients, their caregivers, biomedical scientists, and information technology experts to promote patient engagement and shared decision making.

Suggested Readings

Bashshur, R. L., Shannon, G. W., Krupinski, E. A., Grigsby, J., Kvedar, J. C., Weinstein, R. S., et al. (2009). National telemedicine initiatives: Essential to healthcare reform. Telemedicine and e-Health, 15, 600–610. This paper discusses cost-benefit tradeoffs associated with telemedicine within the context of large-scale efforts promoting health care reform in the United States.

Chi, N. C., & Demiris, G. (2015). A systematic review of telehealth tools and interventions to support family caregivers. Journal of Telemedicine and Telecare, 21, 37–44. This review focuses on telehealth applications targeting either solely the family caregiver of a patient or the dyad (caregiver and patient) examining the impact of telehealth on caregiver outcomes. Six categories of telehealth interventions for caregivers were identified: education, consultation (including decision support), psychosocial/cognitive behavioral therapy, social support, data collection and monitoring, and clinical care delivery. Studies demonstrate caregiver satisfaction as well as reduction of caregiver anxiety and burden.

Reed, M. E., Parikh, R., Huang, J., Ballard, D. W., Barr, I., & Wargon, C. (2018). Real-time patient-provider video telemedicine integrated with clinical care. New England Journal of Medicine, 379, 1478–1479. Kaiser Permanente Northern California began offering telemedicine visits enabling patients to use videoconferencing on a mobile phone, computer or tablet to communicate with their physicians. In this study 210,383 video visits conducted over three years (involving 2796 primary care providers and 152,809 patients) were examined. The telemedicine visits extended established patient–physician relationships and led to high levels of patient satisfaction (93% of surveyed patients responded that the video visit met their needs).

Wechsler, L. R., Demaerschalk, B. M., Schwamm, L. H., Adeoye, O. M., Audebert, H. J., Fanale, C. V., Hess, D. C., Majersik, J. J., Nystrom, K. V., Reeves, M. J., Rosamond, W. D., Switzer, J. A., & American Heart Association Stroke Council; Council on Epidemiology and Prevention; Council on Quality of Care and Outcomes Research. (2017). Telemedicine quality and outcomes in stroke: A scientific statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke, 48, e3–e25. This is an updated systematic evidence-based review of scientific data examining the use of telemedicine for stroke care delivery. Published studies are categorized according to their level of certainty and class of evidence.

Questions for Discussion

1. Telehealth has evolved from systems designed primarily to support consultations between clinicians to systems that provide direct patient care. This has required changes in hardware, user interfaces, software, and processes. Discuss some of the changes that must be made when a system designed for use by health care professionals is modified to be used directly by patients.
2. There are still some challenges regarding reimbursement for telemedicine services. Imagine that you are negotiating with an insurance carrier to obtain reimbursement for a store-and-forward telemedicine service that you have developed. The medical director of the second insurance payer states: "Telemedicine seems like 'screening' rather than a mechanism for delivering health care. This is because you are simply using technology to identify patients who need to be referred to a real doctor, rather than providing true medical care. Therefore, we should only reimburse a very small amount for these screening services." In your opinion, is this a legitimate argument? Explain.
3. Using telehealth systems, patients can now interact with multiple health care stakeholders and monitor multiple aspects of their health generating large data sets. In order to inform timely and

tailored interventions based on data generated both during clinical encounters and outside the clinical setting, many propose for telehealth enabled patient generated health data to be directly integrated into the Electronic Health Record. Discuss both challenges and opportunities for such an approach.

4. A significant barrier to widespread telehealth adoption has been limited validation studies demonstrating that its diagnostic accuracy is comparable to that of traditional in-person medical care. Do you feel this is a realistic goal, given the extremely large number of potential disease states and clinical scenarios that may require validation studies? Are there alternate scenarios that could lead to telehealth becoming accepted as standard medical practice? Explain.
5. Home telehealth often requires interpretation of data collected directly by patients, which may create challenges because of concerns about accuracy, as well as challenges from a data management perspective because of the large volume of incoming data. Describe possible approaches toward addressing these challenges involving accuracy and data management.

References

Afrin, J. N., & Critchfield, A. B. (1997). Low-cost telepsychiatry for the deaf in South Carolina. *Proceeding of the AMIA Fall Symposium*, 901.

Arora, S., Geppert, C. M., Kalishman, S., Dion, D., Pullara, F., Bjeletich, B., Simpson, G., Alverson, D. C., Moore, L. B., Kuhl, D., & Scaletti, J. V. (2007). Academic health center management of chronic diseases through knowledge networks: Project ECHO. *Academic Medicine, 82*, 154–160.

Assimacopoulos, A., Alam, R., Arbo, M., et al. (2008). A brief retrospective review of medical records comparing outcomes for inpatients treated via telehealth versus in-person protocols: Is telehealth equally effective as in-person visits for treating neutropenic fever, bacterial pneumonia, and infected bacterial wounds. *Telemedicine Journal of e-Health, 14*, 762–768.

Bashshur RL, Shannon GW, Krupinski EA, Grigsby J, Kvedar JC, Weinstein RS, et al. (2009). National telemedicine initiatives: Essential to healthcare reform. Telemedicine and e-Health, 15, 600–610.

Bird, K. T. (1972). Cardiopulmonary frontiers: Quality health care via interactive television. *Chest, 61*, 204–205.

Brechner, R. J., Cowie, C. C., Howie, L. J., Herman, W. H., Will, J. C., & Harris, M. I. (1993). Ophthalmic examination among adults with diagnosed diabetes mellitus. *JAMA: The Journal of the American Medical Association, 270*, 1714–1718.

Buchman, T., Coopersmith, C. M., Meissen, H., Grabenkort, W. R., Bakshi, V., Hiddleson, C. A., & Gregg, S. R. (2017). Innovative interdisciplinary strategies to address the intensivist shortage. *Critical Care Medicine, 45*, 298–304.

Cavalleranno, A. A., Cavallerano, J. D., Katalinic, P., Blake, B., Conlin, P. R., Hock, K., Tolson, A. M., Aiello, L. P., & Aiello, L. M. (2005). Joslin vision network research team. *American Journal of Ophthalmology, 139*, 597–604.

Centers for Medicare & Medicaid Servicers. Medicaid Services. Medicare Telemedicine Health Care Provider Fact Sheet. (2020). https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet. Accessed 5 May 2020.

Chang, J., Savage, S. J., & Waldman, D. M. (2017). Estimating willingness to pay for online health services with discrete-choice experiments. *Applied Health Economics and Health Policy, 15*, 491–500.

Chen, J., Sun, D., Yang, W., Liu, M., Zhang, S., Peng, J., & Ren, C. (2018). Clinical and economic outcomes of telemedicine programs in the intensive care unit: A systematic review and meta-analysis. *Journal of Intensive Care Medicine, 33*, 383–393.

Croghan, S. M., Carroll, P., Reade, S., Gillis, A. E., & Ridgway, P. F. (2018). Robot assisted surgical ward rounds: Virtually always there. *Journal of Innovation in Health Informatics, 25*, 982.

Cuadros, J., & Bresnick, G. (2009). EyePACS: An adaptable telemedicine system for diabetic retinopathy screening. *Journal of Diabetes Science and Technology, 3*, 509–516.

Dawadi, P. N., Cook, D. J., & Schmitter-Edgecombe, M. (2016). Automated cognitive health assessment from smart home-based behavior data. *IEEE Journal of Biomedical and Health Informatics, 20*, 1188–1194.

Demiris, G., & Hensel, B. K. (2008). Technologies for an aging society: A systematic review of "smart home" applications. *Yearbook of Medical Informatics*, 33–40.

Downes, M. J., Mervin, M. C., Byrnes, J. M., & Schuffham, P. A. (2017). Telephone consultations for general practice: A systematic review. *Systematic Reviews, 6*(1), 128. https://doi.org/10.1186/s13643-017-0529-0.

Ficarra, V., Novara, G., Fracalanza, S., et al. (2009). A prospective, non-randomized trial comparing robot-assisted laparoscopic and retropubic radical prostatectomy in one European institution. *BJU International, 104*, 534–539.

Fisher, E., Hasselberg, M., Conwell, Y., Weiss, L., Padrón, N. A., Tiernan, E., Karuza, J., Donath, J., & Pagán, J. A. (2017). Telementoring primary care clinicians to improve geriatric mental health care. *Population Health Management, 20*, 342–347.

Fortis, S., Sarrazin, M. V., Beck, B. F., Panos, R. J., & Reisinger, H. S. (2018). ICU telemedicine reduces interhospital ICU transfers in the veterans health administration. *Chest, 154*, 69–76.

Foster V., Briceno-Garmendia C. (2010). Africa's infrastructure: a time for transformation. Washington, DC: The International Bank for Reconstruction and Development / The World Bank.

Franken, E. A., Berbaum, K. S., Marley, S. M., Smith, W. L., Sato, K., Kao, S. C., & Milam, S. G. (1992). Evaluation of a digital workstation for interpreting neonatal examinations. A receiver operating characteristic study. *Investigative Radiology, 27*, 732–737.

Heath, B., Salerno, R., Hopkins, A., et al. (2009). Pediatric critical care telemedicine in rural underserved emergency departments. *Pediatric Critical Care Medicine, 10*, 588–591.

Hripcsak, G., Cimino, J. J., & Sengupta, S. (1999). WebCIS: Large scale deployment of a web-based clinical information system. *Proceedings of the Annual AMIA Symposium, 1999*, 804–808.

Hudson, H. E., & Parker, E. B. (1973). Medical communication in Alaska by satellite. *The New England Journal of Medicine, 289*, 1351–1356.

Jain, S., Khera, R., Lin, Z., Ross, J. S., & Krumholz, H. M. (2020). Availability of telemedicine services across hospitals in the United States in 2018: A cross-sectional study. *Annals of Internal Medicine, 30*, M20–1201. https://doi.org/10.7326/M2-1201.

Jayakody, A., Bryant, J., Carey, M., Hobden, B., Dodd, N., & Sanson-Fisher, R. (2016). Effectiveness of interventions utilizing telephone follow up in reducing hospital readmission within 30 days for individuals with chronic disease: A systematic review. *BMC Health Services Research, 16*(1), 403. https://doi.org/10.1186/s12913-016-1650-9.

Judson, T. J., Odisho, A. Y., Neinstein, A. B., Chao, J., Williams, A., Miller, C., et al. (2020). Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. *Journal of the American Medical Informatics Association*, ocaa051. https://doi.org/10.1093/jamia/ocaa051.

Kane B., Sands D.Z. (1998). Guidelines for the clinical use of electronic mail with patients. The AMIA Internet Working Group, Task Force on Guidelines for the Use of Clinic-Patient Electronic Mail. J Am Med Inform Assoc, 5(1):104–11.

Kent, H. (2001). Hands across the ocean for the world's first trans-Atlantic surgery. *CMAJ: Canadian Medical Association Journal, 13*, 1374.

Koreen, S., Gelman, R., Martinez-Perez, M. E., et al. (2007). *Ophthalmology, 114*, e59–e67.

Lee, J. Y., Du, Y. E., Coki, O., Flynn, J. T., Starren, J., & Chiang, M. F. (2010). Parental perceptions toward digital imaging and telemedicine for retinopathy of prematurity management. *Graefe's Archive for Clinical and Experimental Ophthalmology, 248*, 141–147.

Levine S.R., Gorman M. (1999). "Telestroke": the application of telemedicine for stroke. Stroke, 30(2):464–9.

Lilly, C. M., Cody, S., Zhao, H., et al. (2011). Hospital mortality, length of stay, and preventable complications among critically ill patients before and after tele-ICU reengineering of critical care processes. *JAMA: The Journal of the American Medical Association, 305*, 2175–2183.

Liu, L., Stroulia, E., Nikolaidis, I., Miguel-Cruz, A., & Rios Rincon, A. (2016). Smart homes and home health monitoring technologies for older adults: A systematic review. *International Journal of Medical Informatics, 91*, 44–59.

Lhotska, L., Doležal, J., Adolf, J., Potůček, J., Křížek, M., & Chbani, B. (2018). Personalized monitoring and assistive systems: Case study of efficient home solutions. *Studies in Health Technology and Informatics, 249*, 19–28.

Mackinnon, A. D., Billington, R. A., Adam, E. J., Dundas, D. D., & Patel, U. (2008). Picture archiving and communication systems lead to sustained improvements in reporting times and productivity: Results of a 5-year audit. *Clinical Radiology, 63*, 796–804.

Mann, D. M., Chen, J., Chunara, R., Testa, P. A., & Oded, N. (2020). COVID-19 transforms health care through telemedicine: Evidence from the field. *Journal of the American Medical Informatics Association, 27*(7), 1132–1135. https://doi.org/10.1093/jamia/ocaa072.

Marcin, J. P., Nesbitt, T. S., Kallas, H. J., et al. (2004). *Journal of Pediatrics, 144*, 375–380.

Mazurek, M. O., Brown, R., Curran, A., & Sohl, K. (2017). ECHO autism. *Clinical Pediatrics, 56*, 247–256.

Meyer, B. C., Raman, R., Hemmen, T., et al. (2008). Efficacy of site-independent telemedicine in the STRokEDOC trial: A randomized, blinded, prospective trial. *Lancet Neurology, 7*, 787–795.

Murphy, R. L. H., Block, P., Bird, K. T., & Yurchak, P. (1973). Accuracy of cardiac auscultation by microwave. *Chest, 63*, 578–581.

Rajalakshmi, R., Arulmalar, S., Usha, M., et al. (2015). Validation of smartphone based retinal photography for diabetic retinopathy screening. *PLoS One, 10*(9), e0138285. https://doi.org/10.1371/journal.pone.0138285.

Reiner, B. I., Siegel, E. L., & Hooper, F. J. (2002). Accuracy of interpretation of CT scans: Comparing PACS monitor displays and hard-copy images. *AJR: American Journal of Roentgenology, 179*, 1407–1410.

Reeder, B., Meyer, E., Lazar, A., Chaudhuri, S., Thompson, H. J., & Demiris, G. (2013). Framing the evidence for health smart homes and home-based consumer health technologies as a public health intervention for independent aging: A systematic review. *International Journal of Medical Informatics, 82*, 565–579.

Ricci, M. A., Caputo, M., Amour, J., et al. (2003). Telemedicine reduces discrepancies in rural trauma care. *Telemedicine Journal and E-Health, 9*, 3–11.

Richter, G. M., Williams, S. L., Starren, J., Flynn, J. T., & Chiang, M. F. (2009). Telemedicine for retinopathy of prematurity diagnosis: Evaluation and challenges. *Survey of Ophthalmology, 54*, 671–685.

Rosenfeld, B. A., Dorman, T., Breslow, M. J., et al. (2000). Intensive care unit telemedicine: Alternate paradigm for providing continuous intensivist care. *Critical Care Medicine, 28*, 3925–3931.

Saffle, J. R., Edelman, L., Theurer, L., et al. (2009). Telemedicine evaluation of acute burns is accurate and cost-effective. *The Journal of Trauma, 67*, 358–365.

Sarkar I.N., Starren J. (2002). Desiderata for personal electronic communication in clinical systems. J Am Med Inform Assoc, 9(3):2009–16.

Starr P. (1982). The Social Transformation of American Medicine. New York: Basic Books.

Starren, J., Hripcsak, G., Sengupta, S., Abbruscato, C. R., Knudson, P. E., Weinstock, R. S., & Shea, S. (2002). Columbia University's informatics for diabetes education and telemedicine (IDEATel) project: Technical implementation. *Journal of the American Medical Informatics Association: JAMIA, 9*(1), 25–36.

Steventon, A., Bardsley, M., Billings, J., Dixon, J., Doll, H., Hirani, S., Cartwright, M., Rixon, L., Knapp, M., Henderson, C., Rogers, A., Fitzpatrick, R., Hendy, J., Newman, S., & Whole System Demonstrator Evaluation Team. (2012). Effect of telehealth on use of secondary care and mortality: Findings from the whole system demonstrator cluster randomised trial. *BMJ: British Medical Journal, 344*, e3874.

Su, G. L., Glass, L., Tapper, E. B., Van, T., Waljee, A. K., & Sales, A. E. (2018). Virtual consultations through the veterans administration SCAN-ECHO project improves survival for veterans with liver disease. *Hepatology, 68*(6), 2317–2324.

Swigert, T. J., True, M. W., Sauerwein, T. J., & Dai, H. (2014). U.S. air force telehealth initiative to assist primary care providers in the management of diabetes. *Clinical Diabetes, 32*, 78–80.

Takahashi, P. Y., Pecina, J. L., Upatising, B., Chaudhry, R., Shah, N. D., Van Houten, H., Cha, S., Croghan, I., Naessens, J. M., & Hanson, G. J. (2012). A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. *Archives of Internal Medicine, 172*, 773–779.

United States Department of Health & Human Services. (2020). *Notification of enforcement discretion for telehealth remote communications during the COVID-10 nationwide public health emergency*. https://www.hhs.gov/hipaa/for-professionals/specialtopics/emergency-preparedness/notification-enforcement-discretiontelehealth/index.html. Accessed 5 May 2020.

United States Federal Communications Commission. 2019. *Broadband deployment report (2019)*. https://docs.fcc.gov/public/attachments/FCC-19-44A1.pdf. Accessed 30 May 2020.

Wechsler, L. R., Demaerschalk, B. M., Schwamm, L. H., Adeoye, O. M., Audebert, H. J., Fanale, C. V., Hess, D. C., Majersik, J. J., Nystrom, K. V., Reeves, M. J., Rosamond, W. D., Switzer, J. A., & American Heart Association Stroke Council; Council on Epidemiology and Prevention; Council on Quality of Care and Outcomes Research. (2017). Telemedicine quality and outcomes in stroke: A scientific statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke, 48*, e3–e25.

Wootton, R., & Bonnardot, L. (2015). Telemedicine in low-resource settings. *Frontiers in Public Health, 3*, 3. https://doi.org/10.3389/fpubh.2015.00003.

Patient Monitoring Systems

Vitaly Herasevich, Brian W. Pickering, Terry P. Clemmer, and Roger G. Mark

Contents

This chapter is adapted from an earlier version in the third and fourth edition authored by Reed M. Gardner and M. Michael Shabot.

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_21

Learning Objectives

After reading this chapter, you should be able to answer these questions:

1. What is patient monitoring and why is it used?
2. What patient parameters do bedside physiological monitors track?
3. What are the major problems with acquisition and presentation of monitoring parameters?
4. In addition to bedside physiologic parameters, what other information is fundamental to the care of acutely ill patients?
5. Why is real-time computerized decision support potentially more beneficial than monthly or quarterly quality-of-care reporting?
6. What technical and social factors must be considered when implementing real-time data acquisition and decision support systems?

21.1 What is Patient Monitoring?

Life, from the physiologic standpoint requires supplying oxygen to tissues to address the metabolic needs for the purpose of fueling mitochondrial respiration in cells. When this cycle is broken humans become critically ill. That physiologic cycle could be controlled through oxygenation and perfusion monitoring but there are no direct methods existing to measure mitochondrial respiration. All modern monitoring methods are proxy methods for such processes. In hospitals, and especially in the intensive care unit (ICU), patient monitoring becomes critical for control and optimization of hemodynamic, ventilation, temperature, nutrition, and metabolism of the human body.

Measurement of patient physiologic parameters such as heart rate (HR), heart rhythm, arterial blood pressure (ABP), respiratory rate, and blood-oxygen saturation, have become common during the care of the hospitalized and, especially, critically ill patients. When accurate and prompt decision making is crucial for effective patient care, bedside monitors are used to collect, display, and store physiologic data. Increasingly, such data are collected by noninvasive sensors connected to patients in ICUs, neonatal ICUs, operating rooms (ORs), labor and delivery suites, emergency departments, and other hospital care units where patient acuity is increased.

We often think of a patient monitor as something that watches for, and warns about, serious or life-threatening events in patients, and provides guidance for care of the critically ill. Such systems must include continuous observations of a patient's physiologic measurements and the assessment of the function of attached life support equipment. Such monitoring is important in detecting life-threatening conditions and guiding management decision making, including when to make therapeutic interventions and to assess the effect of those interventions.

In this chapter, we discuss the use of computers in collecting, displaying, storing, and interpreting clinical data, making therapeutic recommendations, and alarming and alerting. In the past, most monitoring data (called vital signs) were in the form of HR and respiratory rate, blood pressure (BP), and body temperature. However, today's ICU monitoring systems are able integrate data from bedside monitors and devices, as well as data from many sources outside the ICU. Although the material presented here deals primarily with patients who are in ICUs, the general principles and techniques are also applicable to other hospitalized patients and electronic medical records (EMRs). Patient monitoring is performed extensively for diagnostic purposes in the emergency department or for therapeutic purposes in the OR. Techniques that initially were only used in the ICU such as bedside monitors are now used routinely on general hospital wards and in some situations even by patients in their homes.

21.1.1 Case Report

This case report provides a perspective on the problems faced by the team caring for a critically ill patient.

A 27-year-old man is injured in an automobile accident and has multiple chest and

head injuries. His condition was stabilized at the scene of the accident by skilled paramedics using a portable computer-based electrocardiogram (ECG) and pulse oximeter, and he is quickly transported to a trauma center. Once in the trauma center, he is connected via noninvasive sensors to a computer-based bedside monitor that displays physiologic signals, including his HR and rhythm, arterial oxygen saturation, and BP. Radiographic and magnetic resonance imaging provide further information for care.

Because of the severe chest injury, the patient has difficulty breathing, so he is connected to a computer-controlled ventilator that has both therapeutic and monitoring functions and he is transferred to the ICU. Because of the head injury a bolt is placed in a hole drilled through his skull and a fiber optic sensor is inserted to continuously measure intracranial pressure with another computer-controlled monitor. Blood is drawn and clinical chemistry and blood gas tests are promptly performed by the hospital laboratory. Results of those tests are displayed to the ICU team as soon as they are available. With intensive treatment, the patient survives the initial threats to his life and he now begins the long recovery process. Figure 21.1 shows a nurse at the patient's bedside surrounded by a bedside monitor, infusion pumps, a ventilator, and other devices.

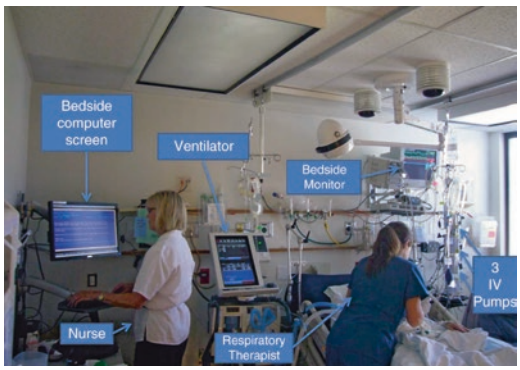


Fig. 21.1 Overall view of an ICU patient's room. Shown is a nurse standing at the bedside computer screen (*left*), a ventilator (*center*), and a respiratory therapist suctioning the patient (*right*). The patient is connected to the ventilator, bedside monitor (*upper right*), and to three IV pumps (*lower right*). ICU indicates intensive care unit, IV intravenous

Unfortunately, a few days later, the patient is beset with a problem common to multiple trauma victims—he develops a major **nosocomial hospital-acquired infection**, sepsis, and acute respiratory distress syndrome (ARDS). Multiple organ failure follows. As a result, antibiotics and electrolytes are required for treatment and are dispensed via intravenous (IV) pumps. The quantity of information required to care for the patient has increased dramatically; with monitor data excluded, critically ill patients produce in average 10 times more clinical data than patients on the typical general hospital unit (Fig. 21.2) (Herasevich et al. 2012).

Decision-making for patients with complex rapidly changing acute conditions requires much data in addition to monitoring data (Fig. 21.3) (Bradshaw et al. 1984).

Increased data flow, rapid changes in the patient's state, and a multidisciplinary approach involving different care providers required the next generation of data management clinical systems outside of distributed, often not connected, multiple patient monitors and data from hospital EMRs.

In ▶ Sect. 21.6 ("Monitoring and Advanced Information Management in ICUs") we describe two integrated patient monitoring systems: the HELP system used by Intermountain Healthcare's Hospitals and the AWARE system developed at Mayo Clinic Hospitals. Both of these systems integrate diverse clinical data for complex decision making.

21.1.2 Patient Monitoring

Careful monitoring and alerting care team about changes in a patient's physiologic status is a vital part of diagnostic and therapeutic processes. There are at least three categories of patients who need physiologic monitoring:

1. Patients with compromised physiologic regulatory systems; eg, a patient whose respiratory system is suppressed by a drug overdose or during anesthesia
2. Patients who are currently stable but with a condition that could suddenly change to become life threatening; eg, a patient who

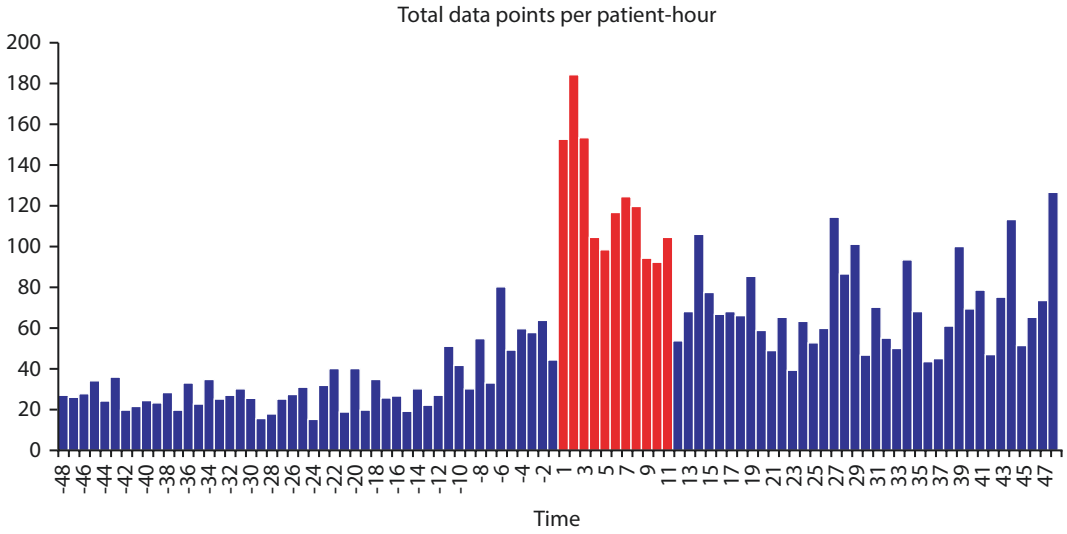


Fig. 21.2 Average number of total clinical data points per patient hour, excluding vitals, before and after admission to the ICU (time zero). Bars in red indicate first hours in ICU. Y-axis indicated number of data points

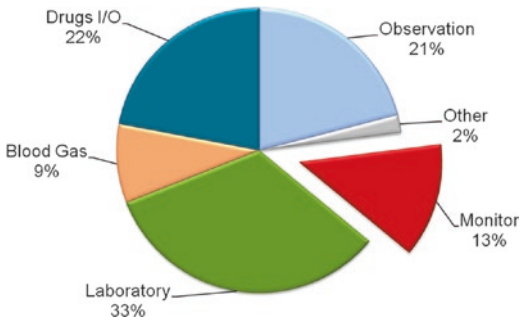


Fig. 21.3 Six main data categories and their relative distribution used in clinical decision making in trauma shock intensive care unit. I/O indicates Intake and Output

has findings indicating an acute myocardial infarction (heart attack) or immediately after open-heart surgery, or a fetus during labor and delivery

3. Patients in a critical physiologic state; eg, patients with multiple trauma or septic shock like in our case study

Clinical monitoring has evolved over the time and has a tendency to change from intermittent to continuous and from invasive to noninvasive methods. Also, there is a trend for monitoring systems to become multipurpose and integrate multiple parameters, including nonmonitoring data. Current stand-alone bedside monitors have data storage and are capable of capturing multiwaveform/multiparameter information with advanced alerting functionality.

In general, patient monitoring could be divided to five groups (Fig. 21.4):

1. Hemodynamic monitoring
2. Respiratory monitoring
3. Neuromonitoring
4. Metabolic monitoring
5. Specialty monitoring

Not all monitoring techniques needed for every patient or are widely available. Also, using invasive methods should take into account the risk of complications.

In terms of technologic classification, monitoring could be divided by three general parts:

1. Vital signs monitoring
2. Diagnostic monitoring
3. Specialized or disease-specific monitoring

21.2 Historical Perspective and the Measurement of Vital Signs

The earliest foundations for acquiring physiologic data occurred at the end of the Renaissance period. In 1625, Santorio

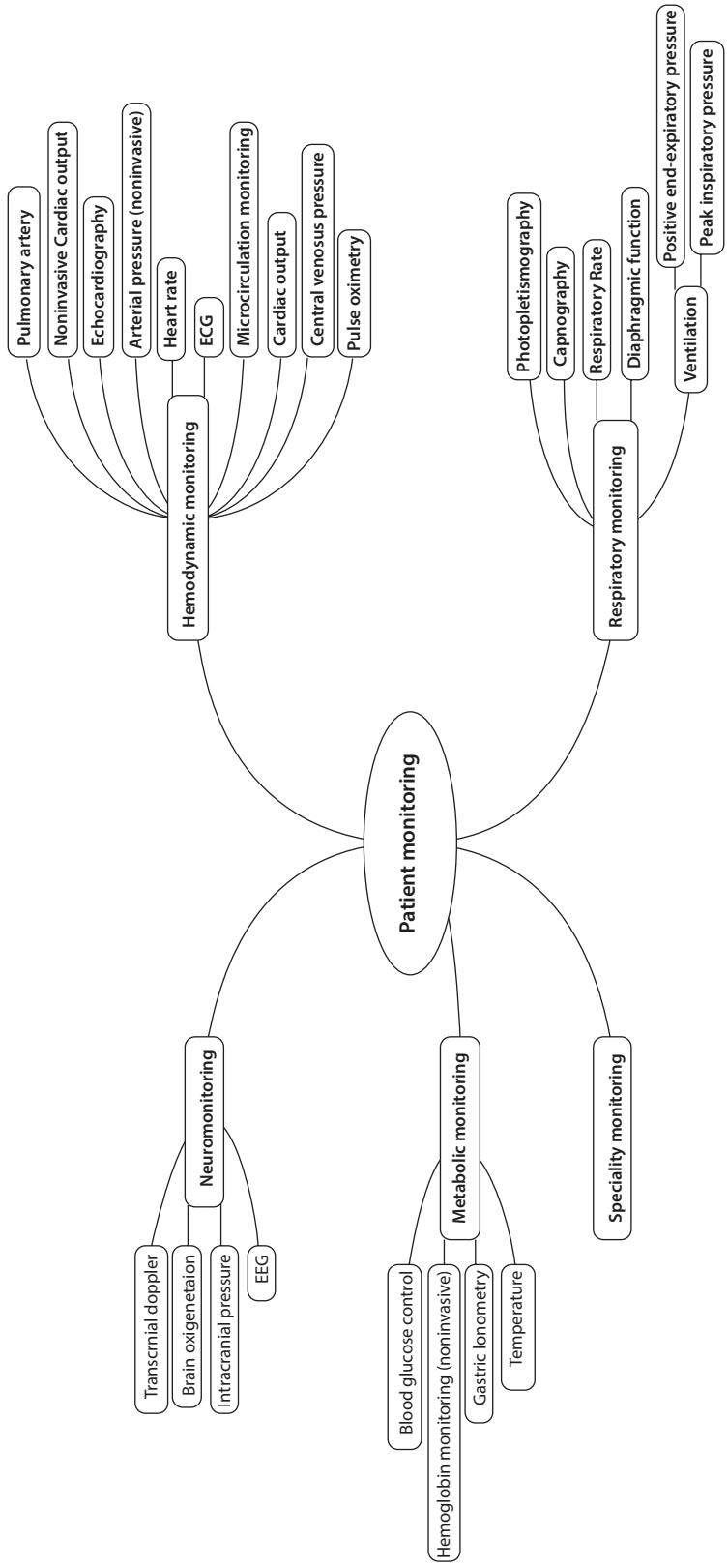


Fig. 21.4 Different types of patient physiologic monitoring. ECG indicates electrocardiogram, EEG electroencephalogram

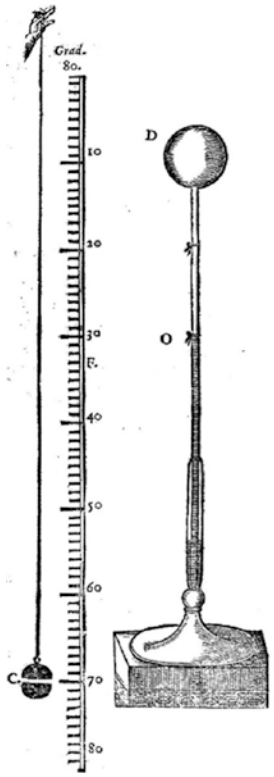


Fig. 21.5 Pulsilogium at center to measure HR and thermoscope (right) to measure body temperature. HR indicates heart rate. (From Sanctorius (1626))

Santori, who lived in Venice, Italy, published his methods for measuring body temperature with the spirit thermometer and for timing the pulse (HR) with a pendulum (Fig. 21.5). The principles for both devices had been established by Galileo Galilei, a close friend. Galileo worked out the uniform periodicity of the pendulum by timing the movement of the swinging chandelier in the Cathedral of Pisa and comparing that to his own pulse rate. The results of this early biomedical-engineering collaboration, however, were ignored. The first scientific report of the pulse rate did not appear until English physician Sir John Floyer published "Pulse-Watch" in 1707. The first published course of fever for a patient was plotted by Ludwig Taube in 1852.

In 1896, Scipione Riva-Rocci introduced the sphygmomanometer (BP cuff), which permitted the fourth vital sign, systolic BP, to be measured. A Russian physician, Nikolai Korotkoff, applied Riva-Rocci's cuff with a stethoscope developed by the French physician Rene Laennec, which allowed the measurement of both systolic and diastolic arterial pressure. Harvey Cushing, a preeminent US neurosurgeon of the early 1900s, predicted the need for and later insisted on routine ABP monitoring in the OR. At the same time, Cushing also raised the following questions, which are still being asked today:

1. Are we collecting too much data?
2. Are the instruments used in clinical medicine too accurate?
3. Would not approximated values be just as good?

Cushing (1903) answered his own questions by stating that vital sign measurements should be made routinely and that their accuracy was important. In 1903, Willem Einthoven devised the string galvanometer for displaying and quantifying the ECG, for which he was awarded the 1924 Nobel Prize in physiology. The ECG has become an important adjunct to the clinician's inventory of tests for both acutely and chronically ill patients. Continuous measurement of physiologic variables has become a routine part of the monitoring of critically ill patients.

At the same time that advances in monitoring were made, major changes in the therapy of life-threatening disorders were also occurring. Prompt quantitative evaluation of measured physiologic and biochemical variables became essential in the decision-making process as physicians applied new therapeutic interventions. For example, it is now possible, and in many cases essential, to use ventilators when a patient cannot breathe independently, for cardiopulmonary bypass equipment when a patient undergoes open-heart surgery, hemodialysis when a patient's kidneys fail, and IV nutritional and electrolyte support when a patient is unable to eat or drink.

Since the 1920s, the four vital signs—temperature, respiratory rate, HR, and ABP—have been recorded in all patient charts and became the standard vital signs. In recent years a fifth vital sign, oxygen saturation, was added as a routine measurement.

21.3 Development of ICUs

Care of critically ill patients requires prompt and accurate decisions so that life-protecting and life-saving therapy can be appropriately applied. Because of these requirements, ICUs have become widely established in hospitals. Such units use computers almost universally for the following purposes:

1. To acquire physiologic data frequently or continuously, such as BP
2. To acquire information from remote data-producing systems to remote locations, eg, laboratory and radiology departments
3. To store, organize, and report patient information
4. To integrate, organize and correlate data from multiple sources
5. To provide clinical alerts and advisories based on multiple sources of data
6. To function as an automated decision support tool that health professionals may use in planning the care of critically ill patients
7. To measure the severity of illness for patient classification purposes
8. To analyze the outcomes of ICU care in terms of clinical effectiveness and cost effectiveness

Until about 1960, if patients had severe cardiac events, there were few treatment options available for physicians to provide care for them. As a consequence, many patients who had life-threatening acute cardiac or pulmonary problems died. However, in the early 1960's two major medical care treatment modalities were developed that provided treatment for previously fatal situations. Development of closed chest cardiopulmonary resuscitation (CPR); (Kouwenhoven et al. 1960) and closed chest defibrillation (Zoll et al. 1956; Lown et al. 1962) provided means for delivering life-saving treatment. Because of availability of these treatments, the demand for continuous monitoring of high-risk patients escalated. Hospitals began to cluster patients with complex disorders together into new organizational units, called ICUs, beginning in the early 1960s. Some of the earliest units were coronary care units where patients were cared for after myocardial infarctions or other acute, life-threatening cardiac events.

Surgical ICUs had their beginnings in the late 1950s when postoperative patients were kept in the recovery rooms for extended time periods after cardiac or other high-risk surgery for close observation. Initially these recovery rooms did not have the benefit of cardiac monitoring. However, as more sophisticated monitoring became available, special units were created and designated as surgical ICUs or thoracic ICUs.

ICUs proliferated rapidly during the late 1960s and early 1970s. The types of units included coronary, thoracic surgery, surgical, medical, shock-trauma, burn, pediatric, neonatal, respiratory, and other multipurpose medical or surgical units. Today there are more than six million patients admitted each year into adult, pediatric, and neonatal ICUs in the United States alone. In the past three decades, the demand for ICU services in the United States has risen dramatically. Because of the complexity of care and the increased acuity of these patients, the need for specialized nursing care has increased dramatically. In a typical non-ICU acute patient care situation, one nurse may be responsible for the care up to six patients. However, because of the observations and care that these acutely ill ICU patients require, intensive care nurses typically are assigned one to three patients.

The average life expectancy is rising and estimates of the US population aged over 65 years (who use ICUs disproportionally more than the rest of the population) is estimated to increase by 50% by 2020 and 100% by 2030, thus continually increasing demand (Kelley et al. 2004; Groves et al. 2008) for ICU-level care.

21.4 Development of Bedside Monitors

A signature feature of each of these early ICUs was the bedside monitor (◘ Fig. 21.6). The original bedside monitors were used primarily to acquire and display the ECG. The first modern bedside monitor, ICU 80, was

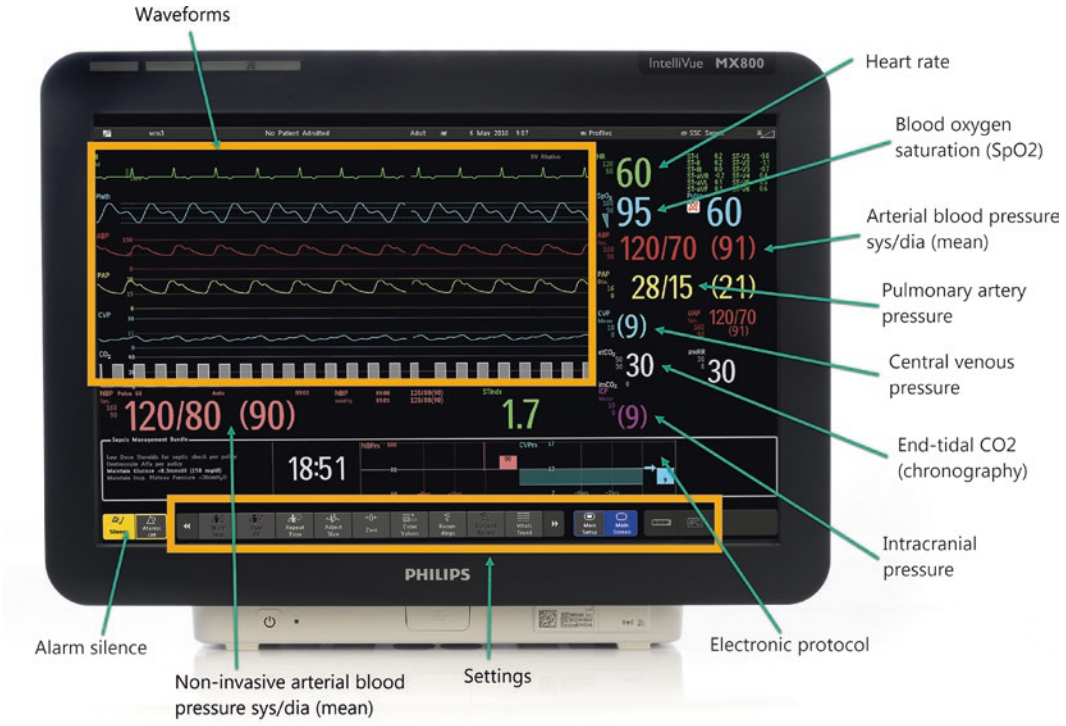


Fig. 21.6 Waveforms on typical bedside monitor. Displays from the monitor show the real-time beat-by ECG, pulse oximeter, and arterial, pulmonary artery, and central venous blood pressures and their derived measures. (Courtesy of Royal Philips, with permission)

introduced by the Nihon Kohden company from Japan in 1967 and was as large as the patient bed.

As a result of the detailed ECG information provided by the new patient monitors, treatment for serious cardiac arrhythmias (heart rhythm disturbances) and cardiac arrest (abrupt cessation of heartbeat)—major causes of death after myocardial infarctions (heart attack)—became possible. Mortality rates from 1960 to 1970 were about 35%, dropped to about 23% between 1970 and 1980 and to about 20% between 1980 and 1990. During the 1990s reperfusion of the coronary arteries became common and mortality rate dropped to about 5% (Braunwald 1988; Rogers et al. 2000).

In the 1960s, bedside monitors were built using analog computer technologies. These systems amplified the ECG signal and displayed the results on an oscilloscope. Such systems required nurses or technicians to watch the oscilloscope to determine if there was a cardiac arrest or other life-threatening cardiac rhythm. Soon after these analog systems were developed, methods for generating high- and low-HR alarm thresholds were included. The alarms were usually audible and very annoying. Unfortunately, since the beginning of the use of these alarms, the false-positive rate has far exceeded the true positive rate. As a result, many times alarm systems for bedside monitors are ignored or turned off. The problem of alert fatigue in hospitals is still considered a challenging technological problem and contributes to poor patient outcomes including deaths.

Teams from several cities in the United States in the 1960s introduced computers into the ICU to assist in physiologic monitoring, beginning in Los Angles with Shubin and Weil (1966) followed by Warner et al. (1968) in Salt Lake City. These investigators had several objectives:

1. To increase the availability and accuracy of the physiologic data
2. To compute derived variables that could not be measured directly
3. To increase patient-care efficacy
4. To allow display of the time trends of the patient's physiologic data, and
5. To assist in computer-aided decision making

Each of these teams developed their applications on large **mainframe computer systems**, which required large computer rooms and trained staff to keep the system operational 24 hours a day. The computers used by these developers cost over $200,000 each in 1965 dollars. During that time, other researchers were attacking more specific challenges in patient monitoring. For example, Cox (1972) at Barnes Hospital in St. Louis, developed algorithms to analyze the ECG for heart rhythm disturbances in real-time. The arrhythmia-monitoring system, which was installed in the Coronary Care Unit (CCU) in 1969, ran on a relatively inexpensive **minicomputer** rather than a mainframe computer. With the advent of **integrated circuits** and **microprocessors**, affordable computing power increased dramatically. What was considered computer-based patient monitoring by these pioneers in the late 1960s and early 1970s is now entirely built into bedside monitors and is considered simply a bedside monitor. Clemmer (2004) provides an important overview of "where we started and where we are now" to summarize the first four decades following the initiation of computers in the ICU.

21.5 Modern Bedside Monitors

The heart and lungs are crucial to normal body function. For example, if the heart stops (cardiac arrest) there is a cessation of normal circulation of the blood. Likewise, if there is a pulmonary arrest there is a cessation of breathing. Each of these situations leads to a reduced delivery of oxygenated blood (hypoxia) to the body, with major physiologic hazards. For example, brain injury will occur if hypoxia is untreated within 5 minutes. As a consequence, detection of either of these situations is required if life-saving treatments are to be administered. The treatment for cardiac arrest is cardiopulmonary resuscitation (CPR), which provides circulatory and pulmonary support. Prompt use of a defibrillator increases the likelihood of reestablishment of a normal rhythm.

The typical bedside monitor can also display the ECG and the arterial waveform (◘ Fig. 21.6).

21.5.1 ECG Signal Acquisition and Processing

The ECG provides a representation of the electrical activity of the human heart and is a very important tool for the diagnosis of disturbances of HR and rhythm. Original monitors allowed physicians and nurses the ability to watch the ECG trace on an oscilloscope. Since ECG signal measured on the skin is very small (1 mV), it is subject to artifacts (noise) caused by such things as patient movement, electrode movement, and electrical power interference. By using sophisticated analog and digital techniques and presenting data from multiple leads, the quality and reliability of the ECG signals monitored has improved dramatically (Weinfurt 1990; Gregg et al. 2008). At the same time, the demand for improved quality of the ECG signal and an increase in the number and types of parameters has increased. Initially, the ECG signal was processed to obtain HR and basic rhythm (periodicity of the beat) while today's monitors can detect signals from artificial heart pacemakers, complex arrhythmias, **myocardial ischemia** and disturbances in the conduction of electrical signals through the heart muscle.

Two types of computerized ECG analysis are in common use today:

1. The 12-lead ECG is typically performed in a physician's office or in the hospital. Usually a technician brings a recording device to the patient's bedside and attaches the leads, and records the signal during a

short interval while the patient is lying quietly in a supine position. From this 12-lead ECG, a wide variety of ECG diagnoses are made. Computer processing of these ECG signals taken at that moment in time has become the definitive practical option for ECG interpretation. Automated ECG analysis has become widespread in clinical practice since the mid-1980s although, in most hospitals, cardiologists will also read them to confirm the automated findings. Automated ECG analysis is quite accurate, especially in normal individuals, but disagreements with cardiologists are seen and may be clinically important (Guglin and Thatai 2006; Bogun et al. 2004). On the other hand, cardiologists are not perfect either (Clark et al. 2010)!

Today, physicians expert in ECG interpretation from multiple professional organizations such as the American Heart Association and the Electrocardiographic Society have come to consensus and established standards designed to improve computerized ECG interpretation. In biomedical informatics terminology, these experts have developed the knowledge base for diagnostic ECG interpretation. The detailed pattern recognition and signal processing does not need to occur in real-time. Thus the 12-lead ECG processing can be more sophisticated than with the requirements of real time monitoring situations (Gregg et al. 2008).

2. Continuous, real-time monitoring is required while the patient is in the ICU. Because of patient movement, caregiver activities such as administering medications, bathing and the like, the amount of artifact generated poses important challenges to real-time monitoring. To minimize these effects, filtering of the acquired ECG signal is performed. This filtering slightly distorts the ECG but at the same time makes it possible to process the signals on a beat-by-beat basis. Although standards for interpretation of ECG monitoring are more recent than those for 12-lead monitoring, they are now becoming more common and sophisticated (Drew and Funk 2006; Funk et al. 2010). The clinical experts who are establishing the knowledge base now include critical-care nurses, cardiologists, anesthesiologists, and thoracic surgeons (Crossley et al. 2011).

ECG processing in today's vendor-supplied bedside monitors continues to improve and become more reliable. Sophisticated pattern recognition and signal processing techniques are used to allow extraction of key parameters in real-time while adding the ability to measure the utility of new physiologic parameters (Crossley et al. 2011). Investigators have created publically available databases of ECG waveforms and other physiologic signals as well as other important data from actual patients to allow validation of these monitoring systems (Saeed et al. 2011; Burykin et al. 2011).

21.5.2 ABP Signal Acquisition and Processing

Accurate and continuous monitoring of ABP requires insertion of a catheter into an artery. Once the catheter is successfully inserted into an artery, the catheter is connected, via a length of sterile fluid-filled tubing, to a stopcock with a continuous flush device and a factory calibrated disposable BP transducer (Gardner 1996). The BP transducer is then connected to an amplifier and the pulsatile signal it detects is displayed on the screen of the bedside patient monitor. With the advent of inexpensive, disposable, accurate pressure transducers, the quality and accuracy of ABP monitoring has improved dramatically. However, two sources of inaccuracy of the ABP signal still depend on medical staff set-up and validation: (1) zeroing is the process by which the monitor is informed when a port on the stopcock is opened to the atmosphere at mid-heart level – thus becoming the point from which pressure is measured; and (2) since the ABP signal contains pulsatile characteristics with frequencies up to 20 Hz that must be transmitted from the artery through the plumbing system to the transducer, the

dynamic response characteristics must be optimized. Optimization is typically done by doing a fast flush test (by pushing sterile saline through the tubing) to optimize the system by removing blood and very tiny air bubbles that can dramatically distort the ABP waveform and result in erroneous measures of systolic and diastolic pressure.

At least two types of artifacts in the ABP signal are commonly observed. If a patient rapidly moves or a care giver bumps the tubing, a pressure artifact is generated and transmitted to the transducer and displayed. In addition, when the clinical staff draws arterial blood for laboratory tests, they typically turn off the stopcock connected to the transducer and draw blood through the tubing, causing an immediate loss of the pulsatile ABP signal. The pressure sensed by the transducer then typically rises to that found in the pressurized flush solution. Thus, continuous vigilance on the part of nurses and other care givers is needed for the arterial catheter and monitoring systems to be properly maintained. As a historical note, the continuous flush device was developed over 50 years ago to prevent arterial catheters from clotting and to allow one of the pioneering computerized monitoring systems to become more reliable (Gardner et al. 1970). Since that time, investigators have developed computerized methods to minimize these "human caused artifacts" (Li et al. 2009; Gorges et al. 2009). Unfortunately, these strategies have seldom been implemented into commercially available bedside monitors.

Since the early 1900s, efforts have been made to estimate **cardiac output** from the pulsatile pressure in the arterial system by multiplying the HR with estimates of stroke volume (the volume of blood ejected from the heart during a single contraction) made from the pressure waveform. Warner and his colleagues at Mayo Clinic published some early work in 1953 (Warner et al. 1953) on the topic and followed up again in 1983 further substantiating the feasibility of the method. However, Cundick and Gardner (1980) showed that the widely varying mean BPs found in critically ill patients adversely affected the reliability of the method. Since that early work, multiple publications and commercially available devices using the pulse-pressure method have appeared. The issue is still active, with such publications as Chen et al. (2009), Sun et al. (2009), and Gardner and Beale (2009). Other estimates of stroke volume and cardiac output have been made from determining the bioreactance – a measure of the degree of phase-shift in the electrical signal—across the chest. This method shows promise of being a rather simple, continuous and noninvasive method for measuring cardiac output (Keren et al. 2007).

Investigations have made assessments of delta pulse pressure, which measures the variability of the peak to peak arterial pressure pulse signal across the breathing cycle to make an estimate of a patient's fluid balance. The supposition is that if there is larger variability in this delta pulse pressure marker the patient may require fluid administration (Deflandre et al. 2008).

It is clear that future methods that process available physiologic signals will be applied to enhance and improve the availability of important measures of cardiac function, a key parameter for making treatment decisions used by critical care caregivers.

21.5.3 Pulse Oximeter Signal Acquisition and Processing

One of the most common technologic devices used in hospitals today is the pulse oximeter; continuous monitoring of oxygen saturation became a standard of practice in the early 1990s (Brown et al. 1990). The pulse oximeter sensing device is usually placed on a finger and measures oxygen saturation and pulse rate, or HR (Clark et al. 2006). The modern device works by shining red and infrared light generated by two light emitting diodes through the tissue. With each arterial pulse there is a variation in the light as it passes through the tissue and is sensed by a light-sensitive photodiode on the opposite side. The more oxygenated the blood is, the more red light is transmitted, with less infrared light passing through. By calibrating these devices, reasonably accurate estimates of oxygen saturation (SpO_2) can be determined. Although the pulse oximeter is convenient and easy to use, it has several

important limitations, including motion artifact, when the patient moves, and other physiologic considerations such as anemia, low perfusion state and low peripheral skin temperature. If the blood flow to the hand gets disturbed, by perhaps squeezing the arm during BP with a sphygmomanometer, the blood flow to the hand is interrupted and the pulsatile BP signal required for the pulse oximeter is no longer available. The pulse oximeter is one of ICU monitoring devices with most false alarms (Malviya et al. 2000).

21.5.4 Bedside Data Display and Signal Integration

While colorful and dynamic, the displays on the bedside monitor can be complex. Figure 21.6 shows a typical bedside monitor display as one example.

In 2018 Medtronic led the US patient monitoring device market by capturing the largest segment of pulse oximetry monitoring after acquiring Covidien, the previous leader of pulse oximetry monitoring. With one of most extensive product lines, Philips Healthcare is leading multiparameter vital sign monitoring, wireless telemetry, and fetal/neonatal monitoring markets.

The other major manufacturers in the patient monitoring market are: GE Healthcare, Masimo, Edwards Lifesciences, Mindray Medical, Natus Medical, Welch Allyn, Omron Healthcare, Honeywell Life Sciences, Nihon Kohden, Spacelabs Healthcare, St. Jude Medical, Nonin Medical and Boston Scientific. Each vendor of bedside monitors has made a "best effort" at displaying the variety of physiologic signals derived. In most cases this consists of three channels: ECG, ABP, and pulse oximetery. Additional important physiologic parameters can be derived from these signals, as noted in Table 21.1.

Today's bedside monitors still present both waveforms and derived parameters in

Table 21.1 Bedside physiologic monitoring capabilities

Modality	Transducer	Frequency		Additional parameters		
ECG	Chest electrodes	Continuous	Heart rate	Heart rhythm	Complete ECG waveforms	Pacemaker signal
Arterial blood pressure invasive	Catheter & blood pressure transducer	Continuous	Heart rate	Systolic, diastolic & mean pressure	Estimates of cardiac output	Pulse pressure variation & fluid loading
Pulse oximeter	Finger probe	Continuous	Arterial oxygen saturation	Heart rate		
Temperature	Skin sensor	Continuous	Temperature			
Respiration	Chest belt	Continuous	Respiratory rate			
Bioreactance	Electrodes	Continuous	Cardiac output	Heart rate	Stroke volume	
Arterial blood pressure non-invasive	Inflatable cuff	Intermittent	Heart rate	Systolic, diastolic and mean pressure		

Abbreviation: *ECG* electrocardiogram

a single-sensor-single-indicator format. That is, for each individual sensor attached to the patient there is a single indicator, a waveform with derived values presented on the screen (Drews et al. 2008). One of the simplistic consequences of this display strategy is that each indicator is treated as if it had come from a different patient. For example, if ECG, ABP, and pulse oximeter signals were displayed, they would each have the capability of determining HR. Thus, three different HR measures might be displayed. Although there are physiologic reasons for such differences, the most common situation is that the HR should be an integrated assessment of the three signals since artifact is a far more common event than the unusual conditions that would cause the differences in HR. Studies suggest that there are better methods for designing hemodynamic monitoring displays (Doig et al. 2011; Drews et al. 2008).

A more important problem relates to the integration of data from multiple bedside devices. Two examples will illustrate the problem:

1. The patient's pulse oximeter has shown a recent increase of SpO_2. However, the bedside monitor has no knowledge that the respiratory therapist has increased the FIO_2 from 30% to 40% on the ventilator.
2. The patient's HR has recently increased from a dangerously low value of 45 beats per minute to 72 beats per minute. Unfortunately, the bedside monitor has no way of knowing that a nurse has increased the drip rate of a cardioactive medication.

Patients in today's ICUs can have 50 or more electronic devices attached (Mathews and Pronovost 2011). Many of these electronic devices were developed by independent companies and do not easily interface or communicate with each other. However, even though the larger monitoring companies have purchased several of the specialty monitoring companies, problems still exist although it was understood more than three decades ago, and standards for bedside data interchange (CEN ISO/IEEE 11073) (Gardner et al. 1989, 1991) were developed. The Medical Information Bus (MIB) is the simple term used to designate CEN ISO/IEEE 11073. So, why has the MIB been a commercial failure to this point? There are multiple reasons; unfortunately, the MIB standard was designed during the time when serial communications via RS-232 was the norm; there were no Universal Serial Bus (USB) interfaces or convenient wireless devices (Wi-Fi or Bluetooth) at the bedside. Furthermore, each vendor of bedside devices and ICU data management systems would like to be the "data integrator" (for a price) and thus has little incentive to adhere to standards that would allow other vendors to compete for the integrator role. The business model apparently has not worked (Kennelly and Gardner 1997; Mathews and Pronovost 2011).

In spite of the lack of interface standards, the group at Intermountain Healthcare, LDS Hospital (Salt Lake City, Utah) has been actively interfacing ventilators, IV pumps, and similar devices for over three decades (Dalto et al. 1997; Vawdrey et al. 2007). Another effort is The Medical Device "Plug-and-Play" Interoperability Program at Partners HealthCare (Boston, Massachusetts) – ▶ http://www.mdpnp.org. The open source result is OpenICE, implementation framework of an Integrated Clinical Environment.

21.5.5 Challenges of Bedside Monitor Alarms

Care of the critically ill is complex and challenging. Most of these patients have medical problems or injuries that are life threatening. They might have heart problems that within minutes could result in sudden death, or they might have breathing problems that require mechanical ventilation to maintain life. As a consequence, each of these situations requires intense minute-by-minute observation with real-time, continuous physiologic monitoring. For those conditions, the requirement for record keeping, monitoring, and alarming is intense.

There are clear expectation that bedside physiologic monitors, ventilators, IV pumps,

and similar devices attached to patients should provide true and valid alarms and that caregivers will be promptly notified and provide the needed care immediately for those patients (Kowalczyk 2011). On the other hand, a report from the New England Journal of Medicine outlines 24 electronic requirements for classification of a hospital as having a comprehensive electronic record system (Jha et al. 2009), yet recording of data from bedside physiologic monitoring systems with their alarming systems and data gathering from other bedside devices such as ventilators and IV pumps were not even mentioned.

So, currently there is a curious and inexplicable set of expectations being generated for care of the critically ill patients. As a consequence, there are niche vendors who have built their own data gathering and recording systems and nurse charting systems; in some cases these systems include simple interfaces to allow them to acquire laboratory data and perhaps data from the administrative admissions process. They may even include bedside computers or displays to allow care givers to have access to such things as radiographic images, dictated reports, and others. However, these systems are stand-alone devices and do not typically provide interfaces to *transmit* their physiologic data to the hospital's EMR.

In the past, the number of physiologic signals that can and are being monitored has grown. With each signal and derived parameter that is added there is typically a high and low alarm added to warn the clinical staff of actual or impending patient crisis. Alarms may be highlighted on the bedside monitor's screen by using a color change or flashing indicators. Most alarms also generate a sound.

Imhoff and Kuhls (2006) noted from 1.6 to 14.6 alarms for each ICU patient each hour; up to 90% of those alarms were false. Alarm overload is clearly a significant issue in ICU monitoring; from clinical informatics professionals working in the ICU is needed to minimize the number of false alarms. Just noting the titles of several editorials and articles should be informative:

1. Alarms in the intensive care unit: How can the number of false alarms be reduced? (Chambrin 2001)
2. Monitoring the monitors – beyond risk management (Thompson and Mahajan 2006)
3. Alarms and human behavior: Implications for medical alarms (Edworthy and Hellier 2006)
4. Alarms in the intensive care unit: Too much of a good thing is dangerous: Is it time to add some intelligence to alarms? (Blum and Tremper 2010)
5. Intensive care unit alarms – How many do we need? (Siebig et al. 2010)

Figures 21.7 and 21.8 give examples of the complexity of determining whether an alarm is true or false based on two life-threatening conditions. Alarms for ventricular tachycardia are shown in Fig. 21.7. Figure 21.7a shows a true ventricular tachycardia alarm condition while Fig. 21.7b shows a false ventricular tachycardia condition. Figure 21.7b has only a few seconds of ECG artifact, which causes the bedside monitors' alarm detection system to issues an alarm.

Arterial hypotension alarms are shown in Fig. 21.8. Figure 21.8a shows a true arterial hypotension alarm condition while Figure 21.8b shows a false condition. If the monitor or human observer only watches the ABP signal, the two conditions appear similar. However, by simultaneously following the ECG signal, the human observer will note that for some unknown reason the ABP signal displays a false representation of the patient's pulsatile BP. The unknown reason is likely related to the catheter and tubing parts of the arterial monitoring system. Alerting the clinical staff to examine the catheter and transducer system is certainly appropriate.

Biomedical informatics specialists, biomedical engineers, and bedside monitor vendors have recently renewed their efforts to reduce false alarms and improve the relevance of existing alarms. Most of the false alarms are caused by noise or artifacts in the primary signals. To help minimize these problems, two examples are used to illustrate the challenges and opportunities to improve bedside alarms.

1. After observing over 200 hours of alarms from bedside monitors and ventilators in an adult medical ICU, Gorges and his

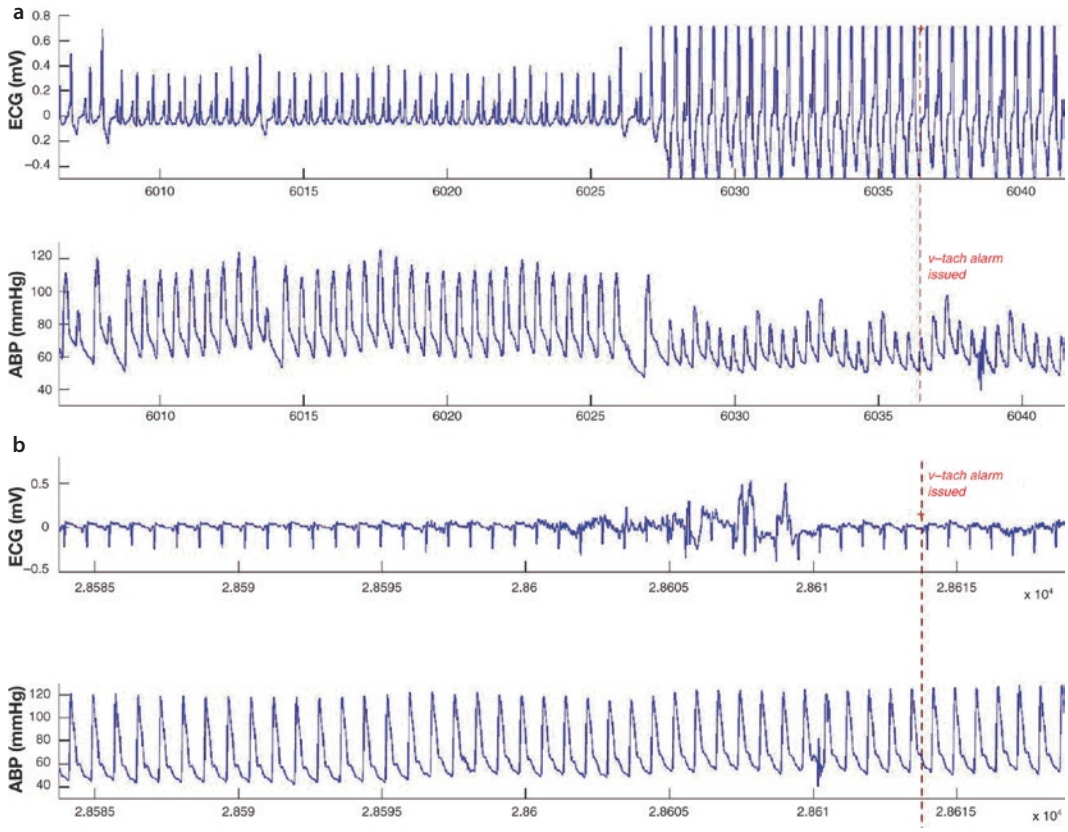


Fig. 21.7 Ventricular tachycardia alarm conditions. **a** A true alarm; note that the ventricle is still pumping but that the arterial pulse pressure is dramatically reduced. **b** A false alarm caused by artifact in the ECG signal; note the ABP waveform is stable during the same time interval. ABP indicates arterial blood Pressure, ECG electrocardiogram, v-tach ventricular tachycardia

colleagues (2009) used the data recorded to recommend a two-step process that would dramatically reduce the number of false alarms. The first step was to add a 19 second delay into the alarming system. That step by itself reduced the number of alarms by 67%. They then noted that by having some method for automatically detecting when a patient was being suctioned, repositioned, given oral care or being washed, there would be a further 13% reduction of ineffective alarms. By using these just these two methods, almost 80% of the false alarms could be eliminated.

2. Using multiple signals to derive identical measures should be an effective method of reducing false alarms (Herasevich et al. 2013). As will be noted in Fig. 21.7, there are five signals that can be used to derive HR: ECG 1, ECG 2, ECG 3, ABP, and pulse oximeter. Since the probability of all those signals having an artifact is smaller than any single physiologic signal, smart alarm algorithms that are more robust should be possible. Two investigators have developed and tested such algorithms (Zong et al. 2004; Poon 2005). The Zong pressure alarm algorithm reduced false alarms from 26.8% to 0.5%. Poon found that the usual HR and rhythm alarm system produced 65.4% false alarms, while an algorithm that integrated multiple signals generated only 31.5% false alarms. Two other findings from the Poon study were also encouraging. By merely delaying the alarms by 10 seconds there was a 60% reduction in false alarms. In addition, he found that default settings for high and

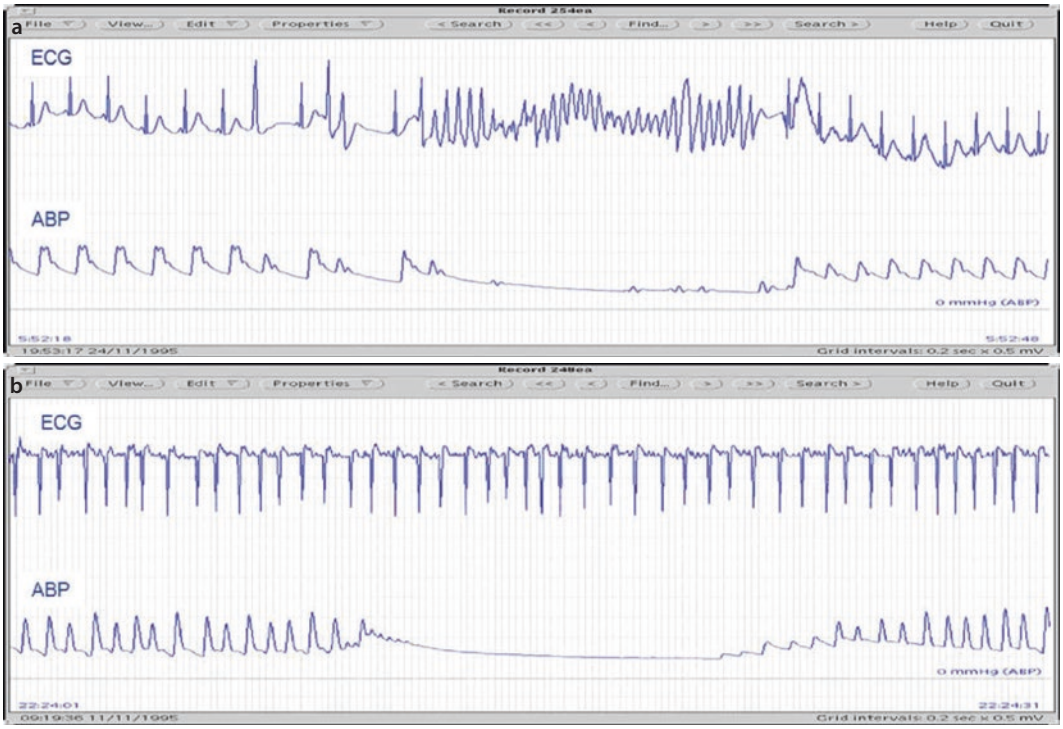


Fig. 21.8 Arterial hypotension alarm conditions. **a** A true alarm; note the normal ventricular beats followed by ventricular fibrillation that renders the heart unable to generate an effective blood pressure. **b** A false alarm; note for some nonphysiologic reason the arterial pressure signal loses its pulsatile characteristics and then eventually it returns. ABP indicates arterial blood pressure, ECG electrocardiogram

low HR alarms were not optimized to prevent false alarms. For example, if a patient had an average HR of 65 beats per minute and the default low HR alarm was 60 beats per minute, there was an increased likelihood of false low HR alarms. Several bedside monitor vendors now provide these more sophisticated alarm algorithms in their newest monitors.

Still other informatics specialists have found different strategies to provide more accurate ABP and cardiac arrhythmia alarm rates (Aboukhalil et al. 2008; Zhang and Szolovitz 2008). Having electronic archives of physiologic waveforms that are publically available should permit development of even better smart alarm algorithms, which should lead to a reduced number of false alarms generated by bedside monitors (Saeed et al. 2011; Burykin et al. 2011).

21.5.6 Biomedical Sensors

One hundred years after the introduction of traditional vital signs they are still in use despite them not describing what exactly happens with patients. Current noninvasive technologies for rapid physiologic function do exist, but have not replaced traditional measurements. Heart function could be measured by stroke volume, arterial oxygenation by pulse oximetry, and ventilation using capnography.

21.5.7 Strategies for Incorporating Bedside Monitoring Data into an Integrated Hospital EMR

Three general strategies are currently used to transfer bedside monitoring data into the hospital's EMR.

The first is the simplest and still widely in use: nurses observe data presented on the bedside monitor screen and manually key-in the observations into an integrated EMR. As simple as this may be to implement, such manual data collection strategy is inefficient, error prone up to 15% when documented on paper and then typed into EMR (Wager et al. 2010), and does not collect representative data gathered by the bedside monitor.

The second strategy used by ICU information systems is to acquire vital sign data directly from the bedside monitoring system's network by using an HL7 feed (see ▶ Chap. 7). The information is automatically gathered by the ICU information system and nurses have the option of either accepting or modifying the data. In typical clinical settings, nurses perform the selection and transfer of bedside monitoring data from the ICU information system to the EMR about once an hour. These ICU information systems typically retain the high frequency bedside monitoring data and can achieve near-real-time computerized decision support. In many cases, the nurse's notes are also entered into the ICU information system – generally once per shift – and some summary vital sign information may find its way into those notes. Physician progress notes are also entered into ICU information systems in a similar fashion. Unfortunately, data in the ICU information system may never find its way into the hospital's EMR. For these systems, the ICU data are usually archived separately. As a consequence, these data cannot be used for real-time decision making by the hospital's EMR.

The third strategy is to have the ICU information system or the hospital's EMR system automatically transfer vital sign data from the bedside monitoring system to the EMR. Most systems that automatically gather data with this strategy take a median of the vital sign data over a 15 minute time interval to smooth the data (Warner et al. 1968; Gardner et al. 1991; Vawdrey et al. 2007). This strategy provides real-time data for computations and computerized decision support for the hospital's EMR and is the preferred strategy.

There are opportunities to improve the automated data gathering from bedside monitors, especially if the false alarm rate can be minimized. In addition to acquiring 15-minute median data, one may wish to detect bedside alarms and record data in the intervals just before and just after these alarms. Thus, there is still opportunity for informaticians to make major improvements in both data recording and bedside monitoring alarms.

21.6 Monitoring and Advanced Information Management in ICUs

21.6.1 Early Pioneering in ICU Systems

As electronic information gathered in ICUs and hospitals started growing, problems with bedside monitors, alarms, and data integrations become more apparent. The next step in information management began 50 years ago at LDS Hospital where a team of people developed what was known as the HELP (Health Evaluation Through Logical Processing) System (Pryor et al. 1983; Kuperman et al. 1991; Gardner et al. 1999).

HELP was the first hospital information system to collect patient data needed for clinical decision making and at the same time incorporate a medical knowledge base and inference engine to assist the clinician in making decisions. The HELP system has been operational at LDS Hospital since 1967. The system initially supported a heart catheterization laboratory and a postoperative open heart ICU. Initially only physiologic data were acquired from the bedside monitors. Nursing note charting promptly followed with ability to chart medications ordered and given, including IV drip rates. Soon, it became apparent that much of the data needed to care for these critically ill patients came from the clinical laboratory and other sites such as radiology. As a consequence, multiple modules were added to the HELP system to support the ICUs.

Clinical decision making in the ICU is complex. Physicians, nurses, respiratory therapists, pharmacists, and others evaluate

Table 21.2 Data used for ICU decision making and their sources

Data types	%	Data source
Clinical & blood-gas laboratories	42	Laboratory interfaces
Drug I/O IV	22	Nurse charting & IV pump interface
Observations	21	Nurse charting & physician notes
Physiologic data	13	Bedside monitor interface
Other	2	

Adapted from Bradshaw et al. (1984) (See Fig. 21.3)
Abbreviations: *ICU* intensive care unit, *I/O* Intake and Output, *IV* intravenous

each patient using different types and modalities of data. In 1984, a study was performed to identify what data were used by the critical care team to make clinical care decisions (Bradshaw et al. 1984). The investigators were surprised to find that data from the physiologic monitor accounted for only 13% of the data used to make treatment decisions. Table 21.2 outlines the data types evaluated with the percentage of time that each type of data was used to make a care decision. Many of the data came from automated instruments in the laboratory, but a large number came from nurse observations and actions that were manually charted into the computerized record.

However, as described earlier, the physiologic monitor serves a very crucial function during life-threatening situations such as cardiac arrest. The observations showed the crucial need for a fast and reliable laboratory interface *and* the importance of data that came from nurse charting. Knowing what drugs the patient was receiving, when those drugs were given, and the types and administration rates of IV medications were crucial to clinical decision making.

21.6.2 Recent Advances in ICU Clinical Management Systems

We are now almost two decades after the Institute of Medicine's, "To Err is Human" report (Kohn et al. 2000) and still medical error remains one of the leading causes of poor outcome of hospitalized patients (Landrigan et al. 2010). Despite high hopes, the current generation of EMR has made things worse, particularly in acute settings where information overload poses a major challenge to timely, evidence-based patient care (Han et al. 2005; Pickering et al. 2012). When caring for unstable patients, providers often have only a few minutes to wade through medical records before making critically important decisions. Clinical decision-making is often hindered by patient information that is difficult to access and use in modern EMRs, which increases the potential for error and delays treatment.

Beginning in 2009 a group of clinicians, clinical informatics specialists, and computer programmers at Mayo Clinic (Rochester, Minnesota) developed Ambient Warning and Response Evaluation System (AWARE) system to address those challenges. AWARE is a data assimilation, communication, workflow, and decision support tool that has the ability to enhance EMR experience (Ahmed et al. 2011). The system is configured to allow a more rapid assessment of patient's clinical data, freeing time to focus on other important patient needs. AWARE has been designed, tested, and validated to foster the best clinical practice by specifically addressing the key clinical components known to improve patient outcomes (Olchanski et al. 2017).

A suite of tools is designed to address patients' needs and focus on acute patient-centered problems rather than organized around specific database or clinical services.

The main components of AWARE could be divided to five domains (Fig. 21.9):

1. Dashboards/viewers. Included are Multi-patient (MPV) and Single-patient viewers (SPV) to reduce information overload by

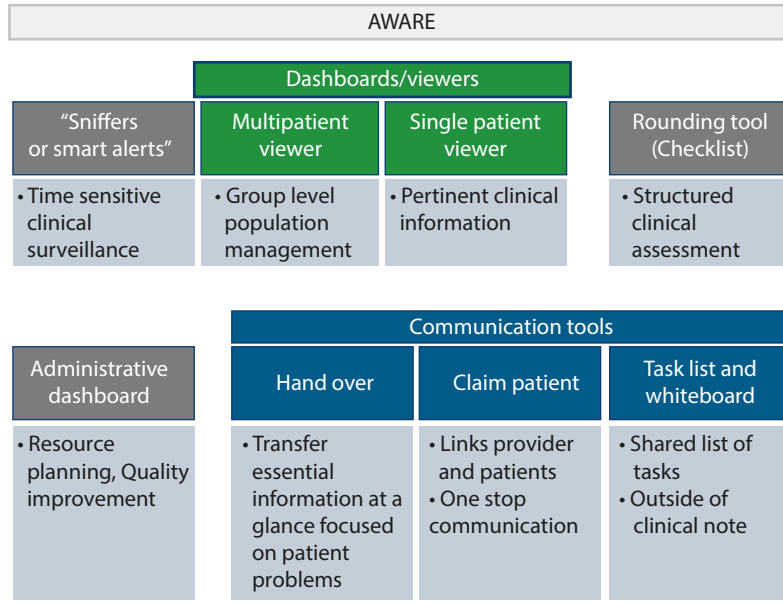


Fig. 21.9 Essential elements and logical structure of the Ambient Warning and Response Evaluation System (AWARE) clinical management system

facilitating real-time access to key information needed for timely medical and interventional decision making at the point of care.

2. Sniffers or rule-based smart alerts. Continuously survey both the patient condition and provider actions detecting potential mismatches and preventing potential errors before they occur.
3. Communication tools. AWARE whiteboard, task-list, readiness for discharge and claim patient functions facilitate communication between team members and during transitions of care, thereby preventing common errors of communication omission.
4. Checklist/ rounding tool. Designed to assists providers in developing and executing a coordinated daily plan of care. The easy-to-use interface minimizes clerical burden while simultaneously assuring adherence to patient-centered best care practices and regulatory requirements.
5. Administrative dashboard. Feedback and reporting tool enables easy access to quality improvement metrics and patient outcomes for administrators and oversight groups, which facilitates rapid-cycle management changes based on continuous feedback.

21.6.3 Acquiring the Data: Quality and Timeliness

A fundamental part of any computerized decision support system, just as with any human clinical decision support system, is the acquisition of data. Clinicians develop observational, interpersonal, and technical skills as they collect accurate patient data. Likewise, a computerized decision support system depends on high-quality, timely data. In many ICUs today, much medical data still continues to be entered into computerized patient records as scanned PDF files or in a structured and coded formats while others (such as the progress note) are be stored in a free text format (either handwritten or typed) (Celi et al. 2001; Pickering et al. 2010; Ahmed et al. 2011; Hripcsak et al. 2011). As noted in ▶ Chap. 8, natural language processing of free text to obtain coded and structured information has seen great improvement over the past decades; however, the process is still far from perfect and all processing is delayed until data is entered to EMR. Such delays is limiting factor to use such free text data in real time monitoring systems.

As designers of clinical monitoring systems look at acquiring and entering clinical data they must decide:

1. *Who* should enter the data: automated acquisition from electronic instruments (such as the bedside monitor) versus manual entry using a keyboard, bar code reader, touch screen, voice input, or some similar method.
2. *When* to enter the data: accurate ICU decision making often requires data to be acquired in a timely manner, sometimes within 1 minute of an event to make a timely decision.
3. *Where to enter the data*: this automated data will naturally be acquired from the bedside monitor or instrument located at the bedside; manual data entry should optimally occur at the bedside as well.
4. *How* data should be collected: methods should take into account the occurrence of artifacts in the patient data; many EMR systems allow nurses to review and validate bedside vital sign data minutes to hours after they are collected, although this process does not meet the requirement for real-time data collection and can lead to "human" and computerized decision support errors (Nelson et al. 2005; Vawdrey et al. 2007)
5. *How much* data to collect: this is particularly an issue with systems such as bedside monitors that *can* generate an HR, systolic and diastolic BP value for each heartbeat, resulting in hundreds of thousands of values per day; except for special situations to use in automatic decision support systems, the collection of such intensive data in regular EMR is inappropriate.

The process of developing and implementing the systems for acquiring data involves not only technology, but adapting that technology to the human users; training those users to properly use the new system is complex and difficult. Consequently, developers and adopters of such systems should plan for and be prepared for challenges that may take years to implement and optimize system.

Despite modern protocols like HL7 and FHIR there are still major problems with acquiring ICU data to EMR either automatically or manually (Gardner et al. 1989, 1991; Dalto et al. 1997; Nelson et al. 2005; Vawdrey et al. 2007). Data from bedside monitors, ventilators and IV pumps should be acquired automatically with a real-time technology. Data thus acquired is timely and by appropriate signal processing methods can be validated (Dalto et al. 1997; Vawdrey et al. 2007; Ahmed et al. 2011; Lilly et al. 2011). Changes in ventilator settings such as FIO_2 may only be present for a few minutes, but blood-gas measurements taken during that time interval will be misinterpreted if only manual electronic charting is used. Similar interpretation errors were found to occur with IV pump drip rate charting when manual charting methods were compared to automated acquisition. Gathering accurate, representative, and timely computerized ICU data requires attention to detail and careful planning to assure its quality. Bedside charting systems have the ability to capture near real-time data from bedside devices, but the presentation layer usually showed normalized and averaged values.

21.6.4 Presentation of Data

Once data have been collected, their quality verified, and the results stored, one must decide how the data should be presented. Currently, most data are presented on a colorful screen. However, some care givers will still prefer a paper copy. Still others will prefer to view these reports on their smart phones or other mobile devices. For ICU patients, it is clear that specialized reports must be developed. The traditional method of segmented reporting (separate reports for laboratory data, vital signs reports, medication lists, etc.) has proven inadequate (Clemmer 2004; Ahmed et al. 2011). The ICU group at Mayo Clinic has developed and tested an ICU rounding tool (Pickering et al. 2010). Thus, one can see there is value in the integration of and presentation of data. As of this writing, there is probably not a single ICU summary report that will satisfy all ICU users. Thus, such reports will require special effort for each institution and perhaps even each

ICU within that institution. For example, the report generated for a thoracic ICU is unlikely to be identical to that required by the neonatal ICU. Accomplishing such tasks typically requires 6 months or more, with continuous ongoing effort to update the report as new data are acquired and caregivers needs evolve.

21.6.5 Establishing the Decision Rules and Knowledge Base

Deciding on the decision rules that should be installed in a computerized ICU decision support system is difficult. Health care is currently driven by implementing evidence-based protocols. However, few of these protocols have been computerized. The long-standing work with the HELP system and some exciting work done at Mayo Clinic and at the University of Massachusetts are exceptions (Clemmer 2004; Morris 2000; East et al. 1992; Ahmed et al. 2011; Lilly et al. 2011). Using a consensus process to develop treatment decisions is essential. However, generating a consensus is a tedious, difficult, and slow process. At the moment, the consensus process involving all the clinical caregivers in the ICU is the best approach, as rules developed by individuals are often not widely accepted or used. However, in some departments there may be trusted clinical leaders who become the "local expert." Developing the rules for clinical decision support is complex and those rules are always subject to change. Development of appropriate rules can take up to 6 months and the rules will need to be continuously reviewed and updated (Gardner 2004; Ahmed et al. 2011; Lilly et al. 2011).

21.6.6 Clinical Charting Systems: Nurses, Pharmacists, Physicians, Therapists

The major portion (43%) of the data used at LDS Hospital for decision-making during ICU rounds came from clinical notes and data charted by nurses and other clinicians (Bradshaw et al. 1984). In a more recent study in the ICU at Mayo Clinic, the team found that as they developed AWARE system, they required similar data content (Pickering et al. 2010).

At LDS Hospital, the computerized nurse charting module allows nurses to enter patient care tasks, qualitative and quantitative data, and a patient's response to therapy (Willson 1994; Willson et al. 1994; Nelson et al. 2005). In addition, nurses interact with a pharmacy module to chart all given medications including IV drip rates (Pryor 1989; Kuperman et al. 1991).

Soon after the nurse charting was implemented at LDS Hospital, respiratory therapists chose to enter their qualitative and quantitative ventilator data and care given to patients (Andrews et al. 1985; Gardner 2004). The motivation for the online charting was to provide clinicians with access to timely and accurate data to make patient care decisions. In addition, these data could be used to implement protocol-controlled ventilator weaning systems (East et al. 1992; Morris 2001).

To optimize the performance of routine care deemed essential for ICU patient recovery, computerized reminders were generated (Oniki et al. 2003). For example, 1 of the goals of the reminders was to provide assistance in determining the required level of sedation while avoiding oversedation. By providing the computerized reminders to nurses, charting deficiencies were reduced by 40% and the number of deficiencies at the end of the shift was improved. To optimize care provided by the reminders, real-time charting was required. However, during a quality improvement process, it was determined that 29% of the medication errors that should have been prevented by online nurse charting were still present. A careful evaluation revealed that the actual nurse charting workflow was different than that envisioned by the system planners. Instead of charting the given medication using a bedside terminal, nurses administered the medication and then at some later time, at the central nursing station, charted that the medication had been given. Consequently, errors were occurring. After careful training and feedback with the nursing staff, the real-time charting rate increased from 40% to 75% and remained at that level a year later. This example shows that having computerized decision

support systems in place without having real-time data entry was ineffective. Conceptually, one could make the same logical observation if the ICU were operating as a tele-ICU, as discussed later in this chapter.

For generations, nurses and other caregivers who have used conventional paper records have had the notion that if their paper chart was up-to-date at the end of the shift then they had met their requirements for good patient care. Clearly, the above example shows that such a strategy is flawed. However, it is interesting that even today reports are being made about charting and use of data for end-of-shift nursing care exchanges and patient handovers, suggesting that the EMR still may not be real-time (Hripcsak et al. 2011; Collins 2011). Collins (2011) found that clinicians preferred oral communications compared to EMR documentation and stated that the perceptions that the EMR was a shift behind might have only been a manifestation of the lack of real-time charting by nurses and acquisition of real-time data from bedside monitors in their ICU.

An early survey of nurses and physicians use of the HELP clinical expert system was conducted in 1994 (Gardner and Lundsgaarde 1994). The investigators were encouraged by a positive response from both nurse and physician users who appreciated having the data available with interpretation and alerting features provided by the HELP system. At the time the survey was conducted, ICU charting and decision support was a major feature of the HELP system. It is exciting to note that other institutions have begun to assess factors related to acceptance of an EMR in critical care (Carayon et al. 2011). The Carayon study showed that ease of use, as well as data presentation strategies, were major determinants of acceptability of their system.

21.6.7 Automated Data Acquisition From All Bedside Devices

Computer systems that support ICU patients are tightly integrated and data are automatically gathered and stored, primarily in a coded format so that real-time computerized decision support can be used. Figure 21.10 shows a schematic of the HELP system at LDS Hospital and Fig. 21.11 shows AWARE system at Mayo Clinic as examples of such systems. Based on the data available to ICU management systems from these multiple data sources, its computerized decision support system makes and displays suggestions for optimum care for the specific problems such as sepsis and acute respiratory distress syndrome. The system provides audible and visual alerts for life-threatening situations. In addition, the system organizes and reports the large amount of data so that the medical team can make prompt and reliable treatment decisions. The patient's physicians are automatically alerted about life-threatening laboratory and other findings.

Much of the information required for ICU patient care comes from underlying laboratories and devices that automatically acquire data. In the upper right hand corner of the HELP system diagram, data from the ventilator, IV pumps, and the bedside monitor are noted. While most of the physiologic bedside monitor vendors now acquire ECG, BP, and pulse oximetry data, they do not provide access to data from ventilators or information from IV pumps. As a consequence, data from these devices must be obtained by developing hardware and software interfaces (Gardner et al. 1991; Dalto et al. 1997; Kennelly and Gardner 1997; Vawdrey et al. 2007). Based on those studies, it is clear that automatically collecting data from all of these devices in real-time is more timely and accurate than manually charted data collected by nurses or respiratory therapists. Although data from these devices can contain artifacts, methods for minimizing those artifacts have been implemented in operational systems.

Initially the MIB standard CEN ISO/IEEE 11073 was designed to help gather data from bedside devices, but has not been widely implemented (Mathews and Pronovost 2011). Fortunately, battery power and wireless communications with IV pumps are now widely available. By using wireless technology, inter-

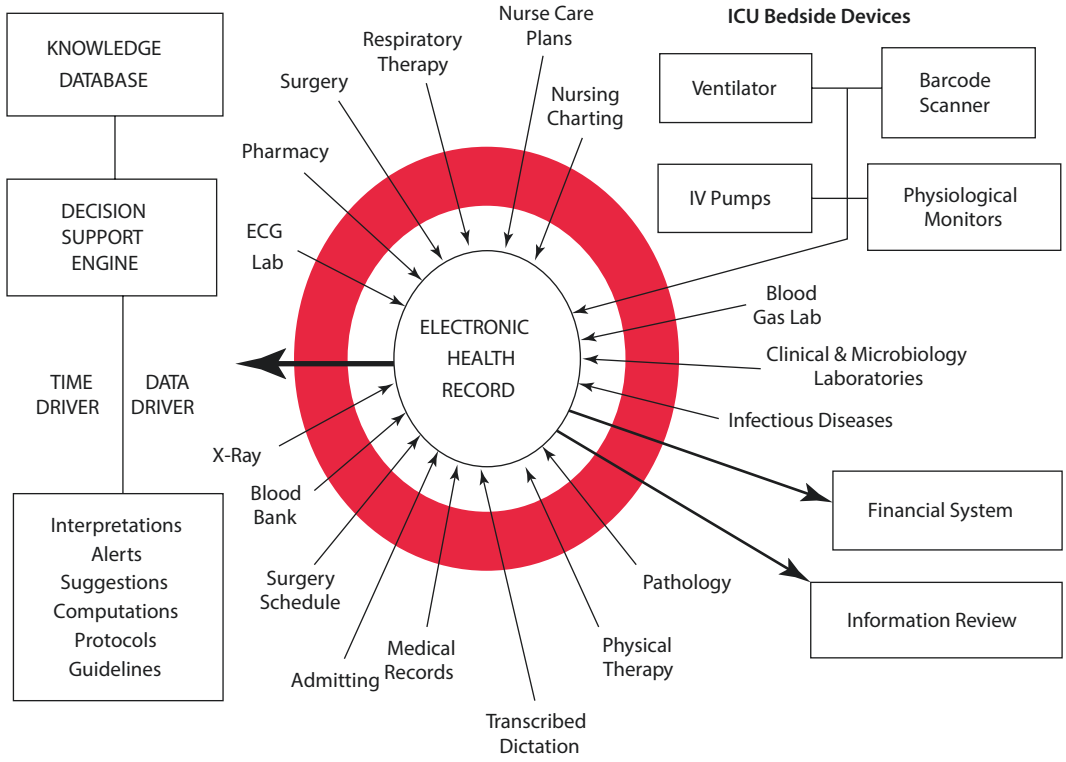


Fig. 21.10 Diagram of HELP the System Used by Intermountain Healthcare's Hospitals. At the *center* is the database for the electronic medical record (*EMR*). Data from a wide variety of clinical and administrative sources flow into the EMR. As the data flows into the EMR, the Data Driver capabilities of the HELP Decision Support Engine (*red circle*) are activated. In addition, Time Driver decisions are also made. Shown schematically, in the *upper right* hand corner of the diagram are blocks representing ICU bedside devices including the physiologic monitor, ventilator, IV pumps and barcode scanner. ECG indicates electrocardiogram, HELP Health Evaluation Through Logical Processing, ICU intensive care unit, IV intravenous

faces with the IV pumps are fast, mobile, and easy for nurses to implement and tangled wires are no longer an issue. In addition, communications with device can be carried throughout the hospital – in the OR, while on transport, and in the ICU.

Although early studies of nurses and therapists showed that computerized charting took longer than manual charting, it is almost certain with automated acquisition available today that charting takes less time and is more accurate. As a consequence, in institutions that have historically collected IV pump and ventilator data automatically, there is a commitment to collect data from every bedside monitoring device. These include measures of urine output, fluid drainage, and similar measures.

21.6.8 Rounding Process: Single Patient Viewer

The care of critically ill patients in the ICU requires collaboration among a diverse team of very competent caregivers to achieve the best care (Clemmer et al. 1998). The teamwork and communications is required in this complex care process.

The rounds activity at LDS Hospital is an example of the collaborative process. Figure 21.12 shows the clinical care team during rounds. There are physicians, house officers, advanced practice clinicians, nurses, pharmacists, respiratory therapists, dieticians, case managers, and others who gather each day to assess each patient and make key care

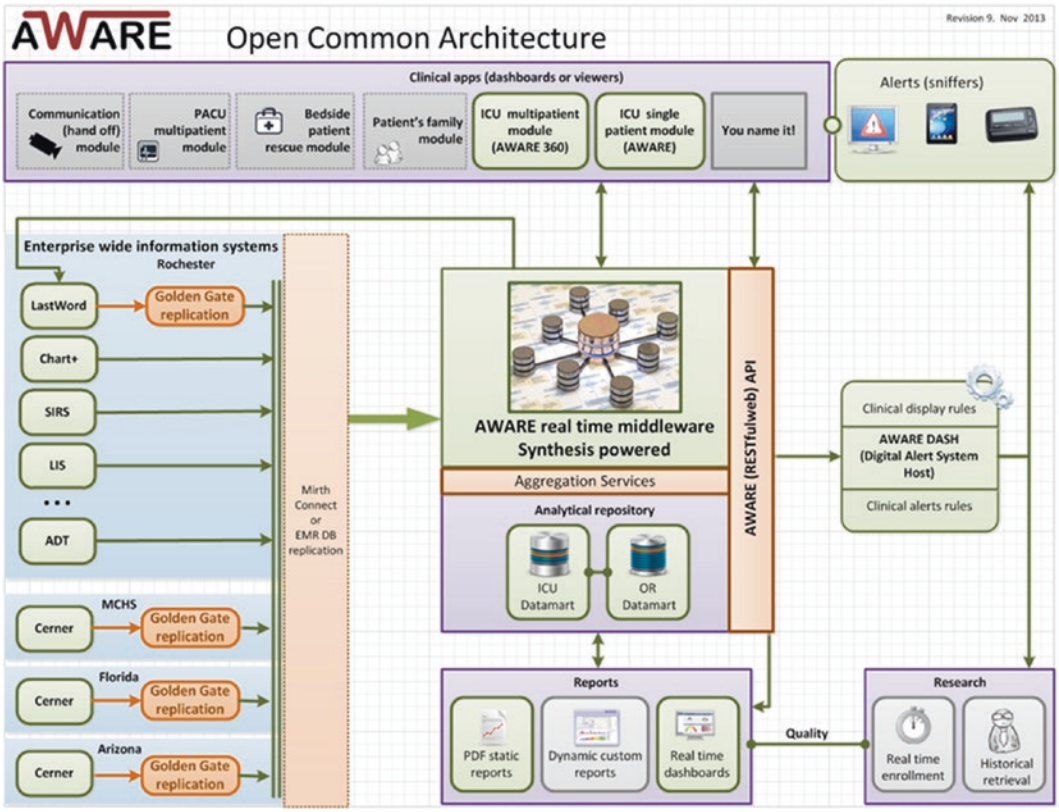


Fig. 21.11 Diagram of AWARE system developed and used at Mayo Clinic. The core component is middleware database that cached data from hospital EMR systems (left). It utilized number of approaches such as DB stored procedures and HL7. Synthesis is Mayo Clinic home-grown API layer and DB. Data from discharged patients archived and stored in Datamart repository. Historical data is used for administrative reporting and cohort identification for research. The alerting module (right) is used for clinical sniffers and research enrollment in clinical trials. On the top are schematically shown clinical applications or viewers that use data from middleware to generate user interfaces. ADT indicates; API,; apps applications, AWARE Ambient Warning and Response Evaluation System, EMR DB Electronic Medical Record Database, HL7 Health Level Seven, ICU intensive care unit, LIS Laboratory Information System, MCHS Mayo Clinic Health System, OR operating room, PACU Post Anesthesia Care Unit, SIRS Surgical Information System

decisions. The rounds leader is usually a physician, but each team member is considered an equal partner, providing key information (most of it stored in the computer record) and given the opportunity to discuss their interpretation and make recommendations about the patient's care.

Over decades, the social process of conducting these rounds has created a very open and cooperative environment. The purpose of rounds is to reduce errors from human factors, to give structure to the evaluation, and to make sure all sides of the decision process are considered as each member considers the decisions from their point of view.

The information from the computer system is organized to support the process. The computerized record is *the* patient record. Information from other sources such as radiographic images and free-text reports are also readily available (Gurses and Xiao 2006).

As a single patient viewer, AWARE extracts data relevant to the treatment of ICU patients and presents it to the provider in a systems-based information package (Fig. 21.13). AWARE content has been selected through the systematic observation of frontline provider information needs and profiling of provider data utilization patterns. The user interface has been optimized to sin-

Fig. 21.12 ICU Rounds room at LDS Hospital in Salt Lake City. The computerized ICU rounds report is displayed by a projector on the wall to physicians, a nurse practitioner, medical students, a respiratory therapist, a pharmacist, and a patient's family member. An important laboratory result is highlighted in red by the rounds director. Note several laptops *and* paper notes used by each of the participants. ICU indicates intensive care unit

gle screen without scrollbars and can be used an enhancement to the bedside monitor during treating critically ill patients.

User interface and organization of data elements on the screen was determined by considering how experts incorporate information into decision-making mental models. Reference ranges for laboratory abnormalities in AWARE are adjusted for critically ill patients based on expert consensus. All this information about reference ranges, alerts, and type of information represented on the interface is embedded in AWARE rules. It is part of DB and does not utilize any third-party rules engine. This approach decreased false-positive alerts without affecting the fraction of false-negative alerts (unchanged sensitivity and negative predictive value) (Kilickaya et al. 2014).

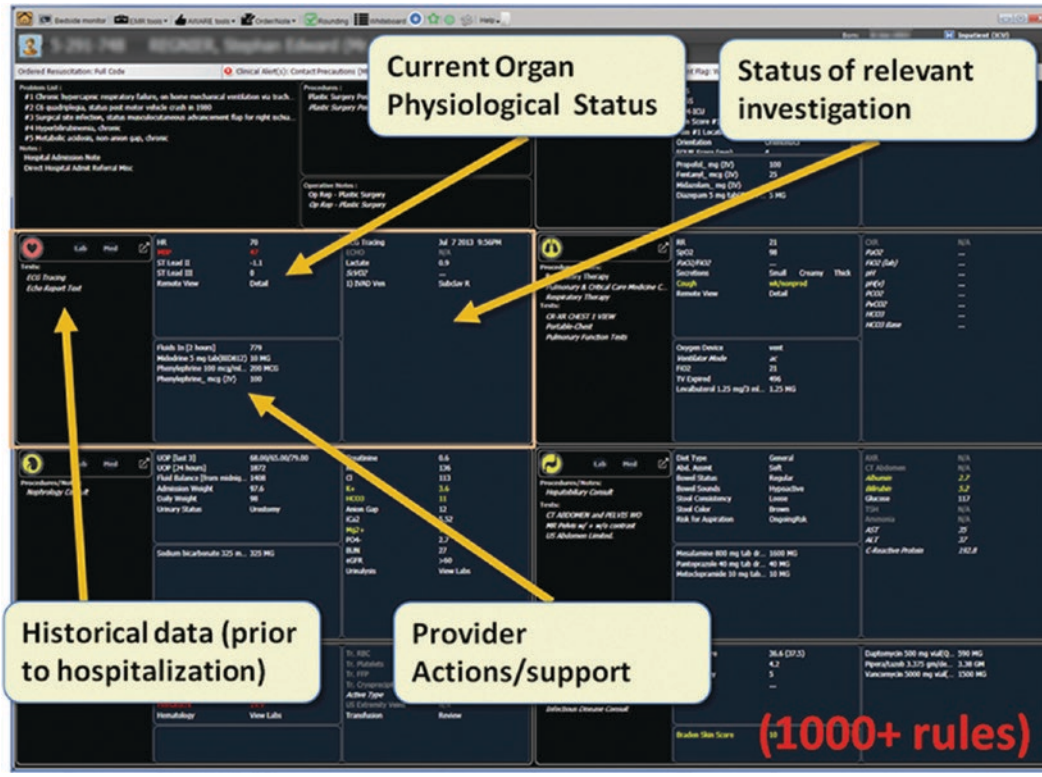


Fig. 21.13 Overview of data on the AWARE single patient viewer organized around human organs and systems. Each block represents four elements. Reading from left the key organizational elements are clinical context pulled from the patient problem list, procedure, medications, and consults lists prior to the current hospitalization. System identifying icons with color-coded status (red, required urgent intervention; yellow, abnormality; white, normal physiology and investigations). Middle top is physiologic status of organ displays current values for key variables. Middle bottom is organ supports, displays the critical care interventions which are supporting the current physiologic status. Right part of the block displays investigations

The popularization of checklists use in the clinical medicine began with the landmark publication by Dr. Atul Gawande team (Haynes et al. 2009). It has been shown to reduce errors and health care costs, increase compliance with evidence-based practice, and ultimately improve outcomes in critically ill patients (Weiss et al. 2011).

The smart checklist is a component of AWARE (Fig. 21.14). The system automatically detects patient characteristics to configure a patient's specific checklist. For example, if a patient is not on the ventilator, no ventilator-related questions are asked. In simulation, the study checklist significantly reduced provider workload and errors (Thongprayoon et al. 2016). Also, use of the checklist in the ICU was associated with increased number of occupational therapy/physical therapy consults in critically ill patients (Ali et al. 2017).

21.6.9 Collaborative Process, ICU Change-of-Shift, and Handover Issues

Cognitive scientists have taken an interest in, and have studied, the dynamic and distributed work environment in critical care medicine (Patel and Cohen 2008; Patel et al. 2008; Ahmed et al. 2011). They have studied issues such as provider task load, errors of cognition, and performance of clinicians involved in these complex tasks. The change-of-shift and handover times are especially critical and require complex exchanges of information that must occur rapidly and efficiently. These investigators have found that errors can occur during this time because of corruption of information and a failure to transfer crucial care facts. Having the majority of the patient record in electronic form and having that data timely and accurate should allow optimiza-

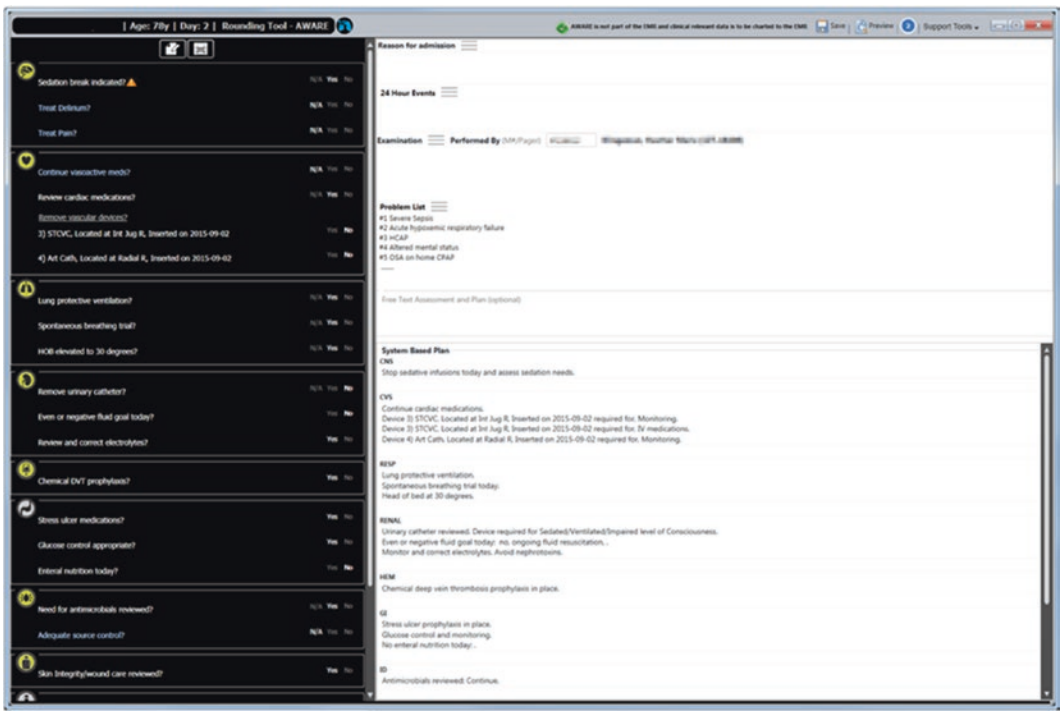


Fig. 21.14 Overview of AWARE checklist. It is a tool to aid during critical care rounds by helping apply best evidence care for every patient, generate meaningful clinical notes, and collect data for precise administrative and quality reports. AWARE indicates Ambient Warning and Response Evaluation System

tion of computerized decision making tools and methods for sharing the patient data. The Rounds Report developed at LDS Hospital three decades ago and recent developments at Mayo Clinic provide laboratory models for better understanding the issues and improving efficiency and eliminating medical errors for ICU patients during shift changes and patient handover times. The AWARE system has incorporated a number of tools that support ICU change of shift and handover processes, including shared tasklist, claim the patient, and handover modules. Each tool is designed to decrease the number of errors including omission, as well as the cognitive load of providers.

One of the core components of AWARE is the multipatient viewer that is a population management tool (Fig. 21.15). The multipatient viewer shows the census and the geographic layout of patient rooms a specific ICU unit. By hovering over or clicking on the icons, clinical information can be accessed quickly. Each patient box has four distinct areas (Fig. 21.16). Workflow or administrative area is on the top and this includes the Task list and Checklist icons as well as the room number and readiness for discharge to the floor in addition to problem list and current medications icons. The second row is alerting area with a purple circle with a line across that gives providers immediate knowledge that a patient's code status is Do not Resuscitate/Do Not Intubate. The middle section of the box contains patient information such as the patient's name, medical record number, age, number of days in the ICU, and teams caring for the patient. The lower section of the box contains the clinical information with the organ icons. These icons exist for the central nervous, cardiovascular, respiratory, gastrointestinal, renal, and hematologic systems. There is also an icon representing infectious diseases with relevant data including white blood cell counts and microbiologic specimen results. Organ icons are color coded the same as on single patient viewer. Second from bottom row has organs support icons. All icons were tested for recognition and recall by clinical providers (Litell et al. 2012). By clicking on a patient box it will be launched into the Single Patient Viewer.

The AWARE system was extensively tested. In cluster randomized trials the time spent on preround data gathering decreased from 12 to 9 minutes per patient before and

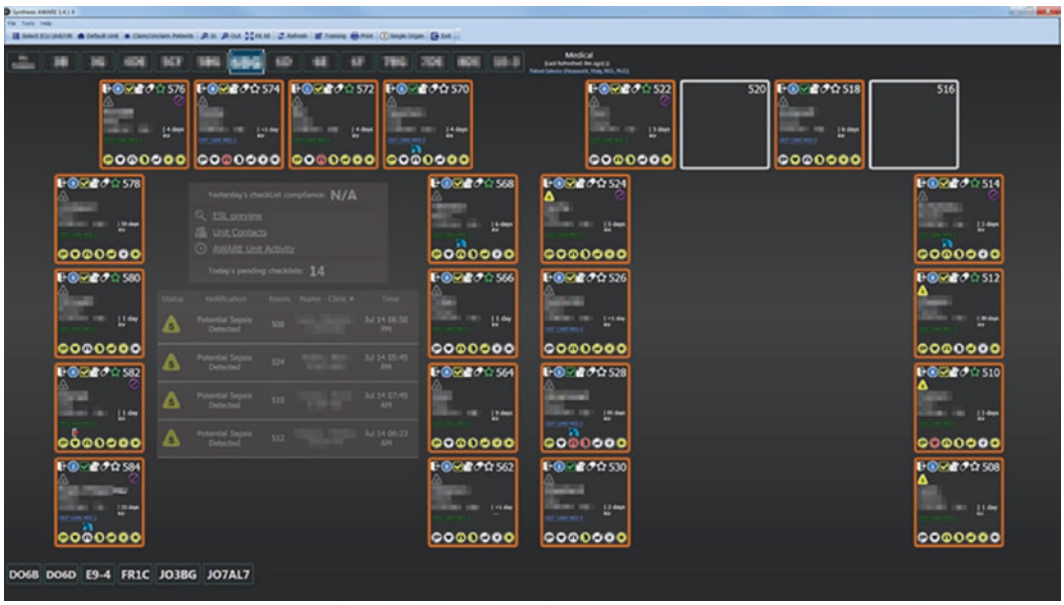


Fig. 21.15 Overview of multipatient viewer. It is population management tool where boxes represent geographical view of patient rooms in the care unit

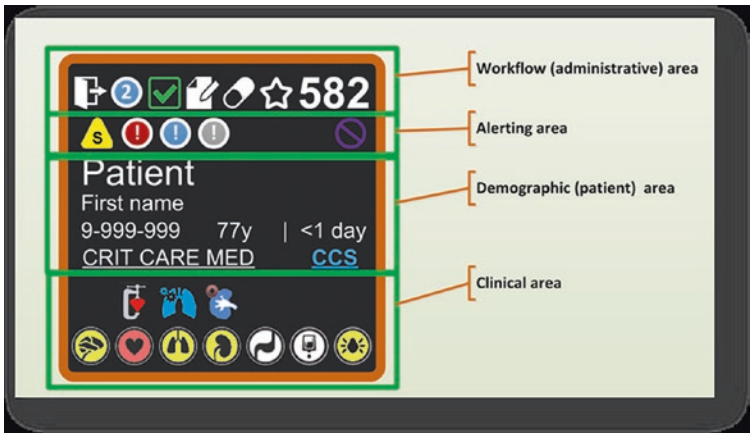


Fig. 21.16 Patient box on the multipatient viewer includes four groups of icons described in the text

after AWARE implementation (Pickering et al. 2015). In another study, tool usage was associated with a 50% decrease in ICU length of stay, 37% in-hospital length of stay, and total charges for hospital stay decreased by 30% in a post-AWARE cohort (by $43,745 after adjusting for patient acuity and demographics) (Olchanski et al. 2017).

21.7 Computerized Decision Support and Alerting

In addition to the alarms from bedside monitors, there are many other types of alerts and decision support tools that can be helpful for the care of hospitalized patients. A sampling of the types of decision support mechanism that have been reported is provided below to give the reader a sense for the breadth of capabilities that have been applied in intensive care as well as other care settings of hospitals. Key to the application of such computerized decision support tools is having access to an integrated, real-time, accurate, and coded EMR. Most of the examples noted are from the HELP system (Gardner et al. 1999). A key function of the HELP system is that the computerized decision support system is activated when new patient data are added to the patient's database, the process is called data-driven decision making. An example would be when the Po2 is put into the medical record an instruction is given to the respiratory therapist to modify the FIO2 or PEEP (Positive end-expiratory pressure) accordingly. Some functions of the HELP system, such as alerts, require that computerized decision support be activated at specific times and that process is called time-driven decision making. An example would be to remind the nurse the next glucose check is due when the patient is on an insulin drip, or instructing the computer to automatically calculate today's APACHE (Acute Physiology, Age, Chronic Health Evaluation) score and update all the reports at 06:00 AM.

In the past, commercially available EMR/ICU systems did not have convenient methods for programming and execution of computerized decision support rules. However, recent surveys by Sitting and Wright have shown that more and more commercial vendor systems have improved capability for providing clinical computerized decision support (Sittig et al. 2011; Wright et al. 2011). Once computerized decisions are made, they must be used to notify clinicians so that the feedback can be used to more effectively care for patients. The most common notification method is presentation on the computer screen when a clinician is interacting with the computer in some task such as order entry or charting. However, the issues of *how* to notify and *who* to notify are much more challenging (Tate et al. 1995; Shabot 1995). Further, verifying that such feedback results in the appropriate care is becoming ever more important.

Research continues on identifying the most efficient and effective notification methods.

Just as with the false alarms generated by bedside monitors, alarm feedback from computer systems must present timely and accurate recommendations with a minimum number of false alarms.

21.7.1 Laboratory Alerts

During the developmental period of the HELP system in the 1980s, it became apparent that on occasion life-threatening laboratory results were not being acted upon promptly. On acute care nursing floors, the initial alert response time averaged from 5.1 to 58.2 hours (Bradshaw et al. 1989). By posting alerts on computer terminals on nursing floors, the average response time was reduced to 3.6 hours. Then a flashing light, similar to those found on road maintenance vehicles, was installed on each nursing floor. The average response time then decreased to 6 minutes but the light was very annoying to the nursing staff (Bradshaw et al. 1989). When a sophisticated nurse paging system was set up that paged the particular nurse caring for the patient with the laboratory alert and required nurses to acknowledge the alerts the new pager system was equally effective and less annoying to other patients and staff (Tate et al. 1995). Similar work was done by Shabot (1995) at Cedars-Sinai Hospital in Los Angeles using a Blackberry pager. Since that time, wireless communications technology has improved dramatically and a variety of even better feedback mechanisms are now available. However in a study by Harrison et al. (2017), the alert acknowledgement rate from the severe sepsis alert system was significantly better with traditional paging system.

21.7.2 Ventilator Weaning Management and Alarm System

Weaning patients from ventilators was one of the first applications of a computerized expert system to routine patient care at LDS Hospital. As a result of the nurse and respiratory therapist charting described earlier, it was possible to develop and test computerized ventilator management protocols. Patient therapy was controlled by protocol 95% of the time and 90% of the protocol instructions were followed by clinicians. Several of the computerized instructions not followed were due to ventilator charting errors. Patients cared for with the computerized protocol had required less positive pressure in the ventilator system, and physiologic measures were disturbed less. The investigators concluded that such protocols could make the ventilator weaning processes "less mystifying, simpler, and more systematic" (East et al. 1992). Since that early work, several other investigators have implemented similar ventilator weaning algorithms.

In the process of implementing automated charting of ventilator parameters at LDS Hospital (Vawdrey et al. 2007), it became clear that critical ventilator alarms were being missed. As discussed earlier, alarm sounds emitted from ventilators were blended with bedside monitor alarm sounds. As a consequence, when a patient became disconnected from a ventilator the alarms could be missed (Evans et al. 2005). Once this situation was recognized, an enhanced notification system was implemented. Figure 21.17 illustrates a ventilator disconnect alarm presented on the patient's bedside display and on every other computer display in the same ICU. The efficacy and user acceptance of the new alarm system has enhanced patient safety and allowed documentation of this important clinical event.

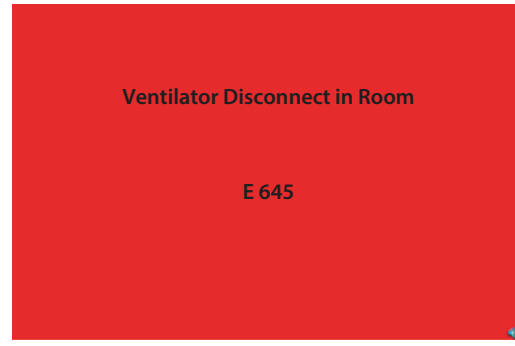


Fig. 21.17 Ventilator disconnect alarm. This alert is for the patient in Room E645 but it is displayed on every computer screen in the intensive care unit

21.7.3 Adverse Drug Event Detection and Prevention

Detection and prevention of **adverse drug events** (ADEs) has been a long-term goal of caregivers, the World Health Organization, and the U.S. Food and Drug Administration (Classen et al. 1991). Physicians, pharmacists, and informatics specialists at LDS Hospital developed a computer-based ADE monitor that detected a variety of triggers in the EMR that could indicate potential ADEs, such as sudden medication stop orders, medication antidote ordering, and specific abnormal laboratory and physiologic results. Pharmacists followed up on each ADE alert and each was verified and categorized. During an 18-month period, 36,653 hospitalized patients were monitored and 731 true ADEs occurred in 648 patients, 701 were classified as moderate or severe. Only 92 of the ADEs were identified by traditional voluntary reporting methods. Using this knowledge, the investigators developed methods for preventing ADEs. An example is the nurse charting work of Nelson et al. (2005).

Classen and colleagues (1997) followed up their earlier surveillance system for ADEs. They found that the attributable length of stay and costs of hospitalization for ADEs were substantial. If a patient had an ADE there was an increased length of stay of 1.74 days, an increased cost of $2,013, and an increased risk of death of 1.88 (Classen et al. 1997).

Even with the enhanced computerized methods for detecting, preventing and monitoring ADEs, there is still room for improvement (Petratos et al. 2010). In a studyover 200,000 medication alerts in an electronic prescribing system found more than 90% of drug alerts were overridden by physicians (Isaac et al. 2009). Critically ill patients are particularly susceptible to ADEs due to their unstable physiology, complex therapeutic medications, and the large percentage of IV medications (Hassan et al. 2010). Better systems must be developed and implemented to prevent ADEs.

21.7.4 IV Pump and Medications Monitoring

IV medication administration occurs in 90% of hospitalized patients; virtually every ICU patient is connected to an IV pump to receive fluids, nutrients, and medications. Although so-called smart pumps have been developed to minimize errors, those pumps are not yet integrated with the EMR and, as a result, are not capable of helping to prevent IV administration errors. Evans and associates (2010) at LDS Hospital have used cabled or wireless IV pumps integrated with the HELP system to enhance notification of IV pump programming errors. The medication charting system can detect and provide real-time alerts whenever an initial or potential pump rate programming error occurs. A set of 23 high-risk medications are monitored by the HELP system. Whenever IV pump flow rate for one of these medications is outside the acceptable range, a visual alert such as that shown in Fig. 21.18 is presented on the bedside display and on all other computer displays in the same ICU. Over a 2-year period, they found that there were alerts on 4% of the initial or dose rate changes or about 1.4 alerts per day.

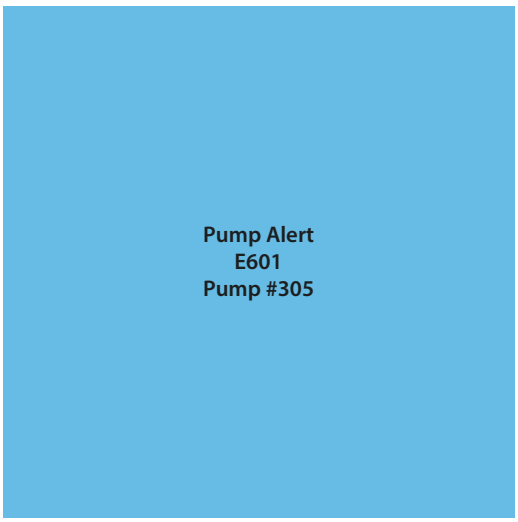


Fig. 21.18 Intravenous pump alert. This alert is for pump 305 located in Room E601, but it is displayed on every computer screen in the intensive care unit

Of those alerts, 14% were found to have prevented potential patient harm.

Clearly the monitoring and alerting system for ICU patients involves quite a different process and strategy than the usual bedside monitoring alarms. However, by having the integrated clinical record and the computerized decision support system available, these investigators have made major advances in minimizing ADEs and providing higher quality patient care.

21.8 Remote Monitoring and Tele-ICU

Tele-ICU is defined as the provision of care to critically ill patients by health care professionals located remotely. Tele-ICU clinicians use audio, video, and electronic links to assist the bedside caregivers in monitoring patients to help provide best practice and to help with the execution of optimized patient care plans. These types of systems have the potential of improving patient outcomes by having shorter response times to bedside monitor alarms and to abnormal laboratory values, initiating life-saving therapies, providing best practice more frequently, and providing expertise to smaller or remote ICUs where subspecialists are not readily available (Lilly et al. 2011). Historically, tele-ICU concepts date back to the mid-1980s, but it was not until the early 2000s that there was a dramatic increase in the use of such systems (Breslow 2007).

Tele-ICU has built on the concepts of computerized patient monitoring discussed earlier in this chapter. The real-time, EMR is fundamental to making tele-ICU care practical. The clinical information system is one of the keys to allow clinicians not physically present in the ICU to be able to suggest appropriate care. Enhanced bedside data acquisition and alarm systems, as well as clinical decision support systems (such as those described above) are required if remote clinicians are to provide practical and effective care for patients located in multiple remote ICUs (Rosenfeld et al. 2000; Celi et al. 2001; Breslow 2007, and Lilly et al. 2011). Table 21.3 gives an overview of the differences between a typical ICU with no electronic record compared with a tele-ICU.

Table 21.3 Comparison of typical ICU care processes with Tele-ICU care processes

Typical no EMR ICU	Tele-ICU
Bedside monitor alarms	Physiologic trend alerts
	Abnormal laboratory value alerts
	Review of response to alerts
	Off-site team rounds
Daily goal sheet	Electronic detection of nonadherence
	Real-time auditing
	Nurse manager audits
	Team audits
Telephone case review initiated by house staff or affiliate practitioner	Workstation review initiated by intensivists including EMR, imaging studies, interactive audio and video of patient, integrated with nurse and respiratory therapist, and assessment of responses to therapy

Adapted from Lilly et al. (2011)
Abbreviation: *EMR* electronic medical record, *ICU* intensive care unit

Recent findings of the impact of tele-ICU are encouraging and exciting. Patients receiving such care have lower hospital and ICU mortality and shorter hospital and ICU lengths of stay. Measures of adherence to best care practices are increased and complication rates are decreased (Lilly et al. 2011). However, the investigators pointed out that they had to implement major process and culture changes in their reengineering activities to make their system work (Lilly et al. 2011). An editorial accompanying the Lilly article outlines challenges still to be studied and understood about tele-ICU (Kahn 2011a). Since many changes were made from the typical ICU to the tele-ICU intervention, simply add-

ing better electronic data recording, electronic physiologic surveillance, and computerized decision support may have provided the same benefit, independent of the telemedicine feature. Informatics specialists clearly have exciting opportunities to improve care of critically ill patients and answer important process and intervention questions.

21.9 Predictive Alarms and Syndrome Surveillance

One key factor for medical errors is information overload (Kohn et al. 2000). False alerts from electronic and monitoring systems continue contribute to problem. Ideal alerts should only be issued when events are clinically significant, undetected by providers, and there is opportunity for correction of the underlying problem (actionable alert) (Norris and Dawant 2001). There are a number of approaches to reduce false alerts. Machine learning and surveillance methods have been developed to assist clinician decision makers in the care of complex care situations (Lee and Mark 2010). The work of Lee and colleagues at Harvard/Massachusetts Institute of Technology presents a methodology that has great promise (Lee et al. 2010). These investigators used machine learning to see if they could use pattern recognition approaches to predict impending hypotension in ICU patients. Using the high-resolution vital sign trends from the MIMIC II (Multiparameter Intelligent Monitoring in Intensive Care) Database, they trained their system to predict impending hypotension. Although the results were not perfect, they were able to identify patients at higher risk for developing hypotensive episodes within the subsequent 2 hours, thus alerting busy clinicians to be vigilant to impending events. Intelligent rule-based alerts, or sniffers developed at Mayo Clinic by Herasevich and his associates (2009, 2010, 2011), used near real-time Multidisciplinary Epidemiology and Translational Research in Intensive Care Data Mart to detect high-risk syndromes in patients and to alert clinicians if therapy has not yet been started. These investigators have provided excellent recommendations for development and use of large databases to allow better understanding of the complexities of patients who are critically ill. Advances in machine learning techniques showed promising results in predicting life-threatening situations (Evans et al. 2015) and death (Johnson and Mark 2018).

21.10 Opportunities for Future Development

Throughout this chapter, we have discussed many challenges and opportunities that remain in the field of patient monitoring systems. There are still important possibilities in the development of better and more effective bedside monitoring systems, especially in the area of maximizing true alarms and minimizing false alarms. Integrating clinical data from a broad variety of hospital and personal records is still challenging and important. Being able to apply computerized decision support systems to warn of life-threatening situations or advise care givers about optimum patient treatment strategies is still a relatively new aspect of health care. Development of patient care protocols and then having them be executable by computers, especially for ICU patients, is also an exciting field of endeavor.

Since the early 1950s, when physicians began to understand control system theory, there has been a fascination with having control systems that closed the loop without the need for any human intervention. Implantable defibrillators and pacemakers are examples closed-loop devices. The publication in 1957 started the idea of automating mechanical ventilation (Saxton and Myers 1957).

We still believe that applying informatics in the ICU is a "contact and team sport," that you must be involved at the patient care level and work with the incredibly talented clinical teams to maximize the benefits that biomedical informatics specialists can provide.

21.10.1 Value of Computerized ICU Care Processes

Challenges and opportunities lie in proving the value of health information systems. There have been dramatic improvements in the adoption of EMR in recent years (Gardner RM 2016). A review by Chaudhry and associates (2006) assessing the impact of health information technology on the quality, efficiency, and cost of medical care is illustrative of the challenge. An even more sobering report was presented by Karsh and associates (2010). These investigators suggest that not only is rate of adoption of health information technology low, but such technology may not improve quality of care or reduce costs. Introducing a new EMR system can lead to change in workflow. There are documented increases in revenue after EMR implementation despite of productivity loss (less patient visits) (Howley et al. 2015). Thompson and colleagues (2015) systematic review "Impact of the Electronic Medical Record on Mortality, Length of Stay, and Cost in the Hospital and ICU" were not shown to have a substantial effect on those metrics. However, there was a significant effect of surveillance systems on in-hospital mortality.

21.11 Clinical Control Tower and Population Management

Clinical Control Tower is a newly-developed central alert-screening clinical surveillance system developed at Mayo Clinic. The concept behind the Clinical Control Tower is to serve as a centralized non-life-threatening alert and prediction "cockpit."

This unified screening system is managed by a designated capsule communicator or "CapCom," analogous to the US National Aeronautics and Space Administration ground-based astronaut who maintains contact with astronauts during space missions. The CapCom in the healthcare context is the clinician responsible for screening incoming alerts and notifications. As no alerts have 100% accuracy it is essential to perform initial validation of notifications before activating specific workflows with bedside providers. When the CapCom decides that an alert is valid, he or she communicates "down to the ground" to a bedside clinician and guides them through necessary and recommended tasks. Each step may be captured electronically in the control tower application. Workflow and actions are captured and analyzed using a feedback loop tool. Deviations from intended care processes may be identified. Control Tower is a tool designed to minimize errors and information overload in hospital practice.

As the COVID-19 pandemic hit the USA healthcare system, the Control Tower platform was modified and expanded to address the surveillance needs of hospitalized Covid-19 patients (Fig. 21.19). The system identifies the status of COVID lab tests, COVID results, patient isolation information and MEWS score.

The features required to facilitate reliable monitoring and management of acutely ill patient populations differ significantly from those required to manage single patient encounters. These demands are not easily met by the most common commercially available comprehensive electronic medical records and often require complementary alternative approaches such as the one illustrated in the Control Tower example above.

Suggested Readings

Clemmer, T. P. (2004). Computers in the ICU: Where we started and where we are now. Journal of Critical Care, 19, 204–207. PMID 15648035.

Gardner, R. M. (2009). Clinical decision support systems: The fascination with closed-loop control. IMIA Yearbook of Medical Informatics, 12–21. PMID 19855866.

Greenes, R. A. (Ed.). (2007). Clinical decision support: The road ahead. Burlington: Elsevier Inc.

Harrison, A. M., Park, J. G., & Herasevich, V. (2015). Septic shock electronic surveillance. In Septic shock: Risk factors, management and prognosis (pp. 1–25). New York: Nova Biomedical.

Herasevich, V., Kor, D. J., Subramanian, A., & Pickering, B. W. (2013). Connecting the dots:

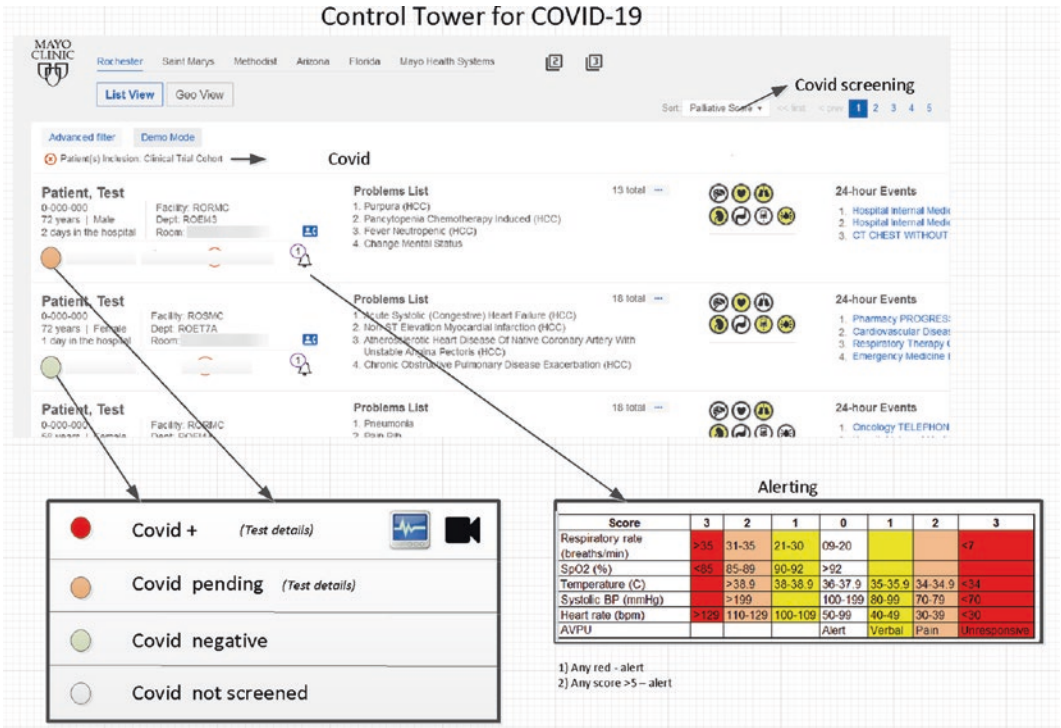


Fig. 21.19 Clinical Control Tower modified for COVID-19 screening

Rule-based decision support systems in the modern EMR era. Journal of Clinical Monitoring Computing, 27(4), 443–448. PMID: 23456293.

Kuperman, G. J., Gardner, R. M., & Pryor, T. A. (1991). HELP: A dynamic hospital information system. New York: Springer.

Morris, A. H. (2001). Rational use of computerized protocols in the intensive care unit. Critical Care, 5, 249–254. PMID 11737899.

Pickering, B. W., Litell, J. M., Herasevich, V., & Gajic, O. (2012). Clinical review: The hospital of the future - building intelligent environments to facilitate safe and effective acute care delivery. Critical Care, 16(2), 220. PMID: 22546172.

Shabot, M. M., & Gardner, R. M. (Eds.). (1994). Decision support systems in critical care. Boston: Springer.

Questions for Discussion

1. Describe how the integration of information from multiple bedside monitoring signals, the pharmacy, and clinical laboratory data can help improve alarm systems used in an ICU.
2. How would you decide whether to buy a standalone ICU patient monitoring system versus an integrated EMR system?
3. How do care providers impact the installation and optimization of real-time data collection and real-time decision support?
4. Perhaps real-time data collection and computerized decision support are not necessary. How would you assess these issues? Is there sufficient literature to validate or disprove your supposition? If not, what is missing?
5. How would you go about selecting the optimum data for monitoring and improving the care of a critically ill patient?
6. How would you optimize a patient monitoring system that you were building or buying to provide the most accurate, timely, and helpful computerized decision support capabilities? Be specific and give literature references to support your optimization plan.

7. If you were the Chief Clinical Information Officer of a large hospital without data from the ICU integrated into your EMR system, what factors would you have to consider to implement such a system AND to apply computerized clinical decision support to optimize such a system? How long do you think it would take to implement such a system?

References

Aboukhalil, A., Nielsen, L., Saeed, M., Mark, R. G., & Clifford, G. D. (2008). Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. *Journal of Biomedical Informatics, 41*, 442–451. PMID 18440873.

Ahmed, A., Chandra, S., Herasevich, V., Gajic, O., & Pickering, B. W. (2011). The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. *Critical Care Medicine, 39*, 1626–1634. PMID 21478739.

Ali, R., Cornelius, P. J., Herasevich, V., Gajic, O., & Kashyap, R. (2017). Effect of daily use of electronic checklist on physical rehabilitation consultations in critically ill patients. *Journal of Critical Care, 38*, 357–361. PMID: 28043721.

Andrews, R. D., Gardner, R. M., Metcalf, S. M., & Simmons, D. (1985). Computer charting: An evaluation of a respiratory care computer system. *Respiratory Care, 30*, 695–707. PMID 10315682.

Blum, J. M., & Tremper, K. K. (2010). Alarms in the intensive care unit: Too much of a good thing is dangerous: Is it time to add some intelligence to alarms? *Critical Care Medicine, 33*, 702–703. PMID 20083933.

Bogun, F., Anh, D., Kalahasty, G., Wissner, E., Cou Serhal, C., Bazzi, R., Weaver, W. D., & Schuger, C. (2004). Misdiagnosis of atrial fibrillation and its clinical consequences. *American Journal of Medicine, 117*, 636–642. PMID 15501200.

Bradshaw, K. E., Gardner, R. M., Clemmer, T. P., Orme, J. F., Thomas, F., & West, B. J. (1984). Physician decision-making: Evaluation of data used in a computerized ICU. *International Journal of Clinical Monitoring and Computing, 1*(2), 81–91.

Bradshaw, K. E., Sittig, D. F., Gardner, R. M., Pryor, T. A., & Budd, M. (1989). Computer-based data entry for nurses in the ICU. *MD Computing, 6*(5), 274–280. PMID 2486506.

Braunwald, E. (1988). Evolution of the management of acute myocardial infarction: A 20th century saga. *The Lancet, 352*, 1771–1774. PMID 9848369.

Breslow, M. J. (2007). Remote ICU care programs: Current status. *Journal of Critical Care, 22*(1), 66–76.

Brown, L. T., Purcell, G. J., & Traugott, F. M. (1990). Hypoxaemia during postoperative recovery using continuous pulse oximetry. *Anaesthesia and Intensive Care, 18*(4), 509–516. PMID: 2268017.

Burykin, A., Peck, T., & Buchman, T. G. (2011). Using "off-the-shelf" tools for a terabyte-scale waveform recording in intensive care: Computer system design, database description and lessons learned. *Computer Methods and Programs in Biomedicine, 103*, 151–160. PMID 21093093.

Carayon, P., Cartmill, R., Blosky, M. A., Brown, R., Hackenberg, M., Hoonakker, P., Hundt, A. S., Norfolk, E., Wetterneck, T. B., & Walker, J. M. (2011). ICU nurses' acceptance of electronic health records. *Journal of the American Medical Informatics Association: JAMIA, 18*(6), 812–819. PMID 21697291.

Celi, L. A., Hassan, E., Marquardt, C., Breslow, M., & Rosenfeld, B. (2001). The eICU: It's not just telemedicine. *Critical Care Medicine, 29*(8 Suppl), N183–N189. PMID 11496041.

Chambrin, M. C. (2001). Alarms in the intensive care unit: How can the number of false alarms be reduced? *Critical Care, 5*, 184–188. PMID 11511330.

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S. C., et al. (2006). Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine, 144*(10), 742–752.

Chen, T., Clifford, G. D., & Mark, R. G. (2009). The effect of signal quality on six cardiac output estimators. *Computers in Cardiology, 36*, 197–200. PMC 2926988.

Clark, A. P., Giuliano, K., & Chen, H. M. (2006). Pulse oximetry revisited "but His O2 was normal". *Clinical Nurssing Specialist CNS, 20*, 268–272. PMID 17149014.

Clark, E. N., Sejersten, M., Clemmensen, P., & MacFarland, P. W. (2010). Automated electrocardiogram interpretation programs versus cardiologists' triage decision making based on teletransmitted data in patients with suspected acute coronary syndrome. *The American Journal of Cardiology, 106*, 1696–1702. PMID 21126612.

Classen, D. C., Pestotnik, S. L., Evans, R. S., & Burke, J. P. (1991). Computerized surveillance of adverse drug events in hospital patients. *JAMA, 266*, 2847–2851. PMID 194452.

Classen, D. C., Pestotnik, S. L., Evans, R. S., Lloyd, J. F., & Burke, J. P. (1997). Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. *Journal of the American Medical Association, 277*(4), 301–306.

Clemmer, T. P. (2004). Computers in the ICU: Where we started and where we are now. *Journal of Critical Care, 19*, 204–207. PMID 15648035.

Clemmer, T. P., Spuhler, V. J., Berwick, D. M., & Nolan, T. W. (1998). Cooperation: The foundation of improvement. *Annals of Internal Medicine, 128*(12 Pt 1), 1004–1009. PMID 9625663.

Collins, F. S. (2011). Reengineering translational science: The time is right. *Science Translational Medicine, 3*(90), 90cm17.

Cox, J., Jr. (1972). Digital analysis of the electroencephalogram, the blood pressure wave, and the electrocardiogram. *Proceedings of the IEEE, 60*, 1137.

Crossley, G. H., Poole, J. E., Rozner, M. A., et al. (2011). The Heart Rhythm Society (HRS)/American Society of Anesthesiologists (ASA) expert consensus statement on the perioperative management arrhythmia monitors. *Heart Rhythm, 8*, 1114–1154. PMID 21722856.

Cundick, R. M., Jr., & Gardner, R. M. (1980). Clinical comparison of pressure-pulse and indicator-dilution cardiac output determination. *Circulation, 62*, 371–376. PMID 6994922.

Cushing, H. (1903). On routine determination of arterial tension in operating room and clinic. *Boston Medical Surgical Journal, 148*, 250.

Dalto, J. S., Johnson, K. V., Gardner, R. M., Spuhler, V. J., & Egbert, L. (1997). Medical information bus usage for automated IV pump data acquisition: Evaluation of use and nurse attitudes. *International Journal of Clinical Monitoring and Computing, 14*, 151–154. PMID 9387004.

Deflandre, E., Bonhomme, V., & Hans, P. (2008). Delta down compared with delta pulse pressure as an indicator of volaemia during intracranial surgery. *British Journal of Anaesthesia, 100*, 245–250. PMID 18083787.

Doig, A. K., Drews, F. A., & Keefe, M. R. (2011). Informing the design of hemodynamic monitoring displays. *Computers, Informatics, Nursing, 29*(12), 706–713. PMID 21412150.

Drew, B. J., & Funk, M. (2006). Practice standards for ECG monitoring in hospital setting: Executive summary and guide for implementation. *Critical Care Nursing Clinics of North America, 18*(2), 157–168. PMID 16728301.

Drews, F. A., Musters, A., & Samore, M. H. (2008). Error producing conditions in intensive care unit. In *Advances in patient safety: New directions and alternative approaches* (Performance and tools) (Vol. 3). Rockville: AHRQ. PMID 21249947.

East, T. D., Bohm, S. H., Wallace, C. J., Clemmer, T. P., Weaver, L. K., Orme, J. F., Jr., & Morris, A. H. (1992). A successful computerized protocol for clinical management of pressure control inverse ration ventilation. *Chest, 101*(3), 697–710.

Edworthy, J., & Hellier, E. (2006). Alarms and human behavior: Implications for medical alarms. *British Journal of Anaesthesia, 97*, 12–17. PMID 16698858.

Evans, R. S., Johnson, K. V., Flint, V. B., Kinder, T., Lyon, C. R., Hawley, W. L., Vawdrey, D. K., & Thomsen, G. E. (2005). Enhanced notification of critical ventilator events. *Journal of the American Medical Informatics Association, 12*, 589–595. PMID 16049226.

Evans, R. S., Carlson, R., Johnson, K. V., Palmer, B. K., & Lloyd, J. F. (2010). Enhanced notification of infusion pump programming errors. *Studies in Health Technology and Informatics, 160*(Pt 1), 734–738. PMID 20841783.

Evans, R. S., Kuttler, K. G., Simpson, K. J., Howe, S., Crossno, P. F., Johnson, K. V., Schreiner, M. N., Lloyd, J. F., Tettelbach, W. H., Keddington, R. K., Tanner, A., Wilde, C., & Clemmer, T. P. (2015). Automated detection of physiologic deterioration in hospitalized patients. *Journal of the American Medical Informatics Association, 22*(2), 350–360. PMID: 25164256.

Funk, M., Winkler, C. G., May, J. L., Stephens, K., Fennie, K. P., Rose, L. L., Turkman, Y. E., & Drew, B. J. (2010). Unnecessary arrhythmia monitoring and underutilization of ischemia and QT interval monitoring in current clinical practice: Baseline results of the practice use of the latest standards for electrocardiography trial. *Journal of Electrocardiology, 43*, 542–547. PMID 20832819.

Gardner, R. M. (1996). Accuracy and reliability of disposable pressure transducers coupled with modern pressure monitors. *Critical Care Medicine, 24*, 879–882. PMID 8706469.

Gardner, R. M. (2004). Computerized clinical decision-support in respiratory care. *Respiratory Care, 49*, 378–386. PMID 15030611.

Gardner, R. M. (2016). Clinical information systems – from yesterday to tomorrow. *Yearbook of Medical Informatics*, (Suppl 1), S62–S75. PMID: 27362589.

Gardner, R. M., & Beale, R. J. (2009). Pressure to perform: Is cardiac output estimation from arterial waveforms good enough for routine use? *Critical Care Medicine, 37*, 337–338. EDITORIAL PMID 19112291.

Gardner, R. M., & Lundsgaarde, H. P. (1994). Evaluation of user acceptance of a clinical expert system. *Journal of the American Medical Informatics Association: JAMIA, 1*(6), 428–438.

Gardner, R. M., Warner, H. R., Toronto, A. F., & Gaisford, W. D. (1970). Cather-flush system for continuous monitoring of central arterial pulse waveform. *Journal of Applied Physiology, 29*, 911–913. PMID 5485368.

Gardner, R. M., Tariq, H., Hawley, W. L., & East, T. D. (1989). Medical information bus: The key to future integrated monitoring. *International Journal of Clinical Monitoring and Computing, 6*, 197–204. PMID 2628508.

Gardner, R. M., Hawley, W. H., East, T. D., Oniki, T., & Young, H. F. W. (1991). Real time data acquisition: Experience with the medical information bus (MIB). *Proceedings of the Annual Symposium on Computer Applications in Medical Care, 15*, 813–817. PMID 1807719.

Gardner, R. M., Pryor, T. A., & Warner, H. R. (1999). The HELP system: Update 1998. *International Journal of Medical Informatics, 54*, 169–182. PMID 10405877.

Gorges, M., Markewitz, B. A., & Westenskow, D. R. (2009). Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesthesia and Analgesia, 108*, 1546–1552. PMID 19372334.

Gregg, R. E., Zhou, S. H., Lindauer, J. M., Helfendbein, E. D., & Giuliano, K. K. (2008). What is inside the electrocardiograph? *Journal of Electrocardiology, 41*, 8–14. PMID 1819652.

Groves, R. H., Jr., Holcomb, B. W., Jr., & Smith, M. L. (2008). Intensive care telemedicine: Evaluating a model for proactive remote monitoring and intervention in the critical care setting. *Studies in Health Technology and Informatics, 131*, 131–146. PMID 18305328.

Guglin, M. E., & Thatai, D. (2006). Common errors in computer electrocardiogram interpretation. *International Journal of Cardiology, 106*, 232–237. PMID 16321696.

Gurses, A. P., & Xiao, Y. (2006). A systematic review of the literature on multidisciplinary rounds to design information systems. *Journal of the American Medical Informatics Association: JAMIA, 13*, 267–278. PMID 16501176.

Han, Y. Y., Carcillo, J. A., Venkataraman, S. T., Clark, R. S., Watson, R. S., Nguyen, T. C., Bayir, H., & Orr, R. A. (2005). Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics, 116*(6), 1506–1512. PMID: 16322178.

Hassan, E., Badawi, O., Weber, R. J., & Cohen, H. (2010). Using technology to prevent adverse drug events in the intensive care unit. *Critical Care Medicine, 38*(6 Suppl), S97–S105. PMID 20502181.

Haynes, A. B., Weiser, T. G., Berry, W. R., Lipsitz, S. R., Breizat, A. H., Dellinger, E. P., Herbosa, T., Joseph, S., Kibatala, P. L., Lapitan, M. C., Merry, A. F., Moorthy, K., Reznick, R. K., Taylor, B., Gawande, A. A., & Safe Surgery Saves Lives Study Group. (2009). A surgical safety checklist to reduce morbidity and mortality in a global population. *The New England Journal of Medicine, 360*(5), 491–499. PMID: 19144931.

Herasevich, V., Yilmaz, M., Khan, H., Hubmayr, R. D., & Gajic, O. (2009). Validation of an electronic surveillance system for acute lung injury. *Intensive Care Medicine, 35*, 1018–1023. PMID 19280175.

Herasevich, V., Pickering, B. W., Dong, Y., Peters, S. G., & Gajic, O. (2010). Informatics infrastructure for syndromic surveillance, decision support, reporting, and modeling of critical illness. *Mayo Clinic Proceedings, 85*, 247–254. PMID 21194152.

Herasevich, V., Afessa, B., & Pickering, B. W. (2011). Sepsis in critically ill patients with trauma. *Critical Care Medicine, 39*, 876–879. PMID 21613830.

Herasevich, V., Litell, J., & Pickering, B. (2012). Electronic medical records and mHealth anytime, anywhere. *Biomedical Instrumentation & Technology, 46*(Suppl 2), 45–48. PMID: 23039776.

Herasevich, V., Kor, D. J., Subramanian, A., & Pickering, B. W. (2013). Connecting the dots: Rule-based decision support systems in the modern EMR era. *Journal of Clinical Monitoring and Computing, 27*(4), 443–448. PMID: 23456293.

Howley, M. J., Chou, E. Y., Hansen, N., & Dalrymple, P. W. (2015). The long-term financial impact of electronic health record implementation. *Journal of the American Medical Informatics Association, 22*(2), 443–452. PMID: 25164255.

Hripcsak, G., Vawdrey, D., Fred, M., & Bostwick, S. (2011). Use of electronic clinical documentation: Time spent and team interactions. *Journal of the American Medical Informatics Association: JAMIA, 18*(1), 112–117.

Imhoff, M., & Kuhls, S. (2006). Alarm algorithms in critical care monitoring. *Anesthesia and Analgesia, 102*, 1525–1537. PMID 16632837.

Isaac, T., Weissman, J. S., Davis, R. B., Massagli, M., Cyrulik, A., Sands, D. Z., & Weingart, S. N. (2009). Overrides of medication alerts in ambulatory care. *Archives of Internal Medicine, 169*(3), 305–311. PMID: 19204222.

Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rap, S. R., Ferris, T. G., Shields, A., Rosenbaum, S., & Blumenthal, D. (2009). Use of electronic health records in U.S. Hospitals. *The New England Journal of Medicine, 360*, 1628–1638. PMID 19321858.

Johnson, A. E. W., & Mark, R. G. (2018). Real-time mortality prediction in the Intensive Care Unit. *American Medical Informatics Association Annual Symposium Proceedings, 2017*, 994–1003. PMID: 29854167.

Kahn, J. M. (2011a). The use and misuse of ICU telemedicine. *JAMA, 305*, 2176–2183. PMID 21576623.

Karsh, B.-T., Weinger, M. B., Abbott, P. A., & Wears, R. L. (2010). Health information technology: Fallacies and sober realities. *Journal of the American Medical Informatics Association: JAMIA, 17*(6), 617–623.

Kelley, M. A., Angus, D., Chalfin, D. B., Crandall, E. D., Ingbar, D., Johanson, W., Medina, J., Sessler, C. N., & Vender, J. S. (2004). The critical care crisis in the United States: A report from the profession. *Chest, 125*, 1514–1517. PMID 15078767.

Kennelly, R. J., & Gardner, R. M. (1997). Perspectives on development of IEEE 1073: The medical information bus (MIB) standard. *International Journal of Clinical Monitoring and Computing, 14*(3), 143–149.

Keren, H., Burkhoff, D., & Squara, P. (2007). Evaluation of a non-invasive continuous cardiac output monitoring system based on thoracic bioreactance. *American Journal of Physiology - Heart and Circulatory Physiology, 293*, H583–H589. PMID 17384132.

Kilickaya, O., Schmickl, C., Ahmed, A., Pulido, J., Onigkeit, J., Kashani, K., Gajic, O., Herasevich, V., & Pickering, B. (2014). Customized reference ranges for laboratory values decrease false positive alerts in intensive care unit patients. *PLoS One, 9*(9), e107930. PMID: 25233485.

Kohn, L. T., Corrigan, J., & Donaldson, M. S. (2000). *Institute of Medicine. To err is human. Building a safer health system* (p. xxi, 287). Washington, DC: National Academy Press. PMID: 25077248.

Kouwenhoven, W. B., Jude, J. R., & Knickerbocker, G. G. (1960). Closed-chest cardiac massage. *JAMA, 173*, 1064–1067. PMID 14411374.

Kowalczyk, L. (2011, February 13). Patient alarms often unheard, unheeded. *Boston Globe*.

Kuperman, G. J., Gardner, R. M., & Pryor, T. A. (1991). *HELP: A dynamic hospital information system*. New York: Springer.

Landrigan, C. P., Parry, G. J., Bones, C. B., Hackbarth, A. D., Goldmann, D. A., & Sharek, P. J. (2010). Temporal trends in rates of patient harm resulting from medical care. *The New England Journal of Medicine, 363*(22), 2124–2134. PMID: 21105794.

Lee, J., & Mark, R. G. (2010). An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care. *Biomedical Engineering, 9*, 62. PMID 20973998.

Lee, J. Y., Du, Y. E., Coki, O., Flynn, J. T., Starren, J., & Chiang, M. F. (2010). Parental perceptions toward digital imaging and telemedicine for retinopathy of prematurity management. *Graefe's Archive for Clinical and Experimental Ophthalmology, 248*, 141–147.

Li, Q., Mark, R. G., & Clifford, G. D. (2009). Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator. *Biomedical Engineering Online, 8*, 13. PMID 19586547.

Lilly, C. M., Cody, S., Zhao, H., et al. (2011). Hospital mortality, length of stay, and preventable complications among critically Ill patients before and after Tele-ICU reengineering of critical care processes. *JAMA, 305*, 2175–2183.

Litell, J. M., Suther, T. W., Ridgeway, C. J., Tiong, I. C., Pickering, B. W., & Herasevich, V. (2012). Representation of organ system domains in a novel critical care EMR interface: Implications for effective partnership between clinicians and design professionals. In *AMIA 2012 proceedings*, p. 1840.

Lown, B., Amarasingham, R., & Neuman, J. (1962). New method for terminating cardiac arrhythmias. Use of synchronous capacitor discharge. *JAMA, 182*, 548–555. PMID 12921298.

Malviya, S., Reynolds, P. I., Voepel-Lewis, T., Siewert, M., Watson, D., Tait, A. R., & Tremper, K. (2000). False alarms and sensitivity of conventional pulse oximetry versus the Masimo SET technology in the pediatric postanesthesia care unit. *Anesthesia and Analgesia, 90*(6), 1336–1340. PMID: 10825316.

Mathews, S. C., & Pronovost, P. J. (2011). The need for systems integration in health care. *JAMA, 305*, 934–935. PMID 21364143.

Morris, A. H. (2000). Developing and implementing computerized protocols for standardization of clinical decisions. *Annals of Internal Medicine, 132*, 373–383. PMID 10691588.

Morris, A. H. (2001). Rational use of computerized protocols in the intensive care unit. *Critical Care, 5*, 249–254. PMID 11737899.

Nelson, N. C., Evans, R. S., Samore, M. H., & Gardner, R. M. (2005). Detection and prevention of medication errors using real-time nurse charting. *Journal of the American Medical Informatics Association: JAMIA, 12*, 390–397. PMID 15802486.

Norris, P. R., & Dawant, B. M. (2001). Closing the loop in ICU decision support: Physiologic event detection, alerts, and documentation. In *Proceedings/AMIA annual symposium AMIA symposium*, pp. 498–502.

Olchanski, N., Dziadzko, M. A., Tiong, I. C., Daniels, C. E., Peters, S. G., O'Horo, J. C., & Gong, M. N. (2017). Can a novel ICU data display positively affect patient outcomes and save lives? *Journal of Medical Systems, 41*(11), 171. PMID: 28921446.

Oniki, T. A., Clemmer, T. P., & Pryor, T. A. (2003). The effect of computer-generated reminders on charting deficiencies in the ICU. *Journal of the American Medical Informatics Association: JAMIA, 10*, 177–187.

Patel, V. L., & Cohen, T. (2008). New perspectives on error in critical care. *Current Opinion in Critical Care, 14*(4), 456–459.

Patel, V. L., Zhang, J., Yoskowitz, N. A., Green, R., & Sayan, O. R. (2008). Translational cognition for decision support in critical care environments: A review. *Journal of Biomedical Informatics, 41*, 413–431.

Petratos, G. N., Kim, Y., Evans, R. S., & Gardner, R. M. (2010). Comparing the effectiveness of computerized adverse drug event monitoring systems to enhance clinical decision support for hospitalized patients. *Applied Clinical Informatics, 1*, 293–303.

Pickering, B. W., Herasevich, V., & Gajic, A. A. (2010). Novel representation of clinical information in the ICU – developing user interfaces which reduce information overload. *Applied Clinical Informatics, 1*, 116–131. PMID: 23616831.

Pickering, B. W., Litell, J. M., Herasevich, V., & Gajic, O. (2012). Clinical review: The hospital of the future - building intelligent environments to facilitate safe and effective acute care delivery. *Critical Care, 16*(2), 220. PMID: 22546172.

Pickering, B. W., Dong, Y., Ahmed, A., Giri, J., Kilickaya, O., Gupta, A., Gajic, O., & Herasevich, V. (2015). The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: A pilot step-wedge cluster randomized trial. *International Journal of Medical Informatics, 84*(5), 299–307. PMID: 25683227.

Poon, K. B. (2005). *Fusing multiple heart rate signals to reduce alarms in adult intensive care unit*. Master of Science Thesis from Department of Medical Informatics, University of Utah.

Pryor, T. A. (1989). Computerized nurse charting. *International Joural of Clinical Monitoring and Computing, 6*, 173–179. PMID 2592844.

Pryor, T. A., Gardner, R. M., Clayton, P. D., & Warner, H. R. (1983). The HELP system. *Journal of Medical Informatics, 7*(2), 87–102.

Rogers, W. J., Canto, J. G., Lambrew, C. T., Tiefenbrunn, A. J., Kinkaid, B., Shoultz, D. A., Frederick, P. D., & Every, N. (2000). Temporal trends in the treatment of over 1.5 million patients with myocardial infarction in the US from 1990 thru 1999: The National Registry of Myocardial Infarction 1, 2 and 3. *Journal of the American College of Cardiology, 36*, 2056–2063. PMID 11127441.

Rosenfeld, B. A., Dorman, T., Breslow, M. J., et al. (2000). Intensive care unit telemedicine: Alternate paradigm for providing continuous intensivist care. *Critical Care Medicine, 28*, 3925–3931.

Saeed, M., Villarroel, M., Reisner, A. T., Clifford, G., Lethman, L. W., Moody, G., Heldt, T., Kyaw, T. H., Moody, B., & Mark, R. G. (2011). Multiparameter intelligent monitoring in intensive care II: A public-access intensive care unit database. *Critical Care Medicine, 39*, 952–960. PMID 21283005.

Sanctorius, S. (1626). *Commentaria in primam Fen primi libri canonis Avicennae* (p. 22). Venice: Sarcina.

Saxton, G. A., Jr., & Myers, G. (1957). A servomechanism for automatic regulation of pulmonary ventilation. *Journal of Applied Physiology, 11*(2), 326–328. PMID: 13475186.

Shabot, M. M. (1995). Computers in the intensive care unit: Was pogo correct? *Journal of Intensive Care Medicine, 10*, 211–212.

Shubin, H., & Weil, M. H. (1966). Efficient monitoring with a digital computer of cardiovascular function in seriously ill patients. *Annals of Internal Medicine, 65*(3), 453–460.

Siebig, S., Kuhls, S., Imhoff, M., Gather, U., Scholmerich, J., & Wrede, C. E. (2010). Intensive care unit alarms – how many do we need? *Critical Care Medicine, 38*, 451–456. PMID 20016379.

Sittig, D. F., Wright, A., Meltzer, S., Simonaitis, L., Evans, R. S., Nichol, W. P., Ash, J. S., & Middleton, B. (2011). Comparison of clinical knowledge management capabilities of commercially-available and leading internally-developed electronic health records. *BMC Medical Informatics and Decision Making, 11*, 13. PMID 21329520.

Sun, J. X., Reisner, A. T., Saeed, M., Heldt, T., & Mark, R. G. (2009). The cardiac output from pressure algorithms trial. *Critical Care Medicine, 37*, 72–80. PMID 19112280.

Tate, K. E., Gardner, R. M., & Scherting, K. (1995). Nurses, pagers, and patient-specific criteria: Three keys to improved critical value reporting. *Proceedings of the Annual Symposium on Computer Applications in Medical Care*, pp. 154–8. PMID 8563258.

Thompson, J. P., & Mahajan, R. P. (2006). Monitoring the monitors – beyond risk management. *British Journal of Anaesthesia, 97*, 1–3. PMID 16769701.

Thompson, G., O'Horo, J. C., Pickering, B. W., & Herasevich, V. (2015). Impact of the electronic medical record on mortality, length of stay, and cost in the hospital and ICU: A systematic review and metaanalysis. *Critical Care Medicine, 43*(6), 1276–1282. PMID: 25756413.

Thongprayoon, C., Harrison, A. M., O'Horo, J. C., Berrios, R. A. S., Pickering, B. W., & Herasevich, V. (2016). The effect of an electronic checklist on critical care provider workload, errors, and performance. *Journal of Intensive Care Medicine, 31*(3), 205–212. PMID: 25392010.

Vawdrey, D. K., Gardner, R. M., Evans, R. S., Orme, J. F., Jr., Clemmer, T. P., Greenway, L., & Drews, F. A. (2007). Assessing the data quality in manual data entry of ventilator settings. *Journal of the American Medical Informatics Association: JAMIA, 14*, 295–303. PMID 17329731.

Wager, K. A., Schaffner, M. J., Foulois, B., Swanson Kazley, A., Parker, C., & Walo, H. (2010). Comparison of the quality and timeliness of vital signs data using three different data-entry devices. *Computers, Informatics, Nursing, 28*(4), 205–212. PMID: 20571371.

Warner, H. R., Swan, H. J., Connolly, D. C., Tompkins, R. G., & Wood, E. H. (1953). Quantitation of beat-to-beat changes in stroke volume from the aortic pulse contour in man. *Journal of Applied Physiology, 5*, 495–507. PMID 13034677.

Warner, H. R., Gardner, R. M., & Toronto, A. F. (1968). Computer-based monitoring of cardiovascular function in postoperative patients. *Circulation, 37*(4 Suppl), II68–II74.

Weinfurt, P. T. (1990). Electrocardiographic monitoring: An overview. *Journal of Clinical Monitoring, 6*(2), 132–138.

Weiss, C. H., Moazed, F., McEvoy, C. A., Singer, B. D., Szleifer, I., Amaral, L. A., Kwasny, M., Watts, C. M., Persell, S. D., Baker, D. W., Sznajder, J. I., & Wunderink, R. G. (2011). Prompting physicians to address a daily checklist and process of care and clinical outcomes: A single-site study. *American Journal of Respiratory and Critical Care Medicine, 184*(6), 680–686. PMID: 21616996.

Willson, D. (1994). Survey of nurse perception regarding the utilization of bedside computers. *Proceedings of the Annual Symposium on Computer Applications in Medical Care*, pp. 553–557. PMID 7949989.

Willson, D., Nelson, N. C., Rosebrock, B. J., Hujcs, M. T., Wilner, D. G., & Buxton, R. B. (1994). Using an integrated point of care system: A nursing perspective. *Topics in Health Information Management, 14*, 24–29. PMID 19134757.

Wright, A., Sittig, D. F., Ash, J. S., Feblowitz, J., et al. (2011). Development and evaluation of comprehensive clinical decision support taxonomy: Comparison of front-end tools in commercial and internally developed electronic health records. *Journal of the American Medical Informatics Association: JAMIA, 18*, 232–242. PMID 21415065.

Zhang, Y., & Szolovitz, P. (2008). Patient-specific learning in real time for adaptive monitoring in critical care. *Journal of Biomedical Informatics, 41*, 452–460. PMID 18463000.

Zoll, P. M., Linethal, A. J., Gibson, W., Paul, M. H., & Norman, L. R. (1956). Termination of ventricular fibrillation in man by externally applied countershock. *The New England Journal of Medicine, 254*, 727–732. PMID 13309666.

Zong, W., Moody, G. B., & Mark, R. G. (2004). Reduction of false arterial blood pressure alarms using signal quality assessment and relationship between the electrocardiogram and arterial blood pressure. *Medical & Biological Engineering & Computing, 42*, 698–706. PMID 15503972.

Imaging Systems in Radiology

Bradley J. Erickson

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_22

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What are the key components needed for Radiological Image interpretation?
- What are the roles of the Radiology Information System (RIS), Picture Archiving and Communication System (PACS), Computer-Aided Diagnosis (CAD), Vendor Neutral Archives (VNAs), and Advanced Visualization Systems (AVS) in a typical medical imaging department?
- How does the DICOM standard differ from HL-7 in the structure of its information model?

22.1 Introduction

In ▶ Chap. 10, we introduced the concept of **digital images** as a fundamental data type that, because of its ubiquity, must be considered in many applications. We furthermore defined biomedical **imaging informatics** as the study of methods for generating, manipulating, managing, extracting, and representing imaging information.

In this chapter, we continue the study of imaging informatics, begun in ▶ Chap. 10, by describing many of the methods for generating and manipulating images and discuss the relationship of these methods to structural informatics. We emphasize methods for managing and integrating images, focusing on how images are acquired from imaging equipment, stored, transmitted, and presented for interpretation. We also focus on how these processes and the image information are integrated with other clinical information and used in the health care enterprise, so as to have an optimal impact on patient care.

We discuss these issues in the context of Radiology, since imaging is the primary focus of that field.[1] Yet imaging is an important part of many other fields as well, including Pathology, Hematology, Dermatology, Ophthalmology, Gastroenterology, Cardiology, Surgery (for minimally invasive procedures especially) and Obstetrics, which often do their own imaging procedures; most other fields that use imaging rely on Radiology and Pathology for their imaging needs.

The distribution of imaging responsibility has given rise to the need of many departments to address issues of image acquisition, storage, transmission, and interpretation. As these modalities have become increasingly digital in form, the development of electronic systems to support these tasks has been needed.

We begin by describing some of the roles of imaging across all of biomedicine, then concentrate on image management and integration in radiology systems, bringing in illustrative examples from other disciplines where appropriate. Many Radiology departments are becoming highly distributed enterprises, with acquisition sites in intensive care units, regular patient floors, emergency departments, vascular services, screening centers, ambulatory clinics, and in affiliated community-based practice settings. Interpretation of images may be in those locations when dedicated onsite radiologists are needed, such as for local interventional procedures. Increasingly, however, high-speed networks are enabling interpretation at sites far from acquisition, either in a central location or in widely distributed locations according to the different capabilities or time zones of the organization. This is possible because image acquisition and interpreta-

1 The name Radiology is itself a misnomer, since the field is involved in using ultrasound, magnetic resonance, optical, thermal, and other non-radiation imaging modalities when appropriate. Radiology departments in some institutions are thus referred to alternatively as Departments of Medical Imaging or Diagnostic Imaging.

tion can be effectively decoupled. Independent imaging centers in a community face some of the same issues and opportunities, although to a lesser degree, so we focus primarily on the distributed medical center-based Radiology department in this chapter.

22.2 Basic Concepts and Issues

22.2.1 Roles for Imaging in Biomedicine

Imaging is a central part of the healthcare process for diagnosis, treatment planning, image-guided interventions, assessment of response to treatment, and prediction of outcome. In addition, it plays important roles in medical communication and education, as well as in research.

22.2.1.1 Detection and Diagnosis

The primary uses of images are for the detection of medical abnormalities, for diagnosing the nature of those abnormalities, and for planning and guiding therapeutic interventions. Detection focuses on identifying the presence of an abnormality, but in the case where the findings are not sufficiently specific to be characteristic of a particular disease, other information is required for actual diagnosis. This is the case, for example, with mammograms, which are often used to screen for breast cancer; once a suspicious lesion is detected, a biopsy procedure is usually required for diagnosis. In other circumstances, the image finding is diagnostic: for example, the finding of focal stenosis or obstruction of an artery during angiography. Most often there is a continuum between detection and diagnosis, with imaging detecting a lesion with some range of confidence, and suggesting some possibilities, known as the **differential diagnosis** (see ▶ Chap. 2).

Diagnosis and detection can be done with a wide variety of imaging procedures. Images produced by visible light can be used by ophthalmologists for retinal photography, but also by dermatologists to view skin lesions, or by pathologists for light microscopy. The visible-light spectrum is also responsible for producing images seen endoscopically, typically captured as video images or sequences(movies). Sound energy, in the form of echoes from internal structures, is used to form images in ultrasound, a modality used heavily in cardiac, abdominal, pelvic, breast, thyroid, testes, and obstetrical imaging. In addition, **Doppler shifts** of sound frequency can measure blood flow velocity in both arteries and veins and newer microvascular imaging methods can measure perfusion in capillary beds. X-ray energy produces radiographic and **computed-tomography** (CT) images of most parts of the body: the differential absorption of X-rays by various tissues produces the varying densities that enable radiographic images to portray normal and abnormal structures. More recent techniques that separate the various energies of the X-ray beam allow for more precise characterization of tissue composition, a simple form being known as dual-energy CT, and more advanced forms known as spectral CT. Emission of radioactive particles by isotopes that are incorporated into various types of molecules are used to produce nuclear-medicine images, which reflect the differential concentration of those molecules in various tissues. **Magnetic-resonance imaging (MRI)** depicts energy fluctuations of certain atomic nuclei—usually hydrogen—when they are aligned in a magnetic field and then perturbed by a radiofrequency pulse. Parameters such as proton density, the rate at which the nuclei return to alignment (T1), the rate of loss of phase coherence after the pulse (T2), diffusion of water, and even the concentration of certain chemicals (MR Spectroscopy) can be measured. These quantities differ in various tissues under normal conditions, with more variations due to disease, thus enabling MRI to distinguish among them. ◘ Figure 22.1 shows some example images. It is also possible to use MRI methods to accentuate and measure the flow of fluids like blood or CSF, known as Magnetic Resonance Angiography (MRA).

22.2.1.2 Assessment and Planning

In addition to being used for detection and diagnosis, imaging is often used to assess a patient's health status in terms of progression of a disease process (such as determination of

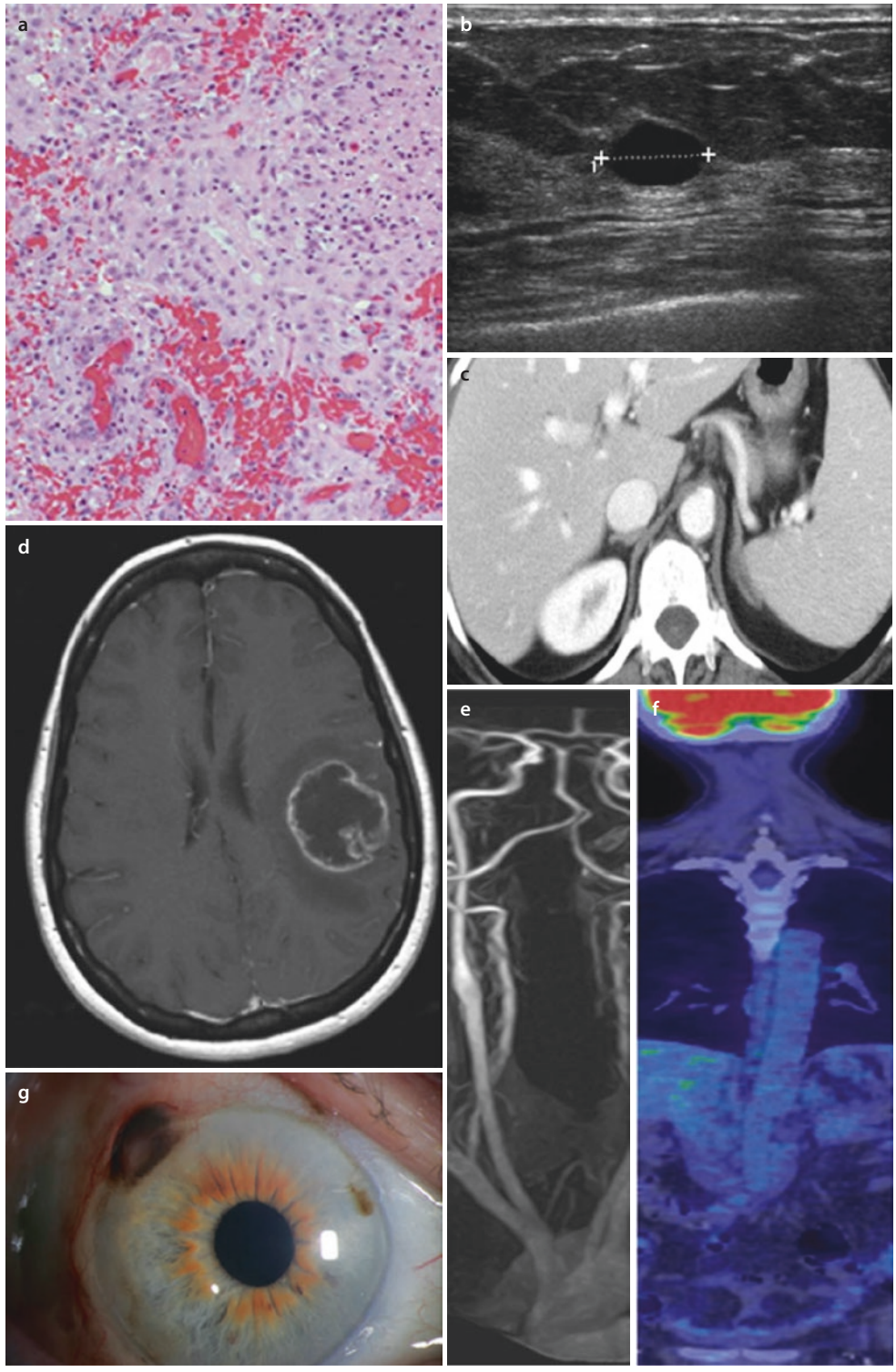


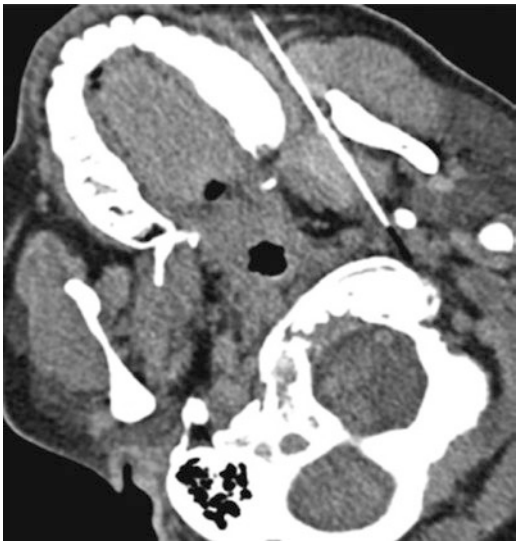
Fig. 22.1 Examples of the types of images discussed in the text. (**a**) is a microscopic image of tissue stained with hematoxylin and eosin; (**b**) is an ultrasound image of the thyroid; (**c**) is a contrast-enhanced CT image of the abdomen; (**d**) is contrast-enhanced MRI of the brain; (**e**) is an MR angiogram of the cervical vessels; (**f**) is an FDG PET-CT image of the upper body; (**g**) is a photgraph of an eye (Dermatopath, US, CDUS, CT, MRI, MRI-DWI)

tumor stage), response to treatment, and estimation of prognosis. One can analyze cardiac status by assessing the heart's size and motion echocardiographically. Similarly, one can use ultrasound to assess fetal size and growth. Computed tomography is used frequently to determine approaches for surgery or for radiation therapy. In the latter case, precise calculations of radiation-beam configuration can be optimized to maximize dose to the tumor while minimizing absorption of radiation by surrounding tissues. This calculation is often performed by simulating multiple radiation-beam configurations and iterating to a best treatment plan. For surgical planning, three-dimensional volumes of CT or MRI data can be constructed and presented for viewing from different perspectives to facilitate determination of the most appropriate surgical approach. More recently, creation of 3D printed models using 3D printers has become very popular for surgical planning and patient education. In some cases, the printed object is even used to guide incisions or implanted as a scaffold for tissue ingrowth.

22.2.1.3 Image-Guided Procedures

Images can provide real-time guidance when virtual-reality (VR) images are superimposed on a surgeon's visual perspective on the appropriate image view in the projection that demonstrates the abnormality, a technique known as augmented reality (AR). With endoscopic and minimally invasive surgery, this kind of imaging can provide a localizing context for visualizing and orienting the endoscopic findings, and can enable monitoring of results of interventions such as focused ultrasound, cryosurgery, or thermal ablation. It is also possible to use intra-operative imaging to update the position and appearance of pre-operative imaging used for procedural planning. ◘ Figure 22.2 shows an example of a CT-guided biopsy of a lesion in the neck. High quality imaging allows precise targeting of small targets such as diseased lymph nodes with little risk of damaging important nearby structures like the aorta or carotid artery.

Improvements in robotics technology and wide-area network capability have enabled minimally invasive procedures to be conducted at a distance (see ► Chap. 20), although it is still practical to do so only in limited settings. Because the abnormality is viewed through a video display, the image source can be physically remote, a technique called **telepresence**. Similarly, the manipulation of the endoscope itself can be controlled by a robotic device that reproduces the hand movements of a remote operator, and can provide **haptic feedback** reproducing the sensations of tissue textures, margins, and resistance. This technology is not too different from the robotic surgery methods that have become quite common today, though the practical limits noted above have limited its use.



◘ **Fig. 22.2** Example of a CT-guided biopsy of a lesion in the neck. High quality imaging allows precise targeting of small targets even near important structures like the carotid artery

22.2.1.4 Communication

Medical decision-making, including diagnosis and treatment planning, is often aided by allowing clinicians to visualize images concurrently with textual reports and discussions of interpretations. Thus, we consider imaging to be an important adjunct to communication and images to be a desirable component of a multimedia electronic medical record. Because medical imaging is an essential element of the practice of medicine, support for transmission and remote image viewing is also a critical component of telemedicine

(▶ Chap. 20). Medical images can also be helpful in doctor-patient communication, to enable the provider to illustrate an abnormality or explain a surgical procedure to a patient (▶ Chap. 11).

22.2.1.5 Education and Training

Images, whether 2D, 3D (either 3 spatial dimensions or 2D plus time), or 4D (3 spatial dimensions plus time) are an essential part of medical education and training because so much of medical diagnosis and treatment depends on imaging and on the skills needed to interpret such images (see ▶ Chap. 24). Case libraries, tutorials, atlases, quiz libraries, and other resources using images can provide this kind of educational support. Three-dimensional printed models of both normal structures and various pathologic conditions are now being routinely created for patient and physician education. The ability to hold the structure in one's hand, and manipulate it have proven very valuable, particularly in cases that are unusual such as congenital deformities.

Taking a history, performing a physical examination, and conducting medical procedures also demand appropriate visualization and observation skills. Training in these skills can be augmented by viewing images and video sequences, as well as through practice in simulated situations. An example of the latter is an approach to training individuals in endoscopy techniques by using a mannequin and video images in conjunction with tactile and visual feedback that correlate with the manipulations being carried out.

As noted in the previous section, patients increasingly expect to understand more about their disease, and patient communications can be more effective by including relevant images. Imaging also has a consumer/patient education benefit, since access to appropriate images can be included along with the provision of instructions and educational materials to patients, whether that is about their disease, the procedures to be performed, required follow-up care, or about healthy lifestyles.

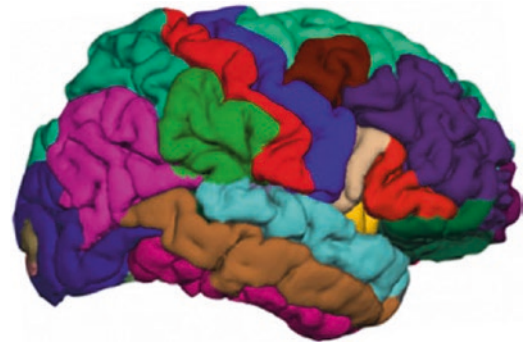


Fig. 22.3 Example of a detailed segmentation of the brain into various anatomic structures by the FreeSurfer package. It uses a combination of image intensities and expected shapes for the brain and substructures to produce its output

22.2.1.6 Research

Imaging is also a critical component of many aspects of research. An example is structural modeling of DNA and proteins, including their 3D and 4D configurations (see ▶ Chap. 9). Images obtained in molecular or cellular biology can show the distributions of fluorescent or radioactively tagged molecules through time or space. The study of **morphometrics**, which is literally the measurement of shape, depends on the use of imaging methods. Figure 22.3 shows an example of a detailed segmentation of the brain into various anatomic structures by the FreeSurfer package.[2] It uses a combination of image intensities and expected shapes for the brain and substructures to produce its output. **Functional mapping**—for example, of the human brain—relates specific sites on images to particular functions. While such quantitative imaging efforts often begin in the laboratory, translation of such quantitative methods is increasingly important to the practice of medicine. Figure 22.4 provides an example of functional mapping of a patient with a brain tumor, where functional mapping is used to identify critical structures, and thus to guide surgical therapy.

2 ▶ http://surfer.nmr.mgh.harvard.edu/ (Accessed 4/27/2018).

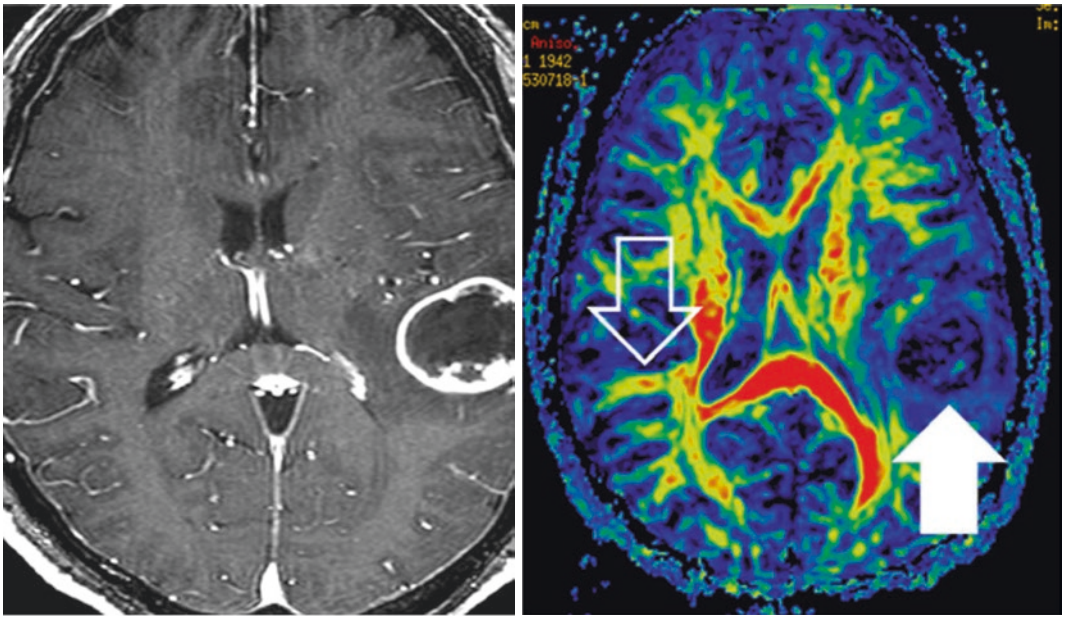


Fig. 22.4 Example of functional mapping of a patient with a brain tumor, where functional mapping is used to identify critical structures, and thus guide surgical therapy

22.2.2 The Radiologic Process and its Interactions

As noted in the introduction, we concentrate in this chapter on the subset of imaging that falls under the purview of **Radiology**. Radiology departments are engaged in all aspects of the healthcare process, from detection and diagnosis to treatment, follow-up and prognosis assessment. Radiology also illustrates well the many issues involved in acquiring and managing images, interpreting them, and communicating those interpretations. Space does not permit us to discuss the other disciplines that utilize imaging, but the processes involved and issues faced which we discuss in the context of radiology, pertain to the other disciplines also. Additional examples are also provided in ▶ Chap. 10. Occasionally, we intersperse examples from other areas, where we wish to emphasize a particular point, and imaging for educational purposes is discussed at length in ▶ Chap. 25.

The primary function of a Radiology department is the acquisition, analysis, and interpretation of medical images but also increasingly, the conduct of minimally invasive image-guided procedures, an area usually referred to as **Interventional Radiology.** Through imaging, healthcare personnel obtain information that can help them to establish diagnoses, to plan or administer therapy, and to follow the courses of diseases or therapies.

Diagnostic studies in the Radiology department are typically performed at the request of referring clinicians, who then use the information for subsequent decision-making. The Radiology department produces the images, and the radiologist provides the primary analysis and interpretation of the radiologic findings. Thus, radiologists play a direct role in clinical problem-solving and in diagnostic-work-up planning for many patients. Interventional radiology and image-guided surgery (if done by the radiologist) are activities in which the radiologist plays a primary role in treatment.

The complete radiologic process (Greenes 1989) is characterized by seven kinds of tasks, each of which involves information exchange, which may be augmented by information technology, as illustrated in Fig. 22.5. The first five tasks occur in sequence, whereas the

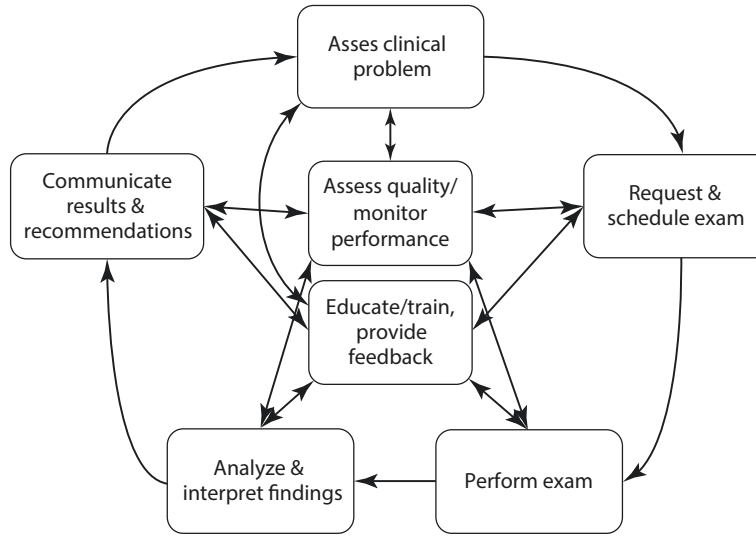


Fig. 22.5 The radiologic interpretation process

final two are done in parallel and are ongoing and support the other five.

1. The process begins with an evaluation by a clinician of a clinical problem and determination of the need for an imaging procedure. Decision-support tools (▶ Chap. 24) are commonly used to help determine if, and what type of, testing should be performed.
2. The procedure is requested and scheduled, the indication for the procedure is stated, and relevant clinical history is made available.
3. The imaging procedure is carried out, and images are acquired. An important step for many types of examinations is the 'protocoling' in which the precise way that the images are acquired are specified by the radiologist. For example, whether oral or IV contrast are administered, the imaging plane(s), slice thickness, and contrast properties are also specified.
4. The radiologist reviews the images in the context of the clinical history and indications for the examination and may measure structures in the images, segment components of the image (e.g. measure volumes such as the left ventricle), or manipulate the images (e.g. create 3D renderings or perform processing such as conversion of a series of images into a new parametric image like a blood volume image). This task actually involves interrelated subtasks: (a) detection of the relevant findings and (b) interpretation of those findings in terms of clinical meaning and significance.
5. The radiologist creates a report and may also directly communicate the results to the referring clinician, as well as making suggestions for further evaluation as needed. In the past, this was free text, but there is increasing use of templates that result in a consistent pattern in the report, or structured reports in which the textual report also exists in a form with codes for all concepts in the textual report. The annotation and markup of images can be very helpful in communicating locations of findings and serves as helpful landmarks for subsequent exams, and for surgical or radiation procedures.
6. Quality control and monitoring are carried out throughout the process, with the aim of improving the foregoing processes. Factors such as patient waiting times, workloads, numbers of exposures obtained per procedure, quality of images (such as ones degraded by patient motion or incorrect acquisition parameters), radiation dose, yields of procedures, incidence of complications, and quality of reports are measured, reported and adjusted to optimize individual and overall quality.

7. Continuing education and training are carried out through a variety of methods, including access to atlases, review materials, teaching-file cases, and feedback of subsequently confirmed diagnoses to interpreting radiologists. Peer review of previously reported cases is now a common expectation or requirement in most radiology practices as well as medical board agencies like the American Board of Radiology.

All these tasks are now, in a growing number of departments, computer-assisted or automated, and most of them involve images in some way. In fact, radiology is one branch of medicine in which even the basic data are usually produced by computers and stored directly in computer memory. Radiology has also contributed strongly to advances in computer-aided instruction (see ▶ Chap. 25), in technology assessment (see ▶ Chap. 13), and in clinical decision support (see ▶ Chap. 24). Speech recognition is commonly used for report creation.

22.2.3 Electronic Imaging Systems

22.2.3.1 Image Acquisition

The first radiographs used an integrated detection, recording and display system—that is, the glass plate (and later, plastic film) served both to detect the X-ray photons, to record them in a permanent form, and also to display the data (with the aid of a lightbox). This integrated arrangement existed for about a century. Today, most radiographs are either (a) recorded in a latent form (i.e., they are not directly visible, but are acquired as an electronic signal on a charged plate) where a 'reader' then scans the plate to create a digital image (known as **computed radiography** or CR) or (b) the photons are directly converted to digital images (known as **digital radiography** or DR). The digital image can then be transmitted and stored like any digital data, using conventional networks and storage systems. The matrix size of the images is variable, ranging from as low as 64 x 64 for some nuclear medicine images, up to 5000 x 4000 picture elements (pixels) for mammograms. The size of typical radiology images and examinations is shown in ◘ Table 22.1. CT, MR, US, NM, and PET all use computer to convert the acquired raw signal to a digital image and thus they are also exist as direct or nearly direct digital modalities.

22.2.3.2 DICOM

The first medical devices to produce digital images routinely were CT scanners, and soon after, MRI scanners. The availability of digital data that represented a three-dimensional

◘ **Table 22.1** Typical sizes for radiology examinations

Modality[a]	Image size (pixels/image)	Images/exam	Exam size (MB)
CR/DR	5000,000	3	29
CT	262,144	500	250
MRI	65,536	500	63
US	262,144	50	25
Mammography	20,000,000	4	153
Interventional/fluoro	1,048,576	50	100
Nuclear medicine	16,384	25	1

Note that there is variability in image size and images per examination, and these numbers should be viewed as very rough estimates. Furthermore, there is a strong trend for both increased image resolution (increasing image size) and more images per examination since the emergence of digital imaging

[a]*CR* computer radiography, *DR* digital radiography, *CT* computed tomography, *MRI* magnetic resonance imaging, *US* ultrasound

image stimulated the field of medical image processing and 3D rendering. An early challenge to such investigations was that the medical device vendors used half-inch tape media for storing the data, but each vendor (and usually each model of scanner) had its own format. Such formats were proprietary, and required each investigator to reverse engineer the format of the tape just to gain access to the data. Although computer networks were used in hospitals at that time, few if any scanners supported network connections.

The need to write all data to tape, and then read it into a different computer using software unique to each scanner resulted in significant unnecessary effort. The need to exchange images efficiently demanded that they be represented in a standard fashion. This need was recognized by the American College of Radiology (ACR) and the National Electrical Manufacturer's Association (NEMA), and led to the development of the ACR/NEMA standard for medical images in 1985. As other imaging devices started to produce digital images, and as the information about the images became richer, the second version was published in 1989. That standard described both a model that described the data, and prescribed a special connector for transferring image data between devices. This was demonstrated at the 1990 Radiological Society of North America (RSNA) conference. Soon after, **TCP/IP** became a widely accepted network standard, and while the ACR/NEMA standard did not describe a method for transferring data over TCP/IP, investigators fairly quickly implemented this, and it worked well. Continued improvements in the information model, as well as extension to medical specialties other than radiology, and standards for storage on physical media like compact disks demanded further revisions. The addition of non-radiology images also demanded a name change, and thus 'ACR/NEMA 3.0' was rebranded as **DICOM**, which stands for Digital Image Communications in Medicine.

To promote adoption, the RSNA commissioned the creation of **Central Test Node (CTN)** software, for demonstrating use of the standard for transmitting images over a local area network at RSNA 1992, followed by increasingly sophisticated versions over the next 2 years. The RSNA also made that software available for free public access as a model for understanding the standard and design of utilities and tools by developers. During the mid 1990's, the RSNA annual meetings hosted a major digital image interoperability demonstration that became progressively more sophisticated and demanding. RSNA and its meetings accordingly facilitated demonstration of the interconnection of vendor products through the Internet, promoted DICOM compatibility as a feature that could be visualized at participating vendor exhibits, and created a model Request For Proposals **(RFP)** for radiology practices and hospitals to use to craft a DICOM requirement as part of the procurement of imaging systems. Later this interoperability testing was done separate from the RSNA annual meeting, and became known as the 'Connectathon'. These efforts turned out to be extremely successful in transforming the marketplace from one that was dominated by proprietary formats to one that was standards-based and interoperable.

DICOM continues to be updated and improved through an international committee process. While it is hard for any standard to be both widely accepted and perfectly up-to-date, the DICOM governance has done a remarkable job of adapting to rapid advances in imaging technology. The governance continues to reflect its roots of combining industry and medical experts who are interested in providing the best technology that can be put into commercial products.

22.2.3.3 Image Transmission, Storage, and Display

Digital image capture provides the opportunity to display and store the images in digital form. In the early days, the size of the images represented a challenge—the amount of data was quite large relative to the capacity of storage devices. As a consequence, there was intense interest in using compression methods that could reduce the amount of storage that was required–as well as increase the speed of network transmission of images. Even with

compression, the amount of storage used for images is quite large relative to non-image data stored in a hospital. A hospital must therefore carefully consider how images are stored. In the early days of **Picture Archiving and Communications Systems (PACS)**, there was little choice about how and where images were stored, because the storage system was tightly integrated with the display and transmission. This was done because the high demands on storage, transmission, and display all required special hardware. As computer technology caught up with medical image sizes, there was less need for specialized versions of networks, archives, and displays.

The military was an early adopter/driver of PACS technology, and released an RFP in the 1990's requiring that 'any image be displayed anywhere on the network within 2 seconds.' To address this requirement, early PACS utilized either uncommon standard technologies or proprietary networking methods to provide high bandwidth transmission. An example of proprietary transmission technology is the PACS developed by LORAL, which leveraged technology developed for its defense applications. Its network was a hybrid of (standard) 10 Mbps **Ethernet**, which provided control signaling, and a (proprietary) unidirectional hub to spoke optical network that had lossless compression built into the network card. The optical network signaled at 100 Mbps, and because it was unidirectional, it routinely realized its nominal speed. Other vendors utilized **FDDI**, which also was an optical fiber network that signaled at 100 Mbps. However, its handling of contention was much less effective, and its performance suffered. Today, standard Ethernet signaling at either 100 Mbps or 1 Gbps can provide adequate performance, as long as reasonable attention is paid to network layout and implementation.

While the display component of PACS drove networking advances in the 1990's, the migration of PACS data from one system to another during an upgrade drove the next major change—the Vendor Neutral Archive, or VNA. In the early days of PACS, updating the system to take advantage of new workstation or network technology meant that the whole system, including the archive, needed to be updated or migrated. Because the data were not stored in a standard format, it was necessary to get the cooperation of the vendor to migrate the data to the new system (which might be from another vendor!). Because workstations rapidly change, but archive contents do not, there was perceived value in separating these two functions (Erickson and Hangiandreou 1998). Today, several companies sell **vendor-neutral archives** (VNAs) that leverage the DICOM standard. These allow a wide variety of image-producing and image-consuming systems to access the archive in a standard fashion. It also enables storing images from outside the radiology department on the same infrastructure. In fact, the use of a VNA is now the norm rather than the exception at any major hospital system because it allows sharing of resources across departments.

Because the image datasets are quite large, there is interest in finding ways to reduce storage requirements. Image compression does exactly this, in one of two ways. There are **lossless compression** methods, which encode redundancy in the image in a way that allows the original to be *exactly* reproduced. **Lossy** (or irreversible) **compression** produces an image that is visually similar to the original. Exactly how similar depends on the algorithm and user-selectable settings that reflect the trade-off between fidelity and compression ratio (the ratio of original size to compressed size). The major challenge is that one cannot select a given setting, reliably get images that are not visibly altered, and also achieve a good compression ratio. Lossless compression methods typically achieve compression ratios of about 2.5, while lossy compression can achieve as much as 40:1 compression without a perceptible or diagnostic loss, *for certain types of images*. While the size of imaging examinations continues to increase, the decrease in storage cost is more rapid, lessening the demand for lossy compression. The use of lossy compression is more widely accepted in non-radiology specialties, such as cardiology and pathology, in part because of the greater uniformity of image characteristics, allowing easy specification of acceptable ratios. A key goal of lossy compression is that

it not have an adverse impact on diagnostic value to the human or to computer aided diagnostic algorithms (Zheng et al. 2000). Thus, lossy methods are usually tuned to the specific diagnostic task so as not to have adverse impacts. In fact, monitoring the performance of CAD as the ratio is changed is one method to select the optimal compression ratio.

Early PACS also required specialized display devices. At the time, standard computer displays were often 640 x 480. Medical images were often more than 2048 pixels in each direction. Liquid crystal technology for large displays was also not developed, meaning that the displays were large cathode ray tubes. These displays were large, heavy, produced much heat, and degraded rather rapidly. Nearly all were monochrome. Imaging also required a higher luminance for detection of subtle gray level distinctions than was available with consumer-grade displays. Today, flat panel technology that meets the demands of radiological interpretation is widely available at reasonable prices. We note here that while consumer grade displays can be used, it is important that quality displays are used with appropriate calibration. A DICOM committee has established display requirements for medical purposes (ACR-NEMA 2006). Medical grade monitors typically have hardware built into the display to perform such calibration in an automated way.

22.2.4 Integration with Other Healthcare Information

22.2.4.1 Radiology Information Systems (RIS)

A **Radiology Information System**, or RIS, is responsible for much of the text information in a radiology department. Core functions of a RIS include capture of the interpretation for a given examination and records the status of an imaging examination. A RIS mayo do more, depending on what other systems are available and preferred for a given situation: some of the functions that might be performed by a RIS include ordering, scheduling, and billing. These functions are usually performed by the HIS or EHR system when available such as in a hospital or large outpatient practice; they are usually performed by the RIS in cases where there is no HIS or EHR (such as many outpatient imaging centers).

22.2.4.2 Speech Recognition

In the past, the RIS would provide a means for a transcriptionist to type the text of the report into the RIS as the radiologist dictated it (either live or via dictation system). Today, the vast majority of radiologists use speech recognition to convert their speech into text. In some cases, the text is immediately reviewed by the radiologist and approved as final. This model has the advantages of rapid **turn-around time**—the time from when the examination is ready to be reported to the time it has a final report available. In this model, a separate application (the speech recognition system) converts the audio to an HL-7 message, which is sent to the RIS along with the final (or other appropriate status). In other cases, a 'correctionist' reviews the text created by the speech recognition system, and corrects it based on listening to the audio file. In this case, the radiologist must then review the text again to make it final, which will degrade turn-around-time.

There are two major advantages to using speech recognition: First, it enables rapid turn-around time. Before speech recognition, turn-around times of 1 week were common, but now, turn-around times of less than 1 hour is common (Hart et al. 2010; Krishnaral et al. 2010; Mattern et al. 1999). This improvement in turn-around time undoubtedly improves the quality of care provided to patients. Second, it reduces staffing for radiology departments or hospitals by reducing the number of transcriptionists/correctionists needed. Of course, some decrease in productivity is commonly observed for radiologists at the time of implementation, which reduces the economic benefit (Langer 2002; Strahan and Schneider-Kolsky 2010).

Over the years, many efforts have attempted to enable radiologists to generate templated or **structured reports** (SR) from a selection of choices in forms, and through use of drop-down entries in the text, macros that produce

predetermined text phrases, and other techniques. Some of these are now used in specific situations, especially where reports have a largely anticipated format and structure, e.g., mammography and obstetrical ultrasound, and macros are used in conjunction with speech recognition approaches for certain "canned" sections or reports. Such reports enable efficient capture of the information for later data analysis, and there are also several reports showing that most referring physicians prefer templated reports because of the consistency in where key information can be found. In fact, some ultrasound devices will send many of the key measurements to the reporting system to 'pre-populate' a structured report with many of the measurements made during image acquisition.

22.2.4.3 Computer-Aided Diagnosis (CAD)

We have described that the interpretation task consists of detection, description, and diagnosis. In some cases, the detection task can be quite challenging, particularly for screening tasks involving mammography and chest X-rays because the incidence is rather low, and the volume is high. Particularly at the end of a long shift, human observers probably have decreased performance due to fatigue. For these cases, computer algorithms that highlight suspicious regions of an image may be useful to assure that important findings aren't missed. Some have called this role 'computer-aided diligence'.

During the first decade of **computer-aided diagnosis (CAD)**, the algorithms searched for specific features that radiologists thought were important, and therefore the value was either minimal (diligence), or provides benefit readers less familiar with the importance of various features (Gur and Sumkin 2006). In the case of mammography, the lack of a clear benefit for experienced readers, combined with the reduction in productivity has resulted in its near complete abandonment, and also its loss of reimbursement. This is a sobering example that computer assistance must be implemented in ways that truly add value.

Another role for CAD is in assisting with diagnosis (the lesion is detected, but unsure if it is cancer or infection). A common application is in determination the nature of lesions on high resolution CT images of the chest. Figure 22.6 shows an image of the output of the experimental algorithm CALIPER, rendered as a 3D image, to show the distribution and change of different degrees of interstitial lung disease in a patient.

More recently, a machine learning technique referred to as 'Deep Learning', has become popular. This technique uses neural networks and derives its 'deep' name because it typically uses many (50+) layers in the network. In addition to having many more layers than traditional neural networks, some forms include several convolutional layers (hence the name convolutional neural networks or CNNs) at the input that learn the features that produce the best output. This means that it learns the best features to use, rather than requiring a human to pre-compute them.

Deep Learning methods have proven very effective for many of the traditional CAD tasks (mammography and chest CT lesion detection) and for classification of lesions. They also excel at automated organ segmentation (U-Nets) that can be useful for many tasks, and this easy access to quantitative data will likely increase the quantitative content in radiology reports.

Perhaps more interesting is that CNNs have also proven effective at predicting molecular properties of tissues using routine images, even in cases where humans are not able to identify any differentiating features. Examples include the prediction of IDH-1 mutation, 1p19q chromosomal status, and MGMT methylation in brain tumors using routine T2-weighted MR.

22.2.4.4 Advanced Visualization

CT and MR scanners provide images that can be thought of as 3D images, even if they are not always truly acquired as 3D, but rather, as a series or 'stack' of 2D images. Some imaging devices can acquire a 3D image directly and repeatedly, thus producing a 4D image (time

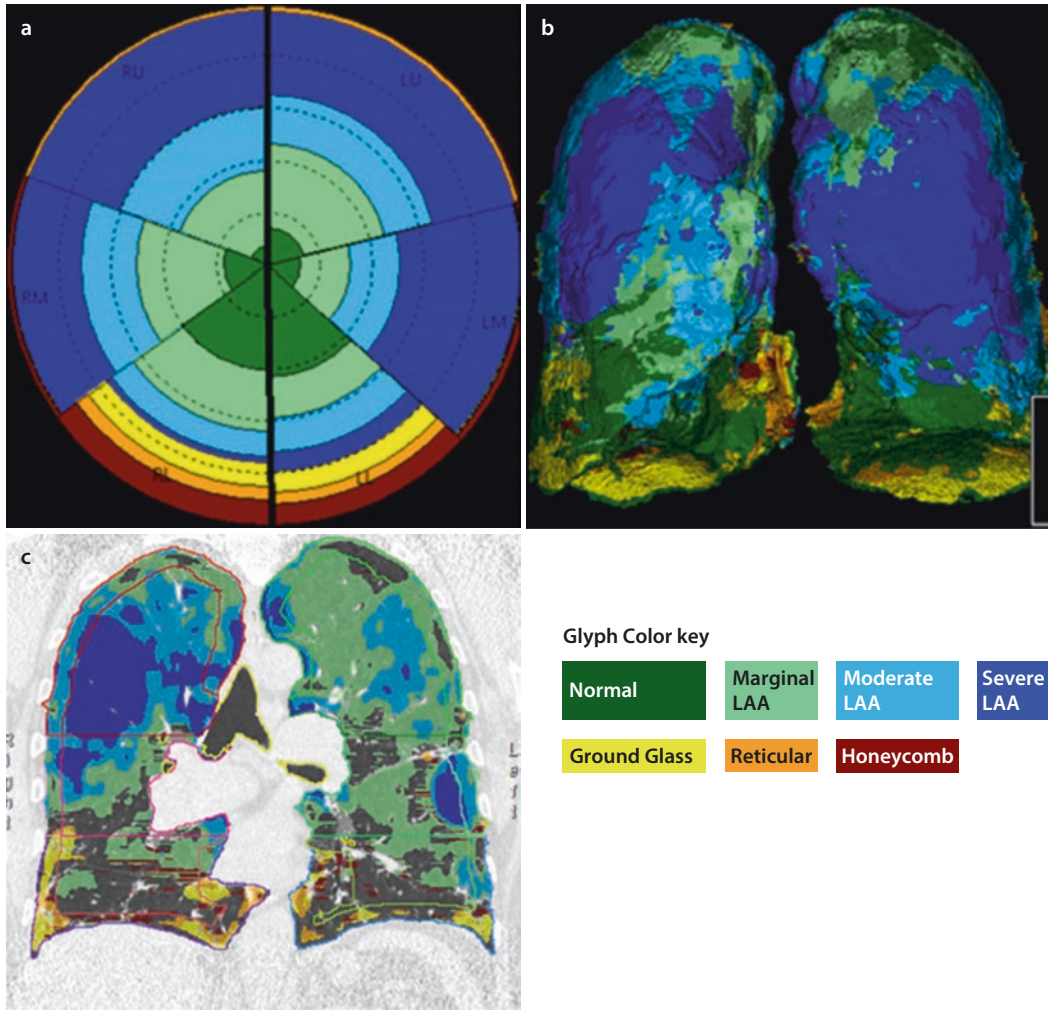


Fig. 22.6 Image of the output of the experimental algorithm CALIPER, rendered as a 3D image, to show the distribution and change of different degrees of interstitial lung disease in a patient. (**a**) shows the overall disease burden of the lungs; (**b**) shows the disease category as a 3D rendering of the lungs; (**c**) shows the disease category on a coronal section of the lungs; (**d**) shows the disease category labels

is the fourth dimension). In particular, cardiac imaging benefits from 4D capability so that the beating heart can be examined throughout the cardiac cycle. Such 3D and 4D data sets are large, and proper demonstration of the important findings requires visualization of the data. For instance, if one wishes to see a skeletal finding using CT, one can set a threshold to select bony structures, and then render it using traditional computer rendering methods. This can be done at multiple time points to produce movies of moving structures.

The great challenge in medical visualization is **segmentation**—deciding whether a volume element (voxel) is a part of the structure of interest or not. In the case of a CT image of bone, segmentation is rather easy. If intravascular contrast is administered during the examination, that can make it fairly straightforward to select vessels (arteries and/or veins depending on the timing). Soft tissue organs like livers, kidneys, and muscles have been more challenging, but are now much more feasible with deep learning techniques. A description of the rendering algorithms and their trade-offs is provided in ▶ Chap. 10.

A recent advance in visualization tools is to have the computation done on a central

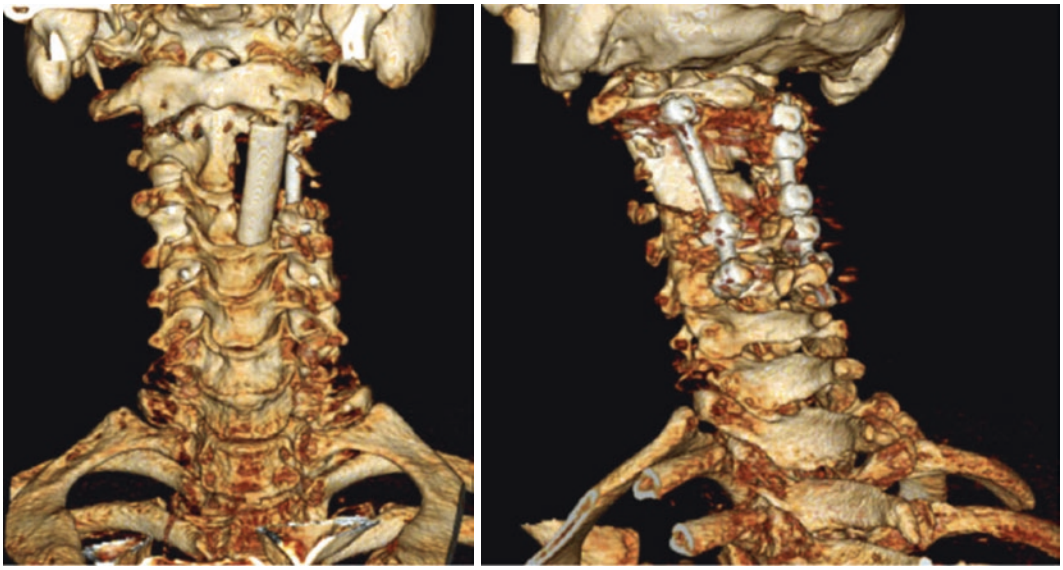


Fig. 22.7 Example advanced visualization of the abdominal aorta, distributed via a web client, which allows vascular surgeons to better plan surgical options in their own office

server, with interactive segmentation and rendering viewed using web browsers. This allows a much larger population of physicians to have access, and can be valuable for surgeons contemplating surgery, as well as for patient education. Figure 22.7 provides an example advanced visualization of the spine, distributed via a web client, which allows surgeons to better plan and review treatment options in their own office, with the patient.

22.2.4.5 Advanced Reporting

While textual reports have served medical practice well for the past century, there are opportunities to improve reporting. Multimedia reports provide a richer representation of the information present in the examination, and might include links from portions of the text report to specific images and locations on the images, moving images ('video'), or audio files such as the heart sounds. In some diseases, it can be important to have specific measurements made, and possibly tracked over time. If these measurements are encoded in a specific way (using Structured Reporting or SR), it will be easier to extract and use that information elsewhere in the medical record, and for other purposes like research. **Lexicons**, such as RadLex, can be helpful in conveying some of the information. There is great interest in routinely collecting more quantitative information from images, because it appears that for an increasing number of diseases, quantitation is receiving increased attention in clinical realms.

An SR is produced when the concepts of a report are represented using coded terminology. There is a DICOM specification for SR, though adoption has been limited. This is because there are currently not efficient user interfaces for creation of structured reports in most areas of radiology. BIRADS is a standardized way to report breast imaging, with an accepted scale for findings. However, those findings are usually not also stored in an encoded format, but rather with highly consistent text. Because BIRADS has truly enabled much better care of patients and research, other areas of imaging are adopting standardized reporting (e.g. TIRADS for thyroid, LIRADS for liver, PIRADS for prostate, just to name a few). The most widely adopted example of SR is probably the sending of specific measurements from ultrasound scanners to reporting systems. Of course, this is not the entire content of the report, and the radiologist usually adds and may edit the SR content. Another example where DICOM SR may be used is for reporting radiation dose, but again, that is not even the main content

of the radiological report. Some would also argue that DICOM SR is more correctly thought of as structured results rather than structure reports.

22.2.4.6 Workflow Management (Including Dashboards)

The ability to monitor and control events in an imaging department is critical to efficient and effective operation. **Dashboards** have been applied in many business arenas to give quick visual displays of **Key Performance Indicators (KPIs)** for that business. Such dashboard technology is now becoming widely used within imaging departments for monitoring such KPIs as report turn-around time, patient waiting time, number of days out to schedule examinations, and revenue days-outstanding. The dashboards give people a quick view of what is happening, and can alert them to problem areas. Figure 22.8 is an example of a radiology dashboard that shows important departmental metrics, including report turn-around time, compliance with notification requirements, and patient waiting times.

Most dashboards provide a mechanism to 'drill down' on a particular measurement. For instance, if the patient waiting time monitor goes 'red', clicking on that indicator light might show the waiting time by location (maybe just one facility is causing the problem), total patient volume (maybe the site is experiencing a spike in patient volume), or examination time (maybe the complexity of examinations is going up). Such information is critical to enabling a timely and effective response to performance that is outside expected service levels.

The popularity of deep-learning-based artificial intelligence (AI) tools for image interpretation has also driven the need for workflow orchestration: the application of workflow management technology to collect data needed by an AI algorithm, assure its execution on the images and other data, and then collect and distribute the results. There

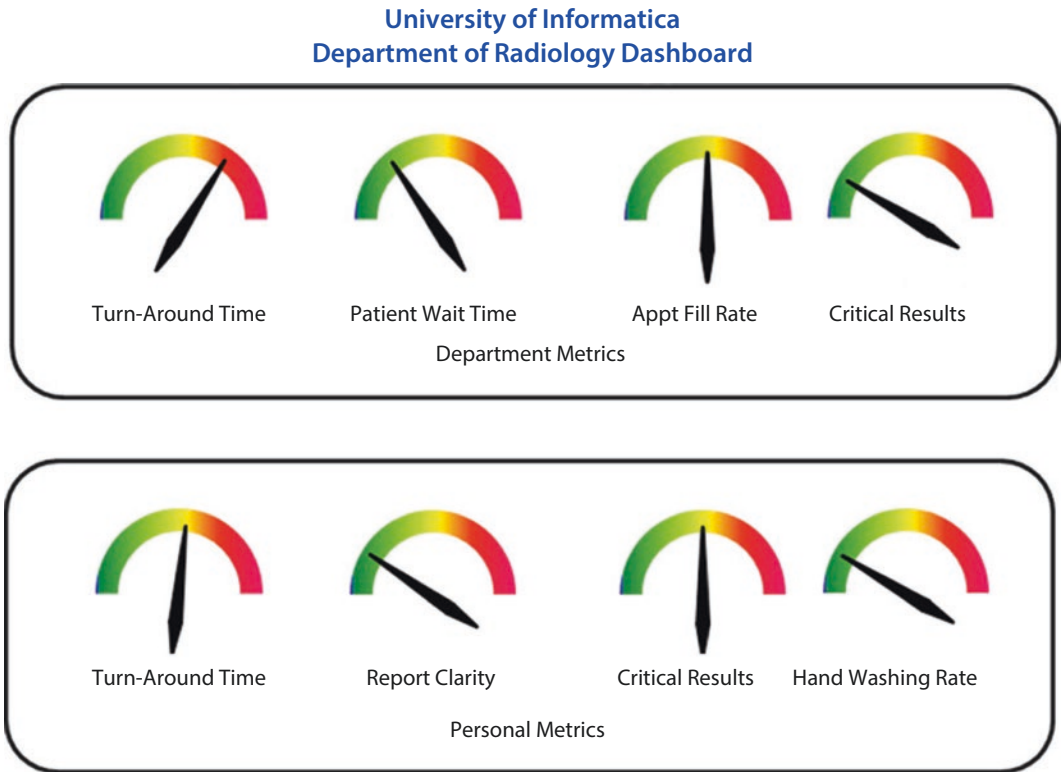


Fig. 22.8 Example of a radiology dashboard that shows important departmental metrics, including report turn-around time, compliance with notification requirements, and patient waiting times

are many reports showing AI can improve efficiency (such as for quantitation of organ or tumor size) and to improve the quality of care in radiology departments.

22.2.4.7 Teleradiology

Teleradiology is the practice of interpreting images at a location that is physically distant from the place where the images are collected. Initially, this referred to transmitting images from the hospital to the radiologist's home in the middle of the night so that the physician did not need to drive in to see the image onsite. While this is still done, it is now common for a hospital to contract with a 'nighthawk' service that will provide these night-time interpretations. A nighthawk service contracts with many hospitals—enough to keep a team of radiologists busy during the night. Having a team continuously operating is usually more efficient, and allows for specialization of image interpretation. Teleradiology is now also practiced during the day to balance clinical workload and to provide specialized interpretation on a routine basis. The technology to rapidly transmit images across large distances is widely available and inexpensive. The greatest challenges to teleradiology are licensing/credentialing issues, especially if films are read across state lines or internationally (radiologists may be licensed to practice where they review the films but not in the location where the patients were located when the images were acquired). Teleradiology is a term that is not heard very often mostly because transmission of images to the best site for interpretation has become routine, and implemented as a part of the PACS network.

22.2.4.8 Enterprise Integration (Including HL7, Decision Support)

Medicine is an information-rich business, and providing access to the relevant information in a timely fashion is critical to success. Integration of systems with the relevant pieces of information is necessary, and in hospitals this is generally done with HL-7 messages (see ► Chap. 7). Both RIS and PACS will typically be able to use HL7 messages, and an increasing number of EHRs are 'image enabled' meaning that they can receive DICOM images, and display them for clinicians.

22.2.4.9 Decision Support

Because medical imaging has been a major component of increases in total healthcare costs, there has been much attention paid to assuring that only necessary examinations are performed. To help assure this, decision-support systems to guide proper ordering have been developed that have been shown to have an impact on utilization of imaging studies (Sistrom et al. 2009). Such systems have been shown to decrease the total number of examinations performed, and in particular to decrease the number of examinations that appear to be unnecessary. In addition to alerting the user to a potentially improper order, educational materials are often provided to help the ordering physician understand when and what imaging examinations might be appropriate for the given indication. In addition, such systems can provide management reports to improve the understanding of ordering practices.

22.3 Imaging in Other Departments

22.3.1 Cardiology

Cardiac imaging has many similarities to radiological imaging, and in many cases is performed either by radiology departments or in conjunction with radiology. The primary imaging modalities for cardiology include echocardiography (ultrasound), catheterization (interventional/vascular, involving fluoroscopy and angiography, i.e., vessel visualization via contrast die administration), MR, CT, and PET. The workflow can be similar, but can be different in those cases where the imaging is performed by the same department and even by the same individual as the person who ordered it. In such cases, there can be less formal ordering, scheduling, and reporting. However, as the imaging is increasingly a part of the general enterprise, such informality will become a greater challenge.

Cardiology has been aggressive in its use of lossy compression. This is primarily because the nature of cardiac imaging is much more stereotyped (allowing better prediction of appropriate compression). The primary focus is the heart, whereas in radiology many different organs with very different appearances can challenge compression methods. The echocardiographic and interventional images are also much more like video—usually being motion-oriented rather than focused on static capture, and that redundancy in content enables effective compression. In fact, the major cardiovascular societies have published and supported the use of specific compression technologies and settings for cardiac imaging (Simon et al. 1994; ACC/ACR/NEMA Ad hoc Group 1995).

Because the focus is primarily on the heart and its function, cardiac imaging is more advanced in structured reporting. The primary variables that are of interest in cardiac imaging (left ventricular volumes, stroke volumes) are of interest in most cases, and are routinely measured and reported. This has driven the acceptance of structured reporting for the common measurements in cardiac imaging—particularly echocardiography.

22.3.2 Obstetrics and Gynecology

Obstetrical and gynecological imaging is rather like cardiac imaging, except that it is nearly always ultrasound imaging. Much like cardiac imaging, there are a well-defined and accepted set of measurements and observations expected for the routine obstetric exam, and as such, structured reports for these findings are widely used. Estimates of fetal gestational age and development are typically automatically computed, on the basis of measurements using well-tested prediction models.

22.3.3 Intraoperative/Endoscopic Visible Light

It is now common to capture still images and video of endoscopic procedures as well as traditional open surgical procedures. These images are valuable for documenting the important findings (or lack thereof) during a procedure. They are also useful for educational purposes, including informing the patient of the findings and procedures carried out. Some surgeons have suggested that medicolegal demands will require routine capture of entire surgical procedures.

Such images can be more graphic and revealing than radiological images, and in some cases have driven the expectation of need for an additional level of privacy protection. In one institution, for instance, all photographic images from the Plastic Surgery department are protected from access—only physician members of the Plastic Surgery department can routinely view those images, with a process for granting temporary access to other care providers (Erickson et al. 2007).

22.3.4 Pathology and Dermatology

Pathology and dermatology have similar needs, except that dermatology includes photographic visible light images of skin lesions. For these purposes, consumer-grade photographs can be sufficient, but transporting those images (usually in JPEG form) to a medical-grade imaging system will usually require an import process. This process will require confidence about the accuracy of patient and site location information. Often, the JPEG images are then 'wrapped' with DICOM information to assure that the connection between a photograph and a patient exists at the file level, rather than via a link to a filename. Image viewers require color capability, but otherwise are not substantially different from what is provided in most radiological image viewers. If images are converted to DICOM, the archive system is usually able to store them without difficulty.

Microscopic images represent a bigger challenge. At this point, there are 2 strategies for the capture of microscope images: the first is **whole slide digitization** in which the entire specimen is digitized; the other is capture of specific views that are of interest. In both cases, though more so in whole-slide imaging, there is a need for multi-resolution viewing.

That is because the workflow is very different. Whole-slide scanning is usually done *prior* to the pathologist reviewing the images, while the spot-capture is done by the pathologist at the time of viewing. In the former case, the computer performs the pan-zoom function, while in the latter case the optical microscope performs that function. In the former, the computer is a diagnostic device, while in the latter, it is used mostly for documentation purposes.

Whole-slide scanning is of greater interest, because it has greater possibilities for improving healthcare delivery by allowing the slides to go to the pathologist. It also represents a much greater challenge, because much larger amounts of data must be stored and the associated computer-based viewing application requires the ability to display low resolution and high resolution images. Such images are typically 1GB in size. Computing low resolution images from high resolution images can be computational expensive.

Another important issue differentiating pathology imaging from radiological imaging is that retrieval of old images is much less common in pathology. Follow-up of disease is biopsy is much less common than with radiological imaging, so comparison with priors is much less common. Recall would usually only occur when there is a medicolegal issue, or possibly in case of disease recurrence or metastasis, where there is a need to compare the older sample with a newer one.

22.4 Cross-Enterprise Imaging

22.4.1 CD Image Exchange

When images were stored on film, sharing images with another hospital required that the films be either physically transferred or copied. Copying a film was labor-intensive and expensive. Therefore, it was standard practice to 'loan' films to other facilities when needed.

With digital images, it is much easier to copy the digital data onto media like a compact disc (CD) and give that to the patient. There are well-accepted standards (DICOM) for how to store the images on a CD. However, it is challenging for most hospitals to use the images on CDs. In some cases, the images can be imported into the PACS, but that can cause confusion about where the study was done, and can be challenging for the RIS to represent this (who ordered the CD exam, and where the report is located). In addition, there are important data integrity issues—as much as 0.1% of CDs have been shown to include images for patients other than the intended patient (Erickson 2011), leading to important health delivery and legal risks.

22.4.2 Direct Network Image Exchange

The problems with CD image exchange noted above, as well as the time delays and costs, have driven many institutions to use internet transfer mechanisms. In cases where there is a high volume and a high level of trust, one can establish **virtual private networks (VPNs)** that allow secure transfer between two institutions. While this allows rapid and low-cost transfer, it still requires confident patient identification, and a method for importing the images into some form of clinical viewer. It also requires a hospital set up an extensive network of VPNs, which is something that can be difficult to secure.

This has led to the creation of the 'image sharing' industry. This industry has developed internet tools that allow images to be securely transferred from one hospital to another. The transport mechanism is usually proprietary, so transfer between hospitals is most efficient if both use the same vendor system. However, nearly all provide a 'Dropbox'-like method where a weblink can be used to either send or receive images. There are also efforts to develop standard interchange methods that are both efficient and secure.

22.5 Future Directions for Imaging Systems

The increasing capabilities of mobile devices and the increasing expectations of ready access to medical professionals have driven

imaging onto mobile devices. At present, the FDA has limited the use of such devices for diagnosis. On the other hand, these devices can be extremely useful for consultation on specific areas of an image, or when therapeutic options are being considered, or for patient communication. As the bandwidth and display qualities improve, these devices will likely play an increasing role in both diagnosis and therapy planning.

Cloud technology is becoming an important technology for image archival and transfer. The ability to leverage efficiencies of scale is an important economic driver that is pushing many smaller imaging providers to use cloud-based storage. The use of cloud methods for image exchange was also described above. The increasing use of computation intensive diagnostic aids has also driven the use of cloud-based CAD tools.

Phenome characterization (see ▶ Chap. 26) is becoming an important aspect of the move to individualizing medicine sometimes referred to as 'precision medicine'. Deep learning is a technology that will likely cause rapid advancement, as it allows both rapid automated segmentation of all major organs in an image, but also characterization of tissue textures. The ability to predict genomic information from such images will likely produce a new generation of 'precision imaging' applications.

Suggested Readings

Birkfelliner, W. (2010). Applied medical image processing: A basic course. CRC Press. As the title says, this is an introductory book, with many excellent explanations and example code (mostly MatLab).

Branstetter, B. F. (2009). Practical imaging informatics. New York: Springer. As its title implies, this book is a practice-oriented book primarily aimed at those responsible for implementing and maintaining a digital imaging practice. The format of the book is an outline with many practical tips from a wide variety of experts.

Dougherty, G. (2009). Digital image processing for medical applications. Cambridge University Press. This is an excellent, practical book on concepts of image processing algorithms used in medical imaging.

Dreyer, K. J. (2006). PACS: A guide to the digital revolution. New York: Springer. This is also a book focused on the practical aspects of implementing and maintaining a digital imaging department. Its format is that of a traditional textbook, and covers a broad range of topics.

Liu, Y., & Wang, J. (2011). PACS and Digital Medicine. Boca Raton: CRC Press. This book goes into greater detail of the technology of PACS, and to a lesser degree RIS and EMR. This is a very good resource for those interested in more details of DICOM and how a PACS can be configured to address specific needs.

Questions for Discussion

1. What are the Pros and Cons of a highly structured technology like DICOM? DICOM has been highly successful in terms of adoption as a standard, and virtually all image communication utilizes it. This differs markedly from some other standards. What are factors that have contributed to this success, and what lessons can be drawn from this in terms of how to promote adoption of standards in the future?
2. If one were to design medical imaging systems today, would the optimal design continue to have PACS and RIS as separate systems, or would they be combined into one system? Should these be separate from the EHR?
3. What are the ways in which radiology reports of examination interpretations can be generated, and what are the advantages and disadvantages of each approach, in terms of ease and efficiency of report creation, timeliness of availability of report to clinicians, usefulness for retrieval of cases for research and education?
4. In these days of high bandwidth and low storage costs, is there still a good reason to use lossy compression in medical imaging? What kinds of trends are likely to affect image growth, as part of the patient's medical record?
5. What are the arguments for maintaining raw rather than compressed data (not

only for imaging data but for compression or summarization of other types of data)?

6. Describe a classification of ways in which image data are used in medical decision making.
7. What are the data management implications of using a separate advanced visualization system for clinicians that is distinct from the PACS used by radiologists for interpretations? What if the radiologists use that system in addition to the PACS?

References

ACC/ACR/NEMA Ad Hoc Group. (1995). American College of Cardiology, American College of Radiology and industry develop standard for digital transfer of angiographic images. *Journal of the American College of Cardiology, 25*, 800–802.

ACR-NEMA. (2006). *Digital imaging and communications in medicine (DICOM) Part 14: Grayscale Standard Display Function*. Rosslyn: National Electrical Manufacturers Association.

Erickson, B. (2011). Experience with importation of electronic images into the medical record from physical media. *Journal of Digital Imaging, 24*(4), 694–699.

Erickson, B., & Hangiandreou, N. (1998). The evolution of electronic imaging in the medical environment. *Journal of Digital Imaging, 11*(3 suppl 1), 71–74.

Erickson, B., Ryan, W., Gehring, D., & Pynadath, A. (2007). Top 10 features clinicians require in an image viewer. *SIIM News* (Summer 2007).

Greenes, R. (1989, June). *The radiologist as clinical activist: A time to focus outward.* Paper presented at the Proc First Internat Conf on Image Management and Communication in Patient Care: Implementation and Impact (IMAC 89), Los Alamitos, CA.

Gur, D., & Sumkin, J. (2006). CAD in screening mammography. *AJR, 187*, 1474. https://doi.org/10.2214/AJR.06.1384.

Hart, J., McBride, A., Blunt, D., Gishen, P., & Strickland, N. (2010). Immediate and sustained benefits of a "total" implementation of speech recognition reporting. *The British Journal of Radiology, 83*(989), 424–427.

Krishnaral, A., Lee, J., Laws, S., & Crawford, T. (2010). Voice recognition software: Effect on radiology report turnaround time at an academic medical center. *AJR. American Journal of Roentgenology, 195*(1), 194–197.

Langer, S. (2002). Impact of speech recognition on radiologist productivity. *Journal of Digital Imaging, 15*(4), 203–209.

Mattern, C., Erickson, B., King, B., & Okryznski, T. (1999). Impact of electronic imaging on clinician behavior in the urgent care setting. *Journal of Digital Imaging, 12*(2 Suppl 1), 148–151.

Simon, R., Brennecke, R., Hess, O., Meier, B., Reiber, H., & Zeelenberg, C. (1994). Report of the ESC task force on digital imaging in cardiology. Recommendations for digital imaging in angiocardiography. *European Heart Journal, 15*, 1332–1334.

Sistrom, C., Dang, P., Weilburg, J., Dreyer, K., Rosenthal, D., & Thrall, J. (2009). Effect of computerized order entry with integrated decision support on the growth of outpatient procedure volumes: Seven-year time series analysis. *Radiology, 251*, 147–155.

Strahan, R., & Schneider-Kolsky, M. (2010). Voice recognition versus transcriptionist: Error rates and productivity in MRI reporting. *Journal of Medical Imaging and Radiation Oncology, 54*(5), 411–414.

Zheng, B., Sumkin, J., Good, W., Maitz, G., Chang, Y., & Gur, D. (2000). Applying computer-assisted detection schemes to digitized mammograms after JPEG data compression: An assessment. *Academic Radiology, 7*(8), 595–602.

Information Retrieval

William Hersh

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_23

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What types of online knowledge-based information are available and useful to clinicians, biomedical researchers, and consumers?
- What are the major components of the information retrieval process?
- What are the major categories of available knowledge-based information?
- How do techniques differ for indexing various types of knowledge-based biomedical information?
- What are the major approaches to retrieval of knowledge-based biomedical information?
- How effectively do searchers use information retrieval systems?
- What are the important research directions in information retrieval?
- What are the major challenges to making digital libraries effective for health and biomedical users?

23.1 Introduction

Information retrieval (IR), sometimes called **search**, is the field concerned with the acquisition, organization, and searching of knowledge-based information (Hersh 2020). Although biomedical IR has traditionally concentrated on the retrieval of text from the biomedical literature, the domain over which IR can be effectively applied has broadened considerably with the advent of multimedia publishing and vast storehouses of images, video, chemical structures, gene and protein sequences, and a wide range of other digital material and artifacts of relevance to biomedical education, research, and patient care. With the proliferation of IR systems and online content, the notion of the library has changed substantially, and new **digital libraries** have emerged (Lindberg and Humphreys 2005).

IR systems and digital libraries historically existed to store and disseminate knowledge-based information. What exactly does that mean? Although there are many ways to classify biomedical information and the informatics applications that use them, in this chapter we will broadly divide them into two categories. **Patient-specific information** applies to individual patients. Its purpose is to document and increasingly analyze for health care providers, administrators, and researchers about the health and disease of a patient. This information historically came from the patient's medical record but now can come from many different sources, including mobile and wearable devices. The second category of biomedical information is **knowledge-based information**. This is information that has been derived and organized from observational or experimental research. In the case of clinical research, this information provides clinicians, administrators, and researchers with knowledge derived from experiments and observations, which can then be applied to individual patients. This information has historically been provided in books and journals but can take a wide variety of other forms, including clinical practice guidelines, consumer health literature, Web sites, and so forth. The distinction between these two types of information is blurred by the growing amount of data that comes from people and is used to derive knowledge.

A basic overview of the IR process is shown in Fig. 23.1 and forms the basis for most of this chapter. The overall goal of IR or search is to find content that meets a person's

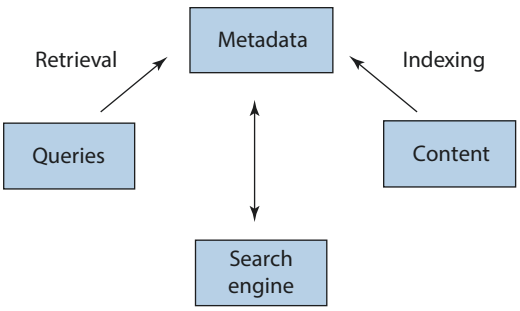


Fig. 23.1 Basic overview of the information retrieval process. Retrieval is made possible via metadata, which is produced via indexing and applied in queries by users. The metadata is used by the search engine, which directs the user to the content

information needs. This is done by posing a *query* to the IR system. A **search engine** matches the query to content items through **metadata**, which is "data about data" that describes the content items (Foulonneau and Riley 2008). There are two intellectual processes of IR. **Indexing** is the process of assigning metadata to content items, while **retrieval** is the process of the user entering his or her query and retrieving content items.

23.2 Evolution of Biomedical Information Retrieval

As with many chapters in this volume, IR has changed substantially over the five editions of this book. In the first edition, this chapter was titled "Bibliographic-Retrieval Systems," reflecting the predominant type of knowledge that was accessible at the time. The second edition saw the emergence of the **World Wide Web (WWW** or **Web)** as a delivery mechanism for knowledge-based information. In the third edition, "Digital Libraries" was added to the chapter name, reflecting that the entire biomedical library and beyond was now part of available online knowledge. The fourth edition reflected the ubiquitous nature of information on computers, smartphones, tablets, and other devices. This fifth edition recognizes that digital data, information, and knowledge have become their primary medium, i.e., even though the world still has plenty of paper, the source of most information in modern times is digital. Essentially all articles, books, patient records, etc. are primarily in digital form and mainly printed for the convenience of reading.

Although this chapter focuses on the use of computers to facilitate IR, methods for finding and retrieving information from medical sources have been in existence for nearly a century and a half. In 1879 Dr. John Shaw Billings created **Index Medicus** to help medical professionals find relevant journal articles (DeBakey 1991). Journal article citations were indexed by author name(s) and subject heading(s) and then aggregated in bound volumes. A scientist or practitioner seeking an article on a topic could manually search the index for the single best-matching subject heading and then be directed to citations of published articles.

The printed Index Medicus served as the main biomedical IR source until 1971, when the **National Library of Medicine (NLM)**[1] unveiled an electronic version, the **Medical Literature Analysis and Retrieval System (MEDLARS)**, which had been cataloging bibliographic records since 1966 (Miles 1982). Because computing power and disk storage were very limited, MEDLARS and its follow-on **MEDLARS Online (MEDLINE)**, stored only limited information for each article, such as author name(s), article title, journal source, and publication date. In addition, the NLM assigned to each article a number of terms from its **Medical Subject Headings (MeSH)** vocabulary. Searching was done by users having to mail a paper search form to the NLM and receiving results back a few weeks later. Only librarians who had completed a specialized course were allowed to submit searches.

As computing power grew and disk storage became more plentiful in the 1980s, full-text databases began to emerge. These new databases allowed searching of the entire text of medical documents. Although lacking graphics, images, and tables from the original source, these databases made it possible to retrieve the full text of important documents quickly from remote locations. Likewise, with the growth of computer networks, end users were now allowed to search the databases directly, though at a substantial cost.

In the early 1990s, the pace of change in the IR field quickened. The advent of the Web and the exponentially increasing power of computers and networks brought a world where vast quantities of medical information from multiple sources with various media extensions were now available over the global Internet (Berners-Lee et al. 1994). In the late 1990s, the NLM made all of its databases available to the entire world for free. Also during this time, the notion of digital libraries developed, with the recognition that the

1 ► https://www.nlm.nih.gov/

entire array of knowledge-based information could be accessed using this technology (Borgman 1999).

Into the twenty-first century, use of IR systems and digital libraries has become ubiquitous. Estimates vary, but among individuals who use the Internet in the United States, over 80% have used it to search for information relevant to their own health or that of an acquaintance (Fox 2011; Taylor 2010). Virtually all physicians use the Internet (Anonymous 2012). Furthermore, access to systems has gone beyond the traditional personal computer and extended to new devices, such as smartphones and tablet devices.

23.3 Knowledge-Based Information in Health and Biomedicine

Knowledge-based information can be subdivided into two categories. Primary knowledge–based information (also called *primary literature*) is original research that appears in journals, books, reports, and other sources. This type of information reports the initial discovery of health knowledge, usually with either original data or reanalysis of data (e.g., systematic reviews, sometimes with meta-analysis). Secondary knowledge–based information consists of the writing that reviews, condenses, and/or synthesizes the primary literature. The most common examples of this type of literature are books, monographs, and review articles. Secondary literature includes the growing quality of patient/consumer-oriented health information that is increasingly available via the Web. It also encompasses opinion-based writing (such as editorials and position or policy papers), clinical practice guidelines, narrative reviews, and health information on Web pages. In addition, it includes the plethora of pocket-sized manuals that were formerly a staple for practitioners in many professional fields. As will be seen later, secondary literature is the most common type of literature used by physicians.

Libraries have been the historical place where knowledge-based information has been stored. Libraries actually perform a variety of functions, including the following:

- Acquisition and maintenance of collections
- Cataloging and classification of items in collections to make them more accessible to users
- Serving as a place where individuals can get assistance with seeking information, including information on computers
- Providing work or study space (particularly in universities)

Digital libraries provide some of the same services, but their focus tends to be on the digital aspects of content.

23.3.1 Information Needs and Information Seeking

Different users of knowledge-based information have differing needs based on the nature of their information need and available resources. The information needs and information seeking of physicians have been most extensively studied. Gorman defined four states of **information need** in the clinical context (Gorman and Helfand 1995):

- Unrecognized need—clinician unaware of information need or knowledge deficit
- Recognized need—clinician aware of need but may or may not pursue it
- Pursued need—information seeking occurs but may or may not be successful
- Satisfied need—information seeking successful

There is a great deal of evidence that the majority of information needs are not being satisfied and that IR applications may help. Among the reasons that physicians do not adhere to the most up-to-date clinical practices is that they often do not recognize that their knowledge is incomplete. While this is not the only reason for such practices, the evidence is compelling. For example, physicians do not always provide patients with most up-to-date care (McGlynn et al. 2003), do not adhere to established guidelines (Diamond

and Kaul 2008), and vary widely in how they provide care (Wennberg 2010).

Studies from the late twentieth century found that when physicians recognize an information need, they were likely to pursue only a minority of unanswered questions. These studies found that physicians in practice had unmet information needs on the order of two questions for every three patients seen and only pursued answers for about 30% of these questions (Covell et al. 1985; Ely et al. 1999; Gorman and Helfand 1995). When answers to questions were actually pursued, these studies showed that the most frequent source for answers to questions was colleagues, followed by paper-based textbooks. Therefore, it is not surprising that barriers to satisfying information needs remain (Ely et al. 2002). Physicians use electronic sources more now than were measured in these earlier studies, with the widespread use of the **electronic health record (EHR)** as well as ubiquity of portable smartphones and tablets, although less research is undertaken in modern times assessing their needs. Another approach to facilitating access to knowledge-based information has been to link it more directly with the context of the patient in the EHR (Cimino and delFiol 2007).

The information needs of other users have been less well-studied. For consumers, surveys found about 80% of all Internet users searched for personal health information (Fox and Duggan 2013). The most common type of search focuses on a specific disease or medical problem (66% of all who have searched), followed by a specific medical treatment or procedure (56%). Consumers also use the Web to search for physicians, health care institutions, and health insurance. Even less studied have been the information needs of researchers, but one recurrent finding is the idiosyncratic nature of their use of IR and other systems (Bartlett and Toms 2005).

23.3.2 Changes in Publishing

The Internet and the Web have had a profound impact on the publishing of knowledge-based information. The technical impediments to electronic publishing of journals have been overcome, such that virtually all scientific journals are published electronically now. A modern Internet connection is sufficient to deliver most of the content of journals. When available in electronic form, journal content is easier and more convenient to access. Furthermore, since most scientists have the desire for widespread dissemination of their work, they have incentive for their papers to be available electronically.

The technical challenges to electronic scholarly publication have been replaced by economic and political ones (Hersh and Rindfleisch 2000; Sox 2009). Printing and mailing, tasks no longer needed in electronic publishing, comprised a significant part of the "added value" from publishers of journals. There is still however value added by publishers, such as hiring and managing editorial staff to produce the journals, and managing the peer review process. Even if publishing companies, as they currently exist today, were to vanish, there would still be some cost to the production of journals. Thus, while the cost of producing journals electronically is likely to be less, it is not zero, and even if journal content is distributed "free," someone has to pay the production costs. The economic issue in electronic publishing, then, is who is going to pay for the production of journals (Sox 2009). This introduces some political issues as well. One of them centers around the concern that much research is publicly funded through grants from federal agencies such as the National Institutes of Health (NIH) and the National Science Foundation (NSF). In the current system, especially in the biomedical sciences (and to a lesser extent in other sciences), researchers turn over the copyright of their publications to journal publishers. The political concern is that the public funds the research and the universities carry it out, but individuals and libraries then must buy it back from the publishers to whom they willingly cede the copyright. This problem is exacerbated by the general decline in funding for libraries.

One solution to this problem has been the emergence of **open-access (OA) publishing**. The premise of the OA model is that the

content is made freely available electronically, with the costs of production covered by other funding sources (Frank 2013). The most common funding source is the research funder, and most funders consider OA publishing costs to be an allowable expense on research grants. While some have expressed concern that OA may give rise to financial incentives for excess publishing, many OA journals have over a decade of experience demonstrating traditional peer review is compatible with OA publishing. Another concern has been the ability of resource-poor scientists to afford OA publishing, although most OA journals have "hardship" policies that allow waiver of the publication fee.

There has been the emergence of two models of OA publishing (Frank 2013). One is OA Gold, where the author (usually through the research funding) pays the cost of production. Publishing charges are typically only a tiny fraction of the overall cost of research, estimated to be about 0.3% (Zerhouni 2004). Two early publishers operating under this model were *Biomed Central* (BMC, now owned by Springer)[2] and the *Public Library of Science* (PLoS).[3] Many commercial publishers now offer authors the ability to publish under OA, and some journals have developed "Open" sister journals, such as JAMA, BMJ, and JAMIA.

A second model is OA Green, where authors are required to deposit the manuscript, either the published manuscript or the last draft of the manuscript prior to typesetting, into public repositories such as *PubMed Central* (PMC).[4] PMC is a repository of life science research that provides free access while allowing publishers to maintain copyright and even optionally keep the papers housed on their own servers. A lag time of up to 6 months is allowed so that journals can reap the revenue that comes with initial publication. The National Institutes of Health (NIH)5 now requires all research funded by its grants to be submitted to PMC, either in the form published by publishers or as a PDF of the last manuscript prior to journal acceptance.[6] Publishers have expressed concern that copyrights give journals more control over the integrity of the papers they publish (Drazen and Curfman 2004). An alternative approach, advocated by non-commercial (professional society) publishers is the *Washington DC Principles for Free Access to Science,*[7] which advocates:

- Reinvestment of revenues in support of science.
- Use of open archives such as PMC as allowed by business constraints.
- Commitment to some free publication, access by low-income countries, and no charges to publish.

One adverse outcome of OA publishing has been the emergence of so-called **predatory journals** (Haug 2013). These journals exist mainly to make money, allowing authors to publish in impressive-sounding titles with little or no peer review. Predatory journals offer inexpensive publishing and send massive emails to academic faculty members offering to publish or even serve on editorial boards. Some authors have exposed the process by writing clearly fake papers (McCool 2017), while others have issued calls for efforts to stop it (Moher and Moher 2016).

Some consider OA to be part of larger "open science," consisting of (Anonymous 2018b):

- Open data – all data collected in research
- Open source – all software code developed and used
- Open methodology – clear and detailed description; availability of all surveys and other tools
- Open peer review – all comments of peer reviewers

The move to predominant electronic publication of science has led many to advocate for growing access to the underlying data from

2 ► https://www.biomedcentral.com/
3 ► https://www.plos.org/
4 ► https://www.ncbi.nlm.nih.gov/pmc/
5 ► https://www.nih.gov/
6 ► https://publicaccess.nih.gov/
7 ► http://www.dcprinciples.org/

research studies. While this has been a standard practice in genomics and related areas for many years (i.e., depositing genome sequences in GenBank as a condition of publication), there have been concerns that have impeded this approach in clinical studies.

The case for requiring publication of scientific data is strong. As most research is taxpayer-funded, it only seems fair that those who paid are entitled to all the data for which they paid. Likewise, all of the subjects were real people who potentially took risks to participate in the research, and their data should be used for discovery of knowledge to the fullest extent possible. In addition, new discoveries may emerge from re-analysis of data. For these reasons, the International Committee of Medical Journal Editors (ICMJE) called for de-identified data from randomized controlled trials to be shared as condition of publication (Taichman et al. 2017). Other have called for data access to be FAIR (finable, accessible, interoperable, and reusable) (Wilkinson et al. 2016).

Some researchers, however, have pushed back on this notion. They argue that those who carry out the work of designing, implementing, and evaluating experiments certainly have some exclusive rights to the data generated by their work. Some also question whether the cost is a good expenditure of limited research dollars, especially since the demand for such data sets may be modest and the benefit is not clear. One group of 282 researchers in 33 countries, the International Consortium of Investigators for Fairness in Trial Data Sharing, notes that there are risks, such as misleading or inaccurate analyses as well as efforts aimed at discrediting or undermining the original research (Anonymous 2016). They also express concern about the costs, given that there are over 27,000 RCTs performed each year. As such, this group calls for an embargo on reuse of data for 2 years plus another half-year for each year of the length of the RCT. Even those who support data sharing point out the requirement for proper curation, wide availability to all researchers, and appropriate credit to and involvement of those who originally obtained the data (Merson et al. 2016).

23.3.3 Quality of Information

In the early days of the growth of the Internet and the Web, another concern was the quality of information available. A large fraction of Web-based health information has always been aimed at nonprofessional audiences. Many lauded this development as empowering those most directly affected by health care—those who consumed it (Eysenbach et al. 1999). Others expressed concern about patients misunderstanding or being purposely misled by incorrect or inappropriately interpreted information (Jadad 1999). Some clinicians also lamented the growing amount of time required to go through stacks of printouts downloaded by patients and brought to their offices. The Web was inherently democratic, allowing anyone to post information. However, this was potentially at odds with the operation of a professional field, particularly one like health care, where practitioners were ethically bound and legally required to adhere to the highest standard of care. Thus, a major concern with health information on the Web is the presence of inaccurate or out-of-date information. An early systematic review of studies assessing the quality of health information found that 55 of 79 studies came to the conclusion that quality of information was a problem (Eysenbach et al. 2002). More recent studies continue to show the variable quality of health information on the Web (Kitchens et al. 2014).

A more recent problem, extending beyond health care, has been the proliferation of "fake news" and "alternative facts," often promulgated by those understanding how to manipulate search engines and social media (Vosoughi et al. 2018; Wenzel 2017). A related area is "adversarial" IR, where the goal is to not retrieve information the user does not want to see or should not see (Castillo and Davison 2011). One major research organization has lamented the larger societal impacts of such "truth decay," lamenting "erosion of civil discourse, political paralysis, alienation and disengagement of individuals from political and civic institutions" (Kavanagh and Rich 2018).

The impact of poor-quality health information is unclear. People were harmed by incorrect and misleading health information long before the emergence of digital information. One well-known self-help expert argued that patients and consumers actually are savvy enough to understand the limits of quality of information on the Web. This view held that patients and consumers should be trusted to discern quality using their own abilities to consult different sources of information and to communicate with health care practitioners and with others who share their condition(s) (Ferguson 2002). Indeed, the ideal situation may be a partnership among patients and their health care practitioners, as it has been shown that patients desire that their practitioners be the primary source of recommendations for online information (Tang et al. 1997).

This concern about quality of information led a number of individuals and organizations to develop guidelines for assessing the quality of health information. One of the earliest and most widely quoted set of criteria was published in JAMA (Silberg et al. 1997). These criteria stated that Web pages should contain the name, affiliation, and credentials of the author; references to the claims made; explicit listing of any perceived or real conflict of interest; and date of most recent update. Another early set of criteria was the *Health on the Net* (HON)[8] codes, a set of voluntary codes of conduct for health-related Web sites. Sites that adhere to the HON codes can display the HON logo. Another approach to insuring Web site quality is accreditation by a third party. URAC (formerly, the *Utilization Review Accreditation Commission*) has a process for such accreditation.[9] The URAC standards cover six general issues: health content editorial process, disclosure of financial relationships, linking to other Web sites, privacy and security, consumer complaint mechanisms, and internal processes required to maintain quality over time.

8 ► https://www.hon.ch/en/

9 ► https://www.urac.org/programs/health-web-site-accreditation

23.3.4 Evidence-Based Medicine

The growing quantity of clinical information available in IR systems and digital libraries requires new approaches to select that which is best to use for clinical decisions. The philosophy guiding this approach is **evidence-based medicine** (EBM), which can be viewed a set of tools to inform clinical decision-making. It allows clinical experience ("art") to be integrated with best clinical science (Guyatt et al. 2014, 2015). Also, EBM makes the medical literature more clinically applicable and relevant. In addition, it requires the user to be facile with computers and IR systems. The process of EBM involves three general steps:

- Phrasing a clinical question that is pertinent and answerable.
- Identifying evidence (studies in articles) that address the question.
- Critically appraising the evidence to determine whether it applies to the patient.

The phrasing of the clinical question is an often-overlooked portion of the EBM process. There are two general types of clinical question: background questions and foreground questions (Guyatt et al. 2014, 2015). **Background questions** ask for general knowledge about a disorder, whereas **foreground questions** ask for knowledge about managing patients with a disorder. Background questions are generally best answered with textbooks and classical review articles, whereas foreground questions are answered using EBM techniques. There are four major foreground question categories:

- Therapy (or intervention)—benefit of treatment or prevention.
- Diagnosis—test diagnosing disease.
- Harm—detrimental health effects of a disease, environmental exposure (natural or man-made), or medical intervention.
- Prognosis—outcome of disease course.

Identifying evidence involves selecting the best evidence for a given type of question. EBM proponents advocate, for example, that randomized controlled trials or a **systematic review** (with or without **meta-analysis**) that

combines multiple trials provide the best evidence for or against particular health care interventions. Likewise, diagnostic test accuracy is best assessed with comparison to a known gold standard in an appropriate spectrum of patients to whom the test will be applied (see ▶ Chap. 3). Questions of harm can be answered by randomized controlled trials when it is ethical to do so; otherwise they are best answered with observational case control or cohort studies. There are checklists of attributes for these different types of studies that allow their critical appraisal and applicability to a given patient in the EBM resources described above.

The original approach to EBM has evolved over time, with less emphasis on critically appraising original evidence and more on synthesized evidence being made readily available to clinicians, usually through electronic sources, including clinical decision support systems (see ▶ Chap. 26) (DiCenso et al. 2009; Hersh 1999). There have also been a number of criticisms of EBM, with arguments that it may ignore clinical expertise (Haynes et al. 2002), patient values (Guyatt et al. 2004), and other ways of "knowing" (Sim 2016). Others express concerns that its methods have been "distorted" (Greenhalgh et al. 2014), "hijacked" by commercial forces and other self-interest (Ioannidis 2016a, 2017), and subverted by the political process (Patashnik et al. 2017). A final criticism is the proliferation of systematic reviews, not all of which may be motivated by objective science (Ioannidis 2016b).

23.4 Content of Knowledge-Based Information Resources

The previous sections of this chapter have described some of the issues and concerns surrounding the production and use of knowledge-based information in biomedicine. It is useful to classify the information to gain a better understanding of its structure and function. In this section, we classify content into bibliographic, full-text, annotated, and aggregated categories, although some content does not neatly fit within them.

23.4.1 Bibliographic Content

The first category consists of **bibliographic content**. It includes what was for decades the mainstay of IR systems: **literature reference databases**. Also called **bibliographic databases**, this content consists of citations or pointers to the medical literature (i.e., journal articles). The best-known and most widely used biomedical bibliographic database is MEDLINE, which contains bibliographic references to all of the biomedical articles, editorials, and letters to the editors in approximately 5000 scientific journals. The journals are chosen for inclusion by an advisory committee of subject experts convened by NIH. At present, over 800,000 references are added to MEDLINE yearly. It contained over 24 million references by the end of 2017.[10]

The current MEDLINE record contains over 60 fields.[11] A clinician may be interested in just a handful of these fields, such as the title, abstract, and indexing terms. But other fields contain specific information that may be of great importance to other audiences. The Supplementary Information (SI) field contains links to records in many different data banks, from clinical trials registries to genomics and other "omics" databases.[12] Even the clinician may, however, derive benefit from some of the other fields. For example, the Publication Type (PT) field can help in the application of EBM, such as when one is searching for a practice guideline or a randomized controlled trial.

MEDLINE records are also assigned other identifiers. The PubMed ID (PMID) is a unique identifier for records in the database. Another identifier in MEDLINE is the PubMed Central ID (PMCID), assigned to records whose articles have been deposited into PMC. MEDLINE also contains an Author Identifier (AUID) field, which allows three possible identifiers, the most common of

10 ▶ https://www.nlm.nih.gov/bsd/index_stats_comp.html

11 ▶ https://www.nlm.nih.gov/bsd/mms/medlineelements.html

12 ▶ https://www.nlm.nih.gov/bsd/medline_databank_source.html

which is the ORCID identifier,[13] a unique identifier for scientific authors (e.g., 0000-0002-4114-5148 for this author). A growing number of journals and other publications make use of the ORCID.

MEDLINE can be accessed without charge via the **PubMed** system,[14] produced by the **National Center for Biotechnology Information (NCBI)**[15] of the NLM. Some other information vendors, such as Ovid Technologies,[16] license the content of MEDLINE and other databases and provide value-added services that can be accessed for a fee by individuals and institutions.

The NLM used to offer a number of other, more focused bibliographic databases, but these have all been folded into MEDLINE. A number of other publishers offer biomedical bibliographic databases. The major non-NLM database for the nursing field is the **Cumulative Index to Nursing and Allied Health Literature (CINAHL)**,[17] which covers nursing and allied health literature, including physical therapy, occupational therapy, laboratory technology, health education, physician assistants, and medical records. Another well-known bibliographic database is **EMBASE**,[18] which is produced by the commercial publisher, Elsevier (Amsterdam, Netherlands). EMBASE contains over 28 million records and covers a superset of journals from MEDLINE. These journals are often important for those carrying out systematic reviews and meta-analyses, which need access to all the studies published across the world.

A second, more modern type of bibliographic content is the **Web catalog**. There are increasing numbers of such catalogs, which consist of Web pages containing mainly links to other Web pages and sites. It should be noted that there is a blurry distinction between Web catalogs and aggregations (the fourth category; see ▶ Sect. 23.4.4, below). In general, the former contains only links to other pages and sites, while the latter include actual content that is highly integrated with other resources. Some well-known Web catalogs include:

- HON Select[19]—a European catalog of quality-filtered, clinician-oriented Web content from the HON foundation.
- Translating Research into Practice (TRIP)[20]—a database of content deemed to meet high standards of EBM.

A specialized registry specific to healthcare was the *National Guidelines Clearinghouse* (NGC). Produced by the Agency for Healthcare Research and Quality (AHRQ), it contained exhaustive information about clinical practice guidelines. In 2018, AHRQ shut down the National Guidelines Clearinghouse (Munn and Qaseem 2018). The original contractor developing the NGC was the nonprofit research firm, ECRI, which developed a new site, ECRI Guidelines Trust.[21]

A final kind of bibliographic-like content consists of **RSS feeds** (originally RDF Site Summary, often dubbed "Really Simple Syndication"), which are short summaries of Web content, typically news, journal articles, blog postings, and other content. Users set up an RSS aggregator, which can be though a Web browser, email client, or standalone software, configured for the RSS feed desired, with an option to add a filter for specific content. There are two versions of RSS (1.0 and 2.0) but both provide:

- Title—name of item
- Link—URL to content
- Description—a brief description of the content

23.4.2 Full-text Content

The second type of content is **full-text content**. A large component of this content origi-

13 ▶ https://orcid.org/
14 ▶ https://pubmed.gov
15 ▶ https://www.ncbi.nlm.nih.gov/
16 ▶ http://ovid.com/site/index.jsp
17 ▶ https://health.ebsco.com/products/the-cinahl-database
18 ▶ https://www.elsevier.com/solutions/embase-biomedical-research
19 ▶ https://www.hon.ch/HONselect
20 ▶ https://www.tripdatabase.com
21 ▶ https://guidelines.ecri.org/

nally consisted of the online versions of books and periodicals. As already noted, just about all traditionally paper-based medical content, from journals to textbooks, is now available electronically. The electronic versions usually have identical content to paper versions but may be enhanced by measures ranging from the provision of supplemental data in a journal article to linkages and multimedia content in a textbook. The final component of this category is the Web site. Admittedly, the diversity of information on Web sites is enormous, and sites may include every other type of content described in this chapter. However, in the context of this category, "Web site" refers to the vast number of static and dynamic Web pages at a discrete Web location.

One of the fields in MEDLINE is the uniform resource locator (URL) for the publisher's full text of the article, allowing linkage directly from the bibliographic database to the full text. This link is active when the PubMed record is displayed, but users may be met by a password screen if the article is not available for free. Many sites allow both access to subscribers or a pay-per-view facility. Many academic organizations now maintain large numbers of subscriptions to journals available to faculty, staff, and students. Other publishers, such as Ovid, provide access within their own password-protected interfaces to articles from journals that they have licensed for use in their systems.

The most common secondary literature source is textbooks, almost all of which are available in electronic form. A common approach with textbooks is bundling them, sometimes with linkages across the bundled texts. An early bundler of textbooks was Stat!-Ref[22] that, like many, began as a CD-ROM product and then moved to the Web. Most other large publishers have now similarly aggregated their libraries of textbooks and other content. Another collection of textbooks is the NCBI Bookshelf,[23] which contains many volumes on biomedical research topics. Initially published by NCBI but now a standalone reference is Online Mendelian Inheritance in Man (OMIM),[24] which is continually updated with new information about the genomic causes of human disease.

Electronic textbooks offer additional features beyond text from the print version. While many print textbooks do feature high-quality images, electronic versions offer the ability to have more pictures and illustrations. They also have the ability to use sound and video, although few do at this time. As with full-text journals, electronic textbooks can link to other resources, including journal references and the full articles. Many Web-based textbook sites also provide access to continuing education self-assessment questions and medical news. Finally, electronic textbooks let authors and publishers provide more frequent updates of the information than is allowed by the usual cycle of print editions, where new versions come out only every few years.

As noted above, Web sites are another form of full-text information. Probably the most effective provider of Web-based health information is the U.S. government. Not only do they produce bibliographic databases, but the NLM, AHRQ, the National Cancer Institute (NCI), Centers for Disease Control (CDC), and others have also been innovative in providing comprehensive full-text information for health care providers and consumers. One example is the popular CDC Travel site.[25] Some of these will be described later as aggregations, since they provide many different types of resources.

A large number of commercial biomedical and health Web sites have emerged in recent years. On the consumer side, they include more than just collections of text; they also include interaction with experts, online stores, and catalogs of links to other sites. Some well-known examples include Mayo Clinic[26] and WebMD.[27] There are also Web sites, either from medical professional societies or companies, which provide information geared toward

22 ► http://statref.com/
23 ► https://www.ncbi.nlm.nih.gov/books
24 ► http://www.omim.org/
25 ► https://wwwnc.cdc.gov/travel
26 ► https://www.mayoclinic.org
27 ► https://www.webmd.com

health care providers, typically overviews of diseases, their diagnosis, and treatment; medical news and other resources for providers are often offered as well.

Other sources of on-line health-related content include encyclopedias, the so-called **body of knowledge** (BOK; the complete set of concepts, terms and activities that make up a professional domain), and **Weblogs** or **blogs**. A well-known online encyclopedia with a great deal of health-related information is Wikipedia,[28] which features a distributed authorship process whose content has been found to reliable (Giles 2005; Nicholson 2006) and frequently shows up near the top in health-related Web searches (Laurent and Vickers 2009). A growing number of organizations have a body of knowledge, such as the *American Health Information Management Association* (AHIMA)[29] Blogs tend to carry a stream of consciousness but often high-quality information is posted within them.

23.4.3 Annotated Content

The third category consists of **annotated content**. These resources are usually not stored as freestanding Web pages but instead are often housed in database management systems. This content can be further subcategorized into discrete information types:

- Image databases—collections of images from radiology, pathology, and other areas
- Genomics databases—information from gene sequencing, protein characterization, and other genomic research
- Citation databases—bibliographic linkages of scientific literature
- EBM databases—highly structured collections of clinical evidence
- Other databases—miscellaneous other collections

A great number of biomedical image databases are available on the Web. Some examples from the NLM include:

- Visible Human Project[30] – collection of three-dimensional representations of normal male and female bodies, consisting of cross-sectional slices of cadavers, with sections of 1 mm thickness in the male and 0.3 mm thickness in the female (Spitzer et al. 1996). Also available from each cadaver are transverse computerized tomography and magnetic resonance images.
- Images from the History of Medicine[31] – online access to images from the historical collections of the NLM.
- Open-I[32] – collection of images from PubMed Central papers (Demner-Fushman et al. 2012).

Many genomics databases are available on the Web. The first issue each year of the journal *Nucleic Acids Research* (NAR) catalogs and describes these databases, and is now available by open access means (Rigden and Fernández 2020). NAR also maintains an ongoing database of such databases, the Molecular Biology Database Collection.[33] Among the most important of these databases are those available from NCBI (Anonymous 2018a). All their databases are linked among themselves, along with PubMed and OMIM, and are searchable via the **NCBI Search** system.[34] More details on the specific content of genomics databases is provided in ▶ Chap. 28.

Citation databases provide linkages to articles that cite others across the scientific literature. The earliest citation databases were the Science Citation Index (SCI) and Social Science Citation Index (SSCI), which are now part of the larger Web of Science (Clarivate Analytics, Philadelphia, PA). Two well-known bibliographic databases for biomedical and health topics that also have citation links include SCOPUS35 and Google Scholar.[36] A

28 ▶ https://en.wikipedia.org/wiki/Main_Page
29 ▶ http://bok.ahima.org/
30 ▶ https://www.nlm.nih.gov/research/visible/visible_human.html
31 ▶ https://www.nlm.nih.gov/hmd/ihm/index.html
32 ▶ https://openi.nlm.nih.gov/
33 ▶ http://www.oxfordjournals.org/nar/database/a/
34 ▶ https://www.ncbi.nlm.nih.gov/search/
35 ▶ https://www.scopus.com/
36 ▶ https://scholar.google.com/

final citation database of note is CiteSeer,[37] which focuses on computer and information science, including biomedical informatics.

EBM databases are devoted to providing annotated evidence-based information. Some examples (all available with through subscription fees) include:

- *Cochrane Database of Systematic Reviews*—one of the original collections of systematic reviews[38]
- *BMJ Best Practice*[39]
- *JAMA Evidence*[40]
- *Up-to-Date*—content centered around clinical questions[41]
- *Essential Evidence Plus*[42]

There is a growing market for a related type of evidence-based content in the form of clinical decision support order sets, rules, and health/disease management templates. Publishers include EHR vendors whose systems employ this content as well as other vendors such as Zynx[43] and Provation.[44]

There are a variety of other annotated content. The *ClinicalTrials.gov* database[45] began as a database of clinical trials sponsored by NIH. After concerns about clinical trials having their protocols altered after the start of trials, ClinicalTrials.gov expanded its scope to be a registry of all clinical trials (DeAngelis et al. 2005; Zarin et al. 2017) and to contain actual results of trials (Zarin et al. 2011). Another important database for researchers is *NIH RePORTER,*[46] which is a database of all grant awards funded by NIH. An additional annotated resource is *DataMed,*[47] which aims to catalog and provide linkage for use of data sets from biomedical research (Ohno-Machado et al. 2017).

23.4.4 Aggregated Content

The final category consists of **aggregations** of content from the first three categories. The distinction between this category and some of the highly-linked types of content described above is admittedly blurry, but aggregations typically have a wide variety of different types of information serving the diverse needs of users. Aggregated content has been developed for all types of users from consumers to clinicians to scientists.

Probably the largest aggregated consumer information resource is **MedlinePlus**[48] from the NLM. MedlinePlus includes all of the types of content previously described, aggregated for easy access to a given topic. MedlinePlus contains health topics, drug information, medical dictionaries, directories, and other resources. Each topic contains links to health information from the NIH and other sources deemed credible by its selectors. There are also links to current health news (updated daily), a medical encyclopedia, drug references, and directories, along with a preformed PubMed search related to the topic.

Another well-known group of aggregations of content for genomics researchers is the **model organism databases**. These databases bring together bibliographic databases, full text, and databases of sequences, structure, and function for organisms whose genomic data have been highly characterized. One of the oldest and most developed model organism databases is the Mouse Genome Informatics resource.[49] More details are provided in ▶ Chap. 28.

23.5 Indexing

As noted at the beginning of the chapter, indexing is the process of assigning metadata

37 ▶ http://citeseerx.ist.psu.edu
38 ▶ https://www.cochranelibrary.com
39 ▶ https://bestpractice.bmj.com/info/evidence-information/
40 ▶ https://jamaevidence.mhmedical.com
41 ▶ https://www.uptodate.com/home
42 ▶ http://www.essentialevidenceplus.com
43 ▶ https://www.zynxhealth.com/
44 ▶ https://www.provationmedical.com/order-set-management
45 ▶ https://clinicaltrials.gov/
46 ▶ http://projectreporter.nih.gov/reporter.cfm
47 ▶ https://datamed.org
48 ▶ https://medlineplus.gov/
49 ▶ http://www.informatics.jax.org/

to content to facilitate its retrieval. Most modern commercial content is indexed in two ways:

1. **Manual indexing**—where human indexers, usually using a controlled terminology, assign indexing terms and attributes to documents, often following a specific protocol.
2. **Automated indexing**—where computers make the indexing assignments, usually limited to breaking out each word in the document (or part of the document) as an indexing term.

Manual indexing is done most commonly with bibliographic databases and annotated content. In this age of proliferating electronic content, such as online textbooks, practice guidelines, and multimedia collections, manual indexing has become either too expensive or outright unfeasible for the quantity and diversity of material now available. Thus there are increasing numbers of databases that are indexed only by automated means. Before covering these types of indexing in detail, let us first discuss controlled terminologies.

23.5.1 Controlled Terminologies

A **controlled terminology** contains a set of terms that can be applied to a task, such as indexing. When the terminology defines the terms, it is usually called a **vocabulary**. When it contains variants or synonyms of terms, it is also called a **thesaurus**. Before discussing actual terminologies, it is useful to define some terms. A **concept** is an idea or object that exists in the world, such as the condition under which human blood pressure is elevated. A **term** is the actual string of one or more words that represent a concept, such as "Hypertension" or "High Blood Pressure". One of these string forms is the preferred or **canonical form**, such as "Hypertension" in the present example. When one or more terms can represent a concept, the different terms are called **synonyms**.

A controlled terminology usually contains a list of terms that are the canonical representations of the concepts. If it is a thesaurus, it contains relationships between terms, which typically fall into three categories:

- Hierarchical—terms that are broader or narrower. The hierarchical organization not only provides an overview of the structure of a thesaurus but also can be used to enhance searching (e.g., MeSH tree explosions that add terms from an entire portion of the hierarchy to augment a search).
- Synonym—terms that are synonyms, allowing the indexer or searcher to express a concept in different words.
- Related—terms that are not synonymous or hierarchical but are somehow otherwise related. These usually remind the searcher of different but related terms that may enhance a search.

The MeSH terminology is used to manually index most of the databases produced by the NLM (Coletti and Bleich 2001). The latest version contains over 28,000 subject headings (the word MeSH uses for the canonical representation of its concepts). It also contains over 90,000 synonyms to those terms, which in MeSH jargon are called **entry terms**. MeSH also contains Supplementary Concept Records, representing 230,000 additional chemicals, drugs, genes, organisms, etc. that indexers encounter in indexing process and map to MeSH headings.

MeSH contains the three types of relationships described previously:

- Hierarchical—MeSH is organized hierarchically into 16 trees, such as Diseases, Organisms, and Chemicals and Drugs
- Synonym—MeSH contains a vast number of entry terms, which are synonyms of the headings
- Related—terms that may be useful for searchers to add to their searches when appropriate are suggested for many headings

The MeSH terminology files, their associated data, and their supporting documentation are available on the NLM's MeSH Web site.[50] There is also a browser that facilitates explo-

50 ► http://www.nlm.nih.gov/mesh/

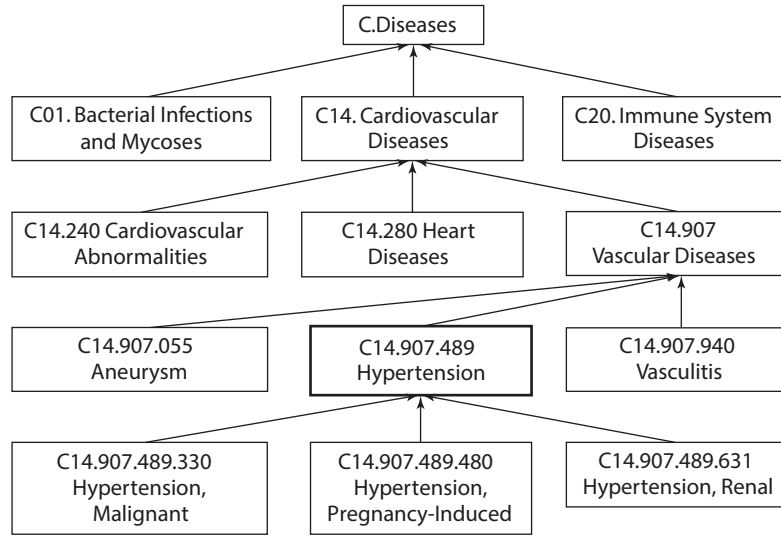


Fig. 23.2 A slice through the Medical Subject Headings (MeSH) hierarchy for "Hypertension" and related terms, showing the location of the term in the C. Diseases. The arrows show links to broader terms in the hierarchy, while the codes give the tree address used internally by the MeSH system. (Courtesy of National Library of Medicine, with permission)

ration of the terminology[51] as well as a tool for mapping text to MeSH terms, called MeSH on Demand.[52] Figure 23.2 shows a slice through the MeSH hierarchy for "Hypertension" and related cardiovascular diseases in the C. Diseases tree.

There are features of MeSH designed to assist indexers in making documents more retrievable. One of these is **subheadings**, which are qualifiers of subject headings that narrow the focus of a term. In Hypertension, for example, the focus of an article may be on the diagnosis, epidemiology, or treatment of the condition. Another feature of MeSH that helps retrieval is **check tags**. These are MeSH terms that represent certain facets of medical studies, such as age, gender, human or nonhuman, and type of grant support. Related to check tags are the geographical locations in one particular part of the MeSH hierarchy (called the "Z tree", because their term codes start with "Z"). Indexers must also include these, like check tags, since the location of a study (e.g., Oregon) must be indicated. Another feature gaining increasing importance for EBM and other purposes is the **publication type**, which describes the type of publication or the type of study. A searcher who wants a review of a topic may choose the publication type Review or Review Literature. Or, to find studies that provide the best evidence for a therapy, the publication type Meta-Analysis, Randomized Controlled Trial, or Controlled Clinical Trial would be used.

MeSH is not the only thesaurus used for indexing biomedical documents. A number of other thesauri are used to index non-NLM databases. CINAHL, for example, uses the **CINAHL Subject Headings**, which are based on MeSH but have additional domain-specific terms added. EMBASE has a terminology called **EMTREE**,[53] which has many features similar to those of MeSH.

One problem with controlled terminologies, not limited to IR systems, is their proliferation. As already described in ▶ Chap. 8, there is great need for linkage across these different terminologies. This was the primary motivation for the **Unified Medical Language System (UMLS)** Project,[54] which was undertaken in the 1980s to address this problem (Humphreys et al. 1998). There are three components of the UMLS Knowledge Sources: the **Metathesaurus**, the **Semantic Network**, and the **Specialist Lexicon**. The Metathesaurus component of the UMLS links parts or all of over 100 terminologies (Bodenreider 2004).

51 ▶ http://www.nlm.nih.gov/mesh/MBrowser.html
52 ▶ https://meshb.nlm.nih.gov/MeSHonDemand
53 ▶ https://www.elsevier.com/solutions/embase-biomedical-research
54 ▶ https://www.nlm.nih.gov/research/umls/

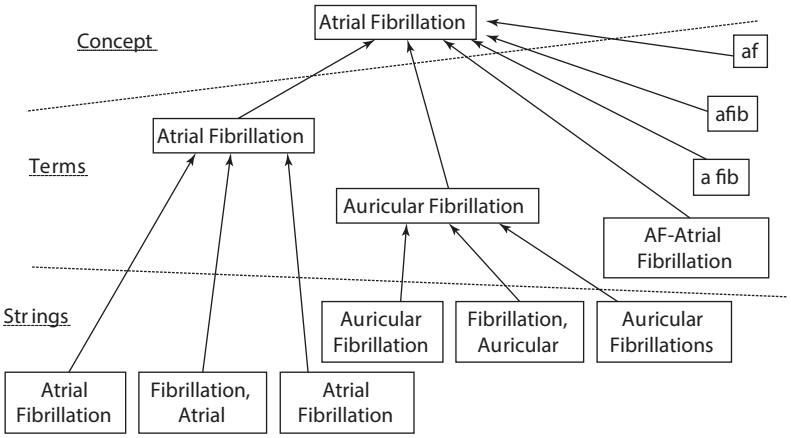


Fig. 23.3 Concepts, terms, and strings for the Metathesaurus concept atrial fibrillation. Each string may occur in more than one vocabulary, in which case each would be an atom. (Courtesy of National Library of Medicine, with permission)

In the Metathesaurus, all terms that are conceptually the same are linked together as a concept. Each concept may have one or more terms, each of which represents an expression of the concept from a source terminology that is not just a simple lexical variant (i.e., differs only in word ending or order). Each term may consist of one or more strings that represent all the lexical variants that are represented for that term in the source terminologies. One of each term's strings is designated as the preferred form, and the preferred string of the preferred term is known as the canonical form of the concept. There are rules of precedence for determining the canonical form, the main one being that the MeSH heading is used if one of the source terminologies for the concept is MeSH.

Each Metathesaurus concept has a single concept unique identifier (CUI). Each term has one term unique identifier (LUI), all of which are linked to the one (or more) CUIs with which they are associated. Likewise, each string has one string unique identifier (SUI), which is likewise linked to the LUIs in which they occur. In addition, each string has an atomic unique identifier (AUI) that represents information from each instance of the string in each vocabulary. Figure 23.3 depicts the English-language concepts, terms, and strings for the Metathesaurus concept atrial fibrillation. (Each string may occur in more than one vocabulary, in which case each would be an atom.) The canonical form of the concept and one of its terms is atrial fibrillation. Within both terms are several strings that vary in word order and case.

The Metathesaurus contains a wealth of additional information. In addition to the synonym relationships between concepts, terms, and strings described earlier, there are also non-synonym relationships between concepts. There are a great many attributes for the concepts, terms, strings, and atoms, such as definitions, lexical types, and occurrence in various data sources. Also provided with the Metathesaurus is a word index that connects each word to all the strings it occurs in, along with its concept, term, string, and atomic identifiers.

23.5.2 Manual Indexing

Manual indexing is most commonly done for bibliographic and annotated content, although it is sometimes for other types of content as well. Manual indexing is usually done by means of a controlled terminology of terms and attributes. Most databases utilizing human indexing usually have a detailed protocol for assignment of indexing terms from the thesaurus. The MEDLINE database is no exception. The principles of MEDLINE indexing were laid out in the two-volume MEDLARS Indexing Manual (Charen 1976, 1983). Subsequent modifications have occurred with changes to MEDLINE, other databases, and MeSH over the years. The major concepts of the article, usually from

two to five headings, are designed as main headings, and designated in the MEDLINE record by an asterisk. The indexer is also required to assign appropriate subheadings. Finally, the indexer must also assign check tags, geographical locations, and publication types. Although MEDLINE indexing is still manual, indexers are aided by a variety of electronic tools for selecting and assigning MeSH terms.

Few full-text resources are manually indexed. One type of indexing that commonly takes place with full-text resources, especially in the print world, is that performed for the index at the back of the book. However, this information is rarely used in IR systems; instead, most online textbooks rely on automated indexing (see ▶ Sect. 23.5.3, below).

Manual indexing of Web content would likewise be challenging. With billions of pages of content, manual indexing of more than a fraction of it is not feasible. On the other hand, the lack of a coherent index makes searching much more difficult, especially when specific resource types are being sought. A simple form of manual indexing of the Web takes place in the development of the Web catalogs and aggregations as described earlier. These catalogs contain not only explicit indexing about subjects and other attributes, but also implicit indexing about the quality of a given resource by the decision of whether to include it in the catalog.

Two major approaches to manual indexing have emerged on the Web that are often complementary. The first approach, that of applying metadata to Web pages and sites, is exemplified by the **Dublin Core Metadata Initiative (DCMI)**[55] (Weibel and Koch 2000). The goal of the DCMI has been to develop a set of standard data elements that creators of Web resources can use to apply metadata to their content. The specification has defined 15 elements, as shown in ◘ Table 23.1. The DCMI was recently approved as a standard by the National Information Standards Organization (NISO) with the designation Z39.85. It is also a standard with the International Organization for Standards (ISO), ISO Standard 15836:2009.

There have been some medical adaptations of the DCMI. The most developed of these is the Catalogue et Index des Sites Médicaux Francophones (CISMeF)[56] (Darmoni et al. 2000). A catalog of French-language health resources on the Web, CISMeF has used DCMI to catalog Web pages, including information resources (e.g., practice guidelines, consensus development conferences), organizations (e.g., hospitals, medical schools, pharmaceutical companies), and databases. The Subject field uses the French translation of MeSH but also includes the English translation. For Type, a list of common Web resources has been enumerated.

While Dublin Core Metadata was originally envisioned to be included in **Hypertext Markup Language (HTML)** Web pages, it became apparent that many non-HTML resources exist on the Web and that there are reasons to store metadata external to Web pages. For example, authors of Web pages might not be the best people to index pages or other entities might wish to add value by their own indexing of content. An emerging standard for cataloging metadata is the **Resource Description Framework (RDF)** (Akerkar 2009).

A second approach to manually indexing content on the Web has been to create directories of content. The first major effort to create these was for use in the Yahoo! search engine,[57] which created a subject hierarchy and assigned Web sites to elements within it. When concern began to emerge that the Yahoo directory was proprietary and not necessarily representative of the Web community at large, an alternative movement sprung up: the Open Directory Project (dmoz.org). Due to increasing growth of the Web, these projects were eventually disbanded.

Manual indexing has a number of limitations, the most significant of which is inconsistency. Funk and Reid (Funk and Reid 1983) evaluated indexing inconsistency in

55 ▶ http://dublincore.org/

56 ▶ http://www.chu-rouen.fr/cismef/

57 ▶ https://www.yahoo.com/

Table 23.1 Elements of Dublin Core Metadata

Element	Definition
DC.title	The name given to the resource
DC.creator	The person or organization primarily responsible for creating the intellectual content of the resource
DC.subject	The topic of the resource
DC.description	A textual description of the content of the resource
DC.publisher	The entity responsible for making the resource available in its present form
DC.date	A date associated with the creation or availability of the resource
DC.contributor	A person or organization not specified in a creator element who has made a significant intellectual contribution to the resource but whose contribution is secondary to any person or organization specified in a creator element
DC.type	The category of the resource
DC.format	The data format of the resource, used to identify the software and possibly hardware that might be needed to display or operate the resource
DC.identifier	A string or number used to uniquely identify the resource
DC.source	Information about a second resource from which the present resource is derived
DC.language	The language of the intellectual content of the resource
DC.relation	An identifier of a second resource and its relationship to the present resource
DC.coverage	The spatial or temporal characteristics of the intellectual content of the resource
DC.rights	A rights management statement, an identifier that links to a rights management statement, or an identifier that links to a service providing information about rights management for the resource

MEDLINE by identifying 760 articles that had been indexed twice by the NLM. The most consistent indexing occurred with check tags and central concept headings, which were only indexed with a consistency of 61–75%. The least consistent indexing occurred with subheadings, especially those assigned to non-central-concept headings, which had a consistency of less than 35%. A repeat of this study in more recent times found comparable results (Marcetich et al. 2004). Manual indexing also takes time. While it may be feasible with the large resources the NLM has to index MEDLINE, it is probably impossible with the growing amount of content on Web sites and in other full-text resources. Indeed, the NLM has recognized the challenge of continuing to have to index the growing body of biomedical literature and is investigating automated and semiautomated means of doing so (Mork et al. 2017).

23.5.3 Automated Indexing

In automated indexing, the indexing is done by a computer. Although the mechanical running of the automated indexing process lacks cognitive input, considerable intellectual effort may have gone into development of the system for doing it, so this form of indexing still qualifies as an intellectual process. In this section, we will focus on the automated indexing used in operational IR systems, namely the indexing of documents by the words they contain.

Some might not think of extracting all the words in a document as "indexing," but from

the standpoint of an IR system, words are descriptors of documents, just like human-assigned indexing terms. Most retrieval systems actually use a hybrid of human and word indexing, in that the human-assigned indexing terms become part of the document, which can then be searched by using the whole controlled term or individual words within it. With the development of full-text resources in the 1980s and 1990s, systems that allowed only word indexing began to emerge. This trend increased with the advent of the Web.

Word indexing is typically done by defining all consecutive alphanumeric sequences between white space (which consists of spaces, punctuation, carriage returns, and other non-alphanumeric characters) as words. Systems must take particular care to apply the same process to documents and the user's query, especially with characters such as hyphens and apostrophes. Many systems go beyond simple identification of words and attempt to assign weights to words that represent their importance in the document (Salton 1991).

Many systems using word indexing employ processes to remove common words or conflate words to common forms. The former consists of filtering to remove **stop words**, which are common words that always occur with high frequency and are usually of little value in searching. The stop word list, also called a **negative dictionary**, varies in size from the seven words of the original MEDLARS stop list (and, an, by, from, of, the, with) to the list of 250–500 words more typically used. Examples of the latter are the 250-word list of van Rijsbergen (1979), the 471-word list of Fox (1992), and the PubMed stop list.[58] Conflation of words to common forms is done via **stemming**, the purpose of which is to ensure words with plurals and common suffixes (e.g., -ed, -ing, -er, -al) are always indexed by their stem form (Frakes 1992). For example, the words cough, coughs, and coughing are all indexed via their stem cough. Both stop word remove and stemming reduce the size of indexing files and lead to more efficient query processing.

A commonly used approach for term weighting is **TF*IDF weighting**, which combines the **inverse document frequency (IDF)** and **term frequency (TF)**. The IDF is the logarithm of the ratio of the total number of documents to the number of documents in which the term occurs. It is assigned once for each term in the database, and it correlates inversely with the frequency of the term in the entire database. The usual formula used is:

$$IDF\left(term\right) = \log \frac{\text{number of documents in database}}{\text{number of documents with term}} + 1 \tag{23.1}$$

The TF is a measure of the frequency with which a term occurs in a given document and is assigned to each term in each document, with the usual formula:

$$TF\left(term,document\right) = \text{frequency of term in document} \tag{23.2}$$

In TF*IDF weighting, the two terms are combined to form the indexing weight, WEIGHT:

$$WEIGHT\left(term,document\right) = TF\left(term,document\right) * IDF\left(term\right) \tag{23.3}$$

Another automated indexing approach generating increased interest is the use of **link-based** methods, fueled by the success of the Google search engine.[59] This approach gives weight to pages based on how often they are cited by other pages. The **PageRank (PR) algorithm** is mathematically complex, but can be viewed as giving more weight to a Web page based on the number of other pages that link to it (Brin and Page 1998). Thus, the home page of the NLM or a major medical journal is likely to have a very high PR, whereas a more obscure page will have a lower PR.

General-purpose search engines such as Google and Microsoft Bing use word-based approaches and variants of the PageRank algorithm for indexing. They amass the content in their search systems by "crawling" the Web, collecting and indexing every object they

58 ► https://www.ncbi.nlm.nih.gov/books/NBK3827/table/pubmedhelp.T.stopwords/

59 ► https://www.google.com/

find on the Web. This includes not only HTML pages, but other files as well, including Microsoft Word, Portable Document Format (PDF), and images.

Word indexing has a number of limitations, including:

- Synonymy—different words may have the same meaning, such as high and elevated. This problem may extend to the level of phrases with no words in common, such as the synonyms hypertension and high blood pressure.
- Polysemy—the same word may have different meanings or senses. For example, the word lead can refer to an element or to a part of an electrocardiogram machine.
- Content—words in a document may not reflect its focus. For example, an article describing hypertension may make mention in passing to other concepts, such as congestive heart failure (CHF) that are not the focus of the article.
- Context—words take on meaning based on other words around them. For example, the relatively common words high, blood, and pressure, take on added meaning when occurring together in the phrase high blood pressure.
- Morphology—words can have suffixes that do not change the underlying meaning, such as indicators of plurals, various participles, adjectival forms of nouns, and nominalized forms of adjectives.
- Granularity—queries and documents may describe concepts at different levels of a hierarchy. For example, a user might query for antibiotics in the treatment of a specific infection, but the documents might describe specific antibiotics themselves, such as penicillin.

► Chapter 9 on Natural Language Processing (NLP) describes automated methods for addressing these limitations.

23.6 Retrieval

There are two broad approaches to retrieval. Exact-match searching allows the user precise control over the items retrieved. Partial-match searching, on the other hand, recognizes the inexact nature of both indexing and retrieval, and instead attempts to return to the user content ranked by how close it comes to the user's query. After general explanations of these approaches, we will describe actual systems that access the different types of biomedical content.

23.6.1 Exact-Match Retrieval

In exact-match searching, the IR system gives the user all documents that exactly match the criteria specified in the search statement(s). Since the **Boolean operators** AND, OR, and NOT are usually required to create a manageable set of documents, this type of searching is often called Boolean searching. Furthermore, since the user typically builds sets of documents that are manipulated with the Boolean operators, this approach is also called **set-based searching**. Most of the early operational IR systems in the 1950s through 1970s used the exact-match approach, even though Salton was developing the partial-match approach in research systems during that time (Salton and McGill 1983). Currently, exact-match searching tends to be associated with retrieval from bibliographic and annotated databases, while the partial-match approach tends to be used with full-text searching.

Typically the first step in exact-match retrieval is to select terms to build sets. Other attributes, such as the author name, publication type, or gene identifier (in the secondary source identifier field of MEDLINE), may be selected to build sets as well. Once the search term(s) and attribute(s) have been selected, they are combined with the Boolean operators. The Boolean AND operator is typically used to narrow a retrieval set to contain only documents with two or more concepts. The Boolean OR operator is usually used when there is more than one way to express a concept. The Boolean NOT operator is often employed as a subtraction operator that is applied to a pair of sets, with the result being the documents found in the first set but not in the second set. Some systems more accurately call this the ANDNOT operator.

Some systems allow terms in searches to be expanded by using the wild-card character, which adds all words to the search that begin with the letters up until the wild-card character. This approach is also called truncation. Unfortunately, there is no standard approach to using wild-card characters, so syntax for them varies from system to system. PubMed, for example, allows a single asterisk at the end of a word to signify a wild-card character. Thus the query word *can** will lead to the words *cancer* and *Candida*, among others, being added to the search.

23.6.2 Partial-Match Retrieval

Although **partial-match searching** was conceptualized very early, it did not see widespread use in IR systems until the advent of Web search engines in the 1990s. This is most likely because exact-match searching tends to be preferred by "power users" whereas partial-match searching is preferred by novice searchers. Whereas exact-match searching requires an understanding of Boolean operators and (often) the underlying structure of databases (e.g., the many fields in MEDLINE), partial-match searching allows a user to simply enter a few terms and start retrieving documents.

The development of partial-match searching is usually attributed to Salton, who pioneered the approach in the 1960s (Salton and McGill 1983). Although partial-match searching does not exclude the use of non-term attributes of documents, and for that matter does not even exclude the use of Boolean operators (e.g., Salton et al. 1983), the most common use of this type of searching is with a query of a small number of words, also known as a **natural language query**. Because Salton's approach was based on vector mathematics, it is also referred to as the **vector-space model** of IR. In the partial-match approach, documents are typically ranked by their closeness of fit to the query. That is, documents containing more query terms will likely be ranked higher, since those with more query terms will in general be more likely to be relevant to the user. As a result this process is called **relevance ranking**. The entire approach has also been called **lexical–statistical retrieval**.

The most common approach to document ranking in partial-match searching is to give each a score based on the sum of the weights of terms common to the document and query. Terms in documents typically derive their weight from the TF*IDF calculation described above. Terms in queries are typically given a weight of one if the term is present and zero if it is absent. The following formula can then be used to calculate the document weight across all query terms:

$$Document\ weight = \sum_{all\ query terms} Weight\ of\ term\ in\ query * Weight\ of\ term\ in document \quad (23.4)$$

This may be thought of as a giant OR of all query terms, with sorting of the matching documents by weight. The usual approach is for the system to then perform the same stop word removal and stemming of the query that was done in the indexing process. (The equivalent stemming operations must be performed on documents and queries so that complementary word stems will match).

A number of other ranking algorithms have been developed over the years. The BM25 approach has been found to be effective for many diverse test collections (Robertson and Walker 1994). Another approach showing efficacy in research has been the use of language modeling techniques (Zhai and Lafferty 2004). More recently, the learning-to-rank approach has found value for machine learning approaches in the relevance ranking process (Li 2011). The large search engines such as Google make use of other information in their results, such as aiming to provide answers to queries that are likely to be questions (e.g., defining words or

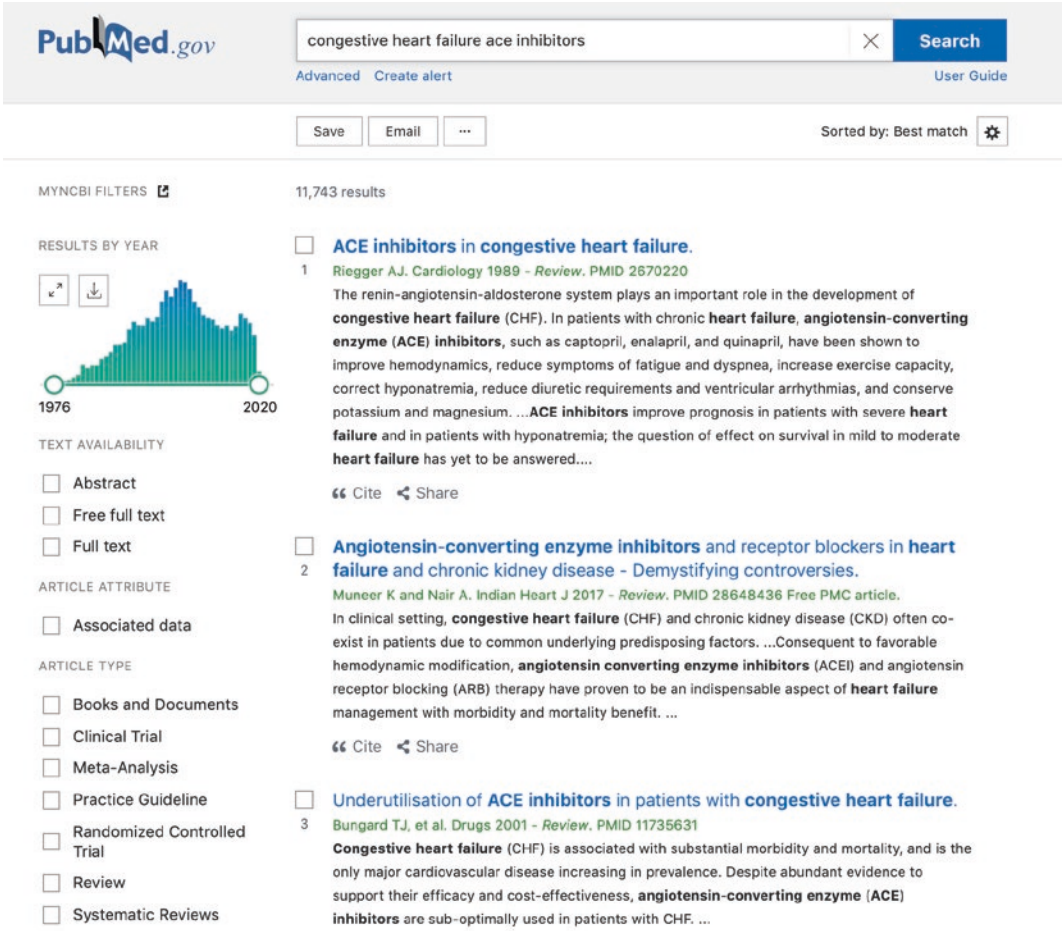


Fig. 23.4 Search results from PubMed, showing query, results, and ability to set limits (on the left side of the window). (Courtesy of National Library of Medicine, with permission)

listing airplane flights) and using other features of Web pages, such as their geographic location (e.g., a user querying for restaurants).

23.6.3 Retrieval Systems

This section describes searching systems used to retrieve content from the four categories previously described in ▶ Sect. 23.4.

As noted above, PubMed is the system at NLM that searches MEDLINE and other bibliographic databases. Although presenting the user with a simple text box, PubMed does a great deal of processing of the user's input to identify MeSH terms, author names, common phrases, and journal names (described in the on-line help system of PubMed). In this automatic term mapping, the system attempts to map user input, in succession, to MeSH terms, journals names, common phrases, and authors. Remaining text that PubMed cannot map is searched as text words (i.e., words that occur in any of the MEDLINE fields). A results screen for a search combining the disease congestive heart failure (CHF) and the angiotensin converting enzyme (ACE) inhibitor class of drugs is shown in Fig. 23.4.

PubMed allows the use of wild-card characters. It also allows phrase searching whereby two or more words can be enclosed in quotation marks to indicate they must occur adjacent to each other. If the specified phrase is in PubMed's phrase index, then it will be searched as a phrase. Otherwise the individual words will be searched. PubMed allows speci-

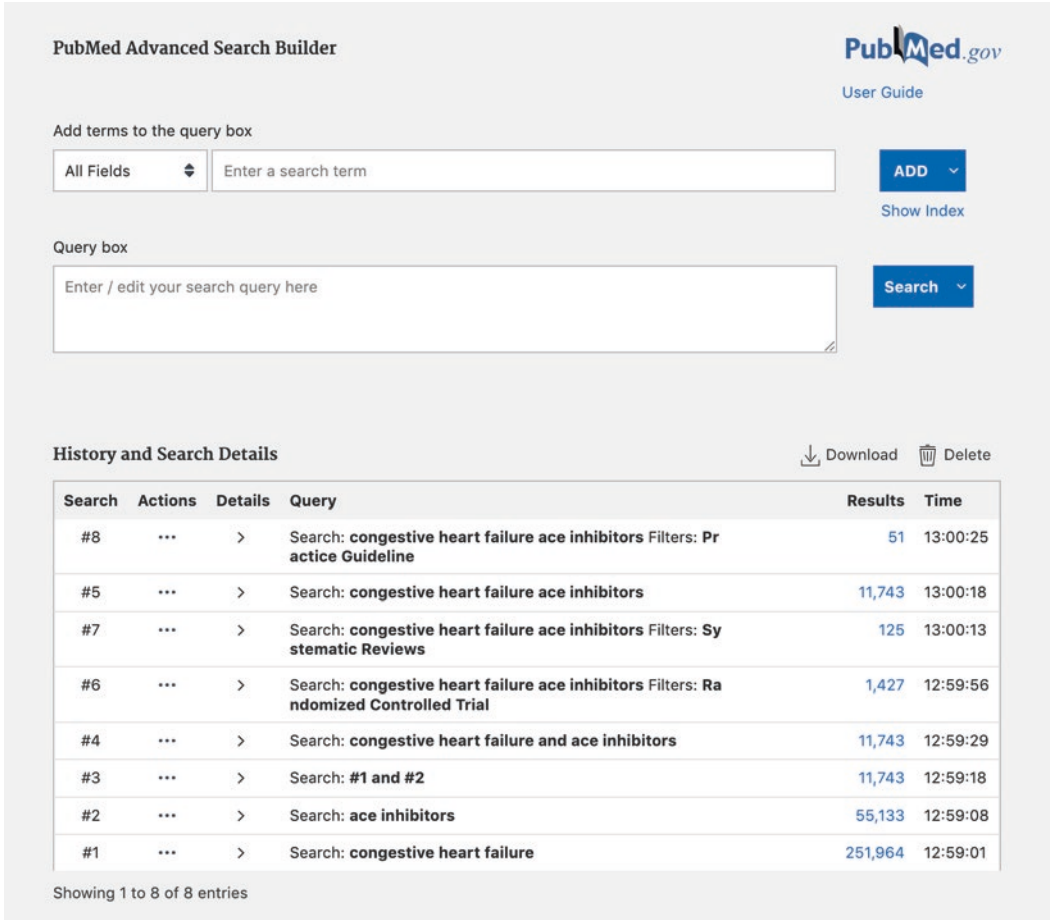


Fig. 23.5 PubMed Advanced Search Builder, showing the use of sets and application of Boolean operators as well as limits. (Courtesy of National Library of Medicine, with permission)

fication of other indexing attributes via "Limits." These include publication types, subsets, age ranges, and publication date ranges. These are accessed from the left-hand side of the results screen, with the most commonly used ones shown and the others accessible by additional mouse clicks.

As in most bibliographic systems, users can also search PubMed by building search sets and then combining them with Boolean operators to tailor the search. This is called the PubMed Advanced Search Builder. Consider a user searching for studies assessing the reduction of mortality in patients with CHF through the use of ACE inhibitors. A simple approach to such a search might be to combine the terms ACE inhibitors and CHF with an AND. The easiest way to do this is to enter the search string congestive heart failure and ACI inhibitors. Figure 23.5 shows the PubMed Advanced Search Builder screen such a searcher might develop. This searcher has limited the output (using some of the limits shown in Fig. 23.4) with various publication types known to contain the best evidence for this question. Also note that the search does not require the "and," as PubMed determines the Boolean operator should be placed there automatically.

Most MEDLINE systems have ranked output sorted by reverse chronological order, based on the notion that the most recent articles have the mostly timely and complete information. PubMed has also featured relevance-ranked output and recently improved its algorithms (Fiorini et al. 2018).

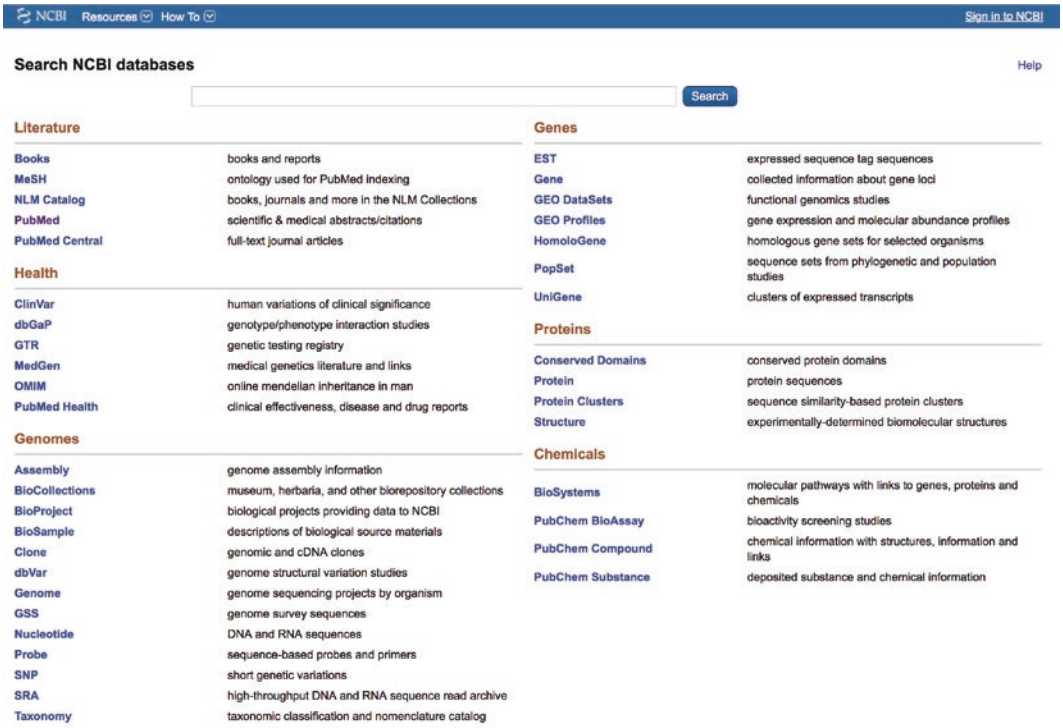


Fig. 23.6 Search screen for NLM NCBI Search, showing the variety of different databases that can be search on the NLM site. (Courtesy of National Library of Medicine, with permission)

PubMed has an additional approach to finding the best evidence, which is through the use of its **Clinical Queries** function,[60] where the subject terms are limited by search statements designed to retrieve the best evidence based on principles of EBM. There are two different approaches. The first uses strategies for retrieving the best evidence for the four major types of clinical questions. These strategies arise from research assessing the ability of MEDLINE search statements to identify the best studies for therapy, diagnosis, harm, and prognosis (Haynes et al. 1994). The second approach to retrieving the best evidence aims to retrieve evidence-based resources that are syntheses and synopses, in particular meta-analyses, systematic reviews, and practice guidelines. The strategy derives in part from research by Boynton et al. (Boynton et al. 1998). When the clinical queries interface is used, the search statement is processed by the usual automatic term mapping and the resulting output is limited (via AND) with the appropriate statement.

A growing number of search engines allow searching over many resources. The general search engines Google, Microsoft Bing, and others allow retrieval of any types of documents they have indexed via their Web crawling activities. Other search engines allow searching over aggregations of various sources, such as NCBI Search,[61] which allows searching over all NLM NCBI databases and other resources in one simple interface, as shown in Fig. 23.6.

23.7 Evaluation

There has been a great deal of research over the years devoted to evaluation of IR systems. As with many areas of research, there is con-

60 ► https://www.ncbi.nlm.nih.gov/pubmed/clinical

61 ► https://www.ncbi.nlm.nih.gov/search/

troversy as to which approaches to evaluation best provide results that can assess searching in the systems they are using. Many frameworks have been developed to put the results in context. One of those frameworks organized evaluation around six questions that someone advocating the use of IR systems might ask (Hersh and Hickam 1998):

1. Was the system used?
2. For what was the system used?
3. How well did they use the system?
4. Were the users satisfied?
5. What factors were associated with successful or unsuccessful use of the system?
6. Did the system have an impact?

While early evaluation studies asked questions about whether IR systems would be used if made available, in modern times their use is ubiquitous. A study by Google and Manhattan Research found that essentially all physicians reported searching on digital devices daily, with most searching resulting in action, such as changing treatment decisions or sharing with a colleague or patient (Anonymous 2012). Similarly, a recent study of internal medicine residents at three sites found that nearly all responding to the survey searched daily, with the most common resource searched being UpToDate (Duran-Nelson et al. 2013). The next most frequent source of information was consultation with attending faculty, followed by the Google search engine, the Epocrates drug reference, and various other "pocket" references. Another study of family medicine resident and attending physicians found universal use of smartphones and tablet devices daily in their practices (Yaman et al. 2016).

Patient and consumer searching of the Web for health information also continues to be reported high. In the most recent update of her ongoing survey of health-related searching, Fox found that 72% of US adult Internet users (59% of all US adults) have looked for health information in the last year (Fox and Duggan 2013). Earlier research found that the most common types of searches done by these users was for a specific disease or medical condition and for a certain medical treatment or procedure (Fox 2011). Three focus groups convened by Mayo Clinic researchers asked consumers about their online searching use and needs, finding that subjects reported searching, filtering, and comparing information retrieved, with the process stopping due to saturation and fatigue (Fiksdal et al. 2014).

Most evaluation research has focused on the third question from the above list, i.e., how well did search systems or their users perform? The rest of this section on evaluation will focus on studies of that question, grouping approaches and studies into those that are system-oriented, i.e., the focus of the evaluation is on the IR system, and those that are user-oriented, i.e., the focus is on the user.

23.7.1 System-Oriented Evaluation

There are many ways to evaluate the performance of IR systems, the most widely used of which are the relevance-based measures of **recall** and **precision**. These measures quantify the number of relevant documents retrieved by the user from the database and in his or her search. Recall is the proportion of relevant documents retrieved from the database:

$$\text{Recall} = \frac{\text{number of retrieved and relevant documents}}{\text{number of relevant documents in database}} \tag{23.5}$$

In other words, recall answers the question, for a given search, what fraction of all the relevant documents have been obtained from the database?

One problem with Eq. (23.5) is that the denominator implies that the total number of relevant documents for a query is known. For all but the smallest of databases, however, it is unlikely, perhaps even impossible, for one to succeed in identifying all relevant documents in a database. Thus most studies use the measure of **relative recall**, where the

denominator is redefined to be the total number of unique, relevant documents identified by one or more searches on the query topic.

Precision is the proportion of relevant documents retrieved in the search:

$$\text{Precision} = \frac{\text{number of retrieved and relevant documents}}{\text{number of documents retrieved}} \tag{23.6}$$

This measure answers the question, for a search, what fraction of the retrieved documents is relevant?

One problem that arises when one is comparing systems that use ranking versus those that do not is that nonranking systems, typically using Boolean searching, tend to retrieve a fixed set of documents and as a result have fixed points of recall and precision. Systems with relevance ranking, on the other hand, have different values of recall and precision depending on the size of the retrieval set the system (or the user) has chosen to show. The problem has been addressed by the development of aggregate measures that combine recall and precision and that account for ranking. One of the most common measures used is **mean average precision (MAP)**, which measures precision at each point that a relevant document is retrieved, and then provides a mean of all the average precision values (Buckley and Voorhees 2005).

Another challenge for IR evaluation occurs with large collections, where every possible item retrieved cannot be judged for relevance. If there is concern that there are large numbers of unjudged documents, the **B-Pref** measure can be used, which only makes use of unjudged documents in its calculations (Buckley and Voorhees 2004). An additional measure increasingly used in **normalized distributed cumulative gain (NDCG)**, which allows differential value of retrieved documents, e.g., a value of 2 for highly relevant and 1 for partially relevant documents (Jarvelin and Kekalainen 2002).

A good deal of evaluation in IR is done via **challenge evaluations**, in which a common IR task is defined and a **test collection** of documents, topics, and **relevance judgments** are developed. The relevance judgments define which documents are relevant for each topic in the task, allowing different researchers to compare their systems with others on the same task and improve them. The longest running and best-known challenge evaluation in IR is the **Text REtrieval Conference (TREC)**,[62] which is organized by the U.S. National Institute for Standards and Technology (NIST).[63] Started in 1992, TREC has provided a testbed for evaluation and a forum for presentation of results. TREC is organized as an annual event at which the tasks are specified and queries and documents are provided to participants. Participating groups submit "runs" of their systems to NIST, which calculates the appropriate performance measure(s). TREC is organized into "tracks" geared to specific interests. A book summarizing the first decade of TREC provides more information on this important IR initiative that is still ongoing (EM Voorhees and Harman 2005).

While TREC has been mostly focused on retrieval of general information sources (e.g., newswire, government documents, Web pages, etc.), there have been a number of tracks over the years devoted to biomedical IR. These tracks tended to reflect areas of biomedicine that were emerging importance. The first TREC track specific to the biomedical domain was the Genomics Track, due to the emergence at the time of the sequencing of human genome and the rise of the area of bioinformatics. A variety of literature retrieval tasks were developing, focused on journal article abstracts (from MEDLINE records) or full text (Hersh and Bhupatiraju 2003; Hersh

62 ► https://trec.nist.gov/
63 ► https://www.nist.gov/

et al. 2004, 2005, 2006, 2007; Hersh and Voorhees 2009; Roberts et al. 2009).

A second track from the biomedical domain aimed to leverage the growing interest in processing medical records text around the onset of the HITECH Act. The Medical Records Track used a collection of de-identified patient records for a task aiming to retrieve patients who might be candidates for clinical studies (Voorhees 2013; Voorhees and Hersh 2012; Voorhees and Tong 2011).

The next biomedical domain track was the Clinical Decision Support (CDS) Track, which aimed to retrieve full-text journal articles from a snapshot of PubMed Central to identify knowledge relevant to diagnosis, testing, or treatment (Roberts et al. 2015, 2016a, b; Simpson et al. 2014). The CDS Track was refined into the Precision Medicine Track, which aimed to retrieve information relevance to the precision medicine paradigm (Roberts et al. 2017).

Another annual challenge evaluation, based in Europe, has been the Conference and Labs of the Evaluation Forum (CLEF, originally known as the Cross-Language Evaluation Forum). Since 2013, one focus of CLEF has been eHealth, with tasks focused not only in IR, but also information extraction and information management.[64] The track has had a patient-centered retrieval task since 2013, a cross-language retrieval task since 2014, and a systematic review task starting in 2017. An additional challenge evaluation emanating from CLEF and focused on image retrieval, which has included a medical image retrieval component, has been ImageCLEF.[65] Some recent overviews of the state of the art of image retrieval have been published (Li et al. 2018; Müller and Unay 2017).

Some system-oriented studies have focused on specific use cases for IR systems. One area gaining a good deal of attention has been reducing the workload of performing systematic reviews, which require high recall to retrieval all possibly relevant studies. This comes at a cost of low precision, so a question is how to reduce the work of systematic reviews by increasing precision while minimally impacting recall. Early work in this area was carried out by Cohen et al. (2009, 2015), demonstrating value for machine learning approaches. A systematic review task was added to CLEF eHealth in 2017 (Kanoulas et al. 2017, 2018, 2019).

A number of researchers have criticized or noted the limitations of relevance-based measures. While no one denies that users want systems to retrieve relevant articles, it is not clear that the quantity of relevant documents retrieved is the complete measure of how well a system performs (Harter 1992; Swanson 1988). Hersh (1994) noted that clinical users are unlikely to be concerned about these measures when they simply seek an answer to a clinical question and are able to do so no matter how many other relevant documents they miss (lowering recall) or how many nonrelevant ones they retrieve (lowering precision). This has led to more focus on user-oriented evaluation.

23.7.2 User-Oriented Evaluation

What alternatives to relevance-based measures can be used for determining performance of individual searches? Some alternatives have focused on users being able to perform various information tasks with IR systems, such as finding answers to questions (Egan et al. 1989; Hersh and Hickam 1995; Hersh et al. 1996; Mynatt et al. 1992; Wildemuth et al. 1995). For several years, TREC featured an Interactive Track that had participants carry out user experiments with the same documents and queries (Hersh 2001). A number of user-oriented evaluations have been performed over the years looking at users of biomedical information.

When end-user retrieval systems first appeared, a number of studies appeared aiming to measure search performance by clinicians. One of the original studies compared the capabilities of librarian and clinician searchers (Haynes et al. 1990). In this study, 78 searches were randomly chosen for replica-

64 ► https://sites.google.com/site/clefehealth/
65 ► https://www.imageclef.org

tion by both a clinician experienced in searching and a medical librarian. The results showed that the experienced clinicians and librarians achieved comparable recall in the range of 50%, although the librarians had better precision. The novice clinician searchers had lower recall and precision than either of the other groups. This study also assessed user satisfaction of the novice searchers, who despite their recall and precision results said that they were satisfied with their search outcomes. A follow-up study noted that different searchers tended to use different strategies on a given topic. The different approaches replicated a finding known from other searching studies in the past, namely, the lack of overlap across searchers of overall retrieved citations as well as relevant ones. Thus, even though the novice searchers had lower recall, they did obtain a great many relevant citations not retrieved by the two expert searchers. Furthermore, fewer than 4% of all the relevant citations were retrieved by all three searchers.

Recognizing the limitations of recall and precision for evaluating clinical users of IR systems, subsequent studies assessed the ability of systems to help students and clinicians answer clinical questions. The rationale for these studies is that the usual goal of using an IR system is to find an answer to a question. While the user must obviously find relevant documents to answer that question, the quantity of such documents is less important than whether the question is successfully answered. In fact, recall and precision can be placed among the many factors that may be associated with ability to complete the task successfully.

The first study using this task-oriented approach compared Boolean versus natural language searching in an online medical textbook (Hersh and Hickam 1995). There was no difference in ability to answer questions with one interface or the other. Most answers were found on the first search to the textbook. For the questions that were incorrectly answered, the document with the correct answer was actually retrieved by the user two-thirds of the time and viewed more than half the time.

Another study compared Boolean and natural language searching of MEDLINE with two commercial products, CD Plus (now Ovid) and Knowledge Finder representing Boolean and natural language searching respectively (Hersh et al. 1996). Sixteen medical students were recruited and randomized to one of the two systems and given three yes/no clinical questions to answer. The students were able to use each system successfully, answering 37.5% correctly before searching and 85.4% correctly after searching. There were no significant differences between the systems in time taken, relevant articles retrieved, or user satisfaction. This study demonstrated that both types of systems could be used equally well with minimal training.

A more comprehensive study looked at MEDLINE searching by medical and nurse practitioner (NP) students to answer clinical questions. A total of 66 medical and NP students searched five questions each (Hersh et al. 2002). This study used a multiple-choice format for answering questions that also included a judgment about the evidence for the answer. Subjects were asked to choose from one of three answers:

- Yes, with adequate evidence.
- Insufficient evidence to answer question.
- No, with adequate evidence.

Both groups achieved a presearching correctness on questions about equal to chance (32.3% for medical students and 31.7% for NP students). However, medical students improved their correctness with searching (to 51.6%), whereas NP students hardly did at all (to 34.7%).

This study also attempted to measure what factors might influence searching. A multitude of factors, such as age, gender, computer experience, and time taken to search, were not associated with successful answering of questions. Successful answering was, however, associated with answering the question correctly before searching, spatial visualization ability (measured by a validated instrument), searching experience, and EBM question type (prognosis questions easiest, harm questions most difficult). An analysis of recall and pre-

cision for each question searched demonstrated a complete lack of association with ability to answer these questions.

Two studies extended this approach in different ways. Westbook et al. assessed use of an online evidence systems and found that physicians answered 37% of questions correctly before use of the system and 50% afterwards, while nurse specialists answered 18% of questions correctly and also 50% afterwards (Westbrook et al. 2005). Those who had correct answers before searching had higher confidence in their answers, but those not initially knowing the answer had no difference in confidence whether their answer turned out to be right or wrong. McKibbon and Fridsma performed a comparable study of allowing physicians to seek answers to questions with resources they normally use (McKibbon and Fridsma 2006) employing the same questions as Hersh et al. (2002). This study found no difference in answer correctness before or after using the search system.

Pluye et al. (Pluye and Grad 2004) performed a qualitative study assessing impact of IR systems on physician practice. The study identified 4 themes mentioned by physicians:

- Recall—of forgotten knowledge
- Learning—new knowledge
- Confirmation—of existing knowledge
- Frustration—that system use not successful

The researchers also noted two additional themes:

- Reassurance—that system is available
- Practice improvement—of patient-physician relationship

More recent studies have focused on searchers using well-known modern IR systems. Kim et al. looked at the ability of internal medicine interns to answer questions starting from Google versus an evidence-based summary resource developed by a local medical library (Kim et al. 2014). Ten questions were given to each subject, with each participant randomized to start in either Google or the summary resource for half of questions. Answers were found for 82% of the questions administered, with no difference between groups in correct answers (58–62% correct) or time taken (136–139 seconds). While those starting in the summary resource mostly found answers in resources that were part of the summary system 93% of the time, those starting with Google found answers in commercial medical portals (25.7%), hospital Web sites (12.6%), Wikipedia (12.0%), US government Web sites (9.4%), PubMed (9.4%), evidence-based summary resources (9.4%), and others (18%). Another study looked at medical students' short-term knowledge when randomized to answer questions in Wikipedia, UpToDate, and a digital textbook, finding the best short-term knowledge acquisition with Wikipedia (Scaffidi et al. 2017).

Koopman et al. assessed factors comprising effective queries and those making them (Koopman et al. 2017). They found that query formulation had more impact on retrieval effectiveness than the particular retrieval systems used. The most effective queries were short, ad-hoc keyword queries and queriers who inferred novel keywords most likely to appear in relevant documents.

Other users of IR systems have been studied beyond clinicians. One study used the TREC Genomics Track 2004 collection to assess the value of MeSH terms for different types of searchers (Liu and Wacholder 2017). The researchers recruited four types of searchers:

- Search Novice (SN) – undergraduates with no formal search training or advanced knowledge in biomedicine
- Domain Expert (DE) – biomedical graduate students
- Search Expert (SE) – library and information science graduate students
- Medical Librarian (ML)

The searchers used a digital library system to search on 20 topics from the original test collection. Searchers assigned to search with MeSH were provided access to a MeSH browser. As with other studies, recall (0.15–0.23) and precision (0.29–0.40) were relatively close across different groups. MeSH terms had little impact upon recall in the four groups, but they were found to substantially

increase precision in search novices (SN and DE) and decrease it in search experts (SE and ML) (recall and precision with MeSH; without MeSH). User characteristics that improved precision were number of undergraduate and graduate biology courses for SN and DE respectively. User characteristics associated with improved recall included having had online search courses and MeSH use experience. Other factors having no association with search results included gender, native language, age, or experience or frequency with database searching.

Another group of searchers that have been studied are consumers. A study from Mayo Clinic analyzed search queries submitted through general search engines but leading users into a consumer health information portal from computers and mobile devices (Jadhav et al. 2014). The most common types of searches were on symptoms (32–39%), causes of disease (19–20%), and treatments and drugs (14–16%). Health queries tended to be longer and more specific than general (non-health) queries. Health queries were somewhat more likely to come from mobile devices. Most searches used key words, although some were also phrased as questions (wh- or yes/no).

An additional study aimed to assess differences in searching between medical experts and lay people (Palotti et al. 2016). This study found that medical experts were more persistent in their interaction with the search engine. They also noted that the main focus of users, both laypeople and professionals, was on disease rather than symptoms.

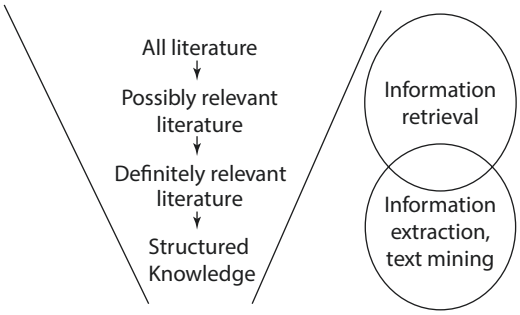


Fig. 23.7 Funnel of knowledge discovery, showing how an information need starts with a search (information retrieval) leading to a large possibly relevant set of literature that is winnowed down to a smaller definitely relevant set (usually by human inspection but with techniques like information extraction and text mining possibly automating the process in the future). Ultimately actionable knowledge is obtained that can be applied by a human or fashioned into, for example, rules for a computer-based decision support system

23.8 Research Directions

The above evaluation research shows that there is still plenty of room for IR systems to improve their abilities. In addition, there will be new challenges that arise from growing amounts of information, new devices, and other new technologies.

There are also other areas related to IR where research is ongoing in the larger quest to help all involved in biomedicine and health—including patients, clinicians and researchers—to better apply knowledge to improve health. Figure 23.7 shows this author's "funnel" by which the user searches all of the scientific literature using IR systems to obtain a set of possibly relevant literature. In the current state of the art, he/she reviews this literature by hand, selecting which articles are definitely relevant and may become "actionable knowledge" that can be acted upon to make better decisions.

Our ability to carry out the activities in the upper part of the funnel, i.e., IR, is much better than those in the lower part. These areas include:

- **Information extraction** and **text mining**—usually through the use of natural language processing (NLP, see ▶ Chap. 8) to extract facts and knowledge from text (K. Cohen and Demner-Fushman 2014). These techniques are often employed to extract information from the EHR, with a wide variety of accuracy as shown in a systematic review (Stanfill et al. 2010).
- **Summarization**—providing automated extracts or abstracts summarizing the content of longer documents (Fiszman et al. 2004; Mani 2001). In recent years, these methods have been applied to text and other data in the EHR (Pivovarov and Elhadad 2015).

- **Question-answering**—going beyond retrieval of documents to providing actual answers to questions, as exemplified by IBM's Watson system (Ferrucci et al. 2010). Watson has been evaluated for answering questions on medical board exams (Ferrucci et al. 2012) as well as making cancer treatment recommendations (Somashekhar et al. 2018).

23.9 Digital Libraries

Discussion of IR "systems" thus far has focused on the provision of retrieval mechanisms to access online content. Even with the expansive coverage of some IR systems, such as Web search engines, they are often part of a larger collection of services or activities. An alternative perspective, especially when communities and/or proprietary collections are involved, is the **digital library**. Digital libraries share many characteristics with "brick and mortar" libraries, but also take on some additional challenges. Borgman (1999) noted that libraries of both types elicited different definitions of what they actually are, with researchers tending to view libraries as content collected for specific communities and librarians alternatively viewing them as institutions or services. Lindberg and Humphreys (2005) laid out a vision in 2005 for libraries 10 years hence, noting that while collections would be virtual and accessed in many diverse ways, other elements of science would stay intact, including journals and the peer review process.

This section provides an overview of key issues of digital libraries, with an orientation toward biomedical libraries.

23.9.1 Functions and Definitions of Libraries

The central function of libraries is to maintain collections of published literature. They may also store unpublished literature in archives, such as letters, notes, and other documents. The general focus on published literature has implications. One of these is that, for the most part, quality control can be taken for granted. Until recently, most published literature came from commercial publishers and specialty societies that had processes such as peer review, which, although imperfect, allowed the library to devote minimal resources to assessing their quality. While libraries can still cede the judgment of quality to these information providers in the Internet era, they cannot ignore the myriad of information published only on the Internet, for which the quality cannot be presumed.

Other functions of libraries besides maintaining collections include cataloging and classification of items in those collections, being a place (even virtual) where individuals could go to get assistance with information seeking, and providing space for work or study, particularly in universities.

The paper-based nature of traditional libraries carried a number of assumptions that are challenged in the digital era. For example, items were produced in multiple copies, freeing the individual library from excessive worry that an item could not be replaced. In addition, items were fairly static, simplifying their cataloging. With digital libraries, this status quo is challenged. There is a great deal of concern about archiving of content and managing its change when fewer "copies" of it exist on the file servers of publishers and other organizations. A related problem for digital libraries is that they do not own the "artifact" of the paper journal, book, or other item. This is exacerbated by the fact that when a subscription to an electronic journal is terminated, access to the entire journal is lost; that is, the subscriber does not retain accumulated back issues, as was taken for granted with paper journals.

23.9.2 Access

Probably every Web user is familiar with clicking on a Web link and receiving an error message that a page cannot found. Digital libraries and commercial publishing ventures need mechanisms to ensure that documents have persistent identifiers so that when the

document itself physically moves, it is still obtainable. The original architecture for the Web envisioned by the Internet Engineering Task Force was to have every **uniform resource locator (URL)**, the address entered into a Web browser or used in a Web hyper-link, linked to a **uniform resource name (URN)** that would be persistent (Sollins and Masinter 1994). The combination of a URN and a URL, a **uniform resource identifier (URI)**, would provide persistent access to digital objects. However, no publicly available resource for resolving URNs and URIs was ever implemented on a large scale.

One approach that has seen widespread adoption by publishers, especially scientific journal publishers, is the **digital object identifier (DOI)**[66] (Paskin 2006). The DOI has recently been given the status of a standard by the NISO with the designation Z39.84. The DOI itself is relatively simple, consisting of a prefix that is assigned by the International DOI Foundation (IDF) to the publishing entity and a suffix that is assigned and maintained by the entity. For example, the DOI for articles from the Journal of the American Medical Informatics Association have the prefix 10.1197 and the suffix jamia.M####, where #### is a number assigned by the journal editors. Publishers are encouraged to facilitate resolution by encoding the DOI into their URLs in a standard way, e.g., ► https://doi.org/10.1197/jamia.M0996 for a paper cited earlier in the chapter (Hersh et al. 2002).

23.9.3 Interoperability

As noted throughout this chapter, metadata is a key component for accessing content in IR systems. It takes on an additional value in the digital library, where there is desire to allow access to diverse but not necessarily exhaustive resources. One key concern of digital libraries is **interoperability** (Besser 2002). That is, how can resources with heterogeneous metadata be accessed? Arms et al. note that three levels of agreement must be achieved in digital libraries:

1. Technical agreements over formats, protocols, and security procedures
2. Content agreement over the data and the semantic interpretation of its metadata
3. Organizational agreements over ground rules for access, preservation, payment, authentication, and so forth

23.9.4 Intellectual Property

Intellectual property issues are a major concern in digital libraries. Intellectual property is difficult to protect in the digital environment because although the cost of production is not insubstantial, the cost of replication is near nothing. Furthermore, in circumstances such as academic publishing, the desire for protection is situational. For example, individual researchers may want the widest dissemination of their research papers, but each one may want to protect revenues realized from synthesis works or educational products that are developed. The global reach of the Internet has required that intellectual property issues be considered on a global scale. The World Intellectual Property Organization (WIPO)[67] is an agency of the United Nations devoted to developing worldwide policies, although understandably, there is considerable diversity about what such policies should be.

23.9.5 Preservation

Another function of libraries of all types is preservation of materials. In paper-based libraries, the goal of preservation was the survival of the physical object, i.e., the book, journal, image, etc. that could become lost, stolen, or deteriorated. Preservation issues in digital libraries are somewhat different. Digital libraries still do need to be concerned with physical survival of the information. Lesk compared the longevity of digital materials (Lesk 2005). He noted that the longev-

66 ► http://www.doi.org/

67 ► http://www.wipo.int/portal/en/index.html

ity for magnetic materials was the least, with the expected lifetime of magnetic tape being 5 to 10 years. Optical storage has somewhat better longevity, with an expected lifetime of 30 to 100 years depending on the specific type. Ironically, paper has a life expectancy well beyond all these digital media. Rothenberg noted that the Rosetta Stone, which provided help in interpreting ancient Egyptian hieroglyphics and has survived over 20 centuries (Rothenberg 1999). He reiterated Lesk's description of the reduced lifetime of digital media in comparison with traditional media, and to note another problem familiar to most long-time users of computers, namely, data can become obsolete not only owing to the medium, but also as a result of data format. Both authors noted that storage devices as well as computer applications, such as word processors, have seen their formats change significantly over the last couple of decades.

The US Library of Congress has devoted considerable effort to digital preservation, documenting its efforts on the Web site.[68] An early digital preservation effort in the US was National Digital Information Infrastructure Preservation Program (NDIIPP) of the Library of Congress, which has now become the National Digital Stewardship Alliance (NDSA)[69] and is housed by the Digital Library Federation (DLF), at the Council on Library and Information Resources (CLIR). Other digital preservation efforts include Portico,[70] a collaboration of publishers, libraries, and government agencies to preserve electronic scholarly content and LOCKSS (Lots of Copies Keep Stuff Safe),[71] which provides libraries with digital preservation tools and support.

68 ► http://www.digitalpreservation.gov/
69 ► https://ndsa.org
70 ► https://www.portico.org/
71 ► https://www.lockss.org/

23.10 Future Directions for IR Systems and Digital Libraries

There is no doubt that considerable progress has been made in IR and digital libraries. Seeking online information is now done routinely not only by clinicians and researchers, but also by patients and consumers. There are still considerable challenges to make this activity more fruitful to users. They include:

- How do we lower the effort it takes for clinicians to get to the information they need rapidly in the busy clinical setting?
- How can researchers extract new knowledge from the vast quantity that is available to them?
- How can consumers and patients find high-quality information that is appropriate to their understanding of health and disease?
- Can the value added by the publishing process be protected and remunerated while making information more available?
- How can the indexing process become more accurate and efficient?
- Can retrieval interfaces be made simpler without giving up flexibility and power?
- Can we develop standards for digital libraries that will facilitate interoperability but maintain ease of use, protection of intellectual property, and long-term preservation of the archive of science?

Suggested Readings

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). Modern information retrieval: The concepts and technology behind search (2nd ed.). Reading: Addison-Wesley. A book surveying most of the automated approaches to information retrieval.

Croft, W., Metzler, D., & Strohman, T. (2009). Search engines: Information retrieval in practice. Boston: Addison-Wesley. A book surveying most of the automated approaches to search engines.

Hersh, W. (2020). Information retrieval: A biomedical and health perspective (4th ed.).

New York: Springer. A textbook on information retrieval systems in the health and biomedical domain that covers state-of-the-art as well as research systems.

Miles, W. (1982). A history of the national library of medicine: The nation's treasury of medical knowledge. Bethesda: U.S. Department of Health and Human Services. A comprehensive history of the National Library of Medicine and its forerunners, covering the story of Dr. John Shaw Billings and his founding of Index Medicus to the modern implementation of MEDLINE.

National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Board on Research Data and Information; Committee on Toward an Open Science Enterprise. (2018). Open science by design – realizing a vision for 21st century research. Washington, DC: National Academies Press. A vision for open science from the National Academies of Medicine, Science, and Engineering.

Questions for Discussion

1. With the advent of full-text searching, should the National Library of Medicine abandon human indexing of citations in MEDLINE? Why or why not?
2. Explain why you think open-access publishing will succeed or not.
3. How would you aggregate the clinical evidence-based resources described in the chapter into the best digital library for clinicians?
4. Devise a curriculum for teaching clinicians and patients the most important points about searching for health-related information.
5. Find a consumer-oriented Web page and determine the quality of the information on it.
6. What are the limitations of recall and precision as evaluation measures and what alternatives would improve upon them?
7. Select a concept that appears in two or more clinical terminologies and demonstrate how it would be combined into a record in the UMLS Metathesaurus.
8. Describe how you might devise a system that achieves a happy medium between protection of intellectual property and barrier-free access to the archive of science.

References

Akerkar, R. (2009). *Foundations of the Semantic Web: XML, RDF & Ontology*. Oxford: Alpha Science International Ltd.

Anonymous. (2012). *From screen to script: The doctor's digital path to treatment*. Retrieved from New York: https://www.thinkwithgoogle.com/research-studies/the-doctors-digital-path-to-treatment.html.

Anonymous. (2016). Toward fairness in data sharing. *New England Journal of Medicine, 375*, 405–407.

Anonymous. (2018a). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research, 46*(D1), D8–D13. Retrieved from https://academic.oup.com/nar/article/46/D1/D8/4621330.

Anonymous. (2018b). *Open science by design – realizing a vision for 21st century research*. Washington, DC: National Academies Press.

Bartlett, J., & Toms, E. (2005). Developing a protocol for bioinformatics analysis: An integrated information and behavior task analysis approach. *Journal of the American Society for Information Science and Technology, 56*, 469–482.

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H., & Secret, A. (1994). The World-Wide Web. *Communications of the ACM, 37*, 76–82.

Besser, H. (2002). The next stage: Moving from isolated digital collections to interoperable digital libraries. *First Monday, 7*, 6. Retrieved from http://www.firstmonday.dk/issues/issue7_6/besser/.

Bodenreider, O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Research, 32*, D267–D270.

Borgman, C. (1999). What are digital libraries? Competing visions. *Information Processing and Management, 35*, 227–244.

Boynton, J., Glanville, J., McDaid, D., & Lefebvre, C. (1998). Identifying systematic reviews in MEDLINE: Developing an objective approach to search strategy design. *Journal of Information Science, 24*, 137–157.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, 30*, 107–117. Retrieved from http://infolab.stanford.edu/pub/papers/google.pdf.

Buckley, C., & Voorhees, E. (2004). *Retrieval evaluation with incomplete information.* Paper presented at the proceedings of the 27th annual international ACM SIGIR conference on research and development in information retrieval, Sheffield, England.

Buckley, C., & Voorhees, E. (2005). Retrieval system evaluation. In E. Voorhees & D. Harman (Eds.), *TREC: Experiment and evaluation in information retrieval* (pp. 53–75). Cambridge, MA: MIT Press.

Castillo, C., & Davison, B. (2011). *Adversarial Web search*. Delft: Now Publishers.

Charen, T. (1976). *MEDLARS indexing manual, part I: Bibliographic principles and descriptive indexing, 1977*. Springfield: National Technical Information Service.

Charen, T. (1983). *MEDLARS indexing manual, part II*. Springfield: National Technical Information Service.

Cimino, J., & delFiol, G. (2007). Infobuttons and point of care access to knowledge. In R. Greenes (Ed.), *Clinical decision support: The road ahead* (pp. 345–371). Amsterdam: Elsevier.

Cohen, K., & Demner-Fushman, D. (2014). *Biomedical natural language processing*. Amsterdam: John Benjamins Publishing.

Cohen, A., Ambert, K., & McDonagh, M. (2009). Cross-topic learning for work prioritization in systematic review creation and update. *Journal of the American Medical Informatics Association, 16*, 690–704.

Cohen, A., Smalheiser, N., McDonagh, M., Yu, C., Adams, C., Davis, J., & Yu, P. (2015). Automated confidence ranked classification of randomized controlled trial articles: An aid to evidence-based medicine. *Journal of the American Medical Informatics Association, 22*, 707–717.

Coletti, M., & Bleich, H. (2001). Medical subject headings used to search the biomedical literature. *Journal of the American Medical Informatics Association, 8*, 317–323.

Covell, D., Uman, G., & Manning, P. (1985). Information needs in office practice: Are they being met? *Annals of Internal Medicine, 103*, 596–599.

Darmoni, S., Leroy, J., Baudic, F., Douyere, M., Piot, J., & Thirion, B. (2000). CISMeF: A structured health resource guide. *Methods of Information in Medicine, 9*, 30–35.

DeAngelis, C., Drazen, J., Frizelle, F., Haug, C., Hoey, J., Horton, R., et al. (2005). Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *Journal of the American Medical Association, 293*, 2927–2929.

DeBakey, M. (1991). The National Library of Medicine: Evolution of a premier information center. *Journal of the American Medical Association, 266*, 1252–1258.

Demner-Fushman, D., Antani, S., Simpson, M., & Thoma, G. (2012). Design and development of a multimodal biomedical information retrieval system. *Journal of Computing Science and Engineering, 6*, 168–177.

Diamond, G., & Kaul, S. (2008). The disconnect between practice guidelines and clinical practice – stressed out. *Journal of the American Medical Association, 300*, 1817–1819.

DiCenso, A., Bayley, L., & Haynes, R. (2009). ACP Journal Club. Editorial: Accessing preappraised evidence: Fine-tuning the 5S model into a 6S model. *Annals of Internal Medicine, 151*(6), JC3–JC2, JC3-3.

Drazen, J., & Curfman, G. (2004). Public access to biomedical research. *New England Journal of Medicine, 351*, 1343.

Duran-Nelson, A., Gladding, S., Beattie, J., & Nixon, L. (2013). Should we Google it? resource use by internal medicine residents for point-of-care clinical decision making. *Academic Medicine, 88*, 788–794.

Egan, D., Remde, J., Gomez, L., Landauer, T., Eberhardt, J., & Lochbaum, C. (1989). Formative design-evaluation of Superbook. *ACM Transactions on Information Systems, 7*, 30–57.

Ely, J., Osheroff, J., Ebell, M., Bergus, G., Levy, B., Chambliss, M., & Evans, E. (1999). Analysis of questions asked by family doctors regarding patient care. *British Medical Journal, 319*, 358–361.

Ely, J., Osheroff, J., Ebell, M., Chambliss, M., Vinson, D., Stevermer, J., & Pifer, E. (2002). Obstacles to answering doctors' questions about patient care with evidence: Qualitative study. *British Medical Journal, 324*, 710–713.

Eysenbach, G., Su, E., & Diepgen, T. (1999). Shopping around the internet today and tomorrow: Towards the millennium of cybermedicine. *British Medical Journal, 319*, 1294–1298.

Eysenbach, G., Powell, J., Kuss, O., & Sa, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the world wide web: A systematic review. *Journal of the American Medical Association, 287*, 2691–2700.

Ferguson, T. (2002). From patients to end users: Quality of online patient networks needs more attention than quality of online health information. *British Medical Journal, 324*, 555–556.

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., et al. (2010). Building Watson: An overview of the DeepQA Project. *AI Magazine, 31*(3), 59–79. Retrieved from http://www.aaai.org/ojs/index.php/aimagazine/article/view/2303.

Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., & Mueller, E. (2012). Watson: Beyond Jeopardy! *Artificial Intelligence, 199-200*, 93–105.

Fiksdal, A., Kumbamu, A., Jadhav, A., Cocos, C., Nelsen, L., Pathak, J., & McCormick, J. (2014). Evaluating the process of online health information searching: A qualitative approach to exploring consumer perspectives. *Journal of Medical Internet Research, 16*(10), e224. Retrieved from http://www.jmir.org/2014/10/e224/.

Fiorini, N., Canese, K., Starchenko, G., Kireev, E., Kim, W., Miller, V., et al. (2018). Best Match: New relevance search for PubMed. *PLoS Biology, 16*(8), e2005343. Retrieved from https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2005343.

Fiszman, M., Rindflesch, T., & Kilicoglu, H. (2004). *Summarization of an online medical encyclopedia.*

Paper presented at the MEDINFO 2004 – proceedings of the Eleventh World Congress on Medical Informatics, San Francisco, CA.

Foulonneau, M., & Riley, J. (2008). *Metadata for digital resources: Implementation, systems design and interoperability*. New York: Neal-Schuman Publishers.

Fox, C. (1992). Lexical analysis and stop lists. In W. Frakes & R. Baeza-Yates (Eds.), *Information retrieval: Data structures and algorithms* (pp. 102–130). Englewood Cliffs: Prentice-Hall.

Fox, S. (2011). *Health topics*. Retrieved from Washington, DC: http://www.pewinternet.org/2011/02/01/health-topics-2/.

Fox, S., & Duggan, M. (2013). *Health online 2013*. Retrieved from Washington, DC: https://www.pewinternet.org/2013/01/15/health-online-2013/.

Frakes, W. (1992). Stemming algorithms. In W. Frankes & R. Baeza-Yates (Eds.), *Information retrieval: Data structures and algorithms* (pp. 131–160). Englewood Cliffs: Prentice-Hall.

Frank, M. (2013). Open but not free – publishing in the 21st century. *New England Journal of Medicine, 368*, 787–789.

Funk, M., & Reid, C. (1983). Indexing consistency in MEDLINE. *Bulletin of the Medical Library Association, 71*, 176–183.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature, 438*, 900–901. Retrieved from http://www.nature.com/nature/journal/v438/n7070/full/438900a.html.

Gorman, P., & Helfand, M. (1995). Information seeking in primary care: How physicians choose which clinical questions to pursue and which to leave unanswered. *Medical Decision Making, 15*, 113–119.

Greenhalgh, T., Howick, J., & Maskrey, N. (2014). Evidence based medicine: A movement in crisis? *British Medical Journal, 348*, g3725. Retrieved from http://www.bmj.com/content/348/bmj.g3725.

Guyatt, G., Cook, D., & Haynes, R. (2004). Evidence based medicine has come a long way. *British Medical Journal, 329*, 990–991.

Guyatt, G., Rennie, D., Meade, M., & Cook, D. (Eds.). (2014). *Users' guides to the medical literature: A manual for evidence-based clinical practice, 3E*. New York: McGraw-Hill.

Guyatt, G., Rennie, D., Meade, M., & Cook, D. (Eds.). (2015). *Users' guides to the medical literature: Essentials of evidence-based clinical practice* (3rd ed.). New York: McGraw-Hill.

Harter, S. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science, 43*, 602–615.

Haug, C. (2013). The downside of open-access publishing. *New England Journal of Medicine, 368*, 791–793.

Haynes, R., McKibbon, K., Walker, C., Ryan, N., Fitzgerald, D., & Ramsden, M. (1990). Online access to MEDLINE in clinical settings. *Annals of Internal Medicine, 112*, 78–84.

Haynes, R., Wilczynski, N., McKibbon, K., Walker, C., & Sinclair, J. (1994). Developing optimal search strategies for detecting clinically sound studies in MEDLINE. *Journal of the American Medical Informatics Association, 1*, 447–458.

Haynes, R., Devereaux, P., & Guyatt, G. (2002). Clinical expertise in the era of evidence-based medicine and patient choice. *ACP Journal Club, 136*(2), A11–A14.

Hersh, W. (1994). Relevance and retrieval evaluation: Perspectives from medicine. *Journal of the American Society for Information Science, 45*, 201–206.

Hersh, W. (1999). "A world of knowledge at your fingertips": The promise, reality, and future directions of on-line information retrieval. *Academic Medicine, 74*, 240–243.

Hersh, W. (2001). Interactivity at the Text Retrieval Conference (TREC). *Information Processing and Management, 37*, 365–366.

Hersh, W. (2020). *Information retrieval: A biomedical and health perspective* (4th ed.). New York: Springer.

Hersh, W., & Bhupatiraju, R. (2003). *TREC genomics track overview.* Paper presented at the The Twelfth Text Retrieval Conference (TREC 2003), Gaithersburg, MD.

Hersh, W., & Hickam, D. (1995). An evaluation of interactive Boolean and natural language searching with an on-line medical textbook. *Journal of the American Society for Information Science, 46*, 478–489.

Hersh, W., & Hickam, D. (1998). How well do physicians use electronic information retrieval systems? A framework for investigation and review of the literature. *Journal of the American Medical Association, 280*, 1347–1352.

Hersh, W., & Rindfleisch, T. (2000). Electronic publishing of scholarly communication in the biomedical sciences. *Journal of the American Medical Informatics Association, 7*, 324–325.

Hersh, W., & Voorhees, E. (2009). TREC genomics special issue overview. *Information Retrieval, 12*, 1–15.

Hersh, W., Pentecost, J., & Hickam, D. (1996). A task-oriented approach to information retrieval evaluation. *Journal of the American Society for Information Science, 47*, 50–56.

Hersh, W., Crabtree, M., Hickam, D., Sacherek, L., Friedman, C., Tidmarsh, P., et al. (2002). Factors associated with success for searching MEDLINE and applying evidence to answer clinical questions. *Journal of the American Medical Informatics Association, 9*, 283–293.

Hersh, W., Bhuptiraju, R., Ross, L., Johnson, P., Cohen, A., & Kraemer, D. (2004). *TREC 2004 genomics track overview.* Paper presented at the The Thirteenth Text Retrieval Conference (TREC 2004), Gaithersburg, MD.

Hersh, W., Cohen, A., Yang, J., Bhupatiraju, R., Roberts, P., & Hearst, M. (2005). *TREC 2005 genomics track overview.* Paper presented at the The Fourteenth Text Retrieval Conference – TREC 2005, Gaithersburg, MD.

Hersh, W., Cohen, A., Roberts, P., & Rekapalli, H. (2006). *TREC 2006 genomics track overview.* Paper presented at the The Fifteenth Text Retrieval Conference (TREC 2006), Gaithersburg, MD.

Hersh, W., Cohen, A., Ruslen, L., & Roberts, P. (2007). *TREC 2007 genomics track overview.* Paper presented at the The Sixteenth Text Retrieval Conference (TREC 2007) Proceedings, Gaithersburg, MD.

Humphreys, B., Lindberg, D., Schoolman, H., & Barnett, G. (1998). The Unified Medical Language System: An informatics research collaboration. *Journal of the American Medical Informatics Association, 5*, 1–11.

Ioannidis, J. (2016a). Evidence-based medicine has been hijacked: A report to David Sackett. *Journal of Clinical Epidemiology, 73*, 82–86.

Ioannidis, J. (2016b). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Quarterly, 94*, 485–514.

Ioannidis, J. (2017). Hijacked evidence-based medicine: Stay the course and throw the pirates overboard. *Journal of Clinical Epidemiology, 84*, 11–13.

Jadad, A. (1999). Promoting partnerships: Challenges for the Internet age. *British Medical Journal, 319*, 761–764.

Jadhav, A., Andrews, D., Fiksdal, A., Kumbamu, A., McCormick, J., Misitano, A., et al. (2014). Comparative analysis of online health queries originating from personal computers and smart devices on a consumer health information portal. *Journal of Medical Internet Research, 16*(7), e160. Retrieved from http://www.jmir.org/2014/7/e160/.

Jarvelin, K., & Kekalainen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems, 20*, 422–446.

Kanoulas, E., Li, D., Azzopardi, L., & Spijker, R. (2017). *CLEF 2017 technologically assisted reviews in empirical medicine overview.* Paper presented at the Working Notes of CLEF 2017 – Conference and Labs of the Evaluation Forum, Dublin, Ireland.

Kanoulas, E., Li, D., Azzopardi, L., & Spijker, R. (2018). *CLEF 2018 technologically assisted reviews in empirical medicine overview.* Paper presented at the Working Notes of CLEF 2018 – Conference and Labs of the Evaluation Forum, Avignon, France.

Kanoulas, E., Li, D., Azzopardi, L., & Spijker, R. (2019). *CLEF 2019 technology assisted reviews in empirical medicine overview.* Paper presented at the Working Notes of CLEF 2019 – Conference and Labs of the Evaluation Forum, Lugano, Switzerland.

Kavanagh, J., & Rich, M. (2018). *Truth decay – an initial exploration of the diminishing role of facts and analysis in American public life*. Santa Monica: RAND Corporation.

Kim, S., Noveck, H., Galt, J., Hogshire, L., Willett, L., & O'Rourke, K. (2014). Searching for answers to clinical questions using Google versus evidence-based summary resources: A randomized controlled crossover study. *Academic Medicine, 89*, 940–943.

Kitchens, B., Harle, C., & Li, S. (2014). Quality of health-related online search results. *Decision Support Systems, 57*, 454–462.

Koopman, B., Zuccon, G., & Bruza, P. (2017). What makes an effective clinical query and querier? *Journal of the American Society for Information Science and Technology, 68*, 2557–2571.

Laurent, M., & Vickers, T. (2009). Seeking health information online: Does Wikipedia matter? *Journal of the American Medical Informatics Association, 16*, 471–479.

Lesk, M. (2005). *Understanding digital libraries* (2nd ed.). San Francisco: Morgan Kaufmann.

Li, H. (2011). *Learning to rank for information retrieval and natural language processing*. San Rafael: Morgan & Claypool Publishers.

Li, Z., Zhang, X., Müller, H., & Zhang, S. (2018). Large-scale retrieval for medical image analytics: A comprehensive review. *Medical Image Analysis, 43*, 66–84.

Lindberg, D., & Humphreys, B. (2005). 2015 – the future of medical libraries. *New England Journal of Medicine, 352*, 1067–1070.

Liu, Y., & Wacholder, N. (2017). Evaluating the impact of MeSH (Medical Subject Headings) terms on different types of searchers. *Information Processing & Management, 53*, 851–870.

Mani, I. (2001). *Automatic summarization*. Amsterdam: John Benjamins.

Marcetich, J., Rappaport, M., & Kotzin, S. (2004). *Indexing consistency in MEDLINE.* Paper presented at the MLA 04 Abstracts, Washington, DC.

McCool, J. (2017, April 6). Opinion: Why I published in a predatory journal. *The Scientist*. Retrieved from http://www.the-scientist.com/?articles.view/articleNo/49071/title/Opinion%2D%2DWhy-I-Published-in-a-Predatory-Journal/.

McGlynn, E., Asch, S., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., & Kerr, E. (2003). The quality of health care delivered to adults in the United States. *New England Journal of Medicine, 348*, 2635–2645.

McKibbon, K., & Fridsma, D. (2006). Effectiveness of clinician-selected electronic information resources for answering primary care physicians' information needs. *Journal of the American Medical Informatics Association, 13*, 653–659.

Merson, L., Gaye, O., & Guerin, P. (2016). Avoiding data dumpsters – toward equitable and useful data sharing. *New England Journal of Medicine, 374*, 2414–2415.

Miles, W. (1982). *A history of the national library of medicine: The nation's treasury of medical knowledge*. Bethesda: U.S. Department of Health and Human Services.

Moher, D., & Moher, E. (2016). Stop predatory publishers now: Act collaboratively. *Annals of Internal Medicine, 164*, 616–617.

Mork, J., Aronson, A., & Demner-Fushman, D. (2017). 12 years on – is the NLM medical text indexer still

useful and relevant? *Journal of Biomedical Semantics, 2017*(8), 8. Retrieved from https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-017-0113-5.

Müller, H., & Unay, D. (2017). Retrieval from and understanding of large-scale multi-modal medical datasets: A review. *IEEE Transactions on Multimedia, 19*(9), 17099710. Retrieved from https://ieeexplore.ieee.org/document/7984864/.

Munn, Z., & Qaseem, A. (2018). Disappearance of the National Guideline Clearinghouse: A huge loss for evidence-based health care. *Annals of Internal Medicine, 169*, 648–649.

Mynatt, B., Leventhal, L., Instone, K., Farhat, J., & Rohlman, D. (1992). *Hypertext or book: Which is better for answering questions?* Paper presented at the Proceedings of Computer-Human Interface 92.

Nicholson, D. (2006). *An evaluation of the quality of consumer health information on Wikipedia* (Capstone). Oregon Health & Science University, Portland, OR. Retrieved from http://www.ohsu.edu/dmice/people/students/theses/2006/upload/Nicholson_CapstoneFinal06.pdf.

Ohno-Machado, L., Sansone, S., Alter, G., Fore, I., Grethe, J., Xu, H., et al. (2017). Finding useful data across multiple biomedical data repositories using DataMed. *Nature Genetics, 49*, 816–819.

Palotti, J., Hanbury, A., Muller, H., & Kahn, C. (2016). How users search and what they search for in the medical domain – understanding laypeople and experts through query logs. *Information Retrieval Journal, 19*, 189–224.

Paskin, N. (2006). *The DOI Handbook*. Oxford: International DOI Foundation.

Patashnik, E., Gerber, A., & Dowling, C. (2017). *Unhealthy politics: The battle over evidence-based medicine*. Princeton: Princeton University Press.

Pivovarov, R., & Elhadad, N. (2015). Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association, 22*, 938–947.

Pluye, P., & Grad, R. (2004). How information retrieval technology may impact on physician practice: An organizational case study in family medicine. *Journal of Evaluation in Clinical Practice, 10*, 413–430.

Rigden, D., & Fernández, X. (2020). The 27th annual Nucleic Acids Research database issue and molecular biology database collection. *Nucleic Acids Research, 48*, D1–D8. Retrieved from https://academic.oup.com/nar/article/48/D1/D1/5695332.

Roberts, P., Cohen, A., & Hersh, W. (2009). Tasks, topics and relevance judging for the TREC Genomics Track: Five years of experience evaluating biomedical text information retrieval systems. *Information Retrieval, 12*(1), 81–97.

Roberts, K., Simpson, M., Voorhees, E., & Hersh, W. (2015). *Overview of the TREC 2015 clinical decision support track.* Paper presented at the The Twenty-Fourth Text REtrieval Conference (TREC 2015) Proceedings, Gaithersbug, MD.

Roberts, K., Demner-Fushman, D., Voorhees, E., & Hersh, W. (2016a). *Overview of the TREC 2016 clinical decision support track.* Paper presented at the The Twenty-Fifth Text REtrieval Conference (TREC 2016) Proceedings, Gaithersburg, MD.

Roberts, K., Simpson, M., Demner-Fushman, D., Voorhees, E., & Hersh, W. (2016b). State-of-the-art in biomedical literature retrieval for clinical cases: A survey of the TREC 2014 CDS track. *Information Retrieval Journal, 19*, 113–148.

Roberts, K., Demner-Fushman, D., Voorhees, E., Hersh, W., & Bedrick, S. (2017). *Overview of the TREC 2017 precision medicine track.* Paper presented at the The Twenty-Sixth Text REtrieval Conference (TREC 2017) Proceedings, Gaithersburg, MD.

Robertson, S., & Walker, S. (1994). *Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval.* Paper presented at the Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland.

Rothenberg, J. (1999). *Ensuring the longevity of digital information*. Retrieved from http://www.clir.org/pubs/archives/ensuring.pdf.

Salton, G. (1991). Developments in automatic text retrieval. *Science, 253*, 974–980.

Salton, G., & McGill, M. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.

Salton, G., Fox, E., & Wu, H. (1983). Extended Boolean information retrieval. *Communications of the ACM, 26*, 1022–1036.

Scaffidi, M., Khan, R., Wang, C., Keren, D., Tsui, C., Garg, A., et al. (2017). Comparison of the impact of Wikipedia, UpToDate, and a digital textbook on short-term knowledge acquisition among medical students: Randomized controlled trial of three web-based resources. *JMIR Medical Education, 3*(2), e20. Retrieved from http://mededu.jmir.org/2017/2/e20/.

Silberg, W., Lundberg, G., & Musacchio, R. (1997). Assessing, controlling, and assuring the quality of medical information on the Internet: Caveat lector et viewor – let the reader and viewer beware. *Journal of the American Medical Association, 277*, 1244–1245.

Sim, I. (2016). Two ways of knowing: Big data and evidence-based medicine. *Annals of Internal Medicine, 164*, 562–563.

Simpson, M., Voorhees, E., & Hersh, W. (2014). *Overview of the TREC 2014 clinical decision support track.* Paper presented at the The Twenty-Third Text REtrieval Conference Proceedings (TREC 2014), Gaithersburg, MD.

Sollins, K., & Masinter, L. (1994). *Functional requirements for uniform resource names*. Retrieved from http://www.w3.org/Addressing/rfc1737.txt.

Somashekhar, S., Sepúlveda, M., Puglielli, S., Norden, A., Shortliffe, E., RohitKumar, C., et al. (2018). Watson for Oncology and breast cancer treatment recommendations: Agreement with an expert multidisciplinary tumor board. *Annals of Oncology, 29*, 418–423.

Sox, H. (2009). Medical journal editing: Who shall pay? *Annals of Internal Medicine, 151*, 68–69.

Spitzer, V., Ackerman, M., Scherzinger, A., & Whitlock, D. (1996). The visible human male: A technical report. *Journal of the American Medical Informatics Association, 3*, 118–130.

Stanfill, M., Williams, M., Fenton, S., Jenders, R., & Hersh, W. (2010). A systematic literature review of automated clinical coding and classification systems. *Journal of the American Medical Informatics Association, 17*, 646–651.

Swanson, D. (1988). Historical note: Information retrieval and the future of an illusion. *Journal of the American Society for Information Science, 39*, 92–98.

Taichman, D., Sahni, P., Pinborg, A., Peiperl, L., Laine, C., James, A., et al. (2017). Data sharing statements for clinical trials: A requirement of the International Committee of Medical Journal Editors. *Annals of Internal Medicine, 167*, 63–65.

Tang, P., Newcomb, C., Gorden, S., & Kreider, N. (1997). *Meeting the information needs of patients: Results from a patient focus group.* Paper presented at the Proceedings of the 1997 AMIA Annual Fall Symposium, Nashville, TN.

Taylor, H. (2010). *"Cyberchondriacs" on the Rise? Those who go online for healthcare information continues to increase*. Retrieved from Rochester, NY: http://www.harrisinteractive.com/vault/HI-Harris-Poll-Cyberchondriacs-2010-08-04.pdf.

van Rijsbergen, C. (1979). *Information retrieval*. London: Butterworth.

Voorhees, E. (2013). *The TREC Medical Records Track.* Paper presented at the Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics, Washington, DC.

Voorhees, E., & Harman, D. (Eds.). (2005). *TREC: Experiment and evaluation in information retrieval*. Cambridge, MA: MIT Press.

Voorhees, E., & Hersh, W. (2012). *Overview of the TREC 2012 Medical Records Track.* Paper presented at the The Twenty-First Text REtrieval Conference Proceedings (TREC 2012), Gaithersburg, MD.

Voorhees, E., & Tong, R. (2011). *Overview of the TREC 2011 Medical Records Track.* Paper presented at the The Twentieth Text REtrieval Conference Proceedings (TREC 2011), Gaithersburg, MD.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*, 1146–1151.

Weibel, S., & Koch, T. (2000). The Dublin Core Metadata Initiative: Mission, current activities, and future directions. *D-Lib Magazine, 6*. Retrieved from http://www.dlib.org/dlib/december00/weibel/12weibel.html.

Wennberg, J. (2010). *Tracking medicine: A researcher's quest to understand health care*. Oxford: Oxford University Press.

Wenzel, R. (2017). Medical education in the era of alternative facts. *New England Journal of Medicine, 377*, 607–609.

Westbrook, J., Coiera, E., & Gosling, A. (2005). Do online information retrieval systems help experienced clinicians answer clinical questions? *Journal of the American Medical Informatics Association, 12*, 315–321.

Wildemuth, B., deBliek, R., Friedman, C., & File, D. (1995). Medical students' personal knowledge, searching proficiency, and database use in problem solving. *Journal of the American Society for Information Science, 46*, 590–607.

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., van der Lei, J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, 160018. Retrieved from https://www.nature.com/articles/sdata201618.

Yaman, H., Yavuz, E., Er, A., Vural, R., Albayrak, Y., Yardimci, A., & Asilkan, O. (2016). The use of mobile smart devices and medical apps in the family practice setting. *Journal of Evaluation in Clinical Practice, 22*, 290–296.

Zarin, D., Tse, T., Williams, R., Califf, R., & Ide, N. (2011). The ClinicalTrials.gov results database--update and key issues. *New England Journal of Medicine, 364*, 852–860.

Zarin, D., Tse, T., Williams, R., & Rajakannan, T. (2017). Update on trial registration 11 years after the ICMJE policy was established. *New England Journal of Medicine, 376*, 383–391.

Zerhouni, E. (2004). NIH public access policy. *Science, 306*, 1895.

Zhai, C., & Lafferty, J. (2004). A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems, 22*, 179–214.

Clinical Decision-Support Systems

Mark A. Musen, Blackford Middleton, and Robert A. Greenes

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_24

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What are the key motivations for clinical decision support?
- What are typical design considerations when building a decision-support system?
- What are some ways in which developers of decision-support systems encode and represent clinical knowledge?
- What are some current standards in the HIT industry that facilitate the construction of decision-support applications?
- What are the current main areas of research and development in clinical decision-support systems?

In this chapter, we discuss information technology aimed at furnishing **clinical decision support** (CDS) – the process that "provides clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care" (Osheroff et al. 2007). CDS systems (often referred to as **CDSS**s) communicate information that takes into consideration the particular clinical context, offering situation-specific information and recommendations. Generally, we think of such systems as reasoning about the clinical situation and presenting their conclusions as recommendations to the user. **Information retrieval** systems that find relevant information from repositories of documents with high relevance to a specific patient and clinical context can also serve in this role, but do not themselves make patient-specific recommendations for care. CDS systems do not directly perform clinical decision making; they provide relevant knowledge and analyses that enable the ultimate decision makers—clinicians, patients, and health-care organizations—to develop more informed judgments; hence the importance of the word "support" in the term "CDS". (Closed-loop systems such as in implantable cardioverter defibrillators and other devices such as insulin pumps are exceptions to this model; decision support for these systems is not covered in this Chapter.) Ideally, CDSSs can be described in terms of the five *right*s that they aim to accomplish: "provide the right information, to the right person, in the right format, through the right channel, at the right point in workflow, to improve health and health-care decisions and outcomes" (Osheroff et al. 2012).

Systems that provide CDS may be considered in terms of three basic categories: (1) those that use information about the current clinical context to retrieve highly relevant online documents, as with so-called "**Infobuttons**" (introduced in ▶ Chap. 14); (2) those that are **intelligent systems** that provide patient-specific, situation-specific alerts, advice, reminders, physician order sets, or other recommendations for direct action; or (3) those that organize and present information in a way that facilitates problem solving and decision making, as in **dashboards**, graphical displays, documentation templates, structured reports, and order sets.

Systems that create order sets offer an ideal example of CDS, not only because they can facilitate decision making by providing an actionable set of recommendations such as a combination of orders, but also because they provide a mnemonic function by gathering together items that should be associated in a particular setting. Order sets also can enhance workflow by providing a means to select a group of relevant activities quickly. Not all CDS systems have the ability to optimize workflow, as we shall discuss. In fact, some CDS systems are subject to concern if they are poorly implemented, as they actually can impede workflow or usability.

This chapter provides a review of computer-based decision aids, emphasizing their role and adoption within the current health-care milieu of the United States while keeping an eye on global trends. It offers some thoughts on the nature of the decision-making process, then it provides a description of current implementation strategies and challenges, and it closes with a discussion of critical research questions

that must be addressed to ensure optimal effectiveness of CDS in clinical practice.

24

24.1 The Nature of Clinical Decision Making

If you ask lay people what the phrase "computers in medicine" means, they often describe a computer program that helps physicians to make diagnoses. Although computers play numerous important clinical roles, people have recognized, from the earliest days of computing, that computers might support health-care workers by helping these people to sift through the vast collections of possible diseases, findings, and treatments.

We can view nearly all the contents of this book as addressing clinical data and clinical decision making. In ▶ Chap. 2, we discussed the central role of accurate, complete, and relevant data in supporting the decisions that confront clinicians and other health-care workers. In ▶ Chap. 3, we described the nature of good decisions and the need for clinicians to understand the proper use of information if they are to be effective and efficient decision-makers. In ▶ Chap. 4 we introduced the cognitive issues that underlie clinical decision-making and that influence the design of systems for decision support. Subsequent chapters have mentioned many real or potential uses of computers to assist with such decision-making. Medical practice *is* medical decision-making, so most applications of computers in health care are intended to have a direct or indirect effect on the quality of health-care decisions. In this chapter, we bring together these themes by concentrating on methods and systems that have been developed specifically to assist health workers in making decisions.

By now, you are familiar with the range of clinical decisions. The classic problem of **diagnosis** (analyzing available data to determine the pathophysiologic explanation for a patient's symptoms) is only one of these. Equally challenging, as emphasized in ▶ Chaps. 3 and 4, is the **diagnostic process**—deciding which questions to ask, tests to order, or procedures to perform, and assessing the value of the results that can be obtained in relation to associated risks or financial costs. Thus, diagnosis involves deciding not only what is true about a patient, but also what data are needed to determine what is true. Even when the diagnosis is known, there often are challenging **management** decisions that test the physician's knowledge and experience: Should I treat the patient or allow the process to resolve on its own? If treatment is indicated, what should it be? How should I use the patient's response to therapy to guide me in determining whether an alternate approach should be tried or, in some cases, to question whether my initial diagnosis was incorrect after all? (In that sense, the response to treatment is also a type of diagnostic test.) Also, when a clinician and a patient are faced with alternative treatments, and they seek help to choose among them, the estimation of the chance for cure or the risk of death or of complications is an important decision-making activity. Lastly, many disease processes evolve over time, and evaluation and change in treatment must evolve with them, resulting in the need for guidelines for management that take into account the temporal aspects of prior states and activities in order to decide what to do next. Decision making also may involve integrating data from multiple providers, treatments, and responses to them, as well as increased use of personal sensors and apps to provide additional data, so that aspects of care coordination over time play an important role. This is beyond our scope, but the topic is addressed in part in ▶ Chaps. 21, 22 and 23.

Biomedicine is also replete with decision tasks that do not involve specific patients or their diseases. Consider, for example, the biomedical scientist who is using laboratory data to help with the design of her next experiment or the hospital administrator who uses management data to guide decisions about resource allocation in his hospital. In addition, new financial models for health-care payment or reimbursement based on value (roughly, patient outcomes and cost of care), rather than on fee-for-service calculations based on individual patient care activities, require that decisions be made on the basis of aggregate

data on outcomes for groups of patients to determine expected norms and to identify outliers. Although we focus on systems to assist with clinical decisions in this chapter, we emphasize that the concepts discussed generalize to many other problem areas as well. In ▶ Chap. 29, for example, we examine the need for formal decision techniques and tools in creating health policies. As we develop databases that can identify patients with specific diseases, with risks of complications, or in need of specific interventions such as screening tests or immunizations (see ▶ Chap. 18), **population management** can be used to provide a form of decision support for groups of patients. Some clinical decision support is also aimed directly at patients, in terms of alerts, reminders, or aids to interpretation of information, especially given increasing use of smartphone apps and connected sensors and devices; techniques for assessing prognosis and risk of alternative strategies should involve shared decision making between providers and patients, which is also an important area of activity.

In this chapter, we focus on decision aids for the provider in particular—the clinician seeing the patient at the point of care. The requirements for excellent decision-making fall into three principal categories: (1) accurate data, (2) pertinent knowledge, and (3) appropriate problem-solving, or clinical reasoning, skills.

The data about a patient must be adequate—both accurate and sufficiently comprehensive to include everything relevant for making an informed decision—but they must not be excessive (see ▶ Chap. 4). Indeed, a major challenge occurs when decision-makers are bombarded with so much information that they cannot process and synthesize the information intelligently and rapidly (see, for example, ▶ Chap. 21). Thus, it is important to know when additional data will confuse rather than clarify and when it is imperative to use tools (computational, visual, or otherwise) that permit data to be summarized for easier cognitive management (see ▶ Chap. 4). Operating rooms and intensive-care units are classic settings in which this problem arises; patients are monitored extensively, numerous data are collected, and decisions often have to be made on an urgent basis.

Access to data from all relevant sources is required but difficult to achieve in practice. Patients may be seen in different venues, including a primary care office (perhaps via a telecare visit), a specialist office, an emergency room, a laboratory or imaging facility, a hospital, an extended care facility, or they may be monitoring themselves at home. Each venue may be using different and incompatible EHRs or data repositories, terminologies, and data representation models. Access to these data and interoperability across data repositories may be limited. Typically, the data available are only those within a health system and its EHR–which may include affiliated office practices, clinics, emergency rooms, and hospitals, or by interaction with an available **health information exchange** (HIE) (see ▶ Chap. 15).

Equally important is the quality of the available data to input to a CDSS. In ▶ Chap. 2, we discussed imprecision in terminology, illegibility and inaccessibility of records, and other opportunities for misinterpretation of data. Similarly, measurement instruments or recorded data may simply be erroneous; use of faulty data can have serious adverse effects on patient-care decisions. Thus, clinical data often need to be validated.

Even good data are useless if we do not have the knowledge necessary to apply them properly. Decision-makers must have broad knowledge of medicine, in-depth familiarity with their area of expertise, and access to pertinent additional information resources. Their knowledge must be accurate, with areas of controversy well understood and questions of personal choice well distinguished from those where a more prescriptive approach is appropriate. Their knowledge must also be current; in the rapidly changing world of medicine, facts decay just as certainly as dead tissue does.

Good data and an extensive factual knowledge base still do not guarantee a good decision; good problem-solving skills are equally important. Decision-makers must know how to set appropriate goals for a task, how to reason about each goal, and how to make explicit

the trade-offs between costs and benefits of diagnostic procedures or therapeutic maneuvers. Skilled clinicians draw extensively on personal experience, and new physicians soon realize that good clinical judgment is based as much on an the ability to reason effectively and appropriately about what to do as it is on formal knowledge of the field or access to high-quality patient data. Thus, clinicians must develop a strategic approach to the selection and interpretation of diagnostic tests, understand ideas of sensitivity and specificity, and be able to assess the urgency of a situation. Similar issues relating to test or treatment selection, in terms of costs, risks, and benefits, must be understood. Awareness of biases and of the ways that they can creep into problem-solving also are crucial (see ▶ Chap. 3). Also, as noted above, patient preferences and concerns must be adequately addressed as part of the decision-making process. Thus, good communication and interaction with the patient is essential. This brief review of issues central to clinical decision-making serves as a fitting introduction to the topic of computer-assisted decision-making: Precisely the same topics are pertinent when we develop a computational tool for CDS. The program must have access to good data, it must have extensive background knowledge encoded for the clinical domain in question, and it must embody an intelligent approach to problem-solving that is sensitive to requirements for proper analysis, appropriate cost–benefit trade-offs, and efficiency.

24.2 Motivation for Computer-Based CDS

Since the 1960s, workers in biomedical informatics have been interested in CDS systems because of a desire both to improve health care and to understand better the process of medical decision-making. Building a computer system that attempts to process clinical data to offer situation-specific advice can provide insight into the nature of medical problem solving and can enable the creation of formal models of clinical reasoning. At the same time, construction of such systems offers obvious societal benefits if the computer programs can aid practitioners in their care of patients and can lead to better clinical outcomes. Although the more academic considerations have provided strong motivation for work in the area of computer-based decision aids over several decades, the recognition of the importance of CDSSs as practical tools has increased markedly in recent years as a result of the inexorable growth in health-care complexity and cost, as well as the introduction of new health-care legislation, regulatory initiatives, and payment incentives aimed at addressing these trends—which have all made the development and broad adoption of CDS technology a priority.

The twenty first century has seen changes in health-care practices that make the development of CDS technology particularly necessary. Computer-based CDS has taken on increasing urgency for four reasons: (1) increasing challenges related to knowledge and information management in clinical practice, thus increasing physician information needs, (2) the ubiquity of electronic medical records and the desire to enhance health care through the communication and integration of the relevant data, (3) the goal of delivering increasingly personalized health-care services—tailored to the patient's preferences for care and to his or her individual genome, and (4) the growing evidence that CDS can not only improve the quality of care delivered but also reduce costs of care. We consider these four reasons in the sections that follow.

We note another factor that is shaping new directions for CDS which we will explore further at the end of this chapter—the trend toward use of personal devices and apps, as well as more portable, distributed laboratory or analytic procedures, to provide a growing range of sources of data to incorporate into decision making. These devices and apps also provide an ability to track, monitor, and intervene early whenever needed, both interacting with the user/patient directly, and also for those conditions warranting it, with the provider.

24.2.1 Physician Information Needs and Clinical Data Management

Modern health care is characterized by an ever-expanding knowledge base of clinical medicine, and by a growing clinical data base describing every patient characteristic from phenotype to genotype (Kohn et al. 2002). Despite the growing amounts of data and knowledge with which physicians need to work, health-care practitioners have seen the average time for a clinical encounter steadily curtailed, particularly in the United States, where the pressures of the prevalent fee-for-service reimbursement system and a concomitant rise in the amount of paperwork required for administrative management and billing continue to squeeze practitioners (Baron 2010). Studies of information needs among physicians in clinical practice have long revealed that unanswered clinical questions are common in ambulatory clinical encounters, with as many as one or two unanswered clinical questions about diagnosis, therapy, or administrative issues arising in every visit (Covell et al. 1985). Prior to the broad adoption of EHRs, in as many as 81% of clinical encounters in ambulatory care, clinicians were found to be missing critical information, with an average of four missing items per case (Tang et al. 1994, 1996). Currently, even with an EHR, providers may face major challenges in accessing relevant information, acquiring a complete picture of the patient's clinical state and history, and knowing what further testing or therapeutic actions are best to take. Prior to broad EHR adoption, studies suggested that as many as 18% of medical errors might be due to inadequate availability of patient information (Leape 1994). Today, conversely, the overabundance of information in the EHR can lead to 'information chaos' and may make it difficult for the clinician to find relevant information for clinical decision making (Melnick et al. 2019; Beasley et al. 2011). The demands for increased information management – in the setting of an ever-expanding clinical knowledge base, and now as more data sources are available and use of EHRs is almost ubiquitous – are primary drivers for the adoption of CDS systems. (See ▶ Chap. 23 for a deeper discussion of physician information needs.)

24.2.2 EHR Adoption and Integration of CDS

The motivations for adoption of EHRs and CDS are affected by the way in which health care is financed and paid for, by the structure and organization of the health-care system of a nation or a region, and by political forces that can create constraints, regulations, and incentives. In this section, we use the United States as an example to demonstrate these influences.

Health-care safety and quality concerns, coupled with a seemingly inexorable rise in health-care costs, have led in the United States to a variety of cost-containment and quality-improvement strategies in recent years. Health-care delivery in the United States is in the midst of a profound transformation, in part due to Federal public policy efforts to encourage the adoption and use of health information technology (HIT). The American Recovery and Reinvestment Act (ARRA) of 2009, and the **HITECH regulations** within it, created incentives for the widespread adoption of health information technologies (Blumenthal 2009; see ▶ Chap. 29). These public policy efforts, while ultimately suffering some reversals for political reasons, are often viewed as a long-term adjunct to current health-care–payment reform efforts in the U.S., and a prelude to additional health-care–delivery redesign, payment reform, and cost containment. As recently as 2012, only 34.8% of physicians in ambulatory practice in the US used a basic or comprehensive electronic medical record (Decker et al. 2012), and only 26.6% of U.S. hospitals used health information technologies in inpatient care-delivery settings (DesRoches et al. 2012), although these numbers have had a rapid upward trajectory.

The ARRA and HITECH policies, and the resulting technology adoption, changed the practice of medicine and clinical care delivery in both beneficial and untoward ways (Sittig and Singh 2011).

To achieve meaningful and effective use of HIT, the software must be viewed as one component of a complex sociotechnical system, in which all elements must work effectively (Institute of Medicine 2011a). The movement toward value-based reimbursement, a more recent U.S. trend, encourages emphasis on wellness, prevention, and early intervention in disease processes, and requires a new level of connectivity, continuity, and coordination of care across multiple venues is gaining in acceptance. A prominent example is the advent of Accountable Care Organizations (ACOs; McClellan 2015), which require emphasis on not only integration of data from all sources (clinical and financial) but also aggregate, population-based data on outcomes, costs, and identification of outliers. These alternative payment models are distributing upside and downside financial risk based upon quality outcomes, patient experience, and costs, and thus call for CDSSs to support both payers and providers.

One of the principal motivations for EHR adoption is to provide an infrastructure with which to improve the quality, safety, and efficacy of health-care delivery. In the past decade, the U.S. government placed considerable emphasis on the adoption of quality measures and quality-reporting requirements as part of meaningful use of HIT (Clancy et al. 2009; Institute of Medicine 2011a). Quality measures, despite their ability to provide feedback that stimulates improved performance by the clinician, are only part of the process needed to make the desired improvements. Prospective, proactive clinical decision support must also be in place. The U.S. government's rules for **meaningful use** of HIT, which became progressively more demanding over a 4–6 year period, required only minimal CDS compliance in phase I and II, but Phase III of the meaningful use regulations in 2016 was intended to increase the mandate for CDS in EHR systems substantially by requiring APIs (application program interfaces) for access to EHR data (Blumenthal and Tavenner 2010; Adler-Milstein et al. 2017). More recently, the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA),[1] the Twenty-first Century Cures Act of 2016 in the United States,[2] and other initiatives have promoted a wide array of additional enhancements and improvements to the use of health information technology in clinical practice. Notably, the Twenty-first Century Cures Act removed from U.S. Food and Drug Administration (FDA) consideration as a medical device software devoted to performing the functions of an EHR, administrative tools, or providing clinical decision support. Significant pushback to the burden on both EHR vendors and on health-care organizations to comply with Meaningful Use Phase III resulted in relaxation of these constraints. Meaningful Use regulations have been superseded in the United States by the Cures Act, promoting more sweeping goals for interoperability and access to quality measures (Sinsky and Privitera 2018; see ► Chap. 29). More recently, political considerations regarding health-care financing in the United States have made the speed of adoption of such measures somewhat uncertain.

From a more worldwide perspective, interoperability of data and of CDS knowledge models are essential for wide adoption, as well as for improving the evidence base from which knowledge is derived.

24.2.3 Precision Medicine

The fundamental model for the practice of medicine has undergone dramatic change in the past century or so. The objectives of clini-

1 MACRA (2015). The Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). Retrieval February 19, 2020: ► https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/value-based-programs/macra-mips-and-apms/macra-mips-and-apms.html

2 Twenty-First Century Cures Act. H.R. 34, 114th Congress (2016). Retrieval February 19, 2020: ► https://www.gpo.gov/fdsys/pkg/BILLS-114hr34enr/pdf/BILLS-114hr34enr.pdf

cal care have shifted radically from the archaic goal of correcting putative imbalances of bodily humors to the scientific understanding of pathophysiology, of mechanisms for eliminating pathogens, and of remedying biological aberrancies. The resulting view of medicine as the application of biological principles was at the core of the report produced by Abraham Flexner (1910) that upended medical education in the early twentieth century and that had led to the **reductionist biomedical model** that prevailed for the rest of that century. More recently, however, George Engel's **biopsychosocial model** (Engel 1977) brought to the fore of clinical care the need to address psychological and social factors in clinical treatment plans in addition to underlying biomedical problems. By the end of the twentieth century, it became increasingly accepted that CDS requires not only the rapid communication of appropriate scientific medical knowledge, but also the adaptation of that knowledge to reflect the psychological and social situation that would temper the application of the knowledge. Added to this complexity is the aging of the population, owing in part to advances in health and health care. The result has been a much higher burden of chronic diseases, multiple diseases, and multiple testing and treatment options, with both their positive and negative consequences that must be balanced—all contributing to the increasing intricacy of care. Indeed, the view of care is so complex for some patients that a major role of CDS is to integrate models of patient state and context (provider, setting, specialization, and activity) to provide selective visualization, analysis, and decision-making support for optimal care management (Greenes et al. 2018).

As a further extension of these trends, the genomic era in which we now live has further increased the need for clinical practice to reflect **precision medicine** and the need to tailor care to individual factors in ways that never before were imaginable (Ginsburg and Willard 2009). Precision medicine is characterized by decision making that may take into account patient personal history, family history, social and environmental factors, along with genomic data and patient preferences regarding their own care (see ▶ Chap. 26). In this approach, clinical decision making is explicitly patient-centered in new ways, bringing the best evidence at the genetic level to bear on many clinical scenarios, while incorporating patient preferences for acquiring and applying genetic information (Fargher et al. 2007; Marcial et al. 2018). The increasing use of genomics in medicine (Chan and Ginsburg 2011) is generating data that outstrip the information and knowledge-processing capabilities of practitioners, and many clinicians feel threatened by the impending tsunami of additional knowledge that they will need to master (Baars et al. 2005). As precision medicine becomes the norm, primary-care and specialist practitioners alike will need to manage their patients by interpreting genomic tests along with myriad other data at the point of care. It is hard to imagine how clinicians will manage to perform such activities without substantial computer-based assistance. Informatics is well suited to support a personalized approach to clinical genomics (Ullman-Cullere and Mathew 2011).

As mentioned earlier, another related change is growing recognition of the importance of promoting optimal health and wellness, not just by treating disease but by encouraging healthy lifestyles, fostering compliance with health and health-care regimens, and carrying out periodic health-risk assessments. Key to prescriptive medicine are tools to support *prospective medicine* (Langheier and Snyderman 2004)—assisting the acquisition of a detailed family history, social history, and environmental history, using personal apps and sensors, providing health-risk assessments, and managing genomic information (Hoffman and Williams 2011; Overby et al. 2010).

24.2.4 Savings Potential with Health IT and CDS

CDS has been shown to influence physician behavior (Colombet et al. 2004; Lindgren 2008; Schedlbauer et al. 2009), diagnostic test ordering and other care processes (Bates and Gawande 2003; Blumenthal and Glaser

2007), and the costs of care (Haynes et al. 2010), and it may have a modest impact on clinical outcomes (Bright et al. 2012). While there is enormous promise for HIT and CDS, their implementation is not without potential peril: HIT poorly designed or implemented, or misused, can generate unintended consequences (Ash et al. 2007; Harrison et al. 2007; Bloomrosen et al. 2011), and introduce new types of medical errors (Institute of Medicine 2011a).

Only a handful of studies have examined the **return on investment** (ROI) for HIT, and even fewer have investigated ROI for decision-support specifically. The value of CDS in terms of ROI is difficult to measure. Isolated studies of various hand-crafted systems in academic centers have shown value (Wang et al. 2003; Kaushal et al. 2006), but adoption elsewhere has often been problematic. Broad adoption has not occurred, for many reasons discussed later in this chapter, including the proprietary nature of systems for CDS and for representation of knowledge, the lack of interoperability of data and knowledge, the mismatch of CDS to workflow, and usability concerns.

Systematic reviews of the scientific literature, such as the one performed by Bright et al. (2012), have not been able to demonstrate an effect of CDS on patient outcomes except in the short term. This finding is not surprising, however, because the time point at which CDS occurs is often long before a final outcome, and many intervening factors may have a greater effect. In the case of CDS for many chronic diseases whose complications ensue over years or decades, it simply may be impractical to continue longitudinal studies of sufficiently long duration to be able to measure meaningful differences in outcome. Nevertheless, economic simulation studies of the potential effect of CDS on chronic diseases have demonstrated benefit in the long term (McClellan 2015; Bu et al. 2007; Adler-Milstein et al. 2007).

Historically, the adoption of CDS technology has been motivated by a virtuous desire to enhance the performance of clinicians when dealing with complex situations. The recent advent of legal, regulatory, and financial drivers, as well as the increasing importance of personalizing medical decision making on the basis of genomic data, now make CDS an essential element of modern clinical practice.

24.3 Methods of CDS

As we have already noted, CDS systems (1) may use information about the current clinical context to retrieve pertinent online documents; or (2) they may provide patient-specific, situation-specific alerts, reminders, physician order sets, or other recommendations for direct action; or (3) they may organize information in ways that facilitate decision making and action. Category (2) largely consists of the various computer–based approaches ("classic" CDS based on the application of intelligent systems) that have been the substrate for work in informatics since the advent of applied work in probabilistic reasoning and artificial intelligence in the 1960s and 1970s. Such systems provide custom-tailored assessments or advice based on sets of patient-specific data. They may follow simple logics (such as algorithms), they may be based on decision theory and cost–benefit analysis, or they may use probabilistic approaches or derive their conclusions on the basis of machine learning from large amounts of data. Some diagnostic assistants (such as DXplain; Barnett et al. 1987) suggest differential diagnoses or indicate additional information that would help to narrow the range of etiologic possibilities. Other systems suggest a single best explanation for a patient's symptomatology. Other systems interpret and summarize the patient's record over time in a manner sensitive to the clinical context (Shahar and Musen 1996). Still other systems provide therapy advice rather than diagnostic assistance (Musen et al. 1996).

CDS systems can achieve their results using a wide variety of computational methods. These approaches include Bayesian probabilistic reasoning (see ▶ Chap. 3), the use of machine learning to make predictions based on large amounts of data (often

from the EHR), performing inference using IF–THEN rules, or by the identification of relevant templates for the clinician to fill in, such as knowledge-based groupings of physician orders (order sets), or some combination of these approaches (James et al. 2013). CDS systems may acquire the data on which they base their recommendations interactively from users or directly from a health information system (or some combination of these approaches). We now discuss the issues that drive CDS system design, and we highlight how these issues are manifest in current clinical decision aids.

24.3.1 Acquisition and Validation of Patient Data

As mentioned in the introduction to this chapter, a prerequisite to any decision-making process is having available all the data that are needed to perform the required actions. As emphasized in ▶ Chap. 2, few problems are more challenging than the development of effective techniques for capturing patient data accurately, completely, and efficiently. You have read in this book about a wide variety of techniques for data acquisition, ranging from keyboard entry, to speech input, to methods that separate the clinician from the computer (such as scannable forms, real-time data monitoring, and intermediaries who transcribe written or dictated data for use by computers).

The problems of data acquisition go beyond entry or extraction from the EHR, or from other repositories, of the data themselves, however. A primary obstacle is that we lack standardized ways of expressing most clinical situations in a form that computers can interpret. As discussed in detail in ▶ Chap. 7, there are several controlled medical terminologies that health-care workers use to specify precise diagnostic evaluations (e.g., the International Classification of Diseases and SNOMED CT), clinical procedures (e.g., Current Procedural Terminology and LOINC codes), drugs (e.g., RxNorm), and so on. Still, there is no controlled terminology that can capture all the nuances of a patient's history of present illness or findings on physical examination. There is no coding system that can reflect all the details of physicians' or nurses' progress notes. Given that much of the information in the medical record that we would like to use to drive decision support is not available in a structured, machine-understandable form, there are clear limitations on the data that can be used to assist clinician decision-making. The prose of progress notes, consultation notes, procedure or operation reports, discharge summaries, and other documents contains an enormous amount of information that never makes it to the coded part of the EHR. Nevertheless, even when computer-based patient records store substantial information only as free-text entries, the data that are also available in coded form (typically, diagnosis codes and prescription data) can be used to significant advantage (van der Lei et al. 1991).

The desire to access information from the EHR that may be available only in text has been a topic of great interest to the CDS community. Some information systems provide options for **structured data entry**, asking clinicians to use fill-in-the-blanks forms or templates on the computer screen to enter patient-related information that otherwise would be entered as part of a prose note. In general, providers have resisted such human–computer interfaces, often finding it restrictive and cumbersome to make selections from predefined menus when they would much rather express themselves more freely in prose. In fact, structured templates or methods for collecting information about a particular problem or finding are themselves often regarded as a form of CDS, in that they provide an organized framework and a mnemonic function. Fortunately, work in **natural language processing** has made major advances in recent years, making it increasingly possible to mine the textual notes of EHRs to identify information that might bear on the CDS process (see ▶ Chap. 8).

24

24.3.2 Decision-Support Methodologies

When designing a CDS system, it is helpful to consider several aspects (Fig. 24.1). We consider these aspects as components interacting with one another, because each requires thought and effort to accomplish, and can be facilitated by developing standard, interoperable approaches for implementing them. Further, they can be independently enhanced over time, and are able to be shared and reused if engineered in a component-based manner. Much CDS can be accomplished by direct implementation and embedding into a clinical system, but the result is a set of one-off implementations that cannot be readily maintained, updated, or shared, especially of concern given that an organization may have hundreds, if not thousands, of CDSS artifacts in operation.

There are five aspects of CDS that, at least in principle, can be viewed as independent of each other, and that work together to produce CDS capability (Greenes 2014). These aspects include: (1) the method of computation or inferencing or, more generally, execution of the CDS function (e.g., targeted information retrieval, hard-coded algorithms, Bayesian estimation, neural-network classification, rule-logic evaluation); (2) the knowledge needed to carry out the function (e.g., prior and conditional probabilities, rule assertions, mathematical formulae, clinical guidelines); (3) the information model that governs how data are provided to the CDSS (i.e., the data needed and the method of encoding, such as specific FHIR data from the clinical setting, laboratory-test results encoded in LOINC, environmental data, or medical facts in specific coding schemes); (4) the type of recommendation to be provided (e.g., prediction,

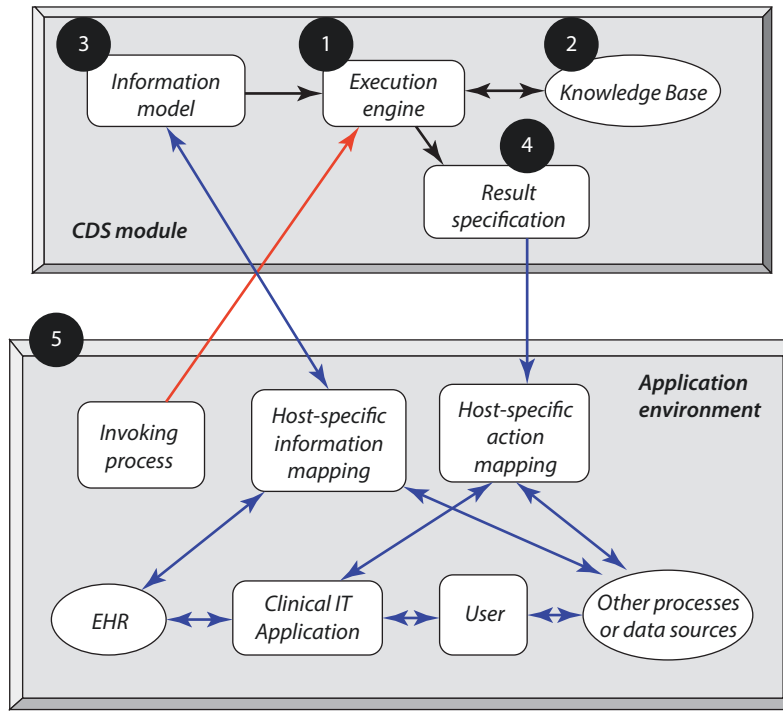


Fig. 24.1 A conceptual model of CDS components. There are five aspects of CDS that work together to produce CDS capability. The aspects are: (1) the method of computation or inferencing; (2) the knowledge needed to carry out the computation; (3) the information model for the data that drive decision making; (4) the type of recommendation to be provided; and (5) how the process interacts with the application environment, including how it is invoked and how the data and recommendations are communicated. (Adapted from Greenes 2014)

action, classification, relevant citations); and (5) how the process interacts with the application environment, including how it is invoked (e.g., a process launched by an event monitor, by the user explicitly, or by being embedded in workflow) and how the data and recommendations are communicated (e.g., provided via a FHIR API or delivered as a popup alert). In this section, we focus primarily on the methods of computation required for CDS (Aspect 1 of ◘ Fig. 24.1), and to a lesser extent on the knowledge needed during computation (Aspect 2). We consider the other elements in subsequent sections.

Aspect 1 was central to much of the early work on CDS development, beginning in the 1960s and largely led by academic centers, which focused on developing and exploring different computational models of CDS in the absence of any health IT infrastructure in which to embed the systems. In recent years, much of CDS has been built by vendors of proprietary systems, giving rise to a range of different approaches, often embedded in those systems, and with methods that are less able to be inspected, formalized, and shared. Some of the trends in health-care delivery that we cited in ▶ Sect. 26.2 are hoped to give rise to increased interoperability and sharing in vendor-developed systems. Even if current commercial products may not manifest the different components of ◘ Fig. 26.1 distinctly, we adopt this conceptual focus in our discussion of various strategies for CDS to clarify the underlying principles.

24.3.2.1 Context-Specific Information Retrieval

Early developers of CDS systems argued that decision support entails more than identifying relevant information that can help a clinician to solve a problem; it was believed that a CDS system has to suggest specifically how the problem should be solved. Thus, simply providing information for the user to read and digest was not considered "real" decision support. Nevertheless, as information-retrieval systems have improved over the years—with better performance characteristics—it is often hard to maintain that such systems do not offer decision support in a genuine sense. Indeed, such systems are now ubiquitous in health care. (▶ Chap. 25 provides a comprehensive discussion of information-retrieval methods.)

The simplest, and perhaps most common, form of CDS uses contextual information from an EHR to perform information retrieval from a database of information about online documents. A person viewing data in an EHR may see selectable icons (**infobuttons**) next to the names of drugs, laboratory tests, patient problems, or other elements of the patient record, or the items themselves may be hyperlinked to an information retrieval engine. Clicking on an infobutton causes the clinical information system to perform a query on the database, providing the user with one or more immediately accessible resources that can offer more information about the item in question. Alternatively, the system may automatically query one or more of those external resources and return the results of the queries for display (Cimino et al. 2002). Clicking on an infobutton next to a drug, for example, might allow the user to access information about customary dosing, side effects, or alternative medications (see ◘ Fig. 14.15 in ▶ Chap. 14). The query that retrieves the links to the documents is tailored based on whatever is next to the infobutton icon on the screen. The query may also take into account contextual information, such as patient-related data, the activity in which the user is engaged, and the role of the user in the health-care enterprise (physician, nurse, patient, and so on).

An **infobutton manager** mediates the queries between the clinical information system and the available information resources. The standards development organization Health Level Seven has created a standard for "context-aware knowledge retrieval," leading to infobutton managers that have been adopted by many commercial EHR vendors.[3]

3 HL7 International (2014). HL7 Version 3 Standard: Context Aware Knowledge Retrieval Application ("Infobutton"), Knowledge Request, Release 2. Retrieval February 19, 2020: ▶ https://www.hl7.org/implement/standards/product_brief.cfm?product_id=208

Infobutton managers need to anticipate how the clinical context might tailor the specific query performed by any given infobutton, so that the result of the query is highly precise and relevant to the situation at hand. Detailing specifically how contextual information might alter the queries performed by each infobutton type can be tedious, and the process requires developers to be adept at second-guessing all the reasons that might cause a user to click on a particular infobutton. Current research concentrates on the development of a Librarian Infobutton Tailoring Environment (LITE; Jing et al. 2015) that promises to aid the authoring of infobutton queries via "wizards" and other user-interface conveniences.

Although infobuttons are unquestionably important knowledge resources, many people would argue that they are not true CDS systems. Infobuttons retrieve relevant information for a user, but they do not explicitly address particular *decisions* that the user needs to make. The possible reasons that a user might click on an infobutton are folded into the query specification at the time that the infobutton is created; at runtime, of course, there is no way for the system to know exactly why the user selected the infobutton. Infobutton managers therefore require sophisticated query capabilities, but they do not need to reason from a clinical situation to a particular recommendation.

When the goal is to generate a situation-specific recommendation regarding diagnosis or therapy, developers need to turn to methods that can perform some kind of inference. The sophistication of the required technique is a function of the kind of inference that is necessary to render a result for the user.

24.3.2.2 Organizing or Grouping Information as a CDS Method

As we have noted earlier, a valuable kind of CDS is to organize information to provide a ready collection of items that need to be considered together (e.g., order sets for particular clinical indications or settings, documentation templates for particular purposes, or structure and formatting of reports). This capability serves not only as a convenience to the user, but also as a mnemonic function (i.e., offering a check list; Gawande 2009).

When patients are admitted to the hospital with a particular condition such as a suspected myocardial infarction or pneumonia, when hospital staff must prepare them for diagnostic procedures or surgery, or when they need to be transferred to another care team or to be discharged home, there often are stereotypical groups of medical orders that physicians tend to request. For example, patients with possible pneumonia often require a chest x-ray examination, the recording of vital signs at certain intervals, the administration of supplementary oxygen, cultures of their sputum or blood, and administration of antibiotics. When patients are admitted to a hospital with the diagnosis of pneumonia, the EHR can automatically suggest to the treating physicians that such a set of orders be considered. Systems that produce such order sets can use the clinical situation to tailor the recommended orders (e.g., the computer may not recommend a chest x-ray examination if the patient has just had one; knowledge that the patient has been placed on an artificial ventilator may trigger a separate set of associated orders for consideration).

Clinicians routinely face many other stereotypical tasks, such as describing the results of a diagnostic procedure such as an imaging study or reporting the sequence of events that took place during a surgical operation. Automated systems that recognize the clinical context can provide a tailored template for the clinician to fill in, helping to ensure that information is provided accurately and completely.

Throughout their professional activities, clinicians constantly must keep in mind large amounts of information when treating patients with even routine medical problems and reporting the results of their work. The use of simple mnemonics such as check lists can be remarkably effective in ensuring that care givers remember everything that they need to do or to report in a given context (Algaze et al. 2016; Alamri et al. 2016; Pageler

et al. 2014). Transforming such check lists into groups of orders to consider, groups of features to note in a diagnostic evaluation, or groups of steps that may have been followed when performing a surgical procedure can help clinicians to remember important details and to improve both the quality of medical care and what clinicians may report about it.

Computationally, systems that offer order sets or templates to clinicians typically assemble predefined information from collections of text strings or database entries. The HL7 organization has issued a standard for creating libraries of order sets. Systems that suggest specific order sets to users by making selections from such a library are generally embedded within the EHRs that call on their services, however. Thus, the specific methods that these systems use tend to be proprietary.

Researchers have experimented with the use of machine-learning techniques to create order sets on the fly based on empirical data. For example, Wang et al. (2018) have built a system that infers groups of orders for particular situations by examining a data warehouse for the individual orders that physicians have administered historically. The system then suggests that the groups of orders that it discerns from the database may be reasonable for clinicians to administer as an ensemble.

24.3.2.3 Hardcoding Clinical Algorithms

From a computational perspective, there is nothing simpler than encoding an algorithm directly in a computer program. In health care, there is generally nothing simpler than defining a decision process in terms of a flowchart. Numerous CDS systems thus have taken problem-specific flowcharts designed by clinicians and encoded them for use by a computer. Although such flowcharts have been useful for the purpose of triaging patients in urgent-care situations and as a didactic technique used in journals and books where an overview for a problem's management has been appropriate, computable interactive flowcharts have been largely rejected by physicians as too simplistic or generic for routine use (Grimm et al. 1975), other than for particular kinds of computations such as drug dose calculations (e.g., for pediatric patients). In addition, the advantage of their implementation on computers has not been clear; the use of simple printed copies of the algorithms generally has proved adequate for clinical care (Komaroff et al. 1974). A noteworthy exception that gained enormous attention in the early 1970s was a computer program deployed in Boston at what was then the Beth Israel Hospital (Bleich 1972); it used detailed algorithmic logic to provide advice regarding the diagnosis and management of acid–base and electrolyte disorders. More recently, such branching-logic and other inference methods approaches have been widely adopted in the administrative information systems that third-party payers use to process requests to pre-certify payment for expensive services such as MRI studies and elective surgery (the HL7 Da Vinci Project is a notable body of work in this area).

Although representing clinical algorithms simply as computer code offers a very direct approach to implementing a CDSS, there are obvious challenges that occur when it becomes necessary to refine or update the program's behavior. Every modification requires reprogramming the system. It may not always be obvious how to reprogram the system to render the desired behavior and making the necessary changes in one part of the program may have unintended consequences when other parts of the program execute. Thus, although it may seem appealing simply to hardcode clinical algorithms, developers of decision support systems generally seek more flexible mechanisms to represent clinical knowledge and to reason about it. We discuss this matter further in ▶ Sect. 24.4.5, and we address future research directions in this area in the Conclusion.

24.3.2.4 Learning from Data

Considerable flexibility is achieved when the CDSS is largely data-driven. The advent of extraordinarily fast computers and the ability to process enormous amounts of data has led to an explosion of interest in the use of

large datasets to learn patterns in the data to support clinical decision making (James et al. 2013). The success of such data-driven methods has led to considerable excitement about the use of "big data" in health care. It is important to appreciate, however, that workers in biomedical informatics benefited from such approaches long before computers could manage the enormous datasets that they process today.

Probabilistic Systems

Attempts to drive computer-based decision support from relationships inferred directly from data began during the earliest days of research in biomedical informatics. A seminal article in 1959 first introduced Bayes theorem and also value theory (later known as utility theory) into health-care decision making (Ledley and Lusted 1959). In the 1960s, workers in the field recognized that they could use computers to apply **Bayes' rule** to determine the posterior probability of diseases based on observations of patient-specific parameters (see ▶ Chap. 3). Such calculations were based on the determination of appropriate probabilistic relationships between findings and diseases by analyzing available datasets. Large numbers of **Bayesian diagnosis programs** have been developed in the intervening years, many of which have been shown to be accurate in selecting among competing explanations of a patient's disease state.

Among the most significant of the early experiments were those of F. T. de Dombal and his associates (1972) in England, who focused on the diagnosis of acute abdominal pain. De Dombal's group used a **naïve Bayesian model** that assumed that there are no conditional dependencies among findings (i.e., a model that makes the inappropriate but convenient assumption that the presence of a finding such as upper abdominal pain never affects the likelihood of the presence of a finding such as lower abdominal pain). Using surgical or pathologic diagnoses as the gold standard, de Dombal's group used sensitivity, specificity, and disease-prevalence data for various signs, symptoms, and test results to calculate, using Bayes' theorem, the probability of seven possible explanations for acute abdominal pain (appendicitis, diverticulitis, perforated ulcer, cholecystitis, small-bowel obstruction, pancreatitis, and nonspecific abdominal pain). To keep the Bayesian computations manageable, the program made the "naïve" Bayesian assumptions of (1) conditional independence of the findings for the various diagnoses, (2) mutual exclusivity, and (3) exhaustiveness of the seven diagnoses (see ▶ Chap. 3).

In one system evaluation (de Dombal et al. 1972), physicians filled out data sheets summarizing clinical and laboratory findings for 304 patients who came to the emergency department with abdominal pain of sudden onset. The data from these sheets provided the attributes that were analyzed using Bayes' rule. Thus, the Bayesian formulation assumed that each patient had one of the seven conditions and it selected the most likely one on the basis of the recorded observations.

In contrast to the clinicians' diagnoses, which were correct in only 65–80% of the 304 cases (with accuracy depending on the individual clinician's training and experience), the program's diagnoses were correct in 92% of cases. Furthermore, in six of the seven disease categories, the computer was more likely to assign the patients to the correct disease category than was the senior clinician in charge of the case.

De Dombal's system began to achieve widespread use—from emergency departments in other countries to the British submarine fleet. Surprisingly, however, the system never obtained the same degree of diagnostic accuracy in other settings that it did where it had initially been deployed—even when adjustments were made for differences in prior probabilities of disease. There are several reasons possible for this discrepancy, which are relevant for all Bayesian CDS systems. The most likely explanation is that there may be considerable variation in the way that clinicians interpret the data that must be entered into the computer. For example, physicians with different training or from different cultures may not agree on the criteria for identification of certain patient findings on physical

examination, such as "rebound tenderness."[4] Another possible explanation is that there are different probabilistic relationships between findings and diagnoses in different patient populations.

Although a naïve Bayesian model may have limitations in accurately modeling a diagnostic problem, a major strength of this approach is computational efficiency. When the findings that bear on a hypothesis are assumed to be conditionally independent, then the order in which the findings are considered in the Bayesian analysis does not matter. The computer starts by considering a given finding, the prior probability of each possible diagnosis under consideration (generally the prevalence of each diagnosis in the population), and the conditional probabilities of the finding (or the absence of the finding) given each diagnosis (or the absence of the diagnosis)—the *sensitivity* and *specificity* of the finding (see the discussion of these concepts in ▶ Chap. 2). The computer then applies Bayes' rule to calculate the posterior probability of each diagnosis given the value of the finding. The computer now is poised to update the probability of each diagnosis given the value of a second finding. The prior probability for each diagnosis in this case is not the prevalence of the diagnosis in the population, however. Having applied Bayes' rule once, we have more information than we had at the start. We can treat the *posterior probability* of each diagnosis given the first finding as the *prior probability* of the diagnosis when we apply Bayes' rule a second time. When it is time to consider a third finding, the posterior probability for each diagnosis after processing the second finding serves as the prior probability for the next application of Bayes rule. The process continues until the value of each finding has been considered. This **sequential Bayes** approach was explored as early as the 1960s for the diagnosis of congenital heart disease (Gorry and Barnett 1968) and has been used in many CDS systems since.

In recent years, the use of naïve Bayesian models has been seriously challenged by the adoption of systems based on Bayesian **belief networks** (see below), which can take advantage of efficient algorithms that overcome the limiting assumptions of naïve Bayesian approaches—albeit at the cost involved in creating a more complex (and more nuanced) model of the underlying probabilistic relationships. In addition, systems that adopt the sequential Bayes approach lack the ability to choose the next test to be applied in a manner that can optimize reasoning. Bayesian systems that use more sophisticated reasoning strategies based on decision analysis (see ▶ Chap. 3) can use utility theory to identify the test that will provide the most useful information given the current state of reasoning. Considerations of cost, discomfort to the patient, and the availability of the test can influence the utility of the particular choice.

Machine Learning

The availability of large biomedical datasets and computers and algorithms that can process huge amounts of data are revolutionizing health care and the life sciences. Machine learning is everywhere in health care, from interpreting radiographic images to predicting utilization of health care services to identifying potential adverse drug events. It is not surprising that decision-support based on machine-learning models is becoming increasingly common in clinical settings. Such systems have been in place for decades, but now they are assuming increasing prominence, given the new opportunities that the renaissance in machine learning is offering all of biomedicine in the era of "big data."

There are a host of **supervised learning techniques** that can determine how data are associated with hypotheses (James et al. 2013), and that consequently can be trained on EHR data to infer conclusions based on some set of input data. For example, the decision-support capabilities of the patient monitoring systems discussed in ▶ Chap. 21 often apply statistical methods to the current data stream to

4 *Rebound tenderness* is pain that is exacerbated when the physician presses down on the abdomen and then suddenly releases, generating a "rebound" when the abdomen returns to its baseline position.

infer corresponding classifications to inform care providers of the patient's current state. Regression analysis or more sophisticated techniques, such as artificial neural networks and support vector machines, when applied to routinely collected patient data, have enabled investigators to develop venerable decision aids such as the APACHE system (Knaus et al. 1991; Zimmerman et al. 2006), which offers prediction models providing prognostic information regarding patients in the ICU (see ► Chap. 21).

Recent work demonstrates the value of applying data-driven techniques to a wide range of clinical problems for which clinicians may benefit from decision support, from assessing newborns in the ICU (Saria et al. 2010) to development of models that can suggest which patients might benefit most from palliative care (Avanti et al. 2018). Scores of start-up companies have emerged in recent years, each hoping that different large datasets and specialized machine-learning techniques will lead to new insights about particular clinical problems in an effort to enhance decision making.

From the beginning, such machine-learning approaches have been criticized when used as the basis for CDS, primarily because of the lack of transparency in how data-driven methods reach their conclusions (Shortliffe et al. 1979). Because the associations between findings and diagnoses are inferred as the system is trained on the data and are not readily available for inspection, such systems cannot offer guidance as to *why* they might reach particular conclusions. This inability to explain the basis for their recommendations is especially important when the recommendations of a system might be overly fitted to the peculiarities of a dataset drawn from a patient population different from the one to which the system is being applied (as may have been the case with de Dombal's system). Because the output of a CDSS based on a machine-learning algorithm must be accepted at face value, there is typically no way to know what biases may exist in the data that trained the system or what clinically relevant intermediary states may have led the system to reach its final conclusion. There is currently intense interest in being able to develop new technologies powered by machine-learning methods that in some measure can explain the basis of their reasoning and that can allow users to assess the likelihood that particular recommendations are appropriate in their clinical setting (Ribeiro et al. 2016).

24.3.2.5 Declarative Representation of Knowledge

In simple Bayesian systems and in those CDSSs based on machine-learning algorithms, data are provided as input into the system, and the output is a classification of the data—often a diagnosis on which to act. Since the 1970s, however, workers in biomedical informatics have pursued the development of decision-support technologies that attempt to encode in a more explicit way how the inputs to the system relate to the outputs. The goal is to encode models—models of reasoning, models of pathophysiology, models of probabilistic relationships, models of the evidence in support of alternative treatment options, and other relevant models—in a way that forms the basis for a system's computation to derive an appropriate recommendation from the input data. Such systems are often built with the objective that the underlying models be examinable and explainable. Often, there is a desire to make those models editable, so that the models easily can be updated in light of new discoveries and new understanding. The unifying idea in this approach is that CDSSs are built with a computer-based representation of the knowledge that drives system behavior. There are many ways to represent knowledge in computers (Musen 2014), and each strategy has different strengths and weaknesses.

Bayesian Belief Networks

Much of the early interest in the naïve, sequential Bayesian approach stemmed from a conviction that it simply was impractical to construct Bayesian systems in which the assumption of conditional independence was lifted: There would be too many probabilities to assess when building the system, and the necessary computation could be intractable.

Work on the use of **belief networks**, however, has demonstrated that it actually is realistic to develop more expressive Bayesian systems in which conditional dependencies are modeled explicitly—often by taking advantage of approximate algorithms for concluding the posterior probabilities that are computationally efficient in most cases. (Belief networks are described in detail in ▶ Chap. 3.) Currently, many modern CDS systems that make recommendations based on probabilistic relationships use belief networks as their primary representation of the underlying clinical situation, and then "solve" the belief network at runtime to calculate the posterior probabilities of the conditions represented in the graph. The use of belief networks is popular because the formalism makes probabilistic relationships perspicuous, overcomes the assumption of conditional independence, and enables the attendant probabilities to be learned from analysis of appropriate data sets (for example, EHR data). The approach has been demonstrated in numerous diagnostic systems, from belief networks that ascertain the status of newborns from data in the neonatal ICU (Saria et al. 2010), systems for differential diagnosis (Shwe et al. 1991; Middleton et al. 1991), to belief networks that offer interpretations of biomedical image data (Kahn et al. 1997).

Because making most decisions in medicine requires weighing the costs and benefits of actions that could be taken in diagnosing or managing a patient's illness, researchers also have developed tools that draw on the methods of **decision analysis**. Decision analysis adds to Bayesian reasoning the idea of explicit decisions and of **utilities** associated with the various outcomes that could occur in response to those decisions (see ▶ Chap. 3). One class of programs for decision-analysis is designed for use by the analysts themselves; such programs are of little use to the average clinician or patient, however (Pauker and Kassirer 1981). A second class of programs uses decision-analysis concepts within systems designed to advise physicians who are not trained in these techniques. In such programs, the underlying decision models generally have been prespecified—either as decision trees that enumerate all possible decisions and all possible ramifications of those decisions or as belief networks in which explicit decision and utility nodes are added, called **influence diagrams** (Shachter 1986).

We say that belief networks and influence diagrams represent knowledge in a **declarative** manner, because a belief network provides an inspectable, editable model of the probabilistic relationships that are relevant to the decision problem under consideration. If a developer wants to designate a new relationship between two entities in the network, then she needs only to augment the model by adding a new edge to the graph that encodes the given network. The network thus provides a transparent mechanism to communicate what the system "knows" about the probabilistic relationships among the entities in the application domain, and changing the network intrinsically changes the behavior of the system when it reasons about those entities. This is different from the case of a hardcoded algorithm or a system based on machine learning, where the knowledge is not readily inspectable or editable in a direct manner.

Rule-Based Approaches

Although belief networks provide a convenient mechanism to encode knowledge about the world in a declarative fashion, they are only one of several alternative frameworks that may be used to drive CDS based on explicit models of the application area. Since the 1970s, workers in medical AI have been exploring the use of methods that emphasize the modeling of rules that describe conclusions that can be reached about the decision problem and the variables that may predicate those conclusions. Often called **knowledge-based systems**, these programs reason about the clinical situation by examining a collection of rules of the form, "If some set of conditions is true, then conclude that something else is true" (◘ Fig. 24.2). Although these rules may be created through machine-learning approaches, they often are built by manually encoding relationships between clinical data and corresponding conclusions that are

Rule 507

IF:

1) The infection that requires therapy is meningitis
2) Organisms were not seen on the stain of the culture
3) The type of infection is bacterial
4) The patient does not have a head injury defect, AND
5) The age of the patient is between 15 years and 55 years

THEN

The organisms that might be causing the infection are Diplococcus-pneuominae and Neisseria-meningitidis

Fig. 24.2 A rule from a rule-based system. Rules are conditional statements that indicate what conclusions can be reached or actions taken if a specified set of conditions is found to be true. This rule, taken from the CDSS known as MYCIN, is able to conclude probable bacterial causes of infection if the five conditions in the premise are all found to be true for a specific patient

offered by experts in the field or by examination of evidence reported in the scientific literature. When a knowledge-based system is encoded using rules, it is referred to as a **rule-based system** (Buchanan and Shortliffe 1984).

Rule-based systems provide an important mechanism for developers to build CDS capabilities into modern information systems. From CDS systems that interpret ECG signals to those that recommend guideline-based therapy, rules provide an extremely convenient means to encode the necessary knowledge. Rule-based systems require a formal language for encoding the rules, plus an interpreter (sometimes called an **inference engine**) that operates on the rules to generate the necessary behavior.

Perhaps the best-known rule-based CDSS is one that was never put into clinical use, but that has served as a prototype for the many rule-based systems that have followed. The program, known as MYCIN, combined a diagnostic component with an advisor component that suggested appropriate management of patients who have infections (Shortliffe 1976). MYCIN's developers believed that straightforward algorithms or probabilistic approaches were inadequate for this clinical problem in which the underlying knowledge was poorly understood and even the experts often disagreed about how best to manage specific patients, especially before definitive bacterial culture results became available. As a result, the researchers were drawn to the use of interacting rules to represent knowledge about organisms that might be causing a patient's infection and the antibiotics that might be used to treat it.

Knowledge of infectious diseases in MYCIN was represented as *production rules* (see Fig. 24.2). A production rule is an IF–THEN conditional statement. The conclusions drawn by one production rule may be used to satisfy the premises of other rules when a system of rules is used for reasoning by an inference engine. MYCIN's power was derived from such rules in a variety of ways:

- The MYCIN program determined which rules to use and how to chain them together to make decisions about a specific case. The MYCIN reasoning program used an approach called **backward chaining**; whenever a rule was being considered and the system did not know whether the condition on the left-hand side of the rule (i.e., the premise) was true, MYCIN would look backward to see whether the knowledge base contained any other rules that, when evaluated, could conclude information that might inform the evaluation of the current rule's premise. (Nearly all contemporary rule-based systems, on the other hand, use an inference method known as **forward chaining**: Whenever a production rule "fires" and the conclusion of that rule is proven to be true, the system looks forward for other rules in the rule base that use the concluded information in their premise, and that therefore also might be able to fire now that the new conclusion is known to be true.)
- The rules often formed a coherent explanation of MYCIN's reasoning—those that applied to the current decision were displayed in response to a user's questions (Fig. 24.3). Although rules were stored in a machine-readable format, English translations could be displayed.

By removing, altering, or adding rules, system developers could modify the program's knowledge structures rapidly, without explicitly reprogramming or restructuring other parts of the knowledge base. The rule base

In this series, how many of the cultures were positive for E.coli (ORGANISM-1)? *(typical system prompt)*

****WHY** *(request from user to determine motivation for question)*

[i.e. WHY is it important to determine the number of cultures which were positive for the organism?] *(system's rephrasing of the WHY question, and answer)*

[1.0] This will help to determine whether there is significant disease associated with this occurrence of the organism.
It has already been established that
[1.1] the culture was taken from a sterile source
[1.2] the number of cultures in the series is greater than 2
Therefore, if
[1.3] the number of cultures which were positive for ORGANISM-1 is greater than 2
Then there is suggestive evidence (.5) that there is significant disease associated with this occurrence of the e.coli (ORGANISM-1).
[RULE049]
****WHY** *(repeated WHY to determine relevance of previous explanation)*

[i.e. WHY is it important to determine if there is significant disease associated with this occurrence of the organism?]

[2.0] This will help to determine a diagnosis for which the infection should be treated.
It has already been established that
[2.1] the infection is not one of: primary bacteremia, meningitis
[2.2] for some current culture of the patient it is true that the site of the culture is blood
Therefore, if
[2.3] there is significant disease associated with this occurrence of the organism
Then
It is definite (1.0) that the diagnosis for which the infection should be treated is secondary-bacteremia
[RULE103]

Fig. 24.3 Two examples of MYCIN's explanation capabilities. User input is shown in boldface capital letters and follows the double asterisks. The system expands each ["WHY"] question (enclosed in square brackets) to ensure that the user is aware of its interpretation of the query

thus offered the kind of inspectability and editability that we discussed when considering belief-network representations of knowledge.

The developers evaluated MYCIN's performance on therapy selection for patients with blood-borne bacterial infections (Yu et al. 1979a), and for those with meningitis (Yu et al. 1979b). In the latter study, MYCIN gave advice that compared favorably with that offered by experts in infectious diseases—results that ushered in enormous excitement about the potential of rule-based systems to offer high-level clinical advice in real-world situations.

The developers of MYCIN had to construct their own syntax for encoding rules and had to program their own inference engine to evaluate the rules. However, now there are many open-source and proprietary **rule engines** that provide custom-tailored editors for writing rules and inference engines that can execute the rules at runtime. For example, JESS is a popular Java-based rule engine that can be licensed from Sandia National Laboratory and that currently is free for academic use. Drools is an open-source rule engine developed by the JBoss community that also has had substantial adoption.

Developers use JESS, Drools, and proprietary rule engines to create CDS systems that contain multiple rules that, as with MYCIN, can chain together to generate conclusions based on a sequence of inference steps. Decision support sometimes requires multiple rules to execute at runtime, together generating a final recommendation that derives from the consequences of the rules chaining off one another.

In most installed information systems, however, rule-based decision support is much simpler and also more limited. Most deployed CDS systems have rules that generally do not chain together, but that are triggered individually, each time that either there is a relevant change to the data in a patient database that should generate an alert, there is a time-related event that should trigger a reminder, or an action is performed (e.g., by a user, when the action is established in the workflow as a triggering event). Each rule examines the state of the database and generates a corresponding action, alert, or a freminder that is usually sent to a particular clinician or to members of the health-care team. Such rules are of the general form: Event – Condition – Action (ON Event, IF Condition, THEN Perform Action) and they are commonly referred to as **ECA rules**.

For example, **Arden Syntax** became an international standard for ECA rules known as **Medical Logic Modules** (MLMs), endorsed by HL7 and ANSI in 1999 (Fig. 24.4). Arden Syntax provides a standard mechanism for declaring the variables about whose values the system will perform its reasoning (values that derive from data in the clinical information system); the conditions that, if true, would predicate specific actions; the actions that should be taken, and the kinds of events that would invoke or trigger the rule. The standard was created with the hope that the informatics community would develop whole libraries of MLMs, all written in Arden Syntax, that could operate in any clinical environment where an information system could interpret the standard format.

A significant obstacle to the sharing of MLMs, however, is that Arden Syntax is, in fact, just a syntax. What is missing from the standard is any notion of the *semantics* of the data on which the MLMs operate. When an MLM executes, the variables that are used in the logic of the rule are bound to values that derive from the patient database of the information system in which the MLMs operate. Arden Syntax specifies that the individual database queries needed to determine the values of the variables should appear within the "curly braces" of variable definitions in the portion of the MLM known as the "data slot" (see Fig. 24.4). What a developer should include within the curly braces depends on the particular schema of the relevant patient database and mechanism for performing queries. EHR information models and the way in which elements are coded differ from system to system. Thus, all system-specific aspects of MLM integration need to be provided within the curly braces. To adapt an MLM for use in a new environment, a programmer needs to consider the variables on which the MLM operates, determine whether those variables have counterparts in the local patient database, and write an appropriate query that will execute at runtime.

The **curly braces problem** is compounded because there may be assumptions regarding the semantics of the variables themselves that may not be obvious to the local implementer: If the MLM refers to serum potassium, should the logic be executed if the original specimen was grossly hemolyzed?[5] If a serum potassium value is not available in the database, but there is a value for a whole-blood potassium, should the MLM be executed using that value instead?[6] If there is no serum potassium value available for today, but there is one from last night, should the logic execute using the most recent value? Decision rules cannot simply be dropped from one system into another and be shared effortlessly; rather, considerable thought, analysis, and computer skill needs to go into writing the appropriate database que-

5 If the red blood cells in a specimen *hemolyze* (burst), they release potassium, which can cause an inaccurate elevation in the measured potassium value.

6 The *serum* is the liquid that is left when the cells are removed from whole blood.

MAINTENANCE:

Title:	Diabetic Foot Exam Reminder;;
Mlmname:	Diabetic_Foot_Exam.mlm;;
Arden:	Version 2.8;;
Version:	1.00;;
Institution:	Intermountain Healthcare ;;
Author:	Peter Haug (Peter.Haug@imail.org) ;;
Specialist:	Peter Haug (Peter.Haug@imail.org) ;;
Date:	2011-11-28;;
Validation:	testing;;

LIBRARY:

Purpose:	Alert for Diabetic Foot Exam Yearly;;
Explanation:	This MLM will send an alert if the patient is a diabetic (diabetes in problem list or discharge diagnoses) and Foot Exam is recorded within the last 12 months.;;
Keywords:	diabetes; Foot Exam;;
Citations:	Boulton AJM, Armstrong DG, Albert SF, Frykberg RG,Richard Hellman, Kirkman MS, Lavery LA, LeMaster JW, Mills JL, Mueller MJ, Sheehan P, Dane K. Wukich DK. Comprehensive Foot Examination and Risk Assessment. Diabetes Care. 2008 August; 31(8): 1679–1685.;;
Links:	http://en.wikipedia.org/wiki/Diabetic_foot_ulcer;;

KNOWLEDGE:

Type:	data_driven;;
Data:	Problem_List_Problem := object [Problem, Recorder]; Problem_List := read as Problem_List_Problem {select problem, recorded_by from Problem_List_Table}; Patient_Dx_Object := object [Dx]; Diabetic_Dx := read as Patient_Dx_Object {ICD_Discharge_Diagnoses}; Foot_Examination := object [Recorder, Observation]; Observation := object [Abnormatlity, Location, Size, Units]; Foot_Exam := read as Foot_Examination latest {select Recorder, Observation.Abnormatlity, Observation.Location, Observation.Size, Observation.Units from PE_Table}; Registration_Event := event { registration of patient }; ICD_for_Diabetes := (250 , 250.0 , 250.1 , 250.2 , 250.3 , 250.4 , 250.5 , 250.6 , 250.7 , 250.8 , 250.9); ;;
Evoke:	Registration_Event;;

Logic:	if (Diabetic_Dx.Dx is in ICD_for_Diabetes or (exist Problem_List and "Diabetes" is in Problem_List.Problem)) then Diabetes_Present := true ;
	endif;
	
	if (Diabetes_Present and exist Foot_Exam and Foot_Exam occurred not within past 12 months) then
	conclude true ;
	endif;
	conclude false ; ;;

Action:	write "Patient is a diabetic with no Diabetic Foot Exam in last 12 months. Please order or perform one.";;

Fig. 24.4 This medical logic module (MLM), written in the Arden syntax, prints a warning for health-care workers whenever a patient who has diabetes is registered for a clinic visit and has not had a documented foot examination in the past year. The evoke slot defines a situation that causes the rule to be triggered; the logic slot encodes the decision logic of the rule; the action slot defines the procedure to follow if the logic slot reaches a positive conclusion. The data slot defines the variables that are to be used by the MLM; the text between curly braces must be translated into queries on the local patient database when the MLM is deployed locally. (Courtesy of P. J. Haug, Intermountain Healthcare, with permission)

ries that go within the curly braces to make such rules operational.

In the case of Arden Syntax, developers write rules to deal with one clinical problem at a time. There may be one MLM to deal with the problem of administering a drug like penicillin to a patient with a history of penicillin allergy; another MLM may report that a patient has a dangerously low serum potassium value. Unlike the rules in MYCIN, MLMs are generally not intended to interact with one another or to be chained together to generate complex inferences. MLMs may be coerced to chain together when one MLM posts to the patient database a value that can trigger another MLM. This mechanism also allows one MLM to set up information in the database that might invoke another MLM in the case of some future event, thus enabling the recommendation of actions that unfold over time, as in the case of many clinical practice guidelines for chronic diseases. Although this approach allows developers to program complex problem-solving behavior, the technique has the same disadvantages that came to light with chaining rule-based systems such as MYCIN: When the rule base grows to a large size, interactions among rules may have unanticipated side effects. Furthermore, when rules are added to or deleted from a previously debugged knowledge base, there may be unexpected system behaviors that emerge as a result (Clancey 1983; Heckerman and Horvitz 1986).

For MLMs to work well in practice, moreover, the rules need to be tailored to the particular clinical environment— triggered by appropriate workflow events, interacting with particular kinds of participants, customizing logic to account for various business and workflow processes, and notifying the user in setting-specific ways. To customize an MLM to account for such considerations requires that it become less portable. Much of the effort required to introduce CDS systems into the health-care enterprise involves precisely such adaptations. To accelerate portability, MLM developers must seek a balance between a generic specification of logic that is widely agreed upon, and site-specific customizations that will facilitate the use of that logic. Achieving the right balance will always remain an elusive target (see also ▶ Sect. 24.5.4).

More General Representations of Knowledge

The variety of approaches for building CDS systems that we have described so far suggests that each method has significant strengths and weaknesses. Hardcoded branching-logic systems can be very easy to build but difficult to update and maintain when the program code becomes complicated. Belief networks and influence diagrams offer precision in probabilistic reasoning when the goal is to make a classification of some clinical phenomena, but they offer limited capabilities if the goal is to generate a plan for medical therapy on the fly or to simulate some biomedical situation. Rule-based systems can help to decompose decision problems into tractable IF–THEN chunks, but they typically are limited in what these chunks can express about a clinical problem. All of these approaches are constrained in their inability to reason about clinical abstractions (e.g., knowing that elevated serum potassium is *a kind of* electrolyte abnormality) and by their inability to make inferences about situations that entail nuance or complexity.

In recent years, developers of CDS systems have become increasingly interested in the use of more general representations of clinical knowledge to provide more sophisticated capabilities for decision support. These interests have paralleled the surge of enthusiasm in the World Wide Web community for the notion of developing **knowledge graphs** that encode facts about the world in a lattice of nodes that represent the kinds of entities in the world and links between them that represent relationships among those entity types (Noy et al. 2019). Unlike belief networks, which typically contain dozens of nodes, knowledge graphs often contain hundreds or thousands or even millions of nodes—or more. Whereas belief networks store numeric, probabilistic relationships among the entities in the graph, knowledge graphs store symbolic, logical connections among the enti-

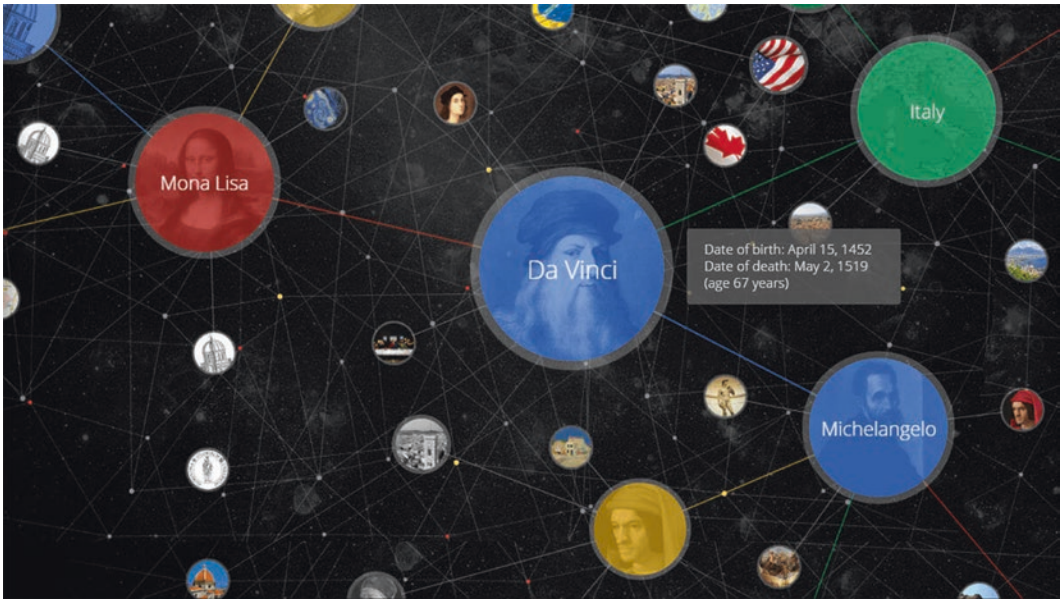


Fig. 24.5 Google depicts its knowledge graph artistically as a vast collection of nodes representing entities in the world, linked to other entities in a rich network. The graph allows Google's search engine to highlight attributes of entities for which users might search, to expand searches by including synonyms, and to collect information about entities related to the subject of a search for presentation to the user. (Source: ▶ https://searchengineland.com/laymans-visual-guide-googles-knowledge-graph-search-api-241935)

ties in addition to various properties of each entity. There currently is no standard way of creating a knowledge graph and there is not even consensus on what features a knowledge graph should support. Nevertheless, virtually all well-known e-commerce sites—from Google to Facebook to Bing to eBay—use a graph-based representation of knowledge to provide key functionality for users hoping to perform a variety of tasks.

Google, for example, has a knowledge graph that comprises more than one billion classes of entities and instances of those entities, and more than 70 billion facts about those entries (Fig. 24.5) (Noy et al. 2019). The graph makes it possible for Google to know that, when a user searches for some abstraction, such as "medications for diabetes," the user may be interested in different *kinds of* medications for diabetes, such as insulin and hypoglycemic drugs. Thus, the knowledge graph causes the search engine to bring up links to related entities when a user performs a search, and the graph helps to disambiguate the user's query when the objective of the search may not be clear.

Many commercial CDS systems are adopting knowledge-graph technology as a component of their software architecture. Because there is no standard knowledge-graph approach, these systems adopt different knowledge-graph formalisms and perform different tasks using their graphs. "IBM Watson," for example, is a name that encompasses a family of CDS systems that have been deployed for a range of clinical domains in recent years. Although these IBM Watson systems have had a variety of capabilities and evolving computational architectures, a common theme has been their use of knowledge graphs to implement precise information retrieval and to assist with natural language understanding. Thus, a knowledge graph can allow an IBM Watson system to interpret a user's natural language query and to locate a document or a fact that is itself stored in a knowledge graph to respond to that query. Indeed, the use of knowledge graphs in general search engines such as Google and Bing provides such technologies with capabilities that make them useful for decision support. When reporting search results, both

Google and Bing display the contents of their knowledge graphs for the indicated item in a "knowledge box" at the upper right of the page. These knowledge boxes often can be very effective in offering specific information to address user queries.

The simplest form of knowledge graph is one that encodes an enumeration of the kinds of entities in an application area and that places them in an abstraction hierarchy, indicating when elements of one entity form a subclass (or superclass) of another. This type of data structure is often referred to as an **ontology**. Ontologies are like **controlled terminologies** (see ▶ Chap. 7), in that they provide a standard means for referring to the types of entities that comprise a domain, but they organize those entities using a graph that makes explicit the semantics of the relationships among the entities. **SNOMED CT** and the **NCI Thesaurus** are examples of commonly used controlled terminologies that are represented using knowledge graphs in a manner that also makes them ontologies.

Ontologies are important in biomedical informatics for their explicit representation of abstraction relationships, thus facilitating interpretation of high-throughput experiments, aiding information retrieval and natural language processing, and indexing data (Bodenreider and Stevens 2006). CDS systems can use ontologies that encode knowledge about different application areas in a manner that aids reuse of that knowledge in new settings and that makes it easy for developers to update the knowledge as their understanding of the application area evolves. Such systems use ontologies (or knowledge graphs more generally) to encode clinical knowledge in a manner that may overcome some of the limitations of the more prevalent CDS architectures. These systems make an explicit distinction between the *static knowledge* of the clinical domain (e.g., knowledge of the specifications entailed by a clinical practice guideline) and the *problem-solving knowledge* needed to apply the static knowledge to a particular patient (e.g., the means to generate specific prescriptions for medications based on the general guideline recommendations and the particular clinical situation that the patient is experiencing). This distinction makes it possible for system builders to address different elements of the knowledge needed to be represented in the computer using tailored approaches and tools (Musen 1998; deClerq et al. 2004).

The ATHENA-CDS system exemplifies this component-oriented approach (Goldstein et al. 2000). ATHENA-CDS is a computer system that is integrated with the HIS that has been used by the U.S. Department of Veterans Affairs (VA), known as VistA.[7] ATHENA-CDS has been installed at several VA medical centers and has remained in continuous use at the Palo Alto VA medical center since the 1990s. ATHENA-CDS offers advice regarding patients who have certain chronic diseases, whose physicians would like to treat those patients in accordance with recognized evidence-based clinical practice guidelines (◘ Fig. 24.6). ATHENA-CDS draws on several electronic knowledge bases, each one constituting a knowledge graph that encodes the knowledge of a particular guideline (e.g., for hypertension, for hyperlipidemia, for diabetes, and so on). Each time that a patient with a relevant diagnosis (e.g., hypertension) is seen in the outpatient clinic, ATHENA-CDS takes as input the corresponding guideline knowledge base and patient-specific data from the VistA EHR and generates as output suggestions to the clinician for treating the patient to ensure that the treatment is consistent with the care that the guideline would recommend. Because the standard documents that define clinical practice guidelines can be long and complicated, it is extremely helpful for the computer to focus the clinician's attention on precisely which interventions should be considered to guarantee that the patient's care is consonant with the medical evidence captured by a given guideline (◘ Fig. 24.7).

ATHENA-CDS was engineered using an approach that separates out static knowledge about the clinical application area from knowledge about problem solving (i.e., knowledge about generating a situation-specific clin-

7 The VA is in the process of replacing VistA with a commercial HIS offered by Cerner.

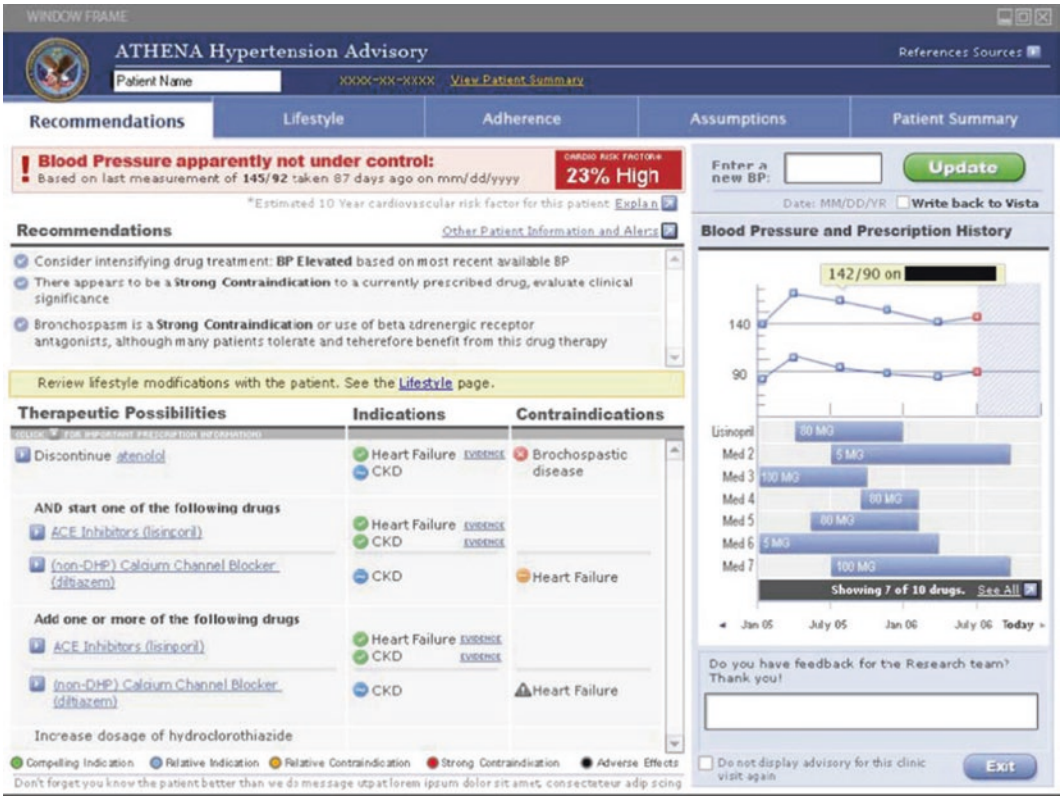


Fig. 24.6 An example of the ATHENA-CDS system interface. ATHENA-CDS provides decision-support for the management of hypertension and several other chronic diseases by using a declarative knowledge base created as an instantiation on a generic guideline ontology. In the screen capture, the provider has entered the patient's most recent blood pressure, and is offered advice about possible alterations in therapy based on the relevant clinical-practice guideline. The screen image depicts only simulated patient data (Courtesy of M. K. Goldstein, VA Palo Alto Healthcare System)

ical recommendation; Musen et al. 1996). To construct ATHENA-CDS, it was necessary first to define an ontology of clinical practice guidelines (Fig. 24.8). The guideline ontology makes it clear that all guidelines must include *eligibility criteria* that indicate which patients should be treated in accordance with the guideline, a *clinical algorithm* that specifies the sequence of treatments recommended by the guideline, and *guideline drugs* that represent all the medications that patients might be given when their provider follows the guideline. Because the guideline ontology is general, the graph does not contain information about any *particular* clinical algorithm, any *particular* eligibility criteria, and so on. The ontology merely states that all guidelines for management of chronic diseases have such characteristics.

Developers of ATHENA-CDS used the Protégé ontology-development system (Musen 2015) to create the ontology of clinical practice guidelines, which constitutes a knowledge graph. The developers then used Protégé to create subgraphs that represent distinct knowledge bases that define how to manage patients in accordance with particular guidelines. The developers created a knowledge base for management of hypertension reflecting the guideline that is used by the VA and the Department of Defense (DOD), supplemented with recommendations from the Joint National Commission on Hypertension (National High Blood Pressure Education Program 2004; Fig. 24.9). They instantiated the ATHENA-CDS guideline ontology to build a knowledge base for management of congestive heart failure based on the guideline

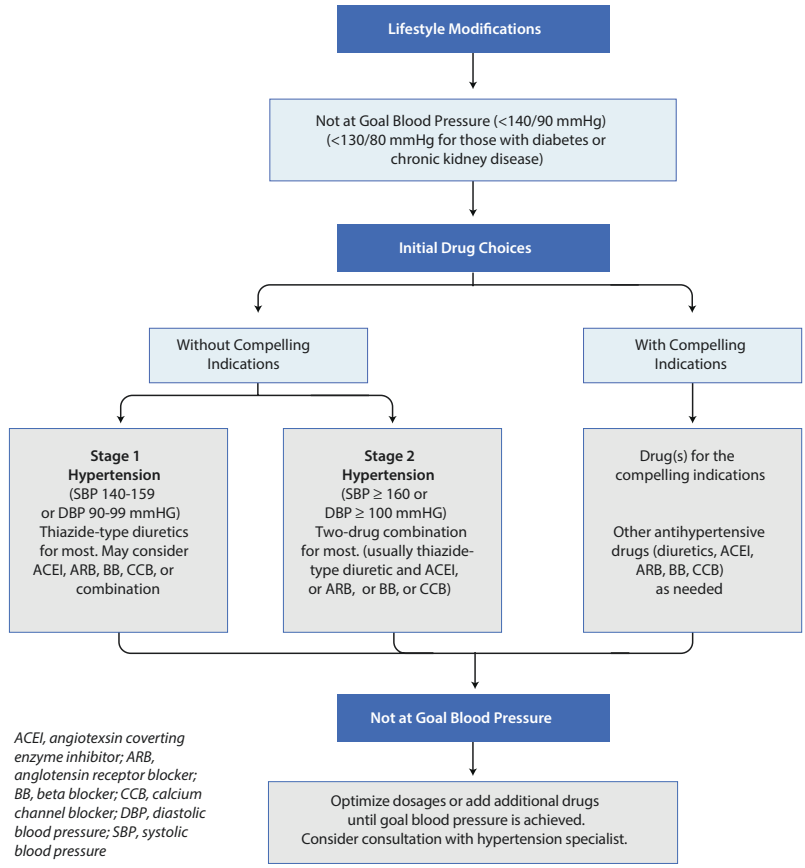


Fig. 24.7 Professional societies, health-care practices, private foundations, and other organizations are all working to capture "best practices" for managing patients in accordance with scientific evidence in terms of clinical practice guidelines. Unfortunately, nearly all these guidelines are published initially as large paper documents. Here is a high-level, paper-based flowchart from the guideline developed by Joint National Commission on Hypertension. The flowchart summarizes detailed recommendations that the guideline document specifies in many pages of text

developed by the American Heart Association and the American College of Cardiology. The developers built a knowledge base for management of chronic pain, based on the guideline promoted by the VA and the DOD (Trafton et al. 2010). Other knowledge bases for guideline-based care of diabetes, hyperlipidemia, and chronic kidney disease were created in a similar manner.

The ontology-driven approach makes it possible to start with a particular ontology (in this case, one for clinical practice guidelines for management of chronic disease) to create multiple knowledge bases, each one instantiating the ontology to specify the knowledge required for particular guidelines. Similarly, the different knowledge bases can be mapped to different problem-solving programs, such that each problem solver automates a different task associated with guideline-based care (therapy planning, eligibility determination, and so on). The ability to "mix and match" knowledge bases and problem solvers offers considerable flexibility, and it enables developers to reuse elements of previous solutions to address new CDS problems that require different domain knowledge or different problem-solving procedures.

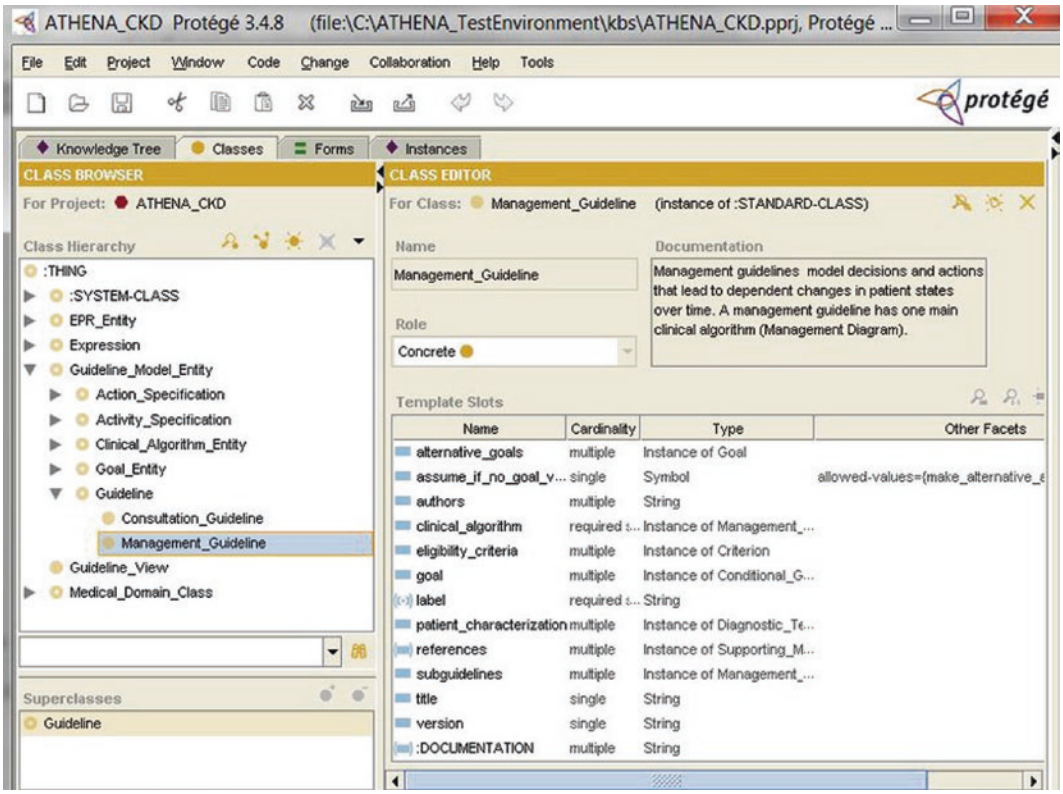


Fig. 24.8 A small portion of the ontology of clinical guidelines used by ATHENA-CDS as entered into the Protégé ontology-editing system. The hierarchy of entries on the left includes entities that constitute building blocks for constructing guideline descriptions. The panel on the right shows the attributes of whatever entity is highlighted on the left. Here, *goal*, *eligibility_criteria,* and *clinical_algorithm*, for example, are attributes of the entity known as *Mangement_Guideline*. The ontology entered into Protégé reflects concepts believed to be common to all guidelines, but does not include specifications for any guidelines in particular. The complete domain model is used to generate automatically a graphical knowledge-acquisition tool, such as the one shown in Fig. 24.7

24.3.3 Coda

There is no standard way to build a decision-support system. Developers need to make choices, based on the nature of the decision task to be performed, the data and knowledge that are available, and the software tools with which they are most familiar. There are a variety of approaches from which to choose, and each one entails different kinds of trade-offs. A CDSS for providing advice about something potentially as complicated as a clinical practice guideline can be hardcoded in software, implemented as a rule-based system, or driven by knowledge-graph technology. In cases when there is no appropriate evidence-based guideline, a CDSS can use probabilistic approaches or machine learning to suggest reasonable treatment decisions. Although interoperability standards such as SMART-on-FHIR are offering the opportunity to embed custom-tailored CDS technology within proprietary Health IT systems (see ▶ Chap. 7), the monolithic nature of most installed systems makes it challenging to achieve this kind of flexibility in the real world. Nevertheless, new IT standards are improving the landscape considerably, as we discuss in the next section.

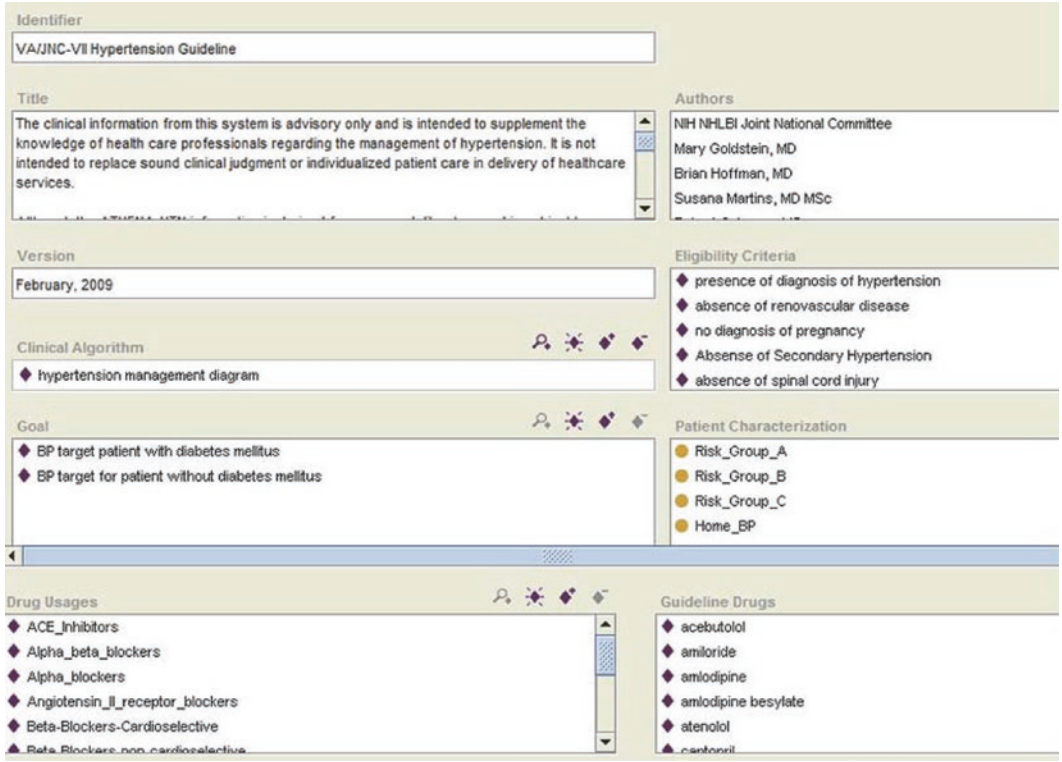


Fig. 24.9 A screen from a Protégé-generated knowledge-acquisition tool for entry of clinical-practice guidelines. The tool is generated automatically from a domain ontology, part of which appears in Fig. 24.7. The entries into the tool specify the knowledge required to treat patients in accordance with the guideline for chronic hypertension adopted by the Department of Veterans Affairs

24.4 Translating CDS to the Clinical Enterprise

Over the past several decades, advanced CDS systems have been developed and deployed in a number of academic medical centers. The technology has subsequently diffused into commercial EHR systems and into routine practice (Chaudhry et al. 2006). The uptake has been greater in medium-to-large hospitals and in medical-center–based networks, including affiliated practices, and has been much less in smaller hospitals, clinics, and independent practices. Although these trends had been sluggish, the "meaningful use" regulations for HIT and other regulatory and incentive-based approaches in the United States as well as elsewhere accelerated the adoption of CDS technology (Blumenthal and Tavenner 2010; Blumenthal 2010).

In general, the CDS systems deployed to date in vendor EHR systems are quite varied, and relatively limited in scope, and their capabilities in knowledge management are also varied (Wright et al. 2009). The greatest uptake has been in the form of simple alerts and reminders, standard physician order sets, CPOE-based prescription templates with dose checks, allergy checks, identification of drug–lab and drug–drug interactions, and some use of info-buttons or access to context-specific knowledge resources. In some specific settings, rule-based systems have been used to drive the intelligent collection of clinical information in a comprehensive, structured clinical documentation form (Schnipper et al. 2008).

A few vendors have been successful at distributing knowledge resources, making available shareable clinical knowledge in the form of drug-interaction databases, order sets for common indications, rule-based knowledge, documentation templates, and information resources for infobutton-based queries (Middleton et al. 1998). More recent research has examined opportunities for creating knowledge repositories to make these resources more readily available in both the public and private sector (Osheroff et al. 2007; Kawamoto et al. 2013).

Despite growing demand, many years of research and development, and the broad adoption of EHRs in recent years, CDS has had relatively limited adoption to date. There are several reasons for the slow uptake of CDS (Wright et al. 2009), which need to be addressed by approaches such as those enumerated in the sections that follow.

24.4.1 Standard Patient Information Model

CDS rules and other problem-solving approaches need to operate on specific patient data with a clear understanding of the patient data model and semantics of the terms. If those data are stored in a proprietary format and with non-standard encodings, then a set of rules needs to be customized to use data in that form, or the data need to be translated to the information model of the CDS rules, quality measure, or other uses. The customization of rules for each EHR (or EHR implantation) has been the prevailing mode, typified by the "curly braces problem" of Arden Syntax rules, described previously. As a result, vendor EHR systems tend to have libraries of rules that operate only in their own systems, using their proprietary data dictionaries and data models. Knowledge sharing across platforms and systems has been limited, and considerable work is required to integrate vendor HIT products with external CDS systems.

One approach is to develop a canonical information model that both the knowledge source systems, and the knowledge artifacts can use. This strategy can facilitate a common understanding of semantics in the data and the semantics of the knowledge artifact, and it alleviates mapping CDSS variables to local data or developing a set of custom rules to access the local data in each and every setting. This approach does not eliminate what some describe as an "irreducible mapping problem" in Health IT, but it does distribute it in an effective manner such that, as data representation evolves more toward a canonical form, we can approach **iso-semantic mappings** of data from data source to knowledge artifact—that is, where the semantics of the source data are identical to the semantics of the data expected in the CDS analytic. This approach has been pursued in the very successful Observational Health Data Sciences and Informatics (OHDSI) project in research informatics (Hripscak et al. 2015), which adopts a Common Data Model for source systems to map to, and for analytics (and federated queries) to run against (Jiang et al. 2017a, b).

As discussed below in ▶ Sect. 26.4.2, in 2012, a **virtual medical record** (vMR) based on the HL7 version 3.0 **Reference Information Model** (RIM) was an initial effort to arrive at the notion of a canonical information model. It was approved by HL7 as a draft standard for trial use for linking dynamically at runtime the arbitrary data elements available in the patient database of an EHR to CDS systems that assume the standard vMR data model for data encoding (Kawamoto et al. 2010). The vMR was designed to serve as an intermediary data model—a canonical form—between proprietary database formats and standards-based CDS systems that developers might plug into any EHR that can make its data available in a vMR-compliant manner. HL7 supported work to map the vMR to standard terminologies and clinical data element definitions. So, for example, data elements in an ECA rule could be referred to in an interoperable manner, rather than, say, the individual mapping of data elements to access methods inside the curly braces of Arden Syntax MLMs (see Rule-Based Systems in ▶ Sect. 26.3.2).

Newer work focuses on the development of the **Quality Data Model** (QDM) to represent clinical data and concepts used in specifying quality measures, and ultimately clinical decision support logic. The QDM is an information model that describes the relationships between patients and clinical concepts in standardized formats. QDM allows the definition of a clinical concept used in a specification to measure quality of patient care via defined data elements, and it provides the vocabulary needed to relate concepts to each other. For example, QDM, at the highest level of abstraction, defines categories such as Medication, Procedure, Condition, and so on. Within a category, the Datatype definition gives the context of the clinical care process being assessed such as "medication - active," or "medication – administered." Further Datatype details are defined with Attributes that can further define the Datatype, or define the expected source for the data. For example, a diagnosis of "Diabetes Mellitus Type I" is an *active diagnosis* datatype, a "Metformin prescription" is a *medication prescribed*, and a "Hemoglobin A1c value" is a *laboratory result*. Each of these elements can be related to terms in controlled terminologies via specification of corresponding code sets. By relating attributes between data elements, the QDM provides a method to construct complex clinical representations both for electronic clinical quality measures and for clinical decision-support logic. Because of this ability to share common logic elements (expressions, value sets, terminology) between quality measure and CDS specifications, QDM has become popular for specifying quality measures among a wide variety of measure developers and CDS implementers (Pathak et al. 2013; Hong et al. 2016).

More recently, the **Fast Healthcare Interoperability Resource (FHIR)** standard from HL7 has emerged from a multi-vendor collaboration in the Argonaut Project (HealthMgt 2015). FHIR is designed specifically for the Web and provides resources and foundations based on common methods and technologies used in the Web (XML, JSON, HTTP, and OAuth) (Bender and Sartipi 2013). FHIR resources are accessed through what is known as a **RESTful API**, which can be defined as one that uses HTTP requests to GET, PUT, POST and DELETE data. FHIR frameworks are built around the concept of resources—basic units of interoperability and modular components that can be assembled into working systems to try to resolve clinical, administrative and infrastructural problems in health care. This capability has partially addressed the heterogeneity of different systems, and it has dramatically improved interoperability between systems.

24.4.2 Adoption of Standard Knowledge-Representation Models

Although Arden Syntax has been an HL7 and ANSI standard since 1999, only a few vendor systems manage their libraries of decision rules using Arden Syntax. Even doing so, of course, the rules still need to be customized for use with vendor-specific patient databases on a tedious, rule-by-rule basis to overcome the curly-braces problem described in ▶ Sect. 24.3.2. HL7 has pursued the development of more interoperable models not only for data query but also for defining elements of ECA rules. A succession of efforts led to a specification for a query language called GELLO (Sordo et al. 2004), and then the Health eDecisions rule formalism,[8] and more recently the Clinical Query Language, discussed below.

24.4.2.1 Standards for Encoding Clinical Guideline Models

Particularly challenging in the standards-development world is creation of a standardized, shared model for representing clinical practice guidelines in a form suitable for execution at run time. The **Guideline Element Model** (GEM; Shiffman et al. 2000)

8 HL7 International (2014). Health eDecisions. Retrieval February 19, 2020: ▶ https://wiki.hl7.org/index.php?title=Health_eDecisions

is an XML mark-up specification that is an American National Standards Institute (ANSI) standard, now in its third revision, that guideline authors can use to annotate their narrative guidelines to identify key elements for both quality assessment and execution. GEM allows authors to demarcate the text that identifies guideline actions or eligibility criteria, and thus can serve an intermediary purpose in work to transform a prose guideline into a computable specification, but the standard does not itself provide a mechanism to translate a marked-up guideline document into a structure that a computer can interpret and execute. Such systems may be criticized as not explicitly representing the underlying ontology of the guideline components to be used at run time.

Other efforts have focused on creation of a guideline ontology, such as the one adopted by ATHENA-CDS (see Fig. 24.8), that can inform the creation of computer-understandable knowledge bases that are able to capture knowledge about specific guidelines. Such knowledge bases then could allow a CDS system to use knowledge about the guideline, data from the EHR, and information concerning patient preferences and available resources to offer situation-specific, guideline-directed advice. As we have noted, an underlying infrastructure known as EON (Musen et al. 1996) drives the ATHENA-CDS system. Other ontology-based approaches have appeared over the years, including GLIF, GUIDE, PRODIGY, Asbru, and GLARE. Peleg and colleagues (2003) compared many of these guideline models, and they showed significant commonalities among them. Despite the large degree of agreement, however, work in this area has not yet led to anything near a standard that is widely adopted. Part of the problem is that there is wide variation in the structure, granularity, and specificity of existing clinical practice guidelines, making it difficult to develop a single comprehensive and yet readily applicable guideline model. Analysis of the use of guidelines also indicates that guidelines themselves are rarely "executed" without considerable adaptation or localization, except in situations such as protocol-driven care (for example, in clinical trials or in very specific procedures such as renal dialysis). ATHENA-CDS thus dispenses with offering specific guideline-based recommendations, and instead suggests to the clinicians when certain treatment options might be "compellingly indicated" or "relatively contraindicated." In highly regimented settings such as the administration of chemotherapy for cancer, however, a CDS system generally would need to be much more "prescriptive" in offering recommendations to clinicians.

24.4.2.2 Standards for Encoding ECA Rules

As we have noted, the ability to share decision rules is hindered in the absence of standards to encode both the logic of the rules (how the IF and the THEN components are to be evaluated) and the clinical data on which the rules depend. Recently, the **Clinical Quality Language** (CQL) has emerged as an expression language that addresses these challenges. CQL is intended to characterize both quality-measure logic and decision-support logic and the data that such logic processes (Odigie et al. 2019). For example, CQL expressions define the expected input data model, library resources that may be called, parameters, value sets, and code sets used in definitions of patient conditions, and additional concepts needed to specify and encode the clinical logic and data types used in a clinical quality measure—or in a CDS rule (Jiang et al. 2015).

Expression languages are commonly used to represent the logic to be used at the presentation layer of an application, and the methods that may be used to interact with standard data models. CQL grew out of prior efforts such as GELLO and the Health eDecisions framework, from work in the U.S. Office of the National Coordinator's Clinical Quality Framework Initiative. It was focused on identifying, defining, and harmonizing standards and specifications that promote integration and reuse between clinical decision support and clinical quality measurement (CQM) knowledge-representation formalisms. It strives to be more clinician friendly and accessible to subject-matter experts. However, the standards used for the electronic representation of CDS and CQM artifacts have not been

developed in consideration of each other, and the domains use different approaches to the representation of patient data and computable expression logic. Harmonization of these approaches now is focused on clearly identifying the various components involved in the specification of quality artifacts, and then establishing as a principle the notion that they should be treated independently (separation of concerns). Broadly, the components of a CQL artifact involve specifying:

- Metadata – Information about the knowledge artifact (whether CQM or CDS) such as its identifier and version, what health topics it covers, supporting evidence, related artifacts, dependencies, etc.
- Clinical Quality Information – The structure and content of the clinical data (data model) involved in the artifact
- Expression Logic – The actual knowledge and reasoning being communicated by the artifact

The CQL specification is an approved HL7 standard endorsed by the U.S. Centers for Medicare and Medicaid Services, the U.S. Centers for Disease Control and Prevention, and others for the representation of either CDS or CQM knowledge artifacts. It is part of the Clinical Quality Framework effort supported by the Standards and Interoperability Framework of the U.S. Office of the National Coordinator for Health IT. As such, it aims to promote the interoperability of knowledge artifacts through standardization of the data models, logic statements, and controlled medical terminology (including value sets). Because CQL is also fully specified in a machine-readable way, it may be programmatically converted into what HL7 calls an *expression logical model* (ELM), which can then be interpreted in an execution environment.

The convergence of a common formalism that accommodates the use of standardized data models, facilitates clinical query and computation through a library of methods, and which is machine interpretable dramatically improves the portability of clinical knowledge artifacts. In addition, this specification is designed to be data model-independent, meaning that CQL has no explicit dependencies on any aspect of any particular data model. Rather, the specification allows for any data model to be used, so long as a suitable description of that data model is supplied.

Current efforts now focus on further standardization of input data models, as well as alignment with the FHIR API standard. For example, the QUICK specification and the Quality Improvement Core (QICore) are being developed concurrently with the CQL specification to ensure that the two specifications interoperate effectively. QUICK is a logical model consisting of clinical objects, attributes, and relationships. QUICK provides a uniform way for clinical decision support and quality measures to refer to clinical data. This initiative began in 2013 with the creation of the Quality Improvement Domain Analysis Model (QIDAM), which drew on the vMR and QDM as sources of requirements. Originally, the QUICK data model was developed entirely independently of FHIR. However, recognizing the broader community focus on FHIR, QUICK was aligned, structurally and semantically, as closely as possible to FHIR. This alignment not only creates a common model for quality and interoperability, but also it will make it easier in the future to leverage other FHIR-related efforts, such as Clinical Document Architecture (CDA) on FHIR, or CQL on FHIR. Authors of future quality measures and clinical decision support artifacts may use QUICK, together with the Clinical Quality Language (CQL), to create interoperable and executable knowledge artifacts, thus dramatically the ability to share computable biomedical knowledge artifacts.

24.4.3 Modes of Deployment of CDS

Even with the emergence of shareable, computable, biomedical knowledge artifacts, one of the key impediments to widespread adoption of CDS, particularly the use of rules

and alerts, is clinician annoyance with pop-ups, messages, emails, and other notifications that interrupt workflow. Ideally CDS systems should be integrated into the organization and presentation of information to facilitate workflow and decision making, by anticipating what information is needed for a decision, pre-fetching it, displaying it in ways that support visualization of trends or relationships, and tying these analyses to care plans or actions that can be offered immediately and quickly selected by the user. Order sets, as stated in the beginning of this chapter, form a good example of use of CDS both to suggest appropriate actions in a given setting and to make it easy to accomplish those actions, by immediately enabling the orders in the set to be entered automatically into the EHR, perhaps with modification.

There is much ongoing research to develop methods for managing the processes of data capture, data presentation, data visualization, and selection of actions, but this work is usually being done outside of vendor EHRs. Given limited interoperability and access to the internals of proprietary systems, this kind of experimentation is now tending to take place in the form of apps and services that operate on externally extracted data (Mandl and Kohane 2012). There is growing support for the notion of SMART apps—Substitutable Medical Applications and Reusable Technologies—that use standard API methods to access electronic health record data, perform clinical inferences external to the EHR, and return insights in a standardized container (whether iFrame, Web page, or free-standing application; Mandel et al. 2016).

24.4.4 Workflow and Setting-Specific Factors

As noted in ▶ Sect. 24.3.2, applications based on single-step situation–action rules are among the most prevalent and useful types of CDS systems. Such systems can be invoked in many contexts to provide either recommendations in real time or reminders or alerts that are processed in batch, based on time-oriented triggers or data-evaluation events. Rules can invoke other knowledge resources—providing new information content, triggering other rules, or offering order sets.

Rule content is ideally based on analysis of clinical evidence, such as recommendations or guidelines emanating from the U.S. Preventive Services Task Force, or from professional-society studies of best practices for specific diseases. The job of formalizing these recommendations into executable logic requires that they be expressed in a specified way, but even having done so, such rules are not typically ready to execute in a particular environment, even if they are expressed in a rule execution language "understood" by an EHR system, and if they refer to the data elements in the EHR in their expected format. The reason the rules are not readily executable is also the reason that rules that work well in one environment are often not able to be successfully deployed elsewhere without substantial modification (even if in the same representation format and if using the same data model).

The reason for the failure is lack of adaptation to what we refer to as *setting-specific factors* (SSFs; Greenes et al. 2010). To work effectively, rules need to integrate well with the clinical setting, workflow, users, application environment, and other factors. These requirements are reflected in how and when the rule should be triggered—on various events such as examination of some element of the EHR, on login to the system, or on the availability of laboratory test results. Rules may also be in developed in the form of reminders that are triggered when the CDSS evaluates on a batch basis a practice's list of patients to be seen on a given day, the patients who have a birthday in a given month, the passage of a specific interval of time since a previous comparison event, and so on. The rules additionally may vary based on the practice setting (e.g., the emergency department, an office practice, or an inpatient unit); particular inclusion or exclusion criteria or threshold

modifications that may be site-specific; how the recommendation should be transmitted (e.g., via electronic mail, popup windows, or sidebar messages); whether the recommendation requires acknowledgment by the recipient; whether it can be overridden; whether the alert should be escalated to supervising clinicians, and so on. Rules that have been custom tailored in such ways by means of executable code naturally are less sharable than are generic rules. Failure to capture the kinds of customizations that are needed, however, makes it time consuming for individual sites to adapt generic medical recommendations to their particular requirements or to capitalize on the experiences of others. What is needed is a way to represent useful experience in terms of SSF combinations that work, without needing to do so at the level of detailed code that is difficult for users to visualize and modify.

24.4.5 Sharing of Best-Practice Knowledge for CDS

The methods described above for sharing computable biomedical knowledge are now gaining momentum (viz., SMART, CQL, FHIR). Historically, it generally has fallen on each health-care organization, user group, or other entity to undertake its own process of identifying and managing the best-practice knowledge it wants to deploy in its CDS systems. Most institutions lack the expertise, or the resources, to accomplish this task. Even having a national or international repository of such knowledge would not preclude the need for customization, but it would certainly make it easier for each health-care entity to start with a trusted source. Over the past decade, the U.S. Agency for Healthcare Quality and Research has funded efforts to create such a public repository, known as CDS Connect (Lomatan et al. 2019). In CDS Connect's repository, the goal is to archive knowledge artifacts represented in the CQL formalism to promote sharing and interoperability. Where such a repository should be hosted, how it might integrate public and private knowledge sources, who would have oversight over it, how its knowledge would be peer reviewed and quality-rated, and how it would be sustained are among the many questions that have not yet been answered, but this is an area of intense research and development. While many health-care organizations continue to perform this kind of knowledge-curation work for their own constituencies, several initiatives show a clear pathway to becoming viable alternatives for knowledge aggregation and dissemination.

24.5 Future Research and Development for CDS

Workers in biomedical informatics have studied problems in assisting with complex decision making for more than half a century. It seems that it is only now, with the very recent adoption of HIT on a widespread basis, that the foundations are finally in place for the rapid advance of CDS technology in clinical settings. Although considerable logistical problems still must be surmounted as outlined in ▶ Sect. 24.4, this is an exciting time in which to study CDS and its translation from the laboratory to the point of care.

24.5.1 Standards Harmonization for Knowledge Sharing and Implementation

Many implementation challenges remain for the broad adoption and effective use of CDS in EHR systems. As mentioned above, one of the most active areas of current research focuses on development of standard approaches to knowledge sharing for CDS. Knowledge sharing may take the form of human-readable artifacts, machine-interpretable artifacts, or executable Web services (Osheroff et al. 2007; Goldberg et al. 2014; Dixon et al. 2013; Kawamoto et al. 2013). A capability for CDS sharing, as well as CDS functionality itself, would be substantially facilitated by the continued develop-

ment and use of common standards designed to serve CDS needs in health care.

As noted, several standards currently exist that are aimed at specific areas of CDS and types of CDS artifacts, or that could be leveraged to benefit CDS. For example, the Clinical Decision Support Consortium, a large collaborative research and development group supported by the U.S. Agency for Healthcare Research and Quality, adopted an enhanced version of the Continuity of Care Document (CCD) to serve as the foundation for input data for multi-institutional trials of CDS technology (Middleton 2009). When taking advantage of the most current standards such as CQL and FHIR, systems developers tend to adopt not only the standards but also arrive at more common implementation approaches or patterns, that promote knowledge sharing and reuse, and which may decrease the implementation burden if EHR vendors accommodate the standardized API and interaction models, as well as the data model expectations of the standards.

24.5.2 Context-based Knowledge Selection

Much of the above effort is aimed at overcoming the non-portability of event-condition-action (ECA) rules, such as alerts and reminders, because of the need to localize and adapt triggering conditions, modes of interaction, and specific workflow processes of sites (Setting-specific factors, or SSFs, as discussed in ▶ Sect. 24.4.4), leading to many variations that must be laboriously designed and tracked.

One of the possible modes for reducing this need for customization and localization is to use the context, state, and activity (CSA) of a CDSS user to automatically identify appropriate knowledge artifacts to be made available. The idea relies on maintaining patient state, user role, expertise, setting, and specific tasks being performed to determine when the knowledge might be pertinent. The latter also requires a rich multi-axial set of metadata for the knowledge artifact repository that allow selection based on user CSA.

Other ways in which context could be used might be to create specialized views or filters of available data, based on CSA. For example, selection of data to be viewed could be based on a user's role and domain expertise. Associations (and, ideally, typed relations) among data items could be highlighted—for example, among problems, findings, and actions. Such relations could also potentially be used to anticipate assessment and plan entries in notes based on data items present or being focused on. Such approaches undoubtedly will only be scalable in generalized web-services implementations. For example, infobutton managers already use context such as user type, app/function being performed, and specific item being evaluated to identify retrieval keys for external knowledge resources. The idea behind the expansion of context as a mode for invocation of CDS more generally is to create a more detailed and continually updated model of context, state, and activity. Such efforts are just getting underway.

24.5.3 Representation Models

To date, standards and related efforts addressing CDS have heavily emphasized specific CDS execution methods and the representation of the clinical context of the patient. For example, as we have noted previously, a variety of frameworks for working with rules, including Arden Syntax, Drools, JESS, CQL, along with several proprietary formats, have worked their way into vendor offerings. This diversity has inhibited the exchange of best-practice knowledge to date, but progress toward more effective knowledge interchange is being made with CQL in particular. The current situation is that both public and private knowledge repositories of knowledge artifacts that address specific pieces of the CDS problem exist (e.g., NLM Value sets, CMS eCQMs, standardized data models, professional society clinical pathways, vendor implementations of algorithms and analytics, and SMART applications) and we see growing adoption given the increased pressures on clinical reasoning and operations in practice.

In addition to reducing unwarranted clinical variability in practice, future work needs to establish a canonical patient information model with a formal ontology, an event model for triggering conditions, an action model for CDS intervention recommendations, a workflow model for appropriately inserting CDS interventions into the routines of clinical practice, a knowledge-representation schema with a standard regular expression language, and, ideally, a measurement standard to assess CDS performance in use.

24.5.4 Externalizing CDS

Standardization of methods for externalized CDS (CDS performed outside of the EHR, in the cloud) is becoming more commonplace due to emerging standards described above (CQL, CQM, QUICK, FHIR). Prior research and development efforts have demonstrated the feasibility of accessing clinical data from the electronic health record via standards-based (viz., FHIR) and proprietary RESTful APIs, and running executable knowledge artifacts – both quality measures, and clinical decision support – on the data in the secure cloud environment and returning insights into the app being used in the clinical workflow (Wright et al. 2015; Dixon et al. 2013). Current work is focused on representing all of the requisite knowledge artifacts required to create a computable practice guideline fully represented in CQL, and on running those knowledge artifacts via FHIR data-access methods. Use of services enables considerable flexibility and breadth of kinds of CDS that can be made available widely, and ongoing work with FHIR resource specifications is improving the ability to gather the data needed and deliver it, and to integrate with clinical workflow processes more smoothly.

We have described the many efforts underway to standardize data models for information exchange, CQL for knowledge representation, and functional integration with EHR workflows using SMART, and FHIR. A deeper functional integration model between externalized services and EHR workflow is provided by **CDS Hooks**. CDS Hooks (Dolin et al. 2018) is a method for enabling an event in the host system to determine when an app (such as provided by SMART-on-FHIR) should be launched programmatically, thus complementing the need to launch apps directly by users. The EHR detects an event such as a physician beginning to write an order, and it then can invoke an external decision-support service. That service can determine what task is being performed and return information in the form of a "card" (a phrase or text snippet containing an inference or assessment, or suggested action) that will be displayed within the EHR. CDS Hooks may also provide a link to an external app. The main downside of these approaches to CDS is that, although they have formal methods for launching apps, they basically have no constraints on what happens within the apps, or support for linking to the intrinsic functionality and workflows of a host EHR. Thus, the methods can result in a proliferation of SMART apps and CDS Hooks modules in an organization's library, with a number of them that may be of limited utility. This situation could cause significant challenges for an enterprise seeking to manage and update its CDS capabilities on a regular basis.

24.5.5 Usability Research and CDS

The use of CDS within EHRs, and health IT in general, have been identified as double-edged swords: technology may provide benefit, but it also may cause considerable harm. Clinician error when using information systems that may result in untoward outcomes and unintended consequences (Karsh et al. 2010; Sittig and Singh 2009) may be an emerging property that is demonstrated only after system implementation or widespread use. Medical errors related to use of Health IT are problematic, not only for clinical and quality of care reasons, but technically, since they may represent a mismatch between the user's model of the task being performed and the model used in a computation (National Research Council (U.S.) Committee on Engaging the Computer

Science Research Community in Health Care Informatics et al. 2009), a mismatch between the application's intended functionality and the resulting action or event (Harrison et al. 2007), or a latent health IT-related error yet to happen (Ash et al. 2007). Excessive alert fatigue can undermine the efficacy of clinical decision support in CPOE (Isaac et al. 2009; Strom et al. 2010), and in other IT functions (Chused et al. 2008), and result in very high user override rates (Shah et al. 2006; van der Sijs et al. 2006; Weingart et al. 2003). Critical research questions need to focus on the potential mismatch between the user's mental model or intent and the application design, use case, or workflow model (Zhang and Walji 2011; Patel et al. 2010). Further attention needs to be given to basic principles of human-factors engineering, such as the use of colors and layout within the application interface. Additional questions remain regarding the ideal design of methods and controls with which a user might interact to choose a medication from a long list, or identify and encode patient problems. More advanced research will enable visualization and decision making by matching problems with care plans, and facilitation of continuity and coordination of care based on underlying CDS rules and guideline-based workflows. Especially challenging is addressing the need for structured data to support clinical decision support and quality reporting, in a manner that is efficient for the end-user. This goal might be achieved by combining structured documentation during data entry, and natural language processing for data abstraction from the clinical narrative. Most important, however, are methods to direct CDS to the right user, at the right time, in the right workflow, with the right level of alerting or intervention, and the right information (Osheroff et al. 2012).

24.5.6 Data-Driven CDS

As we have discussed, a major area of research in informatics concerns methods for deriving knowledge from large data sets using a variety of techniques. With the adoption of health IT broadly, investigators are drawing on large-scale data-mining methods to provide CDS for population monitoring, public health surveillance, and even to offer patient-specific recommendations based on cohort data when there is no specific evidence that could otherwise guide therapy. With the increasing availability of data from diverse sources relevant to patient care, large data sets may be created and used for both discovery of previously unknown associations, and novel clinical predictions (Frankovich et al. 2011; Longhurst et al. 2014). Critical research questions here will include how to define like cohorts of patients, how to structure and frame the index decision, what methods to use to assess the likelihood of alternate prediction scenarios, and how to model and elicit the patient's preferences for each scenario. The Institute of Medicine (2011b) articulated a long-term vision for a Learning Health System, in which clinical and administrative data of all kinds will begin to inform and enhance clinical practice on a national level in a wide variety of ways.

24.6 Conclusions

The future of CDS systems inherently depends on progress in developing useful computer programs and in reducing logistical barriers to implementation. Although ubiquitous computer-based decision aids that routinely assist clinicians in most aspects of their practice are currently the stuff of science fiction, progress has been real and the potential remains inspiring. Early predictions about the effects that such innovations would have on medical education and practice have not yet come to pass (Schwartz 1970), but growing successes support an optimistic view of what technology will eventually do to assist practitioners with the processing of complex data and knowledge. The research challenges have been identified much more clearly, legislative mandates are creating not only new financial incentives but also the practical substrate of increased EHR adoption and convergence toward data interoperability, and the implications for health-science education are much better understood. The basic com-

puter literacy of health professional students can be generally assumed, but health-science educators now must teach the conceptual foundations of biomedical informatics if their graduates are to be prepared for the technologically sophisticated world that lies ahead.

Equally important, we have learned much about what is not likely to happen. The more that investigators understand the complex and changing nature of medical knowledge, the clearer it becomes that trained practitioners of biomedical informatics will always be required as participants in fostering a cooperative relationship between physicians and computer-based decision tools. There is no evidence that machine capabilities will ever equal the human's ability to deal with unexpected situations, to integrate visual and auditory data that reveal subtleties of a patient's problem, to work with patients to incorporate their values and priorities in care plans, or to deal with social and ethical issues that are often key determinants of proper medical decisions. Considerations such as these will always be important to the humane practice of medicine, and practitioners will always have access to information that is meaningless to the machine. Such observations argue cogently for the discretion of health-care workers in the proper use of decision-support tools.

Suggested Readings

Bright, T. J., Wong, A., Dhurjati, R., Bristow, E., Bastian, L., Coeytaux, R. R., Samsa, G., Hasselblad, V., Williams, J. W., Musty, M. D., Wing, L., Kendrick, A. S., Sanders, G. D., & Lobach, D. (2012). Effect of clinical decision-support systems: A systematic review. Annals of Internal Medicine, 157(1), 29–43 This thorough analysis of studies of CDS systems demonstrates that there is good evidence that CDS technology can alter clinician behavior in positive ways, but that evidence that CDS systems can improve long-term patient outcomes is still inconclusive. The paper is also useful for its comprehensive bibliography.

Greenes, R. A. (Ed.). (2014). Clinical Decision Support, 2nd Edition: The Road to Broad Adoption. New York: Elsevier This book offers a comprehensive discussion of the nature of medical knowledge and of information technology to assist with medical decision making. It provides detailed discussions of the computational, organizational, and strategic challenges in the design, development, and deployment of CDS systems.

Institute of Medicine. (2011). Digital infrastructure for the learning health system: The foundation for continuous improvement in health and healthcare. Workshop Series Summary. Washington, DC: The National Academies Press. This monograph summarizes the vision for a national Learning Health System and offers the perspective of a wide range of thought leaders on the work required to achieve that vision.

Ledley, R., & Lusted, L. (1959). Reasoning foundations of medical diagnosis. Science, 130, 9–21 This is the paper that started it all. This classic article provided the first influential description of how computers might be used to assist with the process of diagnosis. The flurry of activity applying Bayesian methods to computer-assisted diagnosis in the 1960s was largely inspired by this provocative paper.

Sittig, D. F., Wright, A., Osheroff, J. A., Middelton, B., Teich, J. M., Ash, J. A., Campbell, E., & Bates, D. W. (2008). Grand challenges in clinical decision support. Journal of Biomedical Informatics, 41(2), 387–392 A rank-ordered list of some of the principal challenges for CDS technology development and implementation, intended "to educate and inspire researchers, developers, funders, and policy makers".

Questions for Discussion

1. Some researchers in medical AI have argued that CDS systems should reason from clinical data in a way that closely matches the reasoning strategies of the very best clinical experts, as such experts are the most clever diagnosticians and the most experienced treatment specialists that there are. Other researchers maintain that expert reasoning, no matter how excellent, is

at some level inherently flawed, and that CDS systems must be driven from the mining of large amounts of solid data. How do you account for the apparent difference between these views? Which view is valid? Explain your answer.

2. Transitioning CDS systems from one clinical setting to another has always been problematic. The Leeds Abdominal Pain System was installed in several major clinical settings, and yet the system never performed as well elsewhere as it had done in Leeds. The Arden Syntax, created expressly to facilitate knowledge sharing across institutions, failed to meet this goal to a significant degree. Why kinds of setting-specific factors make it difficult to transplant decision-support technology from one environment to another? What kinds of research might lead to better methods for knowledge sharing in the future?
3. In one early evaluation study, the decision-support system ONCOCIN provided advice concerning cancer therapy that was approved by experts in only 79% of cases (Hickam et al. 1985). In another study, the HyperCritic CDS system for the management of hypertension offered the same comments that were generated by a panel of experts in only 45% of cases (Van der Lei et al. 1991). Even today, such system performance is fairly typical for computer programs that suggest patient therapy. Do you believe that this performance is adequate for a computational tool that is designed to help physicians to make decisions regarding patient care? What problems might CDS systems encounter as their developers attempt to make the systems more comprehensive in the advice that they offer? Why might it be more difficult for computer systems to offer acceptable recommendations for patient therapy than seems to be the case for diagnosis? What safeguards, if any, would you suggest to ensure the proper use of such systems? Would you be willing to visit a particular physician if you knew in advance that she made decisions regarding treatment that were approved by expert colleagues less than 80% of the time? If you would not, what level of performance would you consider adequate? Justify your answers.
4. A large international organization once proposed to establish an independent laboratory—much like Underwriters Laboratory in the United States—that would test CDS systems from all vendors and research laboratories, certifying the effectiveness and accuracy of those systems before they might be put into clinical use. What are the possible dimensions along which such a laboratory might evaluate decision-support systems? What kinds of problems might such a laboratory encounter in attempting to institute such a certification process? In the absence of such a credentialing system for CDS systems, how can health-care workers feel confident in using a clinical decision aid?
5. There is considerable untapped potential for CDS to help in managing patients with multiple complex conditions. What are the challenges in dealing with such patients, and how can CDS be helpful? What are the features required of an algorithm that might integrate recommendations from the multiple clinical-practice guidelines that a CDS system could apply?
6. CDS is often implemented poorly, resulting in dissatisfaction, if not outright annoyance. What are the human factors that need to be taken into consideration in implementing CDS effectively? Discuss issues and approaches to enhancing usability. What are situations in which graphics and visualization might be used? How can CDS be used to enhance rather than to impede workflow? What are strategies to help avoid unintended consequences of poorly implemented CDS?

References

Adler-Milstein, J., Bu, D., & Pan, E. (2007). The cost of information technology-enabled diabetes management. *Disease Management, 10*(3), 115–128.

Adler-Milstein, J., Embi, P. J., Middleton, B., et al. (2017). Crossing the health IT chasm: Considerations and policy recommendations to overcome current challenges and enable value-based care. *Journal of the American Medical Informatics Association, 24*(5), 1036–1043.

Alamri, Y., Frizelle, F., Al-Mahrouqi, H., et al. (2016). Surgical ward round checklist: Does it improve medical documentation? A clinical review of Christchurch general surgical notes. *ANZ Journal of Surgery, 86*(11), 878–882.

Algaze, C. A., Wood, M., Pageler, N. M., et al. (2016). Use of a checklist and clinical decision support tool reduces laboratory use and improves cost. *Pediatrics, 137*(1), e20143019.

Ash, J. S., Sittig, D. F., Poon, E. G., et al. (2007). The extent and importance of unintended consequences related to computerized provider order entry. *Journal of the American Medical Informatics Association: JAMIA, 14*(4), 415–423. https://doi.org/10.1197/jamia.M2373.

Avanti, A., Jung, K., Harman, S., et al. (2018). Improving palliative care with deep learning. *BMC Medical Informatics and Decision Making, 18*(Suppl 4), 122.

Baars, M. J. H., Henneman, L., & Ten Kate, L. P. (2005). Deficiency of knowledge of genetics and genetic tests among general practitioners, gynecologists, and pediatricians: A global problem. *Genetics in Medicine, 7*(9), 605.

Barnett, G. O., Cimino, J. J., Hupp, J. A., et al. (1987). *Journal of the American Medical Association, 258*(1), 67–74. PMID: 32953167.

Baron, R. J. (2010). What's keeping us so busy in primary care? A snapshot from one practice. *The New England Journal of Medicine, 362*(17), 1632–1636. https://doi.org/10.1056/NEJMon0910793.

Bates, D. W., & Gawande, A. A. (2003). Improving safety with information technology. *New England Journal of Medicine, 348*(25), 2526–2534. https://doi.org/10.1056/NEJMsa020847.

Beasley, J. W., Wetterneck, T. B., Temte, J., et al. (2011). Information chaos in primary care: Implications for physician performance and patient safety. *Journal of American Board of Family Medicine, 24*(6), 745–751.

Bender, D., & Sartipi, K. (2013). HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. *2013 IEEE 26th international symposium on computer-based medical systems (CBMS)*, IEEE, pp. 326–331.

Bleich, H. (1972). Computer-based consultation: Electrolyte and acid-base disorders. *American Journal of Medicine, 53*, 285–291.

Bloomrosen, M., Starren, J., Lorenzi, N. M., et al. (2011). Anticipating and addressing the unintended consequences of health IT and policy: A report from the AMIA 2009 Health Policy Meeting. *Journal of the American Medical Informatics Association, 18*(1), 82–90.

Blumenthal, D. (2009). Stimulating the adoption of health information technology. *New England Journal of Medicine, 360*(15), 1477–1479. https://doi.org/10.1056/NEJMp0901592.

Blumenthal, D. (2010). Launching HITECH. *New England Journal of Medicine, 362*(5), 382–385.

Blumenthal, D., & Glaser, J. P. (2007). Information technology comes to medicine. *New England Journal of Medicine, 356*(24), 2527–2534. https://doi.org/10.1056/NEJMhpr066212.

Blumenthal, D., & Tavenner, M. (2010). The "meaningful use" regulation for electronic health records. *New England Journal of Medicine, 363*(6), 501–504. https://doi.org/10.1056/NEJMp1006114.

Bodenreider, O., & Stevens, R. (2006). Bio-ontologies: Current trends and future directions. *Briefings in Bioinformatics, 7*(3), 256–274. https://doi.org/10.1093/bib/bbl027.

Bright, T. J., Wong, A., Dhurjati, R., et al. (2012). Effect of clinical decision-support systems: A systematic review. *Annals of Internal Medicine, 157*(1), 29–43. https://doi.org/10.1059/0003-4819-157-1-201207030-00450.

Bu, D., Pan, E., Walker, J., et al. (2007). Benefits of information technology-enabled diabetes management. *Diabetes Care, 30*(5), 1137–1142.

Buchanan, B. G., & Shortliffe, E. H. (1984). *Rule-based expert systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading: Addison-Wesley.

Chan, I. S., & Ginsburg, G. S. (2011). Personalized medicine: Progress and promise. *Annual Review of Genomics and Human Genetics, 12*, 217–244. https://doi.org/10.1146/annurev-genom-082410-101446.

Chaudhry, B., Wang, J., Wu, S., et al. (2006). Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine, 144*(10), 742–752.

Chused, A. E., Kuperman, G. J., & Stetson, P. D. (2008). Alert override reasons: A failure to communicate. *Proceeding of the AMIA Annual Symposium,* pp. 111–115.

Cimino, J. J., Li, J., Bakken, S., & Patel, V. (2002). Theoretical, empirical and practical approaches to resolving unmet information needs of clinical information system users. *Proceeding of the AMIA Annual Symposium*, pp. 170–174.

Clancey, W. J. (1983). The epistemology of a rule-based system: A framework for explanation. *Artificial Intelligence, 20*(3), 215–251.

Clancy, C. M., Anderson, K. M., & White, P. J. (2009). Investing in health information infrastructure: Can it help achieve health reform? *Health Affairs, 28*(2), 478–482. https://doi.org/10.1377/hlthaff.28.2.478.

Colombet, I., Bura-Rivière, A., Chatila, R., et al. (2004). Personalized versus non-personalized computerized decision support system to increase therapeutic qual-

ity control of oral anticoagulant therapy: An alternating time series analysis. *BMC Health Services Research, 4*(1), 27. https://doi.org/10.1186/1472-6963-4-27.

Covell, D., Uman, G., & Manning, P. R. (1985). Information needs in office practice: Are they being met? *Annals of Internal Medicine, 103*(4), 596–599.

De Dombal, F. T., Leaper, D. J., Staniland, J. R., McCann, A. P., & Horrocks, J. C. (1972). Computer-aided diagnosis of acute abdominal pain. *British Medical Journal, 1*, 376–380.

Decker, S. L., Jamoom, E. W., & Sisk, J. E. (2012). Physicians in nonprimary care and small practices and those age 55 and older lag in adopting electronic health record systems. *Health Affairs, 31*(5), 1108–1114. https://doi.org/10.1377/hlthaff.2011.1121.

DeClerq, P. A., Blom, J. A., Korsten, H. H., et al. (2004). Approaches for creating computer-interpretable guidelines that facilitate decision support. *Artificial Intelligence in Medicine, 31*(1), 1–27.

DesRoches, C. M., Worzala, C., Joshi, M. S., et al. (2012). Small, nonteaching, and rural hospitals continue to be slow in adopting electronic health record systems. *Health Affairs, 31*(5), 1092–1099. https://doi.org/10.1377/hlthaff.2012.0153.

Dixon, B. E., Simonaitis, L., Goldberg, H. S., et al. (2013). A pilot study of distributed knowledge management and clinical decision support in the cloud. *Artificial Intelligence in Medicine, 59*(1), 45–53.

Dolin, R. H., Boxwalla, A., & Shalaby, J. (2018). A Pharmacogenomics clinical decision support service based on FHIR and CDS Hooks. *Methods of Information in Medicine, 57*(S 02), e115–e123.

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*, 129–136.

Fargher, E. A., Eddy, C., Newman, W., et al. (2007). Patients' and healthcare professionals' views on pharmacogenetic testing and its future delivery in the NHS. *Pharmacogenomics, 8*(11), 1511–1519. https://doi.org/10.2217/14622416.8.11.1511.

Flexner, A. (1910). *Medical education in the United States and Canada: A report to the Carnegie Foundation for the Advancement of Teaching*, Bulletin No. 4., New York City: The Carnegie Foundation for the Advancement of Teaching, OCLC 9795002.

Frankovich, J., Longhurst, C. A., & Sutherland, S. M. (2011). Evidence-based medicine in the EMR era. *The New England Journal of Medicine, 365*(19), 1758–1759.

Gawande, A. (2009). *The checklist manifesto: How to get things right*. New York: Metropolitan Books.

Ginsburg, G. S., & Willard, H. F. (2009). Genomic and personalized medicine: Foundations and applications. *Translational Research, 154*(6), 277–287. https://doi.org/10.1016/j.trsl.2009.09.005.

Goldberg, H. S., Paterno, M. D., Rocha, B. H., et al. (2014). A highly scalable, interoperable clinical decision support service. *Journal of the American Medical Informatics Association, 21*(e1), e55–e62.

Goldstein, M. K., Hoffman, B. B., Coleman, R. W., et al. (2000). Implementing clinical practice guidelines while taking account of evidence: ATHENA, an easily modifiable decision-support system for management of hypertension in primary care. In *Proceedings of the AMIA Fall Symposium* (pp. 300–304). Philadelphia: Hanley & Belfus.

Gorry, G. A., & Barnett, G. O. (1968). Sequential diagnosis by computer. *JAMA, 205*(12), 849–854.

Greenes, R., Bloomrosen, M., Brown-Connolly, N. E., et al. (2010). The Morningside Initiative: Collaborative development of a knowledge repository to accelerate adoption of clinical decision support. *The Open Medical Informatics Journal, 4*, 278–290.

Greenes, R. A., Ed. (2014). Clinical Decision Support: The Road to Broad Adoption. New York, Elsevier.

Greenes, R. A., Bates, D. W., Kawamoto, K., et al. (2018). Clinical decision support models and frameworks: Seeking to address research issues underlying implementation successes and failures. *Journal of Biomedical Informatics, 78*, 134–143.

Grimm, R. H., Shimoni, K., Harlan, W. R., et al. (1975). Evaluation of patient-care protocol use by various providers. *New England Journal of Medicine, 282*(10), 507–511.

Harrison, M. I., Koppel, R., & Bar-Lev, S. (2007). Unintended consequences of information technologies in health care—an interactive sociotechnical analysis. *Journal of the American Medical Informatics Association, 14*(5), 542–549. https://doi.org/10.1197/jamia.M2384.

Haynes, R. B., Wilczynski, N. L., & Computerized Clinical Decision Support System (CCDSS) Systematic Review Team. (2010). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: Methods of a decision-maker-researcher partnership systematic review. *Implementation Science, 5*(1), 12. https://doi.org/10.1186/1748-5908-5-12.

HealthMgt. (2015). HL7 launches Argonaut Project to advance FHIR interoperability standard. *Health Management Technology, 36*(2), 26.

Heckerman, D., & Horvitz, E. (1986). The myth of modularity in rule-based systems for reasoning with uncertainty. In J. Lemmer & L. Kanal (Eds.), *Uncertainty in artificial intelligence 2*. Amersterdam: North Holland.

Hickam, D. H., et al. (1985). "The treatment advice of a computer-based cancer chemotherapy protocol advisor." *Ann Intern Med. 103*(6 (Pt 1)): 928–936.

Hoffman, M. A., & Williams, M. S. (2011). Electronic medical records and personalized medicine. *Human Genetics, 130*(1), 33–39. https://doi.org/10.1007/s00439-011-0992-y.

Hong, N., Li, D., Yu, Y., et al. (2016). A computational framework for converting textual clinical diagnostic criteria into the quality data model. *Journal of Biomedical Informatics, 63*, 11–21.

Hripscak, G., Duke, J. D., Shah, N. H., et al. (2015). Observational Health Data Sciences and Informatics (OHDSI): Opportunities for observational research-

ers. In I. N. Sarkar et al. (Eds.), *Proceedings of Medinfo 2015*. IOS Press. https://doi.org/10.3233/978-1-61499-564-7-574.

Institute of Medicine. (2011a). *Health IT and patient safety: Building safer systems for better care*. Washington, DC: The National Academies Press.

Institute of Medicine. (2011b). *Digital infrastructure for the learning health system: The foundation for continuous improvement in health and healthcare. Workshop Series Summary*. Washington, DC: The National Academies Press.

Isaac, T., Weissman, J., Davis, R., et al. (2009). Overrides of medication alerts in ambulatory care. *Archives of Internal Medicine, 169*(3), 305.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R, (Springer Texts in Statistics)*. New York: Springer.

Jiang, G., Solbrig, H. R., Pathak, J., et al. (2015). Developing a standards-based information model for representing computable diagnostic criteria: A feasibility study of the NQF Quality Data Model. *Studies in Health Technology and Informatics, 216*, 1097.

Jiang, G., Kiefer, R. C., Sharma, D. K., et al. (2017a). A consensus-based approach for harmonizing the OHDSI common data model with HL7 FHIR. *Studies in Health Technology and Informatics, 245*, 887–891.

Jiang, G., Kiefer, R. C., Prud'hommeaux, E., et al. (2017b). Building interoperable FHIR-based vocabulary mapping services: A case study of OHDSI vocabularies and mappings. *Studies in Health Technology and Informatics, 245*, 1327.

Jing, X., Cimino, J. J., & Del Fiol, G. (2015). Usability and acceptance of the librarian infobutton tailoring environment: An open access online knowledge capture, management, and configuration tool for OpenInfobutton. *Journal of Medical Internet Research, 17*(11), e272.

Kahn, C. E., Roberts, L. M., Shaffer, K. A., et al. (1997). Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Computers in Biology and Medicine, 27*, 19–29.

Karsh, B.-T., Weinger, M. B., Abbott, P. A., et al. (2010). Health information technology: Fallacies and sober realities. *Journal of the American Medical Informatics Association, 17*(6), 617–623. https://doi.org/10.1136/jamia.2010.005637.

Kaushal, R., Jha, A., Franz, C., et al. (2006). Return on investment for a computerized physician order entry system. *Journal of the American Medical Informatics Association, 13*(3), 261–266.

Kawamoto, K., Del Fiol, G., Strasberg, H. R., et al. (2010). Multi-national, multi-institutional analysis of clinical decision support data needs to inform development of the HL7 Virtual Medical Record standard. *AMIA Annual Symposium Proceedings, 2010*, 377–381.

Kawamoto, K., et al. (2013). "Key principles for a national clinical decision support knowledge sharing framework: synthesis of insights from leading subject matter experts." *J Am Med Inform Assoc. 20*(1): 199–207.

Knaus, W. A., Wagner, D. P., & Lynn, J. (1991). Short-term mortality predictions for critically ill hospitalized adults: Science and ethics. *Science, 254*, 389–394.

Kohn, L., Corrigan, J., & Donaldson, M. (Eds.). (2002). *To err is human: Building a safer health system*. Institute of Medicine. Washington, DC: National Academy Press.

Komaroff, A., Black, W., & Flatley, M. (1974). Protocols for physician assistants: Management of diabetes and hypertension. *New England Journal of Medicine, 290*, 370–312.

Langheier, J. M., & Snyderman, R. (2004). Prospective medicine: The role for genomics in personalized health planning. *Pharmacogenomics, 5*(1), 1–8. https://doi.org/10.1517/phgs.5.1.1.25685.

Leape, L. L. (1994). Error in medicine. *JAMA: The Journal of the American Medical Association, 272*(23), 1851–1857. https://doi.org/10.1001/jama.1994.03520230061039.

Ledley, R., & Lusted, L. (1959). Reasoning foundations of medical diagnosis. *Science, 130*, 9–21.

Lindgren, H. (2008). Decision support system supporting clinical reasoning process - an evaluation study in dementia care. *Studies in Health Technology and Informatics, 136*, 315–320.

Lomatan, E. A., Meadows, G., Michaels, M., et al. (2019). To share is human! CDS Connect: A growing national repository of shareable, interoperable clinical decision support. *Proceedings of the AMIA Spring Congress*.

Longhurst, C. A., Harrington, R. A., & Shah, N. H. (2014). A 'green button' for using aggregate patient data at the point of care. *Health Affairs, 33*(7), 1229–1235.

Mandel, J. C., Kreda, D. A., Mandl, J., et al. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association, 23*(5), 899–908.

Mandl, K. D., & Kohane, I. S. (2012). Escaping the EHR trap — the future of health IT. *The New England Journal of Medicine, 366*, 2240–2242.

Marcial, L. H., Richardson, J. E., Lasater, B., et al. (2018). The imperative for patient-centered clinical decision support. *eGEMs, 6*(1), 12.

McClellan, M. (2015). Accountable care organizations and evidence-based payment reform. *JAMA: The Journal of the American Medical Association, 313*(21), 2128–2123.

Melnick, E. R., Dyrbye, L. N., Sinsky, C. A., et al. (2019). The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clinic Proceedings*. https://doi.org/10.1016/j.mayocp.2019.09.024.

Middleton, B., Shwe, M. A., Heckerman, D. E., et al. (1991). Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base.

Part II: Evaluation of diagnostic performance. *Meth Inf Medicine, 4*, 256–267.

Middleton, B., Anderson, J., Fletcher, J., et al. (1998). Use of the WWW for distributed knowledge engineering for an EMR: The KnowledgeBank concept. *Proceedings of the AMIA Annual Symposium,* pp. 126–130.

Middleton, B. (2009). "The clinical decision support consortium." *Stud Health Technol Inform. 150*, 26–30.

Musen, M. A. (1998). Domain ontologies in software engineering: Use of PROTÉGÉ with the EON architecture. *Methods of Information in Medicine, 37*(4–5), 540–550.

Musen, M. A. (2014). Knowledge representation. In I. N. Sarkar (Ed.), *Methods in biomedical informatics: A pragmatic approach*. Waltham, MA: Academic Press, Elsevier.

Musen, M. A. (2015). The Protégé project: A look back and a look forward. *AI Matters. Association of Computing Machinery Special Interest Group in Artificial Intelligence, 1*(4), 4–12. https://doi.org/10.1145/2557001.25757003.

Musen, M. A., Tu, S. W., Das, A. K., et al. (1996). EON: A component-based approach to automation of protocol-directed therapy. *Journal of the American Medical Informatics Association, 3*(6), 367–388.

National High Blood Pressure Education Program. (2004). *The Seventh Report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure.* National Heart, Lung, and Blood Institute, National Insitutes of Health.

National Research Council (US) Committee on Engaging the Computer Science Research Community in Health Care Informatics, Stead, W. W., & Lin, H. S. (2009). *Computational technology for effective health care: Immediate steps and strategic directions* (The National Academies Collection: Reports funded by National Institutes of Health). Washington, DC: National Academies Press (US).

Noy, N., Gao, Y., Jain, A., et al. (2019). Industry-scale knowledge graphs: Lessons and challenges. *Communications of the ACM, 62*(8), 36–43.

Odigie, E., Lacson, R., Raja, A., et al. (2019). Fast healthcare interoperability resources, clinical quality language, and systematized nomenclature of medicine-clinical terms in representing clinical evidence logic statements for the use of imaging procedures: Descriptive study. *JMIR Medical Informatics, 7*(2), e13590.

Osheroff, J. A., Teich, J. M., Middleton, B., et al. (2007). A roadmap for national action on clinical decision support. *Journal of the American Medical Informatics Association, 14*(2), 141–145. https://doi.org/10.1197/jamia.M2334.

Osheroff, J. A., Teich, J. A., Levick, D., et al. (2012). *Improving outcomes with clinical decision support: An implementer's guide*. Chicago, IL: HIMSS: Healthcare Information Management and Systems Society.

Overby, C. L., Tarczy-Hornoch, P., Hoath, J. I., et al. (2010). Feasibility of incorporating genomic knowledge into electronic medical records for pharmacogenomic clinical decision support. *BMC Bioinformatics, 11*(Suppl 9), S10. https://doi.org/10.1186/1471-2105-11-S9-S10.

Pageler, N. M., Longhurst, C. A., Wood, M., et al. (2014). Use of electronic medical record-enhanced checklist and electronic dashboard to decrease CLABSIs. *Pediatrics, 133*(3), e738–e746.

Patel, V. L., Arocha, J. F., & Zhang, J. (2010). Medical reasoning and thinking. In K. J. Holyoak & R. G. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 736–754). Oxford: Oxford University Press.

Pathak, J., Bailey, K. R., Beebe, C. E., et al. (2013). Normalization and standardization of electronic health records for high-throughput phenotyping: The SHARPn consortium. *Journal of the American Medical Informatics Association, 20*(e2), e341–e348.

Pauker, S. G., & Kassirer, J. P. (1981). Clinical decision analysis by computer. *Archives of Internal Medicine, 141*(13), 1831–1837.

Peleg, M., Tu, S., Bury, J., et al. (2003). Comparing computer-interpretable guideline models: A case-study approach. *Journal of the American Medical Informatics Association, 10*, 52–68. https://doi.org/10.1197/jamia.M1135.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the Conference on Knowledge Discovery in Databases,* San Francisco, CA.

Saria, S., Rajani, A. K., Gould, J., et al. (2010). Integration of early physiological responses predicts later illness severity in preterm infants. *Science Translational Medicine, 2*(48), 48ra65.

Schedlbauer, A., Prasad, V., Mulvaney, C., et al. (2009). What evidence supports the use of computerized alerts and prompts to improve clinicians' prescribing behavior? *Journal of the American Medical Informatics Association, 16*(4), 531–538. https://doi.org/10.1197/jamia.M2910.

Schnipper, J. L., Linder, J. A., Palchuk, M. B., et al. (2008). "Smart Forms" in an Electronic Medical Record: Documentation-based clinical decision support to improve disease management. *Journal of the American Medical Informatics Association, 15*(4), 513–523. https://doi.org/10.1197/jamia.M2501.

Schwartz, W. B. (1970). Medicine and the computer: The promise and problems of change. *New England Journal of Medicine, 283*(23), 1257–1264.

Shachter, R. D. (1986). Evaluating influence diagrams. *Operations Research, 34*(6), 871–882. https://doi.org/10.1287/opre.34.6.871.

Shah, N. R., Seger, A. C., Seger, D. L., et al. (2006). Improving acceptance of computerized prescribing alerts in ambulatory care. *Journal of the American Medical Informatics Association : JAMIA, 13*(1), 5–11. https://doi.org/10.1197/jamia.M1868.

Shahar, Y., & Musen, M. A. (1996). Knowledge-based temporal abstractions in clinical domains. *Artificial Intelligence in Medicine, 8*(3), 267–298.

Shiffman, R. N., Karras, B. T., Agrawal, A., et al. (2000). GEM: A proposal for a more comprehensive guideline document model using XML. *Journal of the American Medical Informatics Association, 7*(5), 488–498.

Shortliffe, E. H. (1976). *Computer-based medical consultations: MYCIN*. New York: Elsevier/North Holland.

Shortliffe, E. H., Buchanan, B. G., & Feigenbaum, E. (1979). Knowledge engineering for medical decision making: A review of computer-based clinical decision aids. *Proceedings of the IEEE, 67*, 1207–1224.

Shwe, M. A., Middleton, B., Heckerman, D. E., et al. (1991). Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. I. The probabilistic model and inference algorithms. *Methods of Information in Medicine, 30*(4), 241–255.

Sinsky, C. A., & Privitera, M. R. (2018). Creating a "manageable cockpit" for clinicians: A shared responsibility. *JAMA Internal Medicine, 178*(6), 741–742.

Sittig, D., & Singh, H. (2009). Eight rights of safe electronic health record use. *JAMA, 302*(10), 1111–1113.

Sittig, D. F., & Singh, H. (2011). Defining health information technology-related errors: New developments since to err is human. *Archives of Internal Medicine, 171*(14), 1281–1284. https://doi.org/10.1001/archinternmed.2011.327.

Sordo, M., Boxwala, A. A., Ogunyemi, O., et al. (2004). Description and status update on GELLO: A proposed standardized object-oriented expression language for clinical decision support. *Studies in Health Technology and Informatics, 107*, 164–168.

Strom, B. L., Schinnar, R., Aberra, F., et al. (2010). Unintended effects of a computerized physician order entry nearly hard-stop alert to prevent a drug interaction: A randomized controlled trial. *Archives of Internal Medicine, 170*(17), 1578–1583. https://doi.org/10.1001/archinternmed.2010.324.

Tang, P. C., Fafchamps, D., & Shortliffe, E. H. (1994). Traditional medical records as a source of clinical data in the outpatient setting. *Proceedings of the Annual Symposium on Computer Applications in Medical Care*, pp. 575–579.

Tang, P. C., Jaworski, M. A., Fellencer, C. A., et al. (1996). Clinician information activities in diverse ambulatory care practices. *Proceedings of the AMIA Annual Fall Symposium.* pp. 12–16.

Trafton, J., Martins, S., Michel, M., et al. (2010). Evaluation of the acceptability and usability of a decision support system to encourage safe and effective use of opioid therapy for chronic, noncancer pain by primary care providers. *Pain Medicine, 11*(4), 575–585.

Ullman-Cullere, M. H., & Mathew, J. P. (2011). Emerging landscape of genomics in the Electronic Health Record for personalized medicine. *Human Mutation, 32*(5), 512–516. https://doi.org/10.1002/humu.21456.

van der Lei, J., Musen, M. A., van der Does, E., et al. (1991). Comparison of computer-aided and human review of general practitioners' management of hypertension. *Lancet, 338*(8781), 1504–1508.

van der Sijs, H., Aarts, J., Vulto, A., et al. (2006). Overriding of drug safety alerts in computerized physician order entry. *Journal of the American Medical Informatics Association, 13*(2), 138–147. https://doi.org/10.1197/jamia.M1809.

Wang, S. J., Middleton, B., Prosser, L. A., et al. (2003). A cost-benefit analysis of electronic medical records in primary care. *The American Journal of Medicine, 114*(5), 397–403.

Wang, J. K., Hom, J., Balasubramanian, S., et al. (2018). An evaluation of clinical order patterns machine-learned from clinician cohorts stratified by patient mortality outcomes. *Journal of Biomedical Informatics, 86*, 109–119.

Weingart, S. N., Toth, M., Sands, D. Z., et al. (2003). Physicians' decisions to override computerized drug alerts in primary care. *Archives of Internal Medicine, 163*(21), 2625–2631. https://doi.org/10.1001/archinte.163.21.2625.

Wright, A., Sittig, D. F., Ash, J. S., et al. (2009). Clinical decision support capabilities of commercially-available clinical information systems. *Journal of the American Medical Informatics Association, 16*(5), 637–644.

Wright, A., Sittig, D. F., Ash, J. A., et al. (2015). Lessons learned from implementing service-oriented clinical decision support at four sites: A qualitative study. *International Journal of Medical Informatics, 84*(11), 901–911.

Yu, V. L., Fagan, L. M., Wraith, S. M., et al. (1979a). Antimicrobial selection by a computer. A blinded evaluation by infectious disease experts. *Journal of the American Medical Association, 242*(12), 1279–1282.

Yu, V. L., Buchanan, B. G., Shortliffe, E. H., et al. (1979b). Evaluating the performance of a computer-based consultant. *Computer Programs in Biomedicine, 9*(1), 95–102.

Zhang, J., & Walji, M. F. (2011). TURF: Toward a unified framework of EHR usability. *Journal of Biomedical Informatics, 44*(6), 1056–1067.

Zimmerman, J. E., Kramer, A. A., McNair, D. S., et al. (2006). Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Critical Care Medicine, 34*(5), 1297–1310. https://doi.org/10.1097/01.CCM.0000215112.84523.F0.

Digital Technology in Health Science Education

Parvati Dev and Titus Schleyer

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_25

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- How can computers improve the delivery of in-class and self-learning, as well as in-practice learning?
- How can different approaches to learning be implemented using computers?
- How can simulations supplement students' exposure to clinical practice?
- What are the issues to be considered when developing computer-based educational programs?
- What are the significant barriers to widespread integration of computer-aided instruction into the medical curriculum?

25.1 Introduction

The application of digital technology to health science education is a sub-field of biomedical informatics. It includes the application of all aspects of information and computer technology to the content and delivery of education, as well as to research on the improvement and efficacy of education. Healthcare requires constant learning, with its practice in a multidisciplinary team environment in an information-rich world. Digital technology offers new approaches to learning that:

- increase engagement and retention of knowledge,
- allow personalization of knowledge delivery,
- enhance collaboration through connectivity,
- support learning any time and anywhere,
- make available the increasing volume of knowledge,
- support learning of evidence-based clinical practice, and
- enhance research through collection and analysis of large volumes of learner data.

In this chapter, we first discuss approaches to teaching and learning with digital technology. That section includes material about theories of learning, digital technologies for learning environments, and an overview of learner audiences. We then transition to digital learning systems, which includes learning management systems, learning content creation systems, **just-in-time learning** systems and performance support, usability and accessibility, interoperability standards, digital content and assessment of learning. Finally, we discuss future directions and challenges for the application of digital technology in health science education.

25.2 Approaches to Teaching and Learning with Digital Technology

The continual rapid increase in health sciences knowledge requires a shift in learning methods both by the health sciences student as well as the health professional. Decades earlier, memorization and recall of facts were a primary, and sufficient, learning goal. Current learning approaches require learning the basic concepts and methods of a discipline but, in addition, emphasize the ability to integrate knowledge and to solve problems in the context of everyday healthcare.

25.2.1 Theories of Learning

Understanding how digital technology can support learning in the health sciences begins with an appreciation of how people learn. This understanding provides a foundation for thinking about the learning process and how it is shaped by context, purpose, goals, complexity, and the diversity of learners.

At its most basic, learning involves a change in how a learner perceives and understands some part of their world. The term schema is commonly used, in **cognitive science**, to describe the cognitive frameworks the people use to organize the information they have and their beliefs about a particular concept, activity, or experience. Schemas help us quickly assess a situation and act

appropriately. When individuals encounter new information, they try to incorporate it into their existing schema to enhance their understanding.

If new information contradicts the learner's existing knowledge or beliefs, the learner adjusts in one of two ways. They may accept the new information as valid, and modify their schema accordingly. Alternatively, they judge the new information as invalid, unimportant, or irrelevant, and they do not adapt their schema to account for this new information. When strong existing schemas keep individuals from accounting for new, valid information it can be very difficult for them to make changes such as altering unsafe procedures or adopting new safety protocols. To encourage such knowledge and behavioral change, it is important to explicitly address existing misconceptions and support learner motivation to change.

Changes in schema presume that individuals actively construct their own meaning through the interactions they have with information, other people, and the environment around them. This **constructivist** approach argues that individuals are not blank slates and bring their own history and experience to every learning situation. It is also important to recognize that this construction is an inherently social and situated process. Often learners are learning together in classes, groups, or teams. The conversations and shared experiences with others provide interpersonal cognitive and affective context for the learning process, and can shape the direction and scope of knowledge construction.

Digital technology can support a variety of approaches to learning. A common approach is didactic teaching, a one-way transfer of information through lectures, presented online via technologies such as recorded digital video. This approach has the advantage that new, as well as remedial, material can be made available, with additional links to in-depth content. A considerable amount of the available digital content is didactic, though it is usually enhanced with various activities for active learning. Active learning approaches, on the other hand, focus on engaging learners in the learning process by having them interact with the content, with each other, and in reflection on what and how they are learning. Research has shown that instructional approaches that promote active learning consistently outperform transmission-only approaches at a statistically significant level (Freeman et al. 2014).

Flipping classes is a relatively recent strategy for encouraging active learning. Instructors flip the class when they provide the instruction, traditionally delivered through in-class lecture, online and the class time is devoted to active learning which replaces the majority of traditional homework. The homework becomes doing what needs to be done to prepare for the in-person class. Flipped classes are similar to hybrid or blended classes where the seat time that would be used for lecture is focused on active learning instead. Faculty engage with learners through case discussions, problem solving, and deep dives to further understand the content learned outside the classroom.

25.2.2 Digital Technologies for Learning Environments

Much, if not most, of today's learning content is delivered digitally. This is evident in the tools that are used within and outside of the classroom, for on-the-job training, in specialized learning facilities and elsewhere. In the classroom, learning content such as PowerPoint slides, Prezi presentations, websites, **simulations**, games and other digital media are often projected using digital projectors or delivered directly to the devices of learners, such as smartphones, tablets and laptops. Audience interaction methods, such as Web-based surveys, shared digital whiteboards or student/audience response systems, can help learners interact with the learning content, the instructor(s) or fellow learners.

The **classroom technologies** and device ecology accessible to teachers and learners can extend the learning experience seamlessly beyond the classroom. For instance, videoconferencing allows individuals to attend lectures remotely, raise their hand and ask a question. Depending on the teaching style, this remote participation may approximate face-to-face attendance fairly closely or fall

Fig. 25.1 A life-size reconstruction of a digital human is viewed in a horizontal computer screen or "digital table". With finger gestures, learners can identify structures, remove layers of tissue to make vessels, nerves and bones visible, or rotate the body. Additional functions at the edges of the table allow further viewing and measurement functions. Clinical cases, with anatomy reconstructed from radiologic images, such as CT and MRI, allow study of actual cases. (Courtesy of Anatomage Inc., with permission)

far short of it. Collaborative technologies, such as instant messaging, group **chats** or collaborative editing tools, can help bridge physical separation, and enable efficient and effective virtual group work.

Such collaborative work can also happen asynchronously and provide flexibility (within limits) to learner schedules. **Discussion lists**, **messaging boards** and collaboration sites enable episodic, time-independent contributions from and interactions with learners.

Some time ago, using advanced technologies, such as simulations, virtual reality and augmented reality, required a trip to a specialized facility, such as a **simulation center** or "virtual reality cave." However, with devices such as Oculus or HTC Vive, such experiences are now available almost anyplace.

Last, social media such as Facebook, Google Hangouts, Twitter, Snapchat and Instagram have found important uses in education, ranging from real-time updates on projects to sharing experiences on a field-trip.

25.3 Overview of Learner Audiences

As discussed above, education in medicine, nursing, pharmacy, dentistry and other health professions is shifting from a focus on knowledge acquisition to competency-based education (Englander et al. 2013). Educational technologies are well positioned to support this transition, but only if they take the specific context of learners and their goals into account. We therefore discuss learner audiences and their particular needs in the following sections.

25.3.1 Undergraduate and Graduate Health Care Professions Students

Basic science programs in medical schools were among the first to implement technology-supported learning. Visually rich content in anatomy, neuroanatomy and pathology was much more accessible on the computer than through the microscope or via the cadaver dissection room (Fig. 25.1). Excellent 3D learning programs for anatomy are available, such as Netter 3D Anatomy, Primal Pictures, VH Dissector, Anatomage Table,[1] and other products, providing ever more accurate visualization of the human in three dimensions. The use of microscopes in fields such as histology and pathology education has virtually disappeared. Interestingly, in many schools, the use of cadavers has seen a resurgence, both as an important learning tool and as a rite of passage into the health care profession.

1 Netter 3D Anatomy: ▶ http://netter3danatomy.com; Primal Pictures: ▶ https://primalpictures.com; VH Dissector: ▶ https://www.toltech.net; Anatomage: ▶ https://www.anatomage.com

Nursing schools have moved quickly to expand their use of technology in education. For instance, digitally-enhanced physical **mannequins** for simulation of realistic nursing **scenarios** are widely-used learning tools. Many nursing schools that used to share a simulation center with a medical school have found their own demand high enough to require building their own simulation centers.

Dental schools often share a part of their curriculum with medical schools, and as a result use the same or similar learning content. However, they also need specialized anatomical and simulation content for dental and craniofacial topics. 3D software for dental anatomy is used widely in pre-doctoral dentistry. Historically, simulation of dental procedures was practiced using physical objects such as chalk or plastic teeth, a practice that is still widespread. More recently, high fidelity digital simulators have been developed.

For many years, teaching hospitals had patients with interesting diagnostic problems such as "unexplained weight loss" or "fever of unknown origin". This environment allowed for thoughtful "visit rounds," at which the attending physician could tutor the students and house staff, who could then go to the library to research the subject. A patient stayed in the hospital for weeks as testing was pursued and the illness evolved. In the modern era of restricted insurance payments, managed care, and reduced length-of-stay, this opportunity for learning in the hospital environment, has vanished for most junior students. The typical patient in today's teaching hospital is multi-morbid, usually elderly, and commonly acutely ill. The emphasis is on short stays, with diagnostic problems handled on an **outpatient** basis, and diseases evolving at home or in chronic care facilities. Thus, the medical student is faced with fewer diagnostic challenges suited to their level of knowledge, and has little opportunity to see the evolution of a patient's illness over time.

One response of medical educators has been to try to move teaching to the outpatient setting; another has been to use **problem-based learning** and computer-simulated **virtual patients**. *Problem-based learning (PBL)* is a pedagogical approach where each small group of students is given a clinical problem and they engage in discussion to develop an understanding of the problem, identify relevant knowledge, seek required knowledge using online and library research, discuss and challenge each other's interpretations, and settle on a solution to the problem. PBL is widely used in undergraduate clinical learning, teaching students self-directed learning, reflection, and teamwork. An interesting analysis of PBL by a student is available at Chang (2016).

Computer-simulated patients allow a full range of diseases to be presented and allow the learner to follow the course of an illness over any appropriate time period. Faculty can decide what clinical material must be seen and can use the computer to ensure that this core curriculum is delivered. Moreover, with the use of an "indestructible patient," the learner can take full responsibility for decision making, without concern over harming an actual patient by making mistakes. These simulated patients may be fully virtual, or may be computer-enhanced physical mannequins.

25.3.2 Practicing Health Care Providers

Health Sciences education does not stop after the completion of formal training. The science of medicine advances at such a rapid rate that much of what is taught rapidly becomes obsolete, and it has become obligatory for health professionals to be lifelong learners, both for their own satisfaction and, increasingly, as a formal requirement to maintain their professional certification. Therefore, online courses and online certification examinations have become increasingly common for maintenance of certification. An additional advantage of online certification is the automatic tracking of learner performance, and the accompanying automatic generation of certificates and institutional compliance reports.

Health professionals are also required to demonstrate clinical competence through their performance in simulated clinical scenarios. **Advanced Trauma Life Support** and **Advanced Cardiac Life Support** are some of the areas in which clinical competence is dem-

Fig. 25.2 A screenshot of the SimSTAT simulation used for Maintenance of Certification by the American Society of Anesthesiologists. In a simulation of an operating room, the anesthesiologist ventilates the patient who is lying on the operating table. The simulation is viewed by the learner on a computer screen while the learner plays the role of the anesthesiologist. The learner guides the on-screen anesthesiologist to care for the unconscious patient by clicking on desired interactions, such as the equipment in the room or the icons at the bottom of the screen. Through these interactions, the learner can control the level of sedation, give medications, fluids, and gases, and monitor the patient's physiologic status. The inset screen on the top left is a monitor for the patient's vital signs. The inset screen on the top right is the display from the anesthesia machine. (Courtesy of CAE Healthcare, Inc., with permission)

onstrated through actual participation in onsite scenarios. Online simulation of these and other scenarios can be used as preparation for testing in a live, crisis scenario. Some specialties, such as anesthesiology, have developed sufficiently rich online simulations in their specialty, with real time assessment, that they can reduce the requirement for use of live scenarios (Fig. 25.2).

25.3.3 Patients, Caregivers, and the Public

For those outside the health science professions, there is no systematic way to learn how to be a knowledgeable patient or home-based caregiver, or how to effectively communicate with a health care provider. For the provider interacting with the patient, the need to understand the patient accompanies the need to problem-solve. These changes in health care delivery are occurring slowly, and technology, particularly simulation and role playing, will be part of this change (Zaharias 2018).

Meanwhile, healthcare information is widely available to the general public, with the most reliable information being on web sites affiliated with federal library resources such as **Medline Plus**, academic organizations, professional societies or federal health agencies. Online courses, both free and paid, are also available, many from traditional universities or other online education organizations. Interactive learning applications, however, are not widely available to the public though they are able to provide more engaging learning.[2]

2 ▶ https://www.nih.gov/health-information; ▶ https://medlineplus.gov; ▶ https://www.cdc.gov; ▶ https://www.medscape.com; ▶ http://www.diabetes.org; ▶ https://www.heart.org/en/health-topics; ▶ https://my.clevelandclinic.org/health; ▶ https://www.mayoclinic.org

Improving the patient's health literacy has become an important approach to providing higher quality health care. Failure to comply with medication regimes, exercise plans, or hospital discharge instructions are a major cause of return visits to the hospital or clinic. Patient and family caregiver education, leading to better understanding of clinical instructions, could result in a more effective partnership between the patient and the healthcare provider (Nelson 2016) (see ▶ Chap. 11). Online learning resources are one approach to alleviating the compliance problem.

In the next section, we take a close look at how digital learning content is created and delivered.

25.4 Digital Learning Systems

25.4.1 Learning Content Management Systems

A **Learning Content Management System** (LCMS) is a software platform that allows learning content creators, such as faculty, to create, manage, host and track changes in digital learning content. Prior to the development of LCMSs, educational content creators had to assemble several separate and disparate items to create rich, engaging learning content. LCMSs, on the other hand, are one-stop platforms that integrate a wide variety of tools for content creation. They overlap with Learning Management Systems (LMS, described below) in that both support content hosting and delivery. However, LCMSs specialize in tools to create, manage and update content.

Personally created course content may be a web site or a blog, created by a faculty member on any of a range of web site or blog creation tools. Other personal content creation tools include tools for quiz item development, capturing video lectures or demonstrations, and creation of interactive animations or games. The only requirement is that the content files be compatible with the LMS, so that the content can be uploaded to the LMS and deployed to all learners without any need for special integration programming.

Collaborative content creation can be another powerful approach to learning. As opposed to structured content that is created by faculty or a similar creator, collaborative, learner-generated content is informal and created on the fly, for instance in a discussion forum. Structured discussion groups which encourage students to provide their thinking on the discussion questions, and to comment on content from other learners, are useful tools to support learning through active participation, argument, and reflection. Video conferencing tools, such as Zoom and Skype, support real-time discussion and collaboration, and can be content creation tools if there is a repository of the content.

When the content to be created is sufficiently complicated, requires strict adherence to organizational policy, or requires a range of skill sets and significant expense, it becomes necessary to approach it at the level of the enterprise. For example, the **American Heart Association** has developed courses for cardiac life support that are required training in the United States and many other countries.[3] These programs are created by teams, with each person providing a different skill, such as graphic design, programming, or content knowledge, rather than by an informal collaboration of learners with similar skills.

25.4.2 Learning Management Systems

A **Learning Management System** (LMS) is a repository of learning content, an interface for delivering courses and content to learners, and a platform for the course creator or administrator to track learner engagement and performance. From the learner's viewpoint, the LMS provides a single login access to all courses that they may need. Once within

3 American Heart Association's ACLS, BLS, and PALS courses: ▶ https://elearning.heart.org/courses

a course, the learner can access content, such as text, videos, quizzes, games, handouts and assignments. The LMS may include administration features for the learner to select courses, register or join a wait list, pay for each course, and track their grades. It may also include resource sharing and collaboration between learners. From the faculty's point of view, the LMS allows uploading and modification of course content, as well as a dashboard for viewing the performance of learners, groups and classes as a whole. LMS features may include various statistics, such as usage of course components or the performance of the class on individual test items.

Higher education institutions provide a relatively structured curriculum to well-defined learner populations. Their needs typically are served by educational LMS applications such as Blackboard Learn, D2L Brightspace, Moodle, or Canvas (Dahlstrom et al. 2014). Medical centers and corporations often use LMSs that are more suited to corporate needs. Typically, the most important requirement of such an LMS is tracking learner compliance with required training, and export of reports for regulatory and accreditation purposes. There are numerous enterprise-oriented LMSs. Some common ones are Captivate Prime, TalentLMS, and Totara, but market leadership of LMSs changes continually.

25.4.3 Just-in-Time Learning Systems and Performance Support

Learning also occurs outside of formal learning contexts. This "just-in-time learning" happens on demand, for instance when learners need instant information at a critical moment, or something goes wrong and they need to know what to do next. Instant information can provide immediate help, or *performance support*, but can also be considered a learning opportunity. A particularly powerful approach is to make these tools accessible when and where they are likely needed, for instance in online help areas of electronic medical records.

Examples of performance support or just-in-time learning tools include:

- job aids, such as check lists, quick reference cards, and handouts;
- protocols and templates, such as the SBAR (Situation, Background, Assessment, Recommendation and Request) technique for communicating critical information;
- resource or policy documents;
- video and audio recordings, such as brief demonstrations of care procedures, particularly helpful for home-based care providers; and
- animations, simulations, and learning modules that include brief instructions, demonstrations, or explanations.

25.4.4 Interoperability Standards

The education enterprise process includes many parts, such as content, curriculum, LMSs, learner profiles, assessment, certification, and others. These parts are often supported by different tools and platforms that need to work together seamlessly. To enable this seamless environment, tools and platforms need to be interoperable, without the need for custom programming for integration. In this chapter, we discuss some of the interoperability standards in education (see ► Chap. 7).

Historically, the commonly used standard for such learning object interoperability was the Shareable Content Object Reference Model (**SCORM**).[4] It defines how a content object (a course) should be packaged and described, how it should be launched, and how data should be communicated between the LMS and the content package. A SCORM-compatible content object can be published to and played back from any SCORM-compatible LMS. SCORM can report on course completion and time spent. SCORM was last updated in 2009 and, as a standard, has not kept up with changing technology. However, it is still one of the most widely used standards for learning object interoperability.

4 SCORM: ► https://scorm.com/

The newer **xAPI** standard (aka TinCan API)[5] is much more robust in terms of analytics and mobile deployment. The drawback to xAPI is that it does not integrate with older LMSs and tends to be costly to deploy, often requiring professional development assistance.

With the rapid growth in the use of learning management systems in higher and continuing education and workplace learning, it has become critical to have interoperability standards that provide integration of web-based learning objects and applications. Learning Tools Interoperability (LTI)[6] is a standard developed by the IMS Global Learning Consortium to provide a means for seamless and secure pass-through of student credentials and grades between the LMS and the external application. LTI tools are usually web-based applications written in a server-side language which can serve a variety of purposes. These include, but are not limited to, hosting and serving video with quizzing, providing access to interactive learning materials from textbook publishers, allowing learners to create media, use specialized programs, and collaborate in integrated development environments (IDE), whiteboarding and mind mapping applications, or videoconferencing.

Other interoperability standards address various education services. For example, the **Medbiquitous**[7] organization has developed a Curriculum Inventory Data Exchange Standard that is being used by the Association of American Medical Colleges to collect and collate curriculum data from all its medical schools and map these curricula to competency requirements. Other Medbiquitous standards include the Educational Achievement Standard to document learner competency used by numerous medical certification organizations, and the Virtual Patient Standard to enable exchange of virtual patient simulations across institutions.

There has also been an interest in the exchange of education components, specifically learning content modules, between institutions. A number of content collections have been developed, with the most well-known being MERLOT. Standardized descriptions of learning objects, known as Learning Object Metadata, were developed, but exchanging and repurposing individual learning objects did not become commonplace. However, an interesting by-product was a standardized way to describe items in a content collection, to manage a library of learning objects and to track learner use of those objects. Another example of a repository of shared learning resources is the **AAMC**'s MedEdPORTAL[8], which is peer-reviewed and contains both patient cases and clinical scenarios.

25.4.5 Usability and Access

Usability and access are important considerations in developing educational software (see ▶ Chap. 5). About 1 in 50 adults have some form of vision or hearing disability, and need alternate or augmented access to digital learning content. Two standards, Section 508 of federal law (508)[9] and the World Wide Web's Web Content Accessibility Guidelines (WCAG),[10] address use of digital content by people with disabilities.

Section 508 specifies that digital information provided by or to the government must be accessible if there is "no undue burden". In practice, designing accessible online content requires use of techniques available in current web design, such as the "Alt Text" tag for graphics, and indications to make the user interface more visible or audible. Adherence to Section 508 becomes more difficult in more complicated applications, such as 3D **immersive environments** and dynamic simulations,

5 xAPI: ▶ https://xapi.com/overview/
6 LTI: ▶ https://www.imsglobal.org/activity/learning-tools-interoperability
7 Medbiquitous standards: ▶ https://medbiq.org/standards
8 AAMC's MedEdPORTAL: ▶ https://www.mededportal.org
9 Section 508: ▶ https://www.section508.gov/manage/laws-and-policies
10 WCAG: ▶ https://www.w3.org/WAI/standards-guidelines/wcag/

requiring creative solutions to presentation and interface requirements.

WCAG is a set of formal guidelines on how to develop accessible web content. It does not address non-web digital content.

25.5 Digital Content

Digital content, unlike a typical textbook or lecture, can be interactive. Three levels of interactivity are typically possible:

- Level 1: The content includes text, graphics and video but interaction is primarily through clicking to move to the next chunk of content. This level may include simple quizzes such as multiple choice or true/false questions. Much of digital learning content consists of web pages or applications that incorporate this style of expository material.
- Level 2: The content includes multimedia such as audio, video and animations. The interactivity supports simple puzzles and games, like sorting and matching. The cost of development is higher than for Level 1, but the content is more engaging.
- Level 3: The content presented is very rich, including realistic three-dimensional environments and characters, or even immersive virtual reality (Fig. 25.3). Interaction happens through games or simulations, with the content evolving based on the choices and decisions made by the learner.

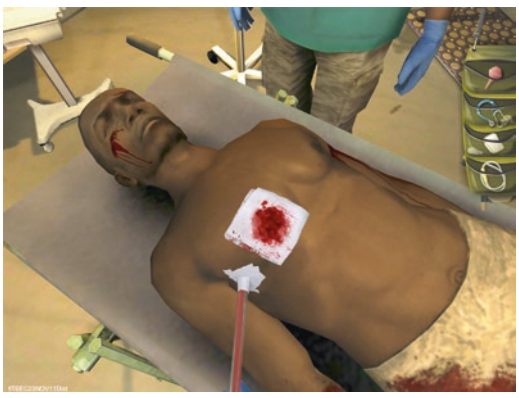


Fig. 25.3 A screenshot of a Level 3 interactive application, BattleCare. The learner can select tools from the medical kit on the right, and drag them onto the simulated patient to clean and compress the wound or to listen to heart and lung sounds. (Courtesy of Innovation in Learning, Inc., with permission)

25.5.1 Text/Image/Video Content

Much of digital learning content consists of web pages or applications that incorporate expository material, using text, graphics and video, and Level 1 interaction. Although much of the focus of computer-based teaching is on the more innovative uses of technology to expand the range of available teaching modalities, computers can be employed usefully to deliver didactic material, with the advantage of the removal of time and space limitations. For example, a professor can choose to record a lecture and to store the digitized video of the lecture as well as related slides and other teaching material, and upload this content to the institution's learning management system (see Section on LMSs.) This approach has the advantage that relevant background or remedial material can also be made available through links at specific points in the lecture. The ease of creating online video lectures has led numerous universities and corporations to provide libraries of recorded lectures for study by learners at their own convenience.

Many refinements have been developed that use technology to optimize the delivery of didactic or expository content. *Microlearning* is the presentation of brief segments of content, typically ranging from 5 to 15 minutes in duration. *Spaced repetition* is the repeated presentation of select content to optimize its retention. *Mastery learning* is a process of testing the learner for competence in a segment of content before they are allowed to progress to the next. The Khan Academy,[11] which includes healthcare content in its catalog, uses brief videos to teach single or small groups of concepts. In this micro-learning approach, the learner can select and complete

11 Khan Academy, Health and Medicine: ▶ https://www.khanacademy.org/science/health-and-medicine

topics with a limited investment of time, and can demonstrate mastery.

Massively Online Open Courses (MOOCs) bring free-to-view, world-class university courses to a global audience. The first major MOOC, Introduction to **Artificial Intelligence**, launched with an astonishing 160,000 subscribers. The structure of the first courses was similar to a typical university course, with lectures released at the same time as they would be taught to an in-person class on campus, along with assignments and final examinations that needed to be turned in on time. Course support was provided by peer support through student discussion groups. Some MOOCs now require fees for certification of completion of courses. Private companies have sprung up providing support to students around selected MOOC, an indication of the ecosystems that develop around interesting technologies. EdX is a MOOC delivery platform launched by the Massachusetts Institute of Technology and Stanford University. Coursera, Udacity and FutureLearn are major private MOOC platforms.[12]

25.5.2 Interactive Content

Teaching programs differ in the degree to which they impose structure on a teaching session. In general, drill-and-practice systems are highly structured. The system's responses to students' choices are specified in advance; students cannot control the course of an interaction directly. In contrast, other programs create an exploratory environment in which students can experiment without guidance or interference. For example, a neuroanatomy teaching program may provide a student with a fixed series of images and lessons on the brainstem, or it may allow a student to select a brain structure of interest, such as a tract, and to follow the structure up and down the brainstem, moving from image to image, observing how the location and size of the structure changes.

Each of these approaches has advantages and disadvantages. Fixed path learning programs ensure that no important fact or concept is missed, but they do not allow students to deviate from the prescribed course or to explore areas of special interest. Conversely, programs that provide an exploratory environment and allow students to choose any actions in any order encourage experimentation and self-discovery. Without structure or guidance, however, students may waste time following unproductive paths and may fail to learn important material, resulting in inefficient learning.

An example is the Tooth Atlas, used in dentistry. Understanding the three-dimensional structure of teeth is important for clinical dentistry. The key instructional objective of the program is to help students learn the complex external and internal anatomy of the variety of teeth in three dimensions. The rich, interactive 3D visualizations show teeth as they would be visually perceived as well as through very high resolution computed tomography scans, radiographs and physical cross-sections. The learners can rotate and section the computed models, and can control the transparency of each structure so as to study inter-relationships. While the visualization is highly exploratory, the embedded pedagogy is very structured, consisting of detailed textual quizzes with multiple-choice answers.

25.5.3 Games

A learning game places learning content within a digital video game. The game play experience engages and entertains the learner

12 Mulgan G, Joshi R. Clicks and mortarboards: how can higher education make the most of digital technology? November 2016. ▶ https://media.nesta.org.uk/documents/higher_education_and_technology_nov16_.pdf

while certain steps in the game instill desired content knowledge. In a learning game, the learning content is embedded in the game. Gamification, on the other hand, has elements such as a score, achievement badges, or a "leader board", to add excitement to an otherwise routine learning experience.[13]

A game has the following components: a goal, such as finding the best treatment for a patient; a setting, such as a three-dimensional rendition of an emergency department; game play, such as the information, tests, procedures, and medications available; and game mechanics, such as accessing game play elements by drawing up medication in a syringe or selecting a medication dosage from a menu. Successful resolution of a clinical problem can give the same satisfaction as an enjoyable game. However, to be considered a game, there need to be challenges, such as conquering "enemies" or accomplishing "quests", evolving clinical problems, or a restricted availability of supplies and personnel, that must be overcome, as well as a clear criterion of success. To be a learning game, actions during game play should result in learning, either by exposure of a nugget of information, by **feedback** from a mentor embedded in the game, or by trying alternative medical approaches to find an effective treatment.

Numerous learning games have been developed for all aspects of healthcare education but the evidence for their efficacy is not clear (Gorbanev et al. 2018). Funding for efficacy research is limited, and is one reason for the paucity of rigorous studies (Reed et al. 2007). Game development that has a clear learning goal, and has been informed by research during the design stage as well as during development of the game play, has been shown to be both engaging and an effective learning tool (Kato et al. 2008) (Box 25.1 and Fig. 25.4a and b).

13 A leader board is a list of players with the highest scores. Players compete to be among the high scorers.

25.5.4 Cases, Scenarios and Problem-Based Learning

The learner is presented with a story that includes a clinical problem. The presentation may be only in text, with text and graphics, in a near-realistic three-dimensional environment, or even in an immersive virtual environment, with correspondingly varying levels of interactivity. The learner's role may be constrained such that the learner knows who they represent, what resources are available, and what problem must be solved. Alternatively, the learner may be required to investigate the situation (examine the patient), define the problem, find any supporting resources (what imaging and laboratory tests are available or what learning resources are at hand) and guide the scenario to an end goal. As the learner proceeds, the scenario evolves on the computer, based on the actions taken and the progress of time.

Prognosis is a case-based program with over 500 cases covering most specialties (Fig. 25.5). Each case begins with a brief story of the clinical presentation. The learner must choose among available tests, diagnoses and treatments, and then receives feedback on the choices made, as well as the preferred or optimal choices. The presentation and interactivity are very simple, and the cases brief, but the engagement provided by the clinical puzzle has made this a popular program among medical students and residents.

An approach that combines the benefits of exploration with the constraint of a linear path through the material is one that breaks the evolving scenario into a series of short vignettes. A situation is presented, information and action options are available, and a decision must be made. Each decision triggers the presentation of the next vignette. This could lead to a branching story line but, usually, the next vignette presents the result of the best actions from the previous vignette. A scenario about a virtual patient could have vignettes that lead the learner through the steps of diagnosis, tests, and the course of treatment. This approach is commonly used in computer-based testing of

25

Box 25.1 "Re-Mission: Fighting Cancer with Video Games" (► http://www.re-mission2.org/)

Re-Mission 2 games help kids and young adults with cancer take on the fight of their lives. Based on scientific research, the games provide cancer support by giving players a sense of power and control, and encouraging treatment adherence. Each game puts players inside the human body to fight cancer with an arsenal of weapons and super-powers, like chemotherapy, antibiotics and the body's natural defenses. The game play parallels real-world strategies used to successfully destroy cancer and win.

Re-Mission 2 games are designed to:

- Motivate young cancer patients to stick to their treatments by boosting self-efficacy, fostering positive emotions and shifting attitudes about chemotherapy. These factors were key drivers of the positive health behavior seen with the original Re-Mission game
- Appeal to a broad audience by offering a variety of gameplay styles; and
- Tap into the popularity of casual games, playable in short bursts or at length, to provide cancer treatment support through fun, engaging play.

The games incorporate key insights from years of scientific studies and qualitative user research with adolescent and young adult cancer patients. An outcomes study showed that the original Re-Mission game improved treatment adherence and boosted self-efficacy in young cancer patients. The *Re-Mission Attitudes Study in the Brain* used fMRI technology to show how interactive gameplay impacts the brain to motivate positive behavior change (Kato et al. 2008).

Fig. 25.4 Screenshots from opening screens of Re: Mission 2. **a** In "Nanobot's Revenge," players use an ever-increasing arsenal of powerful chemo attacks to crush the cancerous forces of the Nuclear Tyrant, firing targeted treatments on a growing tumor to prevent cancer from escaping into the blood stream. **b** In "Nano Dropbot", the player continues to kill cancer cells but also learns to recruit healthy cells in the fight

clinical knowledge where assessment of learner performance would be extremely difficult if the interactions were completely unconstrained.

The ability of the computer to track and store the learner's actions allows post-processing and analysis of the tracked data. An interesting analytic capability is one that compares the performance of novice learners and experts to detect features that define expert information gathering or action sequences. Stevens et al. (1996) compared the information gathering and the conclusions of novices and experts on a set of immunological cases. Using neural nets to process the tracking data, they detected consistent differences in the problem solving approach of novices compared to experts. In particular, novices exhibited considerably more searching and lack of recognition of relevant information, while experts converged rapidly on a common set of information items. The potential of using such expert patterns of performance to educate novice learners has not been widely explored in the health sciences.

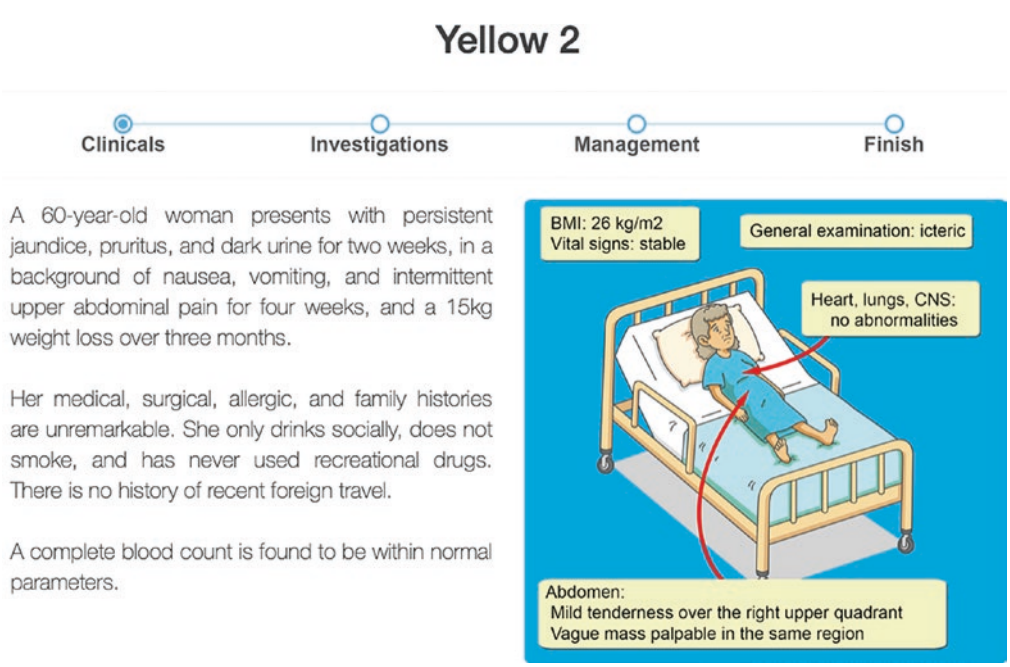


Fig. 25.5 Screenshot from the case-based program, Prognosis. The learner is presented with a summary of the case. The simple graphic presents physical examination information in a similar format for each case. The following screen offers options for laboratory, imaging and other investigation options. The learner selects the management plan and receives summary feedback on the success or failure of the plan. The case ends with a review of the disease background and optimal management, including relevant references. (Courtesy of Medical Joyworks, LLC, with permission)

25.5.5 Simulations – Virtual Patients

Clinical training has been shown to benefit from the use of simulations to engage the learner (Gaba 2004; Aebersold 2018; Jeffries 2005). Learning is most effective when the learner is engaged and actively involved in decision making. The use of a simulated patient presented by the computer can approximate the real-world experience of patient care and focuses the learner's attention on the subject being presented (Huang et al. 2007). The Association of American Medical Colleges has prepared an informational summary of the value of and the issues around issues of using simulation for education.[14]

Talbot et al. (2012) present an analysis of the range of presentation and interactivity available in Virtual Patients. These simulations may be either **static** or **dynamic**. Under the static simulation model, each case presents a patient who has a predefined problem and set of characteristics. At any point in the interaction, the student can interrupt data collection to ask the computer-consultant to display the differential diagnosis (given the information that has been collected so far) or to recommend a data collection strategy. The underlying case, however, remains static. Dynamic simulation programs, in contrast, simulate changes in patient state over time in response to students' diagnostic or therapeutic decisions. Thus, unlike in static simulations, the clinical manifestations of the dynamic simulation can be programmed to evolve as the student works through them. These programs help students to understand the relationships between actions (or inactions), and patients' clinical outcomes. To simulate a patient's response to intervention, the programs may explicitly model underlying physiological processes and use math-

14 ► https://www.aamc.org/download/373868/data/technologynowsimulationinmedicaleducation.pdf

Fig. 25.6 Screenshot from Timeout Training, a mobile training application. This illustrates a time-out dialog between the learner (a resident) and the nurse, prior to initiating a thoracocentesis intervention. The learner selects from one of the presented dialog options. Careful design of dialog options can help in learning nuances of dialog possibilities as well in proper sequencing of a dialog. (Courtesy of Innovation in Learning, Inc., with permission)

ematical models. An example of a dynamic simulation of a virtual patient is SimSTAT (see Fig. 25.2), an operating room simulation that is used by the American Society of Anesthesiologists to train practicing anesthesiologists in the diagnosis and management of crises in the operating room.

Virtual Patients can be as simple as Prognosis, described above, or can be richly complex, simulating a complete encounter with a patient in a clinic or hospital room. Simulation of a conversational interaction with the patient or another character can be an important aspect of learning using a virtual patient (Fig. 25.6).

25.5.6 Simulations – Procedures and Surgery

Procedure trainers or part task trainers have emerged as a new method of teaching, particularly in the teaching of surgical skills. This technology is still under development, and it is extremely demanding of computer and graphic performance. Early examples have focused on endoscopic surgery and laparoscopic surgery in which the surgeon manipulates tools and a camera inserted into the patient through a small incision. In the simulated environment, the surgeon manipulates the same tool controls, but these tools control simulated instruments that act on computer-graphic renderings of the operative field. Feedback systems inside the tools return pressure and other **haptic sensations** to the surgeon's hands, further increasing the realism of the surgical experience.

Commercially available trainers are now in use for many surgical procedures. For example, the 3D Systems company provides a line of Simbionix simulators for training in laparoscopy, endoscopy, and hysteroscopy.[15] Other simulators are now available for all levels of surgery, beginning with training in the basic operations of incision and suturing, and going all the way to robotic surgery.

25.5.7 Simulations – Mannequins

Physical simulations of a patient, in an authentic environment such as an operating room, have evolved into sophisticated learning environments. The patient is simulated by an artificial mannequin with internal mechanisms that produce the effect of a breathing human with a pulse, respiration, and other vital signs (Fig. 25.7). In high-end simulators, the mannequin can be given blood trans-

15 ▶ https://www.3dsystems.com/healthcare/medical-simulators

fusions or medication, and its physiological response changes based on these treatments. These human patient simulators are now used around the world both for skills training and for cognitive training such as crisis management or leadership in a team environment

Fig. 25.7 This plastic mannequin simulates many of the functions of a living patient, including eye opening and closing, breathing, heart rate and other vital signs. Gases, medications, and fluids can be administered to this mannequin, with resulting changes to its simulated vital signs. (Courtesy of Parvati Dev, with permission)

(Fig. 25.8). The environment can represent an operating room, a neonatal intensive care unit, a trauma center, or a physician's office. Teams of learners play roles such as surgeon, anesthetist, or nurse, and practice teamwork, crisis management, leadership, and other cognitive exercises.

A seminal study by Hayden et al. (2014) showed that 50% of nurse clinical training, in the Bachelor's program, could be replaced by training on mannequin simulators. This is particularly important both because of the range of cases that can be presented on the simulator, and because of the difficulty in obtaining clinical training time in hospitals.

25.5.8 Virtual Worlds

An extension of the physical human patient simulator is the **virtual world** simulation, with a virtual patient in a virtual operating room or emergency room. Learners are also present virtually, logging in from remote sites, to form a team to manage the virtual patient. Products such as 3DiTeams and Health TeamSpaces are

Fig. 25.8 Three-dimensional computer-generated virtual medical environments are used to present clinical scenarios to a team of learners. Each learner controls a character in the scenario and, through it, interacts with devices, the patient, and the other characters. Learning goals may include medical goals such as stabilization of the patient, communication goals such as learning to point out a problem to senior personnel, or team goals of leadership and delegation. (Courtesy of Innovation in Learning, Inc. with permission)

Fig. 25.9 A combined image depicting a learner wearing virtual reality glasses, and the scene visible to the learner. The learner feels she is inside the operating room, viewing the procedure. (Courtesy of SimTabs, LLC., with permission)

being used to construct and deliver team training in such virtual medical environments.[16]

25.5.9 Virtual Reality

The use of **virtual reality** glasses, along with spatially accurate sound and virtual hands, creates an immersive experience that surpasses the experience of a three-dimensional world as seen on a computer screen on the learner's desk. The learner feels truly "inside" the experience. The resulting immediacy is so real that it manifests itself through physiologic changes such as a speeding of the learner's pulse and a total lack of awareness of the actual physical surroundings (Fig. 25.9).

There are two types of virtual reality in use at present. One is reality as represented by a completely synthetic three-dimensional environment, within which the learner navigates and acts. The other is represented as a 360° video of a real environment within which the clinical action has been recorded. The video virtual reality is useful for didactic training about procedures, such as new surgical methods, where the learner has a front row view as though they were actually present in the operating room.

Examples of simulations using virtual reality have been demonstrated by many universities and organizations. VR simulations of surgery (https://ossovr.com) and patient examination (https://oxfordmedicalsimulation.com) are in use in medical and nursing curricula. There are a few studies examining the learning efficacy of VR simulations (Kyaw et al. 2019). It is likely that the realism of virtual reality, its "face validity", will result in its use in education even if rigorous efficacy studies are not available.

25.5.10 Augmented Reality

Augmented reality (AR) differs from virtual reality in that a real world is seen through the AR glasses while other information is overlaid on the world by the glasses. Information can be textual, such as heart rate and blood pressure data when looking at a physical mannequin. It can be graphic, such as an open wound seen overlaid on a person on a bed, simulating an injured patient. The AR overlay information changes depending on what is being viewed, creating a world of information on top of the real visible world.

Learning possibilities using AR are endless. A new nurse can walk into an empty operating room and "see" the contents of closets and drawers, thus being trained on the location of OR supplies. A nursing student can see a "pressure sore" evolve on the heel

16 ▶ https://anesthesiology.duke.edu/?page_id=825623, ▶ https://healthteamspaces.simtabs.com

of a real person because of pressure on the heel in the bed. A medical student can "scroll" through the electronic medical record as they talk with a simulated patient.

AR in medical education is in its infancy but its applications are expected to be wide-ranging.

25.5.11 3D Printed Physical Models

Three dimensional (3D) printing is a novel application of printing. Slice data from an object, such as a CT image of a bone, is used to print a layer of solid material, such as plastic. Subsequent image slices are used for printing a cumulative stack of plastic slices until the entire object has been printed. The advantage of sequential printing of slices, instead of carving the shape from a solid block of plastic, is that hollow portions of the original object can be printed as holes in the slice data. A second advantage is that very complex objects can be printed using this technology.

3D models are beginning to be used for learning. For undergraduate students, **cadaver** dissection, **plastinated** specimens, and dried bone have provided the physical specimens. 3D models add a new option. For healthcare practitioners, patient-specific 3D models help in planning procedures, but they also help in educating the patient about their upcoming surgery. In a recent systematic review on surgical planning for congenital heart disease, the authors found that 25% of the studies showed 3D printed models were useful in medical education for healthcare professionals, patients, caregivers, and medical students (Lau and Sun 2018).

25.6 Assessment of Learning

Assessment of student learning compares educational performance with educational goals. Ideally, the content used for assessment resides within a curriculum and an educational program that has a clear set of educational goals. Therefore, student learning is measured (assessed) against these overall goals as well as against the goals of the specific learning module. Without these goals, any assessment is without direction and its purpose may be a mystery to the learner.[17]

Assessment maybe **formative** (guiding future learning and promoting reflection) or **summative** (making a judgment about competence or qualification before being allowed to advance to the next level of study). Assessment can also be used by instructors and education program designers to identify whether the learning content can be improved, and brought closer to the identified learning goals (Epstein 2007).

Digital technology supports rich assessment both because of its ability to present many types of assessment tools, and because of its ability to track learner actions in great detail. Selected assessment methods are presented below.

25.6.1 Quizzes, Multiple Choice Questions, Flash Cards, Polls

Quizzes test the learner's knowledge and, depending on the quiz format, the learner's ability to solve problems. Quizzes can be presented as questions with single or multiple correct answers, or may require sorting and matching two sets of items. Digital technology simplifies the process of preparing, presenting and scoring quizzes, and can make them engaging and fun by adding imagery, animations and game-like success states.

Flash cards that present the question on one side of the card and the answer on the other can also be simulated using technology. The learner types their answer. Through simple word or phrase matching, the learner's answer is matched to the expected answer, and scored based on the level of match achieved. For more complex answers, some level of natural language processing is required.

Polls are particularly useful for an instructor to assess, in real time, the current status of

17 AAHE, ► https://www.oxy.edu/sites/default/files/assets/AAHE9Principles.pdf

learner understanding in the classroom. The poll question, and multiple answers, are displayed on the classroom screen. Each learner selects an answer on a smartphone or on a polling device. The poll responses are immediately collated and presented as a bar graph. If all or most of the learners select the correct answer, the instructor can assume that the topic has been understood. If the responses are distributed over two or more answers, then the instructor can pause to review the topic and clear up learner misconceptions or lack of understanding.

25.6.2 Branching Scenarios

A branching scenario is a structured approach to assessment using simulation. A mini-scenario or a choice of data resources (such as laboratory tests) is presented at each branch point, and the learner chooses one out of a set of available decisions or responses. One or more of the decisions may be correct. Based on each consecutive decision, the learner moves through a branched scenario and achieves a final outcome to the scenario. The learner can be assessed on each decision or on the final outcome. If the same material is presented in a learning mode, the learner may receive feedback about each decision.

25.6.3 Simulations

Simulations for assessment may use standardized patients (actors trained to represent patients), realistic interactive mannequins, on-screen simulations, or simulations presented in virtual reality. In all cases, technology can be used to track learner actions and to assess their performance (Ryall et al. 2016). In all except standardized patients, this tracking is built into the simulation, and can be extracted and analyzed for performance reporting. These more complex, scenario-based simulations, differ from branching simulations in that a large number of decision choices are available to the learner at every moment. Thus the simulation is a more realistic representation of a clinical situation but is also correspondingly more difficult to score for assessment (Dillon et al. 2002).

25.6.4 Intelligent Tutoring, Guidance, Feedback

Intelligent Tutoring Systems *(ITS)* differ from other technology-based learning systems in that they offer individualization of the learning experience based on the learner's performance while using the system (VanLehn 2011). Because of their architecture, continuous assessment of the learner is essential to the operation of ITSs, with **guidance** provided as needed, placing ITSs in the domain of formative assessment. Typically, an ITS is built to replicate one-on-one, personalized tutoring.

Modern ITSs use natural language for dialog between the learner and the tutor (Fig. 25.10). Conversational dialog is more likely to uncover learner misconceptions or gaps in knowledge. As the student responds to the tutor's questions, the response is compared to the expected response using statistical methods that compare the conceptual similarity of the two pieces of text. An example of a conversational tutoring system is Autotutor (Graesser et al. 2004), which has been used for learning domains ranging from physics and mathematics to training nurses for mass casualty triage (Shubeck et al. 2016).

25.6.5 Analytics

Understanding and improving the process and outcome of education requires applying metrics into many facets of the educational process. With digital technology, measurement and resulting data availability has increased steadily. At the same time, educational institutions and businesses are beginning to develop

Fig. 25.10 Screenshot of a conversation with an intelligent simulated tutor. The learner, a first responder, converses with a tutor who uses natural language to guide a student to give detailed answers using their own words. (Courtesy of Innovation in Learning, Inc., with permission)

methods to unlock potential uses of this vast amount of data.

A particularly desirable outcome is personalizing learning to each learner's needs. The many assessment methods described above can be applied to generate a profile of the learner's current knowledge state and to create a detailed list of topics to be learned. To implement such an adaptive system, the content itself must itemized and tagged so that the learner's state can be mapped to the desired learning goal state, and content items can be delivered in an appropriate sequence to achieve optimal learning.

Data analytics can also be applied to individual courses, to identify topics most sought by students, and areas in which testing shows that students consistently fail. Such failure may indicate students' lack of knowledge, but it could also provide a clue to areas in which teaching could be improved.

At the institutional level, analytics is used extensively to match community and business needs to the design of degree and certificate programs by universities. Businesses also use similar analytic approaches to discover knowledge gaps among their workers, and to design programs to develop or upgrade worker skills as changes occur in their industries.

25.7 Future Directions and Challenges

As this chapter has shown, computers have played, and will continue to play, an increasingly important role in health sciences education. How will the rapid change and fluid nature of innovation influence how we use technology in education in the future? As we increasingly "digitize" almost all aspects of our lives, we can expect information technology to continue to weave itself more and more into the essential fabric of how we teach and learn.

How can digital technology help *advance* teaching and learning? Most faculty have embraced, or at least accepted, technology's growing role in education. Students often have higher expectations of technology use than most health sciences schools can fulfill. How computers can help improve education is a key question of interest to faculty and students alike. Faculty members are keenly interested in finding out how technology can help them become better teachers. Students want to know how computers can help them learn more efficiently and effectively. Current trends in digital learning indicate how some of these questions will be answered (Adams

Becker 2017).[18] The following are examples of some of the challenges that we can expect to encounter.

- *Digital content production and verification* remains an ongoing challenge. Digital learning content can range from inexpensive recording and streaming the video of a single lecture to very expensive and time-consuming creation of a rich and dynamic simulation of a disease process. Effective curation and distribution of high-quality content remains a challenge, with some healthcare faculty being reluctant to use content developed at other universities. An emerging trend that may increase use of existing content is to apply the methods of the "flipped classroom" to MOOC-based online courses. In this method, selected online content, from MOOCs or other sources, is assigned for study at home, and group time is used for instructor-led content discussion and problem solving. Such approaches can combine the best of online content with the strengths of classroom teaching by faculty, and it is possible that such hybrid or blended classes will become increasingly common.
- *Learning analytics* is a direct outcome of digital learning content and learning management systems. An immediate challenge is to utilize the available data to improve the healthcare education process at the level of the individual, the course, the curricula, and the institution, and to match this education process to the needs of today's healthcare. A more far-reaching challenge is to use data as evidence to understand what works and why. In particular, we need to understand the best approaches to blended online and face-to-face learning, the uses of collaborative and project-based learning, and the role of simulations and experiential learning.
- ***Real-time feedback***. Significant portions of pre-clinical training in healthcare require use of simulators. With embedded sensing and compute capability, and internet access, these simulators will become capable of real-time monitoring of learner performance. Display of this data on a performance dashboard will allow both learner and teacher to observe flaws in performance and for the teacher to provide appropriate guidance at the time it is needed. With the addition of intelligent tutors that are built into the simulator, the learner can receive needed feedback by using the simulator at any time of the day. Similarly, we can challenge ourselves to understand and implement intelligent, real-time feedback into all aspects of healthcare learning.
- *Artificial intelligence and* ***adaptive learning***. Understanding and engaging with each student's success at the course level is the domain of the individual faculty, and remains a challenge for the application of appropriate digital technology. Implementation of adaptive learning, that is, adapting the presentation of learning content in response to continuous assessment of learner performance, will be an essential next step. We can expect student performance to be tracked, and personalized exercises and assessments presented, so that they can understand their strengths and weaknesses, and can request digital or in-person help they need for success.
- *Learning Management Systems* will see significant evolution. Currently they are narrowly focused on the administration aspects of learning, ensuring that learners are aware of courses needed for their program, delivering course material with the appropriate sequence and timing, and checking when these courses have been completed. In the future, the challenge will be for LMSs to go beyond administration, and to support student learning. In particular, for healthcare education, LMSs will be required to support mastery- and competency-based education, with detailed tracking of concept and skill acquisition.

18 Adams Becker S, Cummins, M, Davis A, Freeman A, Hall Giesinger C, and Ananthanarayanan V. (2017). NMC Horizon Report: 2017 Higher Education Edition. Austin, Texas: The New Media Consortium. ► http://cdn.nmc.org/media/2017-nmc-horizon-report-he-EN.pdf

The topics presented above are only a small selection of the interesting challenges in future healthcare education. Journals such as "Academic Medicine" and "Computers and Education", and websites such as Educause.edu, periodically discuss these and other challenges in more depth.

25.8 Conclusion

Digital learning is widespread in healthcare education and has proven to be both effective and engaging. Digital content ranges from basic web pages to highly immersive interactive 3D virtual spaces. Digital support of learning uses learner tracking to assess performance and to advise and guide the learner towards optimal learning outcomes. Artificial intelligence and adaptive learning methods have the potential to personalize learning, and to provide the institution with detailed understanding of how to support each learner as well as how to align educational approaches with institutional goals. Simulators, for hands-on procedures and for diagnosis and communication, will provide a learning environment that parallels the student's progress through the real clinical environment, providing safe, realistic practice before learners must use that knowledge on real patients. Virtual and augmented reality will make these simulated environments and tools appear and feel realistic, while providing the content scaffolding and mentoring that may not be available in the real clinical environment. Next generation learning management systems will provide the administrative infrastructure to support the student as they progress through their educational program, deliver personalized learning to each student, and offer detailed dashboard information to both faculty and institutional administration.

To realize the full potential of digital learning, there must be significant investment in further development of digital learning technology and content. There must also be effort to develop faculty and staff so that they move beyond simply using technology to understanding how to make each technology elicit the desired learning outcome.

This is an exciting time in digital learning capabilities. It is an even more exciting time to solve the many challenges ahead so as to move towards high performance learning systems.

Suggested Readings

Bligh, D. A. (2000). What's the use of lectures? San Francisco: Jossey-Bass. In this book, the author analyzes the best use of the lecture as a teaching method, and what lectures fail to teach.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How people learn: Brain, mind experience and school. Washington, DC: The National Academies Press. This National Research Council book synthesizes many findings on the science of learning, and explains how these insights can be applied to actual practice in teaching and learning.

NMC Horizon Report: 2017 Higher Education Edition. Available at: https://library.educause.edu/resources/2018/8/2018-nmc-horizon-report. This annual report highlights issues, trends and technologies in education.

Talbot, T. B., Sagae, K., John, B., & Rizzo, A. A. (2012). Sorting out the virtual patient: How to exploit artificial intelligence, game technology and sound educational practices to create engaging role-playing simulations. International Journal of Gaming and Computer-Mediated Simulations, 4(3), 1–19. This paper is a good overview and analysis of the many methods of simulating a patient.

Questions for Discussion

1. In developing effective educational interventions, you are often faced with a choice of instructional methods. Which of the instructional methods listed below would best match the instructional goals listed? Please justify your selection. Note: For some instructional goals, more than one instructional method might be appropriate.

   Instructional Goal

   1. Be able to intubate an unconscious patient
   2. Memorize the terminology used in neuroanatomy

3. Recognize the symptoms of a patient with probable mental illness
4. Describe the pathophysiological process of hypertension
5. Detect histopathologic variations on histology slides

Instructional Method

1. Case-based scenarios that include video
2. Physical simulation with computer-based feedback
3. Didactic material that includes text, images and illustrations
4. Intelligent tutoring system
5. Drill-and-practice program

2. You are developing a software application for interprofessional education to teach participants about managing patients with advanced Type 2 Diabetes. Your audience includes students representative of the clinicians who are typically involved in the care of such patients: primary care physicians, specialists such as ophthalmologists and podiatrists, and nurses. Your software application is focused on the care of individual patients, and you have put together a set of clinical case studies as a basis. How could you leverage current collaborative technologies to help the team manage each case in a way that resembles what they would do in real life?
3. Select a topic in physiology with which you are familiar, such as arterial blood–gas exchange or filtration in the kidney, and construct a representation of the domain in terms of the concepts and sub-concepts that should be taught for that topic. Using this representation, design a teaching program using one of the following methods: (1) a didactic approach, (2) a simulation approach, or (3) a game approach.
4. You are a junior faculty member at a major medical center and you just were appointed director for a course on clinical patient examinations. You decide to check out several sharing sites for curricular material, such as MedEdPORTAL, to try to find relevant teaching materials. What kind of issues/problems would you expect in integrating material from those sites in your course?
5. As Chief of Quality Improvement at the Veterans Administration, you are attempting to improve fairly poor outcomes of patients with Post-traumatic Stress Disorder (PTSD). You would like to develop a computer-based educational tool for patients and caregivers to help them cope with PTSD. Most of the patients and caregivers are quite unfamiliar with the disorder, and health literacy varies widely in your target audience. In conceptualizing your approach, you are focused on the following questions:

   (a) What are the instructional goals of the program?
   (b) What kind of digital content should you use?
   (c) How do you assess baseline knowledge of patients and caregivers about PTSD, and how do you measure knowledge gains after they have used the program?

References

Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., & Ananthanarayanan, V. (2017). *NMC Horizon Report: 2017 Higher Education Edition*. Austin: The New Media Consortium.

Aebersold, M. (2018). Simulation-based learning: No longer a novelty in undergraduate education. *OJIN: The Online Journal of Issues in Nursing, 23*(2), 14. https://doi.org/10.3912/OJIN.Vol23No02PPT39.

Chang, B. J. (2016). Problem-based learning in medical school: A student's perspective. *Annals of Medicine and Surgery, 12*, 88–89.

Dahlstrom, E., Brooks, D. C., & Bichsel, J. (2014). *The current ecosystem of learning management systems in higher education: Student, faculty, and IT perspectives. Research report*. ECAR: Louisville. Available from http://www.educause.edu/ecar.

Dillon, G. F., Clyman, S. G., Clauser, B. E., & Margolis, M. J. (2002). The introduction of computer-based

case simulations into the United States medical licensing examination. *Academic Medicine, 77*(10 suppl), S94–S96.

Englander, R., Cameron, T., Ballard, A. J., Dodge, J., Bull, J., & Aschenbrener, C. A. (2013). Toward a common taxonomy of competency domains for the health professions and competencies for physicians. *Academic Medicine, 88*, 1088–1094.

Epstein, R. M. (2007). Assessment in medical education. *NEJM, 356*(4), 387–396. Accessed at https://www.nejm.org/doi/full/10.1056/NEJMra054784.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences USA, 111*, 8410–8415.

Gaba, D. M., (2004). The future vision of simulation in health care. *BMJ Quality & Safety, 13*, i2-i10.

Gorbanev, I., Agudelo-Londoño, S., González, R. A., et al. (2018). A systematic review of serious games in medical education: Quality of evidence and pedagogical strategy. *Medical Education Online, 23*(1), 1438718.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers, 36*(2), 180–192. Accessed at https://link.springer.com/content/pdf/10.3758%2FBF03195563.pdf, on October 20, 2018.

Hayden, J. K., Smiley, R. A., Alexander, M., Carding-Edgren, S., & Jeffries, P. R. (2014). The NCSBN national simulation study: A longitudinal, randomized, controlled study replacing clinical hours with simulation in prelicensure nursing education. *Journal of Nursing, 5*(Supplement 2), S3–S40.

Huang, G., Reynolds, R., & Candler, C. (2007). Virtual patient simulation at U.S. and Canadian Medical Schools. *Academic Medicine, 82*(5), 446–451.

Jeffries, P.R., (2005). A framework for designing, implementing, and evaluating simulations used as teaching strategies in nursing. Nurs Educ Perspect. 26(2):96–103.

Kato, P. M., Cole, S. W., Bradlyn, A. S., & Pollock, B. H. (2008). A video game improves behavioral outcomes in adolescents and young adults with cancer: A randomized trial. *Pediatrics, 122*(2), e305–e317.

Kyaw, B. M., Saxena, N., Posadzki, P., Vseteckova, J., Nikolaou, C. K., George, P. P., Divakar, U., Masiello, I., Kononowicz, A. A., Zary, N., Tudor Car, L., (2019). Virtual reality for health professions education: Systematic review and meta-analysis by the digital health education collaboration *J Medical Internet Research, 21*(1), e12959.

Lau, I., & Sun, Z. (2018). Three-dimensional printing in congenital heart disease: A systematic review. *Journal of Medical Radiation Sciences, 65*(3), 226–236.

Nelson, R. (2016). Informatics: Empowering ePatients to drive health care reform – part I. *OJIN: The Online Journal of Issues in Nursing, 21*(3), 5. https://doi.org/10.3912/OJIN.Vol21No03InfoCol01.

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *Journal of the American Medical Association, 298*(9), 1002–1009.

Ryall, T., Judd, B. K., & Gordon, C. J. (2016). Simulation-based assessments in health professional education: A systematic review. *Journal of Multidisciplinary Healthcare, 9*, 69–82. https://doi.org/10.2147/JMDH.S92695.

Shubeck, K. T., Craig, S. D., & Hu, X. (2016). Live-action mass-casualty training and virtual world training: A comparison. *Proceedings of the Human Factors and Ergonomics Society, 60*(1), 2103–2107. Available from: https://doi.org/10.1177/1541931213601476. Accessed at https://www.researchgate.net/publication/307944349_Live-action_mass-casualty_training_and_virtual_world_training_A_comparison, on October 20, 2018.

Stevens, R. H., Lopo, A. C., & Wang, P. (1996). Artificial neural networks can distinguish novice and expert strategies during complex problem solving. *Journal of the American Medical Informatics Association, 3*(2), 131–138.

Talbot, T. B., Sagae, K., John, B., & Rizzo, A. A. (2012). Sorting out the virtual patient: How to exploit artificial intelligence, game technology and sound educational practices to create engaging role-playing simulations. *International Journal of Gaming and Computer-Mediated Simulations, 4*(3), 1–19.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197–221.

Zaharias, G. (2018). What is narrative-based medicine?: Narrative-based medicine 1. *Canadian Family Physician, 64*(3), 176–180.

Translational Bioinformatics

Jessica D. Tenenbaum, Nigam H. Shah, and Russ B. Altman

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_26

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- How does translational bioinformatics differ from the more general field of bioinformatics?
- What do T1 and T2 refer to in the context of translational research?
- What is a biomarker, and why is it important in medicine?
- What is precision medicine, and how does it differ from traditional medical practice?
- What is the difference between pharmacokinetics and pharmacodynamics?
- What is the difference between statistical significance and clinical significance?
- How are genomic data being used today in research, clinical care, and consumer health?
- What are some ethical issues surrounding genomic medicine?
- How are ontologies useful in translational bioinformatics?

26.1 What Is Translational Bioinformatics?

▶ Chapter 9 described the field of bioinformatics, or the study of how information from biological systems is represented and analyzed. **Translational Bioinformatics** (**TBI**) is bioinformatics applied to human health and disease. It uses and extends the concepts and methods from bioinformatics to facilitate the translation of biological ("bench") discoveries into actual impact on clinical care ("bedside") and ultimately on population health (Fig. 26.1). Translational bioinformatics lies at the intersection of bioinformatics and clinical informatics, applying informatics methods to increasingly voluminous **omics** data (genomics, transcriptomics, epigenomics, metabolomics, and proteomics data) to improve clinical care and health outcomes through the advancement and practice of **precision medicine** (see ▶ Chap. 28). In this chapter, we describe key concepts and methods in TBI, summarize TBI data-related resources, and introduce the concept of precision medicine, which is enabled by TBI and covered in greater depth in ▶ Chap. 28. We conclude with a discussion of challenges and future directions for the field.

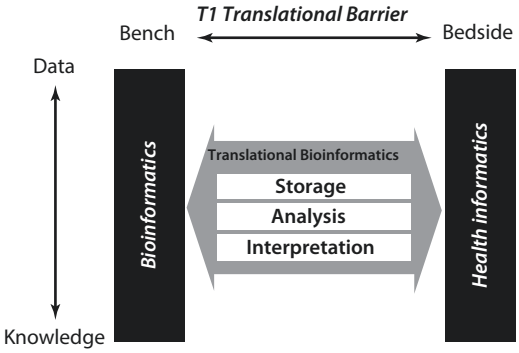


Fig. 26.1 TBI bridges the gap between bioinformatics on the "bench" side of the T1 barrier and health informatics on the "bedside" end of the spectrum. Novel methods for storage, analysis, and interpretation span the spectrum from data to knowledge. (Adapted from Sarkar et al. (2011). Creative Commons CC BY-ND License)

26.1.1 Differences from "Traditional" Bioinformatics

TBI differs from the larger field of bioinformatics in a number of key ways. As described above, the focus of TBI is human health. As such, the discipline centers primarily, though not exclusively, around human data. This fact has a number of implications from an informatics perspective. First, one encounters a range of data management, regulatory, and privacy issues that do not arise in handling data from model organisms such as mice, yeast, or *Escherichia coli*. Laws such as the Health Information Portability and Accountability Act (HIPAA)[1] (see ▶ Chap. 12) dictate how

1 ▶ http://aspe.hhs.gov/admnsimp/pl104191.htm (Accessed 30/11/2012).

patient data must be handled and safeguarded to protect patient privacy. Title 21 of the Code of Federal Regulations Part 11 (21 CFR part 11)[2] mandates how data must be managed if they are to be included as part of a submission to the Food and Drug Administration. In addition, institutional review boards (IRBs) typically require measures to ensure safety and confidentiality of human subjects before they will approve a research protocol. Making complete datasets publicly accessible for a mouse experiment is good scientific citizenship. Making the same type of data accessible for a human study, without approval, could be a serious violation of privacy and confidentiality.

Another difference is that while experimental perturbation through small molecule agonists or antagonists, siRNA, or knockout genes are straightforward and common in yeast or *E. coli*, such approaches would be neither feasible nor ethical in human subjects. This has significant implications for data generation and collection in translational research. **Phase I clinical trials** are the notable exception to this rule, but they are performed only on ostensibly therapeutic agents. They also require a number of preliminary steps, are very expensive, and are performed in a very small number of subjects. Other factors that differentiate research with human subjects include genetic and environmental heterogeneity, which can be controlled in model organisms. Instead, much translational data from human beings comes from *in vitro* experiments on cell lines and observational inquiries regarding factors such as genotype, environmental factors, and outcomes. With so much inherent noise, very large sample sizes are typically required for new discoveries. Novel approaches to data integration, mining, and re-use are thus particularly important in translational research.

2 ► http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=11 (Accessed 30/11/2012).

26.2 The Rise of Translational Bioinformatics

26.2.1 Promise of the Human Genome Project

In January of 2000, two different groups announced that they had fully sequenced the human genome (see ► Chap. 9). The public project, published in *Nature*, was based on multiple individuals (Lander et al. 2001). The other genome, published in *Science*, was a private venture, performed on the DNA of biologist and entrepreneur Craig Venter (Venter et al. 2001). The vision for the human genome was that once all the genes were identified, they could be assigned functional annotations, and we would thus be able to understand what goes wrong when human beings succumb to disease. Additionally, this knowledge would help us to understand exactly which pathways and molecules needed to be targeted in order to prevent or cure disease. Of course, biological reality is not quite so straightforward. To begin with, the "central dogma" of biology (Crick 1970)—DNA is transcribed into mRNA, which is then translated into protein—is overly simplistic. Variations in regulatory regions can affect when the gene is turned on, and to what degree. Most genes have a number of different splice variants, producing a number of different proteins. In addition, proteins undergo post-translational modifications, which impact their structure and function. Finally, additional complexity is added through **epigenetics**, or heritable traits that are not coded for through DNA sequencing alone. An example is methylation of the DNA molecule, which has been shown to affect transcription (Cedar 1988). Despite this, the sequencing of the complete human genome marked a decisive turning point in biomedical research. The parts list had been assembled and researchers could move on to the more interesting aspects of the genome—what each part does, how the parts differ among individuals, and what it all means. The impact

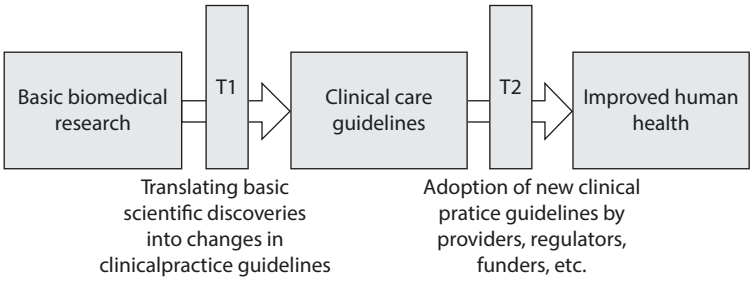


Fig. 26.2 Translational roadblocks along the continuum of biomedical research from scientific discoveries to changes in clinical practice and improvement of human health

this would have on the field of medical informatics was recognized immediately, reflected in the theme for the 2002 AMIA[3] Annual Symposium: "Bio*Medical Informatics: One Discipline" (Tarczy-Hornoch 2007).

26.2.2 What Is Translational Research?

In the early 2000s, there was growing acknowledgement that the population at large, and patients in particular, were not reaping the full benefits of the considerable amount of research money being devoted to scientific discovery. It was recognized that researchers do not do a good job translating their discoveries "from bench to bedside," i.e. bridging biological discoveries in the lab and clinical application of the findings (Lenfant 2003). Two significant roadblocks were initially identified (Fig. 26.2)—one in translating discoveries into clinical care guidelines (dubbed T1 translation), and the other in translating clinical care guidelines into actual practice (T2 translation) (Sung et al. 2003). Additional "T's" have been devised more recently, and definitions refined to be more granular, e.g. splitting out early and late phase trials and knowledge dissemination vs. knowledge application (Waldman and Terzic 2010). In 2004, the National Institutes of Health (NIH) launched the Roadmap for Medical Research, aimed at transforming life science research in the twenty-first century. Biomedical informatics plays a strong role across all three of the major Roadmap themes: New Pathways to Discovery, Research Teams of the Future, and Reengineering the Clinical Research Enterprise (Zerhouni 2006). As part of this Roadmap, the NIH embarked on a major initiative to break down translational barriers through a new funding mechanism known as the Clinical and Translational Science Award (CTSA). This award was aimed at major academic medical centers and their partners with the goal of improving translational research to get treatments to patients quickly.

It was in this context, with newfound attention to translational research, that Butte and Chen coined the term "translational bioinformatics" at the AMIA annual symposium in 2006 in a paper entitled "Finding Disease-Related Genomic Experiments Within an International Repository: First Steps in Translational Bioinformatics" (Butte and Chen 2006). AMIA added TBI as one of its key supported domains and in 2008 held its first annual Summit on Translational Bioinformatics. Later that year, the *Journal of the American Medical Informatics Association* (JAMIA) published a perspective on TBI's "Coming of Age" that enumerated several reasons why the time was right for TBI to come into its own as a field (Butte 2008). In 2009, the editors of the *Journal of Biomedical Informatics* published an explicit Editorial to announce a change in the journal's editorial policy to "focus its bioinformatics attention on innovations in the area of *translational bioinformatics*" (Shortliffe et al. 2009).

3 AMIA is the American Medical Informatics Association, Bethesda, MD: ▶ http://www.amia.org

26.2.3 Precision Medicine as a Driving Force

Precision medicine (and its cousins: personalized medicine, genomic medicine, stratified medicine, individualized medicine, etc.) is health care that is based on an individual's unique clinical, genetic, omic, and environmental profile, in addition to his or her specific values and preferences. In 2004, Lee Hood coined the term "P4 medicine": predictive, personalized, preventive, and participatory (Weston and Hood 2004). Based on an individual's specific risk factors, interventions or changes in lifestyle could be adopted *before* the person falls ill, improving quality of life *and* saving significant costs in health care spending. Armed with this individualized knowledge, patients would be empowered to play an active role in their own health and medical care. Quality medical care has never been one-size-fits-all; precision medicine acknowledges this fact and seeks to change the practice of clinical care accordingly. Precision Medicine is discussed further in ▶ Chap. 28.

26.3 Key Concepts for Translational Bioinformatics

As noted in the definition above, TBI involves the development of novel methods for the storage, analysis, and interpretation of molecular data to guide clinical care. In this section we elaborate on these different levels of informatics methodologies which can be framed as falling along a spectrum from data to knowledge (◘ Fig. 26.1). *Data* represent specific values; at the simplest level, they can be reduced to ones and zeros. In the middle of the spectrum is *information*—ascribing new meaning to the data at hand through analysis. Finally, we arrive at *knowledge*—the ability to interpret information in a specific context, and for that interpretation to guide actions and behavior.

26.3.1 Data Storage and Management

Data are stored at a number of different levels corresponding to different stages along the translational pipeline. At the "bench" end of bench-to-bedside, there is the need to store massive files of raw data generated through omics technologies (Stein 2010). In the case of genome sequencing, these files can be so large that is has been suggested (though not necessarily concluded) that for easily regenerated samples, it might be more cost-effective to discard the raw data and, if necessary, re-sequence at a later time (Hsi-Yang Fritz et al. 2011). For each raw data file type, one can generally choose among several different processing tools or algorithms. Thus, in addition to the raw data, a researcher or core facility may want to store one or more versions of processed data files, still frequently very large in size. In addition to the actual data, experimental **metadata** are needed in order to understand how the data were generated and how they were processed or analyzed. Annotation facilitates both comprehension and data provenance. Unfortunately, that information is rarely standardized, and frequently stored only in the researcher's head, paper notes, or hard drive. Standards and tools such as the Ontology of Biomedical Investigations (OBI) (Brinkman et al. 2010), Minimum Information lists (Taylor et al. 2008) and the Investigation/Study/Assay (ISA) infrastructure (Rocca-Serra et al. 2010) (see ▶ Chap. 9), have been developed to address this issue. Guidelines and best practices were formalized in the "FAIR" framework- making data Findable, Accessible, Interoperable, and Reusable (Wilkinson et al. 2016). A website called ▶ biosharing.org, which had evolved out of the MIBBI initiative (Minimum Information for Biological and Biomedical Investigations), further evolved into ▶ FAIRsharing.org.[4]

4 ▶ https://fairsharing.org/ (Accessed 10/22/2018).

This online resource contains a manually curated collection of data and metadata standards, data repositories, and data sharing policies, as well as the relationships between these entities. For example, it includes a page for the ArrayExpress data repository, with a link to the page for the MINSEQE (Minimal Information about a high throughput SEQuencing Experiment) data standard it adopts, which links to the page for the Journal of Clinical Investigation by which that standard is endorsed.

For translational research purposes, there is also an increasing need to store information related to participant consent. As DNA **Biobanks** (described below) become more common, researchers will have greater access to tissue samples of participants who they did not themselves recruit. It will no longer suffice to have consent information stored on a paper form, locked away in a file drawer. Researchers and biobank administrators will need the ability to know to what each participant has consented, and to perform electronic queries to determine consent status on demand. May John Doe's tissue be used for research beyond the study for which he was enrolled? May the blood collected as a byproduct of care be used for **Genome-Wide Association Studies (GWAS)**? May Jane Doe be contacted for enrollment in a follow-up study? In parallel with work being done to address issues of ethics and governance for this type of data capture and management, researchers are working to develop tools and terminologies to facilitate research permissions management (Obeid et al. 2010; Grando and Schwab 2013). Researchers at the University of California San Diego created iCONCUR, a web-based informed consent tool to enable tiered preferences for use of de-identified data (Kim et al. 2017).

At the bedside end of the translational spectrum, clinicians do not have the time, nor often the training, to analyze the underlying data. They need easy access to what a patient's genotype, protein biomarker pattern, or metabolite profile means, without having to wade through volumes of sequence and biomarker data. Even summary information about test results is not likely to be sufficient; rather, the clinician needs to be provided with *knowledge* of what the test results mean for subsequent treatment decisions. They may also want to know some type of confidence or quality score for the data provided. HL7's Clinical Genomics Workgroup is working to develop an HL7 standard in this area based on HL7's FHIR API (Alterovitz et al. 2015). Incorporating omic data into the EHR (▶ Chap. 14) will not improve clinical care without the incorporation of these data types into clinical guidelines and tools for clinical decision support as discussed in ▶ Chap. 24 (Hoffman 2007).

26.3.2 Biomarkers

Fundamentally, advancements in the ability to analyze and interpret high-throughput molecular datasets advances the discovery of **biomarkers**. The term biomarker has been used for decades, referring to any observation that could be used as an indication of an underlying physiological state. One commonly accepted definition is "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (Atkinson et al. 2001). Exactly what constitutes a biomarker has historically depended in part on what types of observations could be made. Early biomarkers would have included fever, increased respiratory rate, or a rash. As our ability to probe living organisms increased, the domain of biomarkers expanded to the presence or concentration of specific molecules in the blood. For example, increased levels of glucose are indicative of diabetes. Omics-era methodologies give us new types of markers to which we can apply novel analytic methods to anticipate disease and monitor progression. In the genomic era, biomarkers may consist of not just one but many different characteristics, which together give insight into underlying states or processes. Gene expression signatures are a common example of this type of multi-dimensional biomarker.

One important distinction to be made is that of predictive versus mechanistic

biomarkers. Predictive biomarkers are essentially correlative markers of a given observation or outcome. They may or may not be causal for that outcome, but they can assist both clinicians and researchers by anticipating outcomes or suggesting new focus areas for research. Mechanistic biomarkers, on the other hand, can help shed light on what is happening at the molecular level that causes, for example, pathology, disease progression, or sensitivity to a given drug. Understanding a mechanism allows researchers to try to modify it through the activation or inhibition of specific molecules or pathways.

Predictive Biomarkers for Clinical Use

Predictive biomarkers can facilitate decision making in a number of ways. A biomarker indicating poor prognosis might suggest a more aggressive course of therapy than if that biomarker were not present. A signature indicating that lifestyle changes are likely to offer significant benefit to a patient could provide the motivation needed to follow through. For example, a signature indicating that weight loss is likely to improve insulin resistance could identify individuals for whom an intensive lifestyle changes is likely to have the most impact. Shah et al. were able to identify a **metabolomic** profile in subjects who had lost weight that, while independent of the amount of weight lost, was correlated with changes in insulin resistance (Shah et al. 2009b). On the flip side, a signature indicating that lifestyle changes alone are unlikely to confer the desired benefits may suggest that pharmaceutical intervention should be considered as well. Even if a biomarker is in no way actionable *yet*, it can be useful for biomedical research. As an example, osteoarthritis is a debilitating disease that is treated primarily through palliative measures to alleviate symptoms, but for which no disease-modifying therapeutic agents exist. One reason for this is the time and cost required to carry out a clinical trial. Without knowledge of which subjects are likely to progress, studies must enroll large numbers of participants in order to be significantly powered. Identifying biomarkers to predict progression would enable cohort enrichment for individuals in whom disease progression is more likely, thus cutting the total number of subjects required and hence the cost of the trial (Kraus et al. 2011).

Biomarkers that are not clinically actionable may be personally actionable. For example, relapsing-remitting multiple sclerosis (RRMS) is a form of multiple sclerosis in which the patient experiences exacerbations or relapses of neurologic symptoms, followed by periods of partial or complete recovery. If a test could be developed to enable RRMS patients to know in advance if relapses were likely to occur within an upcoming span of weeks or months, it could enable them to make more informed personal or professional decisions, such as when to plan a vacation, or whether to take a new job (Gregory 2011).

One major area for biomarker use is that of pharmacogenomics, described in Sect. 26.5 below. In many cases, a therapeutic gold standard exists, but only a fraction of patients respond to the given therapy. Knowing in advance who is likely not to respond to therapy, or who needs a higher or lower dose than the standard guidelines suggest, can be useful for tailoring therapeutic interventions. Interestingly, while the success of genetic biomarker discovery for common disease has been limited, genotypic biomarkers for response to drugs may be more promising because these variations would not have been selected against through evolution (Cirulli and Goldstein 2010). This may explain why, among published GWAS finding to date, the pharmacogenetic associations tend to have much higher odds ratios than those of genes associated with common diseases.

Molecular Mechanism for Therapeutic Targeting

Biomarkers may also be used for elucidation of disease mechanism which can then enable therapeutic targeting toward a specific molecule or pathway. Comparative analysis of high dimensional molecular signatures in patients versus healthy volunteers, tumors versus normal tissue, responders versus non-responders, etc., can reveal a set of molecules that are differentially expressed among these groups. One can then study those specific molecules

more closely, or the pathways in which those molecules are involved, for example through gene ontology (GO) enrichment (see Sect. 26.6.2) or analysis using a curated pathway database such as Reactome (Fabregat et al. 2018), Ingenuity's IPA, or Thomson Reuter's MetaCore (Nikolsky et al. 2005). These types of tools also help to address a major challenge with pattern detection in high throughput data. Particularly in human data sets where differences are observational and not perturbation-based, it can be difficult, if not impossible, to know what is causal and what is simply correlated. Systems biology, described in ▶ Chap. 9, attempts to address this.

26.4 Biomarker Discovery

One of the most common uses of biomarkers is to categorize samples or patients: cancerous samples versus normal tissues, good versus poor prognosis, bacterial versus viral infection. There are a number of ways to approach this problem, all of which fall under the heading of **supervised learning**. Fundamentally, supervised learning entails taking a set of inputs and corresponding outputs to try to learn a model that will enable one to predict output when faced with a previously unseen input. One is trying to predict one value, the *dependent variable*, based on some number of other values, also called *features* (in computer science), *independent variables* (in statistics), or *risk factors* (in clinical practice). If the dependent variable is categorical, typically one is actually predicting the probability of belonging to one class or the other. For example, one might want to predict whether a person will have a heart attack based on age, race, gender, weight, and cholesterol level. Or, in the context of TBI, one might want to predict the likelihood of a heart attack based on gene expression. Note that this latter approach is useful only if the gene expression signature increases the predictive capabilities beyond that offered by the clinical variables, which are typically easier to collect. Algorithmic approaches to classification and prediction are described in ▶ Chap. 9.

26.4.1 Clinical Relevance Versus Statistical Significance

Statistical significance (typically conveyed via p-values) quantifies whether a difference is reliably *measurable* via a test. With large datasets, most differences detected are statistically significant in the sense that such a difference would not be due to just sampling variation. However, the presence of statistical significance does not guarantee clinical relevance. Clinical relevance is a measure of how valuable information provided by the test is in guiding clinical care. It incorporates not only statistics, but also efficacy, safety, and cost.

A test may be able to predict with 90% precision whether, for example, a patient is likely to respond better to a treatment with unpleasant side effects over another, more innocuous therapy. However, if those side effects would significantly lower the patient's quality of life, then the test, while statistically significant, may not be clinically relevant. Similarly, if the cost of a false negative is very high, for example if a test predicts with 90% precision that a patient will survive without a given intervention, that intervention will likely still be administered. On the other hand, if a test predicts with 90%, or even 100%, precision that a patient is likely to live 1 month longer with a given intervention but the intervention costs $1 million, this highly statistically significant test is still not likely to affect clinical care. Thus, incorporation of molecular data or improvement in an analytic method may make a test's result *statistically* significant while still not affecting clinical practice.

There are various ways to convey a test's "accuracy". The most common metric, which conveys the ability of a test to discriminate two classes, is measured by the Area Under the **receiver operating characteristic** (AU**ROC**) curve (see ▶ Chap. 3), or the **C statistic**. The ideal ROC curve goes straight up the y-axis at x = 0, and then straight across the x-axis at y = 1, giving an AUC of 1. The more reliable a test, the closer it comes to that perfect path. ◘ Figure 26.3 shows hypothetical ROC curves for two tests. Test 2 is a more reliable

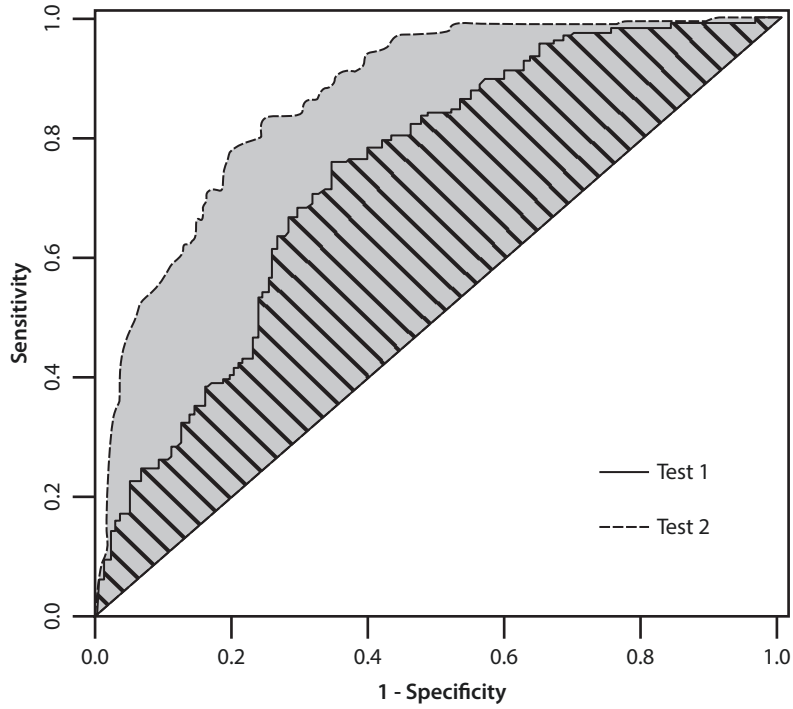


Fig. 26.3 A comparison between two Receiver Operator Characteristic (ROC) curves. The area under the curve (AUC) or C statistic is higher for Test 2 (*gray*) than for Test 1 (*diagonal lines*) to a statistically significant degree, but this increased accuracy does not necessarily imply clinical relevance

test in that it has a statistically significant higher C statistic, but as with the examples above, that may not change any clinical decisions. Much has been written about the limitations of the AUROC, which is not a good measure when the test needs to discriminate between two outcomes where one is very rare (Cook 2007). In such situations, the Area Under Precision Recall Curve (**AUPRC**) may be more meaningful. Instead of 1-specificity on the x-axis and sensitivity on the y-axis, AUPRC plots recall (which is the same as sensitivity) on the x-axis and precision (positive predictive value) on the y-axis. This means it is not skewed by the low absolute number of true positives.

It has been proposed that a better measure than area under a curve is needed for judging the incremental value of novel biomarkers and analytical approaches (Pencina et al. 2008). Alternative methods include **net reclassification improvement** (**NRI**), a measure of the net fraction of reclassifications made in the correct direction using the given biomarker or method over a method without the designated improvement (Steyerberg et al. 2011). This concept is illustrated in Table 26.1. Rows represent the risk level predicted by the hypothetical Test 1 for 1000 subjects, columns represent the risk level predicted by Test 2. Values along the diagonal were predicted to have the same risk by both tests. **Subjects in the black cells (30 + 40 = 70) were correctly reclassified by Test 2 (i.e., the actual rate in parentheses matches the appropriate risk category). Subjects in the light gray cells (10 + 20 = 30) were reclassified incorrectly**. The resulting net

Table 26.1 Hypothetical reclassification of disease risk between two prognostic tests

	Number of individuals (actual rate)			
		Predicted 5-year risk for test 2		
Predicted 5-year risk for Test 1		**0–5%**	**5–20%**	**> 20%**
	0–5%	300 (3%)	20 (2%)	0
	5–20%	30 (3%)	300 (11%)	40 (37%)
	>20%	0	10 (35%)	300 (42%)

reclassification improvement is (70–30)/1000, or 4%.

One final characteristic of a good test is its calibration, the extent to which a test correctly measures absolute risk. That is, do the risk values predicted by the test reflect the actual risk observed in the population. Calibration may differ across the population at different levels of predicted risk, which may in turn affect the test's utility.

26.4.2 Biomarkers for Drug Repurposing

One very promising area for use of biomarkers is in **drug repurposing**, or drug repositioning. That is, identifying existing drugs that may be useful for indications other than those for which they were initially approved. Doing so avoids early clinical trials for toxicity as those have already been performed. A number of different approaches have been used to identify candidates for repositioning. In some cases, overlapping symptoms may suggest a potential match between one disease area and another. In other cases, empirical observation of unexpected positive effects may suggest alternative uses for a given drug. With omic-scale biomarker discovery, it is possible to use underlying molecular pathway signatures to suggest new uses for existing drugs.

One of the prominent early examples of this approach came from the Broad Institute in the form of the "Connectivity Map," a resource intended to enable researchers to identify functional connections between drugs, genes, and diseases (Lamb 2007). The general idea was to identify a gene expression signature in a state of interest, e.g. a disease, and then compare that signature to the gene expression patterns observed upon exposure to a number of different compounds. Correlated signatures suggested pathways that were similarly perturbed between a disease state and an intervention. More importantly, anti-correlated signatures suggested potential utility for a given compound in trying to reverse the underlying molecular mechanisms of a given disease. A similar approach was used by Sirota et al. to identify the anti-ulcer drug cimetidine as a candidate agent to treat lung adenocarcinoma. They were then able to validate this alternate use *in vivo* using an animal model of the disease (Sirota et al. 2011). Their approach is illustrated in Fig. 26.4.

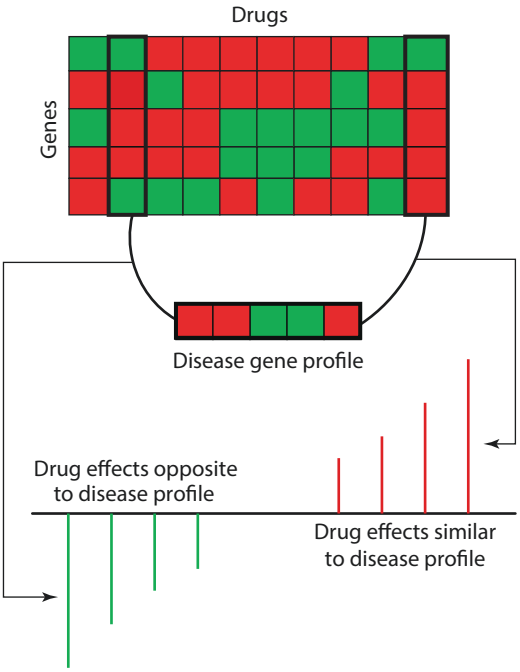


Fig. 26.4 A computational approach to candidate selection for drug repurposing. Sirota et al. first generated genomic signatures representing both diseases and drug exposure. For each disease signature, they compared it to the panel of drug signatures and assigned a drug-disease score based on profile similarity. Drugs whose pattern were most significantly *dis*similar to the disease state were ranked as lead candidates to treat the disease of interest

26.4.3 Genomic Data Resources

Fundamental to advancement in biomarker discovery are high-throughput genomic measurements that have been enabled since the human genome draft was published. Fortunately, the genomic community has been moving toward a culture of data sharing, NIH broadens genomic data-sharing policy (2014) making experimental data via publicly available data repositories (Kaye

et al. 2009, 2014). Resources for genomic data include:

Genetic variation

According to the *Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies* (*GWAS*),[5] genotypic data must be deposited to the NIH database of Genotypes and Phenotypes (dbGaP). Genomic variation data is also available through a number of other online resources—see ► Sect. 9.3 and (Sherry et al. 2001; WTCCC 2007; Altshuler, et al. 2010). The HapMap projects catalog variation over a wide variety of ethnic populations, in order to define the occurrence and frequency of common genetic variations (Rusk 2010). The 1000 Genomes project is taking HapMap further to categorize the occurrence of more rare variations (changes in single DNA bases, as well as insertions/deletions, segmental duplications, and larger scale inversions and translocations) (Via et al. 2010). There are also resources about copy number variations.[6]

dbSNP—**d**ata**b**ase of **S**ingle **N**ucleotide **P**olymorphisms is a publicly available catalog of genome variation (Sherry et al. 2001). Contents primarily represent single nucleotide substitutions, but also include a small number of other types of variation, for example microsatellite repeats and small insertions and deletions (Homerova et al. 2002). The PharmGKB resource specifically annotates genetic variations relevant to drug response (Altman 2007). The increase in exome and genome sequencing has led to powerful resources for assessing human genome variation across diverse populations. Key resources include the Exome Aggregation Consortium (Lek et al. 2016), the Exome Sequencing Project (ESP) (Auer, et al. 2016) and others. There are also pharmacogenomics-specific studies of key pharmacogenes, such as the Pharmacogenetic Research Network Sequence Project (PGRNSeq) (Bush et al. 2016).

Gene expression information

The Gene Expression Omnibus (GEO) (Barrett et al. 2013) contains an extremely large and diverse collection of high throughput gene expression experiments which allow one to evaluate whether a disease (or drug exposure) leads to up- or down-regulation of gene expression (Edgar et al. 2002). Particularly useful examples of gene expression for drug response are the Connectivity Map data set in which gene expression in response to 164 drugs was measured (Lamb et al. 2006). Similarly, the NCI 60 is a set of 60 cancer cell lines that have been exposed to hundreds of drugs in order to determine their sensitivity (Ross et al. 2000). Other efforts have looked at genetic variations that correlate with gene expression in order to associate these genomic regions with the function of the correlated genes (Gamazon et al. 2010; Nicolae et al. 2010). ArrayExpress, developed by EMBL-EBI (European Molecular Biology Laboratory- European Bioinformatics Institute), is a European analog of GEO, containing microarray and sequencing data for functional genomics (Kolesnikov et al. 2015).

Gene associations

The Genetic Association Database (Becker et al. 2004) provides curated information about the results of genetic association studies, including those studies that relate genetic variation to variation in drug response. The Human Genome Mutation Database (HGMD) also provides this information in a highly curated form (Stenson et al. 2009, 2017). dbGaP— database of **G**enotypes **a**nd **P**henotypes is a resource to archive and distribute information about the interaction between genotype and phenotype (Mailman et al. 2007). The PharmGKB resource is devoted entirely to providing information about associations between human genetic variation and drug response phenotypes (Altman 2007). The GWAS Catalog (MacArthur et al. 2017) is a very useful database of genome wide association study (GWAS) hits. ClinVar aggregates genomic variation and their relationship

5 ► http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html (Accessed 12/6/2012).

6 ► http://humanparalogy.gs.washington.edu/structuralvariation/general/intro.html (Accessed 12/3/2012).

to human phenotypes (Landrum et al. 2016). Finally, ClinGen (Clinical Genome Resource) represents a manually curated collection of genetic variants and their clinical relevance (Rehm, et al. 2015).

Genetic pathways

Understanding drug action requires understanding the pathways and networks of drug action and drug metabolism. The PharmGKB provides curated drug pathways for both drug action and drug metabolism, and has links to relevant external pathways created by the National Cancer Institute Pathway Interaction Database (Schaefer et al. 2009), Reactome (Joshi-Tope et al. 2005), and others.

26.5 Pharmacogenomics

Pharmacogenomics, a prominent subtype of biomarker discovery, is the study of how genes and genetic variation influence drug response. The primary challenges for pharmacogenomics are to (1) identify the key genes that influence drug response, and (2) understand how specific variations in these genes modulate drug response. The term **pharmacogenetics** generally refers to drug-gene relationships that are dominated by a single gene, whereas the more general term refers to drug responses that result from a combination of interacting gene products. In this section, we use the word "gene" loosely to refer not only to the DNA coding regions for proteins and RNAs but also the protein and RNA products themselves. In many cases, the gene-drug relationship is really a relationship between the drug and the gene's protein product (or even its non-coding RNA product).

Pharmacogenomics is a prototypical TBI activity because it involves clinical entities such as drugs, diseases, and symptoms as well as molecular entities such as genes, proteins, DNA, RNA, and small molecules. Because drug response is the key phenotype of interest, it is useful to review the basis for drug response. When a drug is administered, there are two distinct genetic "programs" that are relevant. The first is the **pharmacokinetic program** or PK, which describes the absorption, distribution, metabolism and excretion of the drug in the body. Genes implement this program (they encode transporter molecules that move the drug across membranes and the liver enzymes that transform the drug and prepare it for elimination via the kidney or liver) and variation in these genes can lead to a different blood level of drugs or a different timing of these levels. The second is the **pharmacodynamics program** or PD, which describes how the drug works, its protein target, and the mechanism by which it impacts cellular physiology in order to alleviate or cure disease. Genes are clearly also involved in this program (they encode the drug's primary targets, and the other proteins that interact with these targets to create the cellular response to the drug), and variation in these genes can lead to a different response to the drug. In short, PK is "what the body does to the drug", and PD is "what the drug does to the body." The goal of pharmacogenomics is to understand, for every drug, the key PK and PD genes, and which variation impacts their response. This will allow us to realize the vision of using the genome to choose drugs based on maximizing their likely efficacy and minimizing their likely toxicity.

26.5.1 Key Entities and Associated Data Resources

The key computational entities in pharmacogenomics are genes, drugs, and drug-related phenotypes (indications and effects, including side effects). There exist good informatics resources for all of these:

Genes

These are typically specified using the Human Genome Nomenclature Committee (HGNC) standard (Seal et al. 2011). They are typically situated within the genome as a series of exons that are spliced together to create a mature mRNA transcript that is then translated into a protein. This basic concept is made more complex because the strategy for splicing the exons may be variable (alternative splicing)

thereby leading to several proteins, the RNA transcript may be degraded before it is translated, and the proteins may be modified after they are created. There are many resources on the web for gene information, and many aggregators of this information. These create a remarkably powerful network of associations that can be used creatively to make new associations. For example, Fig. 26.5 shows the links on PharmGKB (Pharmacogenomics

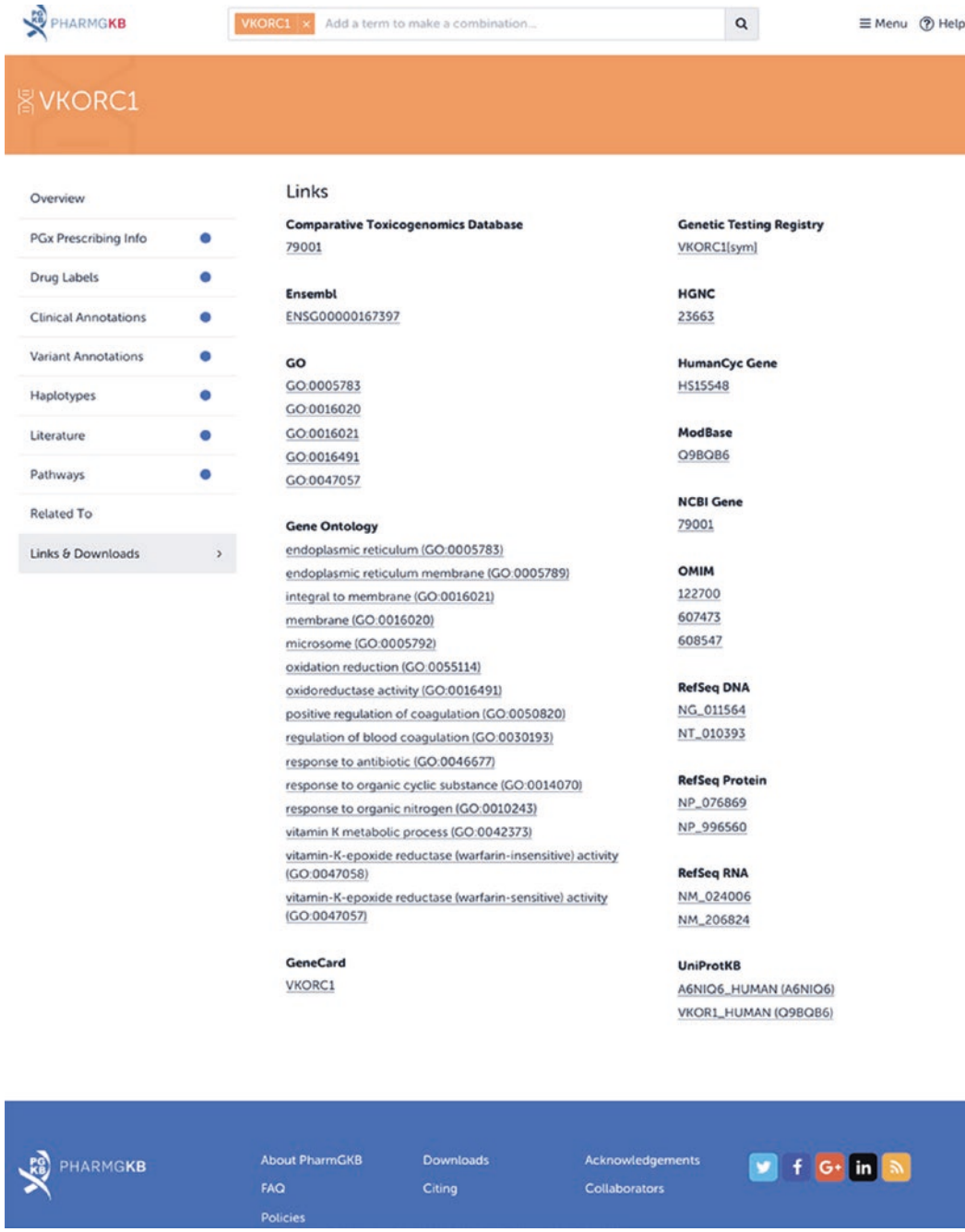


Fig. 26.5 PharmGKB gene pages are organized by tabs and the "Downloads/LinkOuts" tab shown here has links to many other sites with valuable information about human genes. (Copyright PharmGKB, used with permission from PharmGKB and Stanford University)

Knowledge Base) for the drug VKORC1 and includes links to:

- Entrez Gene: summarizes the sequence, variations, homologs across species
- OMIM: provides information about rare genetic diseases involving this gene
- UniProt: provides mapping information to relate this gene to its protein products
- GeneCards: provides aggregated information about function, tissue localization, expression levels, literature references, and more

Drugs and small molecules

RxNorm is a terminology standard for specifying drugs (Parrish et al. 2006). DrugBank provides information about drugs, their targets, pharmacology, uses, and many other characteristics (Knox et al. 2011). There are only around 2000 approved drugs in the United States, and so this list is relatively short. The list of small molecules that are not drugs is much larger and includes the contents of PubChem (Wang et al. 2009, 2010), an NIH-built resource with basic information about the structure, function and literature on small molecules. The Zinc Database (Irwin and Shoichet 2005) lists 13 million commercially available compounds that can be purchased for use in research. Much drug information is contained within the "package insert" that is included in most drug packaging. This is information created by the drug company, but approved by the FDA. The FDA makes these available on a drug information site called DailyMed. For patients, the National Library of Medicine's MedlinePlus resource provides basic drug information as well.

Drug indications and drug effects

Drugs are used to treat particular diseases, and so controlled terminologies of drug indications and drug effects are useful for computational efforts. At the organism level, indications and effects are often diseases (diabetes is an indication) or side effects (hyperglycemia is a side effect). The UMLS and MeSH terminologies are often used to characterize such disease phenotypes (Bodenreider 2004). Of course, other disease terminologies such as SNOMED are also useful (Spackman et al. 1997). For side effects, there are specialized terminologies, including the MedDRA terminology used by the FDA in adverse event reporting (MedDRA replaced a previous terminology called COSTART), and the WHOART (World Health Organization Adverse Reactions Terminology) dictionary for adverse reactions (Brown et al. 1999; Alecu et al. 2006). The SIDER database (Kuhn et al. 2010) provides information mined from drug package inserts about drug indications and side effects. The Anatomic Therapeutic Chemical classification (ATC) from the World Health Organization provides a high level classification of drugs organized hierarchically by the anatomical location of target, the therapeutic category, the pharmacological subgroup, chemical subgroup and precise chemical substance (Miller and Britt 1995).

Pharmacological properties of drugs

There are resources on the web that provide molecular level assay data related to small molecules, including many drugs. The ChEMBLdb resource provides the ability to find targets, binding affinities, inhibition concentrations and information about other drug-oriented assays (Overington 2009). BindingDB also provides binding affinities for small molecules and proteins (Liu et al. 2007). In addition, RxNorm includes an RxClass API for accessing drug classes and members of a given class (Bahr et al. 2017).

Population-based data on drug effects

The FDA maintains information about all reports of adverse events in the FDA Adverse Event Reporting System (FDA AERS). These reports include demographic information, indications for treatment, drugs administered, side effects experienced and a summary of clinical outcomes. They are freely available at the FDA website. The Canadian equivalent system, also making data freely available, is available through the Health Canada website. These data are very noisy and have many confounding variables, but nonetheless can

be useful for discovering "signals" suggesting dangerous side effects or drug-drug interactions (Tatonetti et al. 2011b).

26.5.2 TBI Applications in Pharmacogenomics

The network of data described above is a rich potential source of hypotheses about how genes combine to create drug response, as well as for predicting the particular consequences of genetic variation. This is still a new field, and there remain many opportunities for innovative use of these data. We highlight a few here to illustrate how integration of data can lead to novel discoveries.

■ GWAS to Discover Drug Response Genes

The most straightforward way to associate genes with drug response is to perform a genome-wide association study (GWAS) in which two groups are compared. (See ► Chap. 28 for more on GWAS.) One group (cases) has a drug response of interest (e.g., an adverse event in response to the drug or a particularly good response to it) and the other group (controls) does not have the drug response of interest. It is critical to ensure that the phenotype or response is carefully defined and measured. With each group, DNA is collected and typically 500,000 or 1000,000 **SNPs** (single nucleotide polymorphisms) are measured using microarray technology. Then, for each SNP, an association is measured between the genotypes in cases and controls and the response of interest using a simple statistical test such as the chi-squared test. The SNPs that are most highly associated may represent regions of the genome that are involved in the response. These must be carefully vetted statistically, as there are many potential confounding variables. For example, it is important that the cases and controls are drawn from populations with similar ethnic origin, and that the significance remains after correcting for multiple testing. When one tests 500,000 or 1000,000 hypotheses, adjustments such as Bonferoni correction (see ► Chap. 24) must be made in order to take into account the chance that an association is spurious. If the result is real, then the SNP may be used to identify nearby genes in the region that may be important for the drug response. For example, Shuldiner and colleagues were interested in the ability of the drug clopidogrel to protect patients from cardiovascular events. They found that a polymorphism RS12777823 was associated with a high likelihood of having a cardiovascular event. They noted that this SNP was very close to the metabolizing enzyme CYP2C19, and in particular the "risk" allele for this SNP co-occurred with the CYP2C19*2 variant. Thus, they showed that CYP2C19 is important for the desired effect of clopidogrel, and found a variation of this gene that predicted poor response to the drug in affected patients (Shuldiner et al. 2009).

■ Mining the FDA AERS to Find Drug-Drug Interactions

The FDA Adverse Events database associates multiple drugs with multiple diseases as indications as well as side effects. This database shows promise as a way to find new associations between single drugs and their side effects, as well as multiple drugs and their side effects. As mentioned above, the SIDER database is a "top down" database of side effects derived from the package label of drugs. Another approach to getting good lists of side effects is a more data-driven approach. One way to do this is to look for patterns of side effects associated with certain types of drugs using machine learning. For example, one may analyze the side effects of drugs that alter glucose in order to create a signature of the "typical" profile of side effects associated with a glucose-altering drug. Then, one can search a database of side effects (such as FDA AERS) for other drugs that match this profile. This was done by Tatonetti et al., who created a profile for glucose-altering drugs and found a set of 10 side effects either enriched or deficient (compared to background) in these drugs: hyperglycemia, diarrhea, hypoglycemia, and pain were higher than others, and paresthesia, nausea, pyrexia, abdominal pain, and anorexia were less likely than others (Tatonetti et al. 2011a). Using this pattern,

more than 93% of drugs that are known to alter glucose could be recovered. More interestingly, however, this pattern could be applied to patients on pairs of drugs to search for pairs that altered glucose. A highly correlated combination was the antidepressant paroxetine and the cholesterol medication pravastatin (Tatonetti et al. 2011a). In subsequent validation in three independent EHR systems, large increases in glucose were observed in patients on these two drugs, and in mouse studies of these two drugs, glucose was substantially increased. Thus, the adverse event patterns could be used to create patterns and detect new signals, not specifically reported in the database, but implied by the pattern of other side effects observed.

Mining the Literature to Build a Database of Gene-Drug Associations

Biomedical text can also be an important source of high quality information about the relationships among genes, drugs, and diseases (Garten et al. 2010). High fidelity natural language processing techniques (▶ Chap. 7) can be used to extract information about gene-drug interactions. In some cases, the association between genes and drugs can be inferred simply by their co-occurrence in sentences (Garten and Altman 2009). In these cases, however, there can be many false positives due to sentences in which genes and drugs are mentioned, but are not actually interacting. A more precise method is based on careful parsing of sentences to find subjects and objects that are genes and drugs, and which are related by verbs that connect them (e.g. "CYP2D metabolizes codeine" has CYP2D as the subject, codeine as the object, and the verb "metabolizes" establishes their relationship) (Coulet et al. 2010). The rate of false positives is reduced in this case because more strict criteria are applied before claiming a relationship. These high quality interactions can be chained together to infer new knowledge. For example, drug-drug interactions often occur because two drugs share a common metabolizing gene and that gene becomes saturated in the presence of both drugs, and cannot adequately metabolize both of them. Thus, the observations that "CYP2D metabolizes codeine" and "CYP2D metabolizes metoprolol" might be combined to infer that codeine and metoprolol have a potential drug-drug interaction. There are a large number of similar inferences that could be drawn about the relationships between genes, drugs and diseases given a high quality database of pairwise interactions drawn from the published literature. Of course, some pairwise interactions may be incorrect, and so evidence for interactions should be combined from several sources (including EMR validation, for example) and once predictions are made, they should be embraced only with skepticism. A comprehensive text-based search for relationships between genes, drugs and diseases over all PubMED abstracts yielded a publicly available database of more than 2 million high quality associations (Percha and Altman 2015, 2018).

Using Drug-Target Interactions to Predict New Ones

Another way to find new uses for old drugs is to predict interactions between drugs and new potential targets. Many drugs are designed to interact with a single target based on a detailed understanding of disease pathology. Once the drugs are administered, however, they may not bind only the original target, but they may unexpectedly have effects based on their binding to other targets. Most commonly, these "off target" effects are considered side effects and are avoided. In some cases, however, the "off target" effect may be beneficial in the setting of some other disease. Thus, both for explaining the molecular basis of side effects and for finding new molecular evidence for beneficial novel effects, it is useful to connect drugs to proteins. One way to do this is to build computational and visualization methods for docking a 3D representation of a small molecule into the 3D structure of a target protein. This can be very successful, and has led to the hypothesis that a Parkinson's disease drug may treat tuberculosis (Kinnings et al. 2009)! In that case, the 3D structure of a tuberculosis protein had a pocket that appeared to have high binding potential to a known Parkinson's disease drug, and thus the hypothesis arose that the

Parkinson's drug might inhibit TB growth. These structure-based methods are powerful but limited because we have the 3D structure of only a subset of human proteins. Another approach, therefore, is based on looking for similarities in the list of drugs that have been shown experimentally to bind a protein. In this case, all that is needed are data from chemical assays showing which drugs bind which proteins. These are routinely collected in large screening experiments, and are available at the ChEMBL resource (Heikamp and Bajorath 2011), for example. Given two proteins with two lists of interacting drugs, we can compare the list of drugs to look for commonalities. If there are many commonalities between protein A and protein B, then one might conclude that the drugs that bind protein A may also bind protein B. This was the approach taken in the Similarity Ensemble Approach (SEA) where the list of drugs binding two proteins are compared using a measure of chemical similarity (Keiser et al. 2009). When the chemicals on the two lists are statistically similar (more than would be expected by chance), then the SEA method predicts cross-binding of ligands for the two structures. When this was applied to a large set of proteins, the authors found that the antidepressant fluoxetine (Prozac) had high potential binding to the beta-adrenergic receptor, and this was found experimentally to block the beta-1 receptor—demonstrating that Prozac is a type of beta-blocker!

Identifying Drug Targets Using Side-Effect Similarity

A critical goal in pharmacogenomics is to associate drugs with their target proteins (and thus their coding genes) in order to know where to look for variation that may affect drug response. Determining drug targets can involve a difficult and lengthy experimental program. Thus, it would be very useful to have computational methods for determining targets. One way to do this is to associate drugs to their side effects, and to look for side effect profiles that are similar across drugs. If one drug has a known target, and if another drug has a similar pattern of side effects, then the two drugs may share that target. This is based on the assumption that side effects arise from a few common mechanisms, and so genes involved in this mechanism may be targeted by multiple drugs or drug classes. In one study, Campillos et al. showed that they could create 1018 drug-drug relationships based on shared side effects (Campillos et al. 2008). The side effects were taken from the SIDER database, and the drugs came from a list of 746 marketed drugs. Twenty of these drug-drug relationships were tested experimentally, and 13 of them were shown to bind common targets. Thus, a relatively straightforward association of drugs based on side effects allowed the definition of molecular targets. In related work, Hansen et al. showed that genes could be ranked by their likelihood of interacting with a drug based on looking at the degree of similarity between chemical structure and indications-of-use between the query drugs, and small molecules known to interact with the gene products and their close protein interaction neighbors (Hansen et al. 2009). The pattern of predicted binding of a small molecule to protein off-targets can also yield information about the likely side effect profile for that molecule (Liu and Altman 2015).

The examples we have discussed have several common features: they deal with the basic objects of diseases, drugs, and disease or adverse-event phenotypes; they integrate at least two sources of data to establish new relationships between these basic objects; and they connect clinical entities (drugs and diseases or adverse events) to molecular entities. Such examples represent only a small subset of the types of questions that can be asked with these valuable datasets. The key technical challenges are typically (1) finding adequate gold standards (▶ Chap. 2) to evaluate the success of methods before applying them for novel discoveries; (2) understanding the sources of error and bias so that predictions are as reliable as possible; (3) designing careful statistical tests to ensure that the scoring and estimates of significance are accurate and useful (minimizing false positives, in particular); and (4) identifying and engaging experimental collaborators who can, when appropriate, test the predictions that are made in human or model systems. Recently, it has become clear

that despite their shortcomings, EHRs can be extremely useful for initial validation of hypotheses about connections between drugs and adverse events (Tatonetti et al. 2011a). Gene-drug associations are typically tested in model systems with genes altered in order to reduce or eliminate their normal function, or by looking for covariation in human subjects.

26.5.3 Challenges for Pharmacogenomics

Target Expansion: Molecules to Networks

The emerging field of systems pharmacology is abandoning the view of "one drug, one target" and moving instead toward a view that "the network is the target." That is, the larger network of interacting genes is targeted by a drug at several points, and thus the systemic effects of drugs need to be evaluated in order to understand better the molecular underpinnings of drug response. The challenges to systems pharmacology are similar to the challenges to the more general systems biology: defining the network topology and key players, creating ways to measure parameters, modeling nonlinear responses, and understanding how variation in the basic molecular players impacts the resulting phenotype—in this case drug response phenotypes.

Rare Variants

As whole genome sequencing increasingly provides data about rare variants, the paradigm of looking for common genetic variation that explains variation in drug response will need to be modified. There may be cases when variation in drug response is explained by multiple rare variants rather than one or a few common variants. This is particularly challenging because there will often be insufficient statistics to evaluate rare variants. In some cases, huge population-based studies may provide enough samples, but in other cases even these large cohorts will not have sufficient examples of any rare variant to allow statistical validation. In those cases, we will have to rely on computational techniques to assess the significance of very rare variations.

Computational Methods to Leverage Stem Cell-Based Model Systems and CRISPr Assays

The rise in the use of stem cells will create opportunities for combining direct measurements of cellular response to drugs with systems models of response, whole genome variation, and epigenetic information. As we perfect methods for creating induced pluripotent stem cells and differentiating them into the target tissues, we will be in a position to measure the response to drugs directly on these cells, with identical genetic and perhaps epigenetic backgrounds. Computational methods for analyzing these responses and relating them to the expected response in the patients from whom these cells are derived will be a major challenge in the years ahead. Similarly, the increased availability and ease of generating genome wide CRISPr libraries that allow genes to be knocked out alone and in combination promises to usher in an era of unprecedented data about the effects of genes on drug response (Kweon and Kim 2018).

26.6 Ontologies for Translational Research

In order to apply computational methods for biomarker discovery, one needs a consistent way to refer to genes, diseases, drugs, devices, etc. Several ontologies exist in the biomedical domain, many under active development, that provide the necessary terms for creating consistent annotations—preferably in an automated manner—for the various datasets that are at the core of conducting research in TBI. One primary need in TBI is to identify and refer unambiguously to diseases using one or more disease ontologies. We use the term **disease ontology** to refer to artifacts—terminologies and vocabularies as well as true ontologies—that can provide a hierarchy of parent–child terms for disease conditions. Disease-specific and other clinically-oriented ontologies are discussed in detail in ▶ Chap. 7.

The Ontology for Biomedical Investigations (OBI) was developed as a collaboration among a number of experimental communities around the world in order to represent com-

mon aspects of biological and clinical investigations. It includes broadly applicable terms such as *assay*, as well as more specific terms, such as *transcription profiling by array assay*. It is particularly useful for annotation of experimental metadata, for example to record that a *protein expression profiling assay* was performed on a *blood specimen* (Brinkman et al. 2010).

26.6.1 Ontology-Related Resources for Translational Scientists

The use of ontology-based analyses for TBI, especially disease and drug ontologies as well as analyses using multiple ontologies, is a recent development and the adoption and use of ontologies is likely to accelerate. Several resources are available for researchers who wish to use ontologies in making sense of large-scale datasets. The UMLS, or Unified Medical Language System (see ▶ Chaps. 2 and 7), is a set of files and software that brings together many health and biomedical vocabularies and standards to enable interoperability among computer systems. The UMLS has many uses, including search engine retrieval, data mining, public health statistics reporting, and terminology research. In the field of TBI, the UMLS is a relatively underutilized resource, but that is changing with the increase in the variety of access options (Aronson 2001; Bodenreider 2004; Aronson et al. 2008; Shah and Musen 2008; Aronson and Lang 2010; Mork et al. 2010) and heightened dissemination efforts by the National Library of Medicine.

The National Center for Biomedical Ontology maintains a repository of biomedical ontologies called BioPortal (Musen, et al. 2011) which provides access through both Web pages and Web Services to more than 600 biomedical ontologies and controlled terminologies. Users go to the BioPortal Web site to browse biomedical ontologies and to search for specific ontologies relevant to their work. BioPortal also provides tools such as the Ontology Recommender (Jonquet et al. 2010), which takes as input representative textual data relevant to a domain of interest and returns as output an ordered list of ontologies that would be most appropriate for annotating the corresponding text. By browsing ontologies on BioPortal and using tools such as the ontology recommender, a cancer biologist may find, for example, that although the Gene Ontology offers some terms for annotating her experimental data related to cell division, there are more precise terms in the NCIt. She may discover that the Foundational Model of Anatomy ontology provides terms for consistently naming body parts from which the experimental specimens were obtained, or that the National Drug File – Reference Terminology (NDF-RT) provides the properties of the drugs used in generating the experimental data. BioPortal allows users to navigate ontologies using a tree browser or visualize ontologies as a graph that offer cognitive support for understanding the complexities of large ontologies (◘ Fig. 26.6).

To provide the relationships between terms in two *different* ontologies, BioPortal provides mappings between the terms (Ghazvinian et al. 2009). The mappings can inform the user that the term *Melanoma* in the NCI Thesaurus is related to the term *Malignant*, *Melanoma* in SNOMED-CT and to *Melanoma* in the Human Disease Ontology. These mappings allow users to compare the use of related terms in different ontologies and to analyze how whole ontologies compare with one another (Ghazvinian et al. 2011). In addition to curated mappings from the UMLS metathesaurus, BioPortal enables algorithmic and user-generated mappings as well.

26.6.2 Enrichment Analysis

Enrichment analysis is a statistical method to determine whether, for a set of items, a given concept or value is statistically over-represented compared to what one would expect by chance. For example, informatics-related terms are over-represented in this book compared to what one would expect to find in a random sampling of words from all textbooks. The canonical example of enrich-

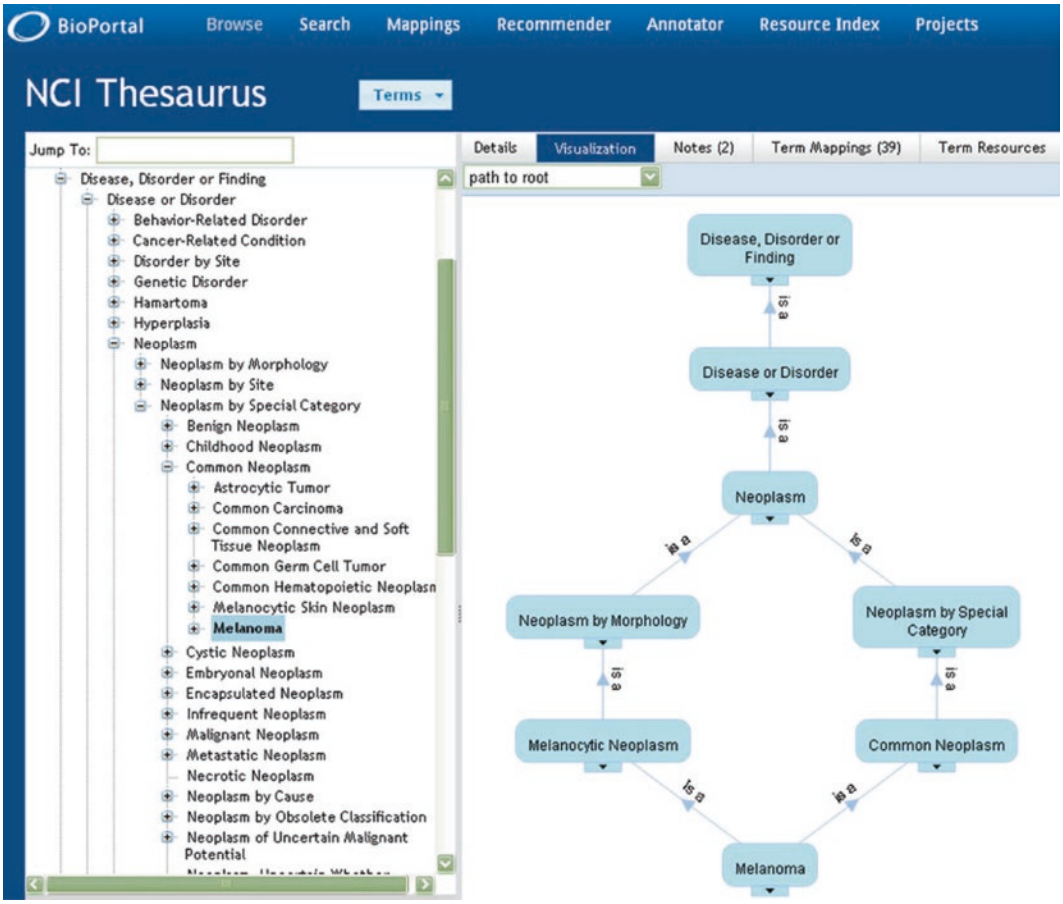


Fig. 26.6 A portion of the National Cancer Institute's thesaurus. The left pane shows a standard tree view for the term 'Melanoma'. The right pane shows a visualization that provides additional context by showing the parent classes of melanoma, all the way to the root node of 'Disease, Disorder or Finding'. The navigation bar just above the graphical visualization provides access to additional information, such as mappings which provide hooks into other disease ontologies that contain the concept Melanoma

ment analysis involves a list of genes differentially expressed in some condition. To determine the biological meaning of such a list, the usual solution is to perform enrichment analysis with the GO (Gene Ontology), which provides terms for consistent naming of the cellular component (CC) of gene products, the molecular functions (MF) they carry out, and the biological processes (BP) in which they participate. Several curation projects use GO terms to annotate gene products from multiple organisms with terms from the three branches (CC, MF, BP) (Camon et al. 2003). These annotations form the basis for enrichment analysis in which we can aggregate the annotating GO concepts for each gene in this list, and arrive at a profile of the biological processes or mechanisms affected by the condition under study. This approach does have certain limitations, for example incomplete annotations for a number of genes, lack of conditional independence between annotations, sensitivity to GO version, and lack of a systematic mechanism to compensate for differing levels of depths in different branches of the ontology hierarchy (Khatri and Draghici 2005; Rhee et al. 2008; Tomczak et al. 2018). Despite this, such analysis is widely popular in the bioinformatics community and has resulted in over 100 tools listed

on the GO website[7] and over 7000 citations to the landmark paper on the Gene Ontology (Ashburner et al. 2000).

Disease and drug ontologies can be used to perform enrichment analysis in a manner similar to GO-based analyses for gene expression data (Subramanian et al. 2005; LePendu et al. 2011a). Just as scientists can ask *Which biological process is over-represented in my set of interesting genes or proteins*, we can also ask *Which disease* (*or class of diseases*) *is over-represented in my set of interesting genes or proteins*? For example, by annotating known protein mutations with disease terms, Mort et al. were able to identify a class of diseases—blood coagulation disorders—that were associated with a lower than expected rate of amino acid substitutions at O-linked glycosylation sites (Mort et al. 2010).

26.7 Natural Language Processing for Information Extraction

Ontologies are also useful in the context of extracting information from a body of text. In-depth methods for natural language processing are discussed in ▶ Chap. 7. Here we describe some applications in the context of translational research.

26.7.1 Mining Electronic Health Records

Researchers have shown that is possible to profile patient cohorts from EHRs using a variety of ontologies including SNOMED CT, MedDRA, and RxNorm (LePendu et al. 2011b). For example, LePendu et al. developed methods to annotate clinical text and methods for the mining of the resulting annotations to compute the risk of having a myocardial infarction on taking Vioxx (rofecoxib) for Rheumatoid arthritis. Subsequently they demonstrated that it is possible to apply annotation analysis methods for detecting drug safety signals using electronic medical records up to 2 years before a drug's recall (LePendu et al. 2013).

Mining EHR data has also been proposed as a solution to the challenge of the large number of subjects that are needed for genome wide association studies (GWAS). Patients are increasingly able to consent to, or in some cases to opt out of, allowing excess biospecimens taken in the course of clinical care to be used in a de-identified fashion for genomic testing. Even for relatively strong genetic effects, GWAS requires thousands of individuals for sufficient statistical power (▶ Chap. 11). For weaker effects, tens of thousands of subjects are likely to be needed. Although the cost of genotyping continues to decrease, recruitment and sample collection for these large numbers is both costly and labor-intensive. Leveraging the health care system and EHRs for research recruitment offers a potential approach to circumvent this problem. Ritchie et al. demonstrated the feasibility of this approach by using EHR data and an associated biobank to replicate a number of previously discovered genotype-phenotype associations (Ritchie et al. 2010).

One major initiative in this area is the eMERGE (Electronic Medical Records and Genomics) Network, whose initial aim was to demonstrate that data captured through routine clinical care are sufficient to identify cases and controls accurately for GWAS (Thorisson et al. 2005). As of 2018, the eMERGE consortium includes eleven institutions with DNA repositories and associated electronic medical record systems. For each site, ontology-based data extraction and natural language processing algorithms are applied to the EHR in order to determine phenotypes such as dementia, cataracts, peripheral artery disease, type 2 diabetes, and cardiac conduction defects. This analysis is performed in a high-throughput, scalable fashion with results compared to a manually curated gold standard in order to determine positive and negative predictive values for cases and controls for the phenotypes in question (Kho et al. 2011). The consortium also looks at cross-institutional algorithm application, ethical, legal, and social issues

7 ▶ http://www.geneontology.org/GO.tools.shtml#alphabet (Accessed 12/3/2012).

around DNA biobanks, and the potential for future incorporation of GWAS findings into clinical care (Liu et al. 2012; Rohrer Vitek et al. 2017; Wan et al. 2017).

These types of EMR-associated biobank resources enable a number of other approaches to data mining. For example, Denny et al. used BioVU at Vanderbilt University to perform what they called a "PheWAS," or a systematic, high-throughput **phenome-wide association scan** (Denny et al. 2010). Instead of measuring whole genomes across thousands of patients in order to find a gene associated with a phenotype in question, they measured only five alleles across thousands of patients and performed enrichment analysis for various diseases based on ICD9 codes. They then were able to reproduce known associations between those genes and certain diagnoses, *and* to generate new hypotheses for associations between these genes and other diagnoses that were statistically enriched for a given genotype. The ability to connect, at a molecular level, diseases that were not previously associated can have implications for therapeutic intervention (Denny et al. 2016).

26.7.2 Dataset Annotations

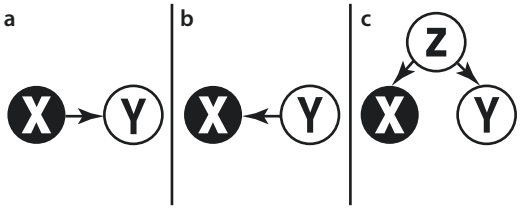
In addition to EHRs, public repositories for omics-scale datasets remain a valuable but underutilized resource for data mining. Upon submission, these datasets typically contain only free-text descriptions. Addressing the lack of annotations, researchers demonstrated that translational analyses are enabled by automatically annotating tissue and gene microarray datasets with ontology terms (Shah et al. 2009a; Doan et al. 2014). Butte et al. employed a crowd-sourcing approach to annotate samples from the Gene Expression Omnibus (Hadley et al. 2017). Such automated annotation approaches have been generalized to create systems that process the free text metadata of diverse database elements such as gene expression data sets, descriptions of radiology images, clinical-trial reports, and PubMed article abstracts to annotate and index them with concepts from appropriate ontologies (Jonquet et al. 2011). Doing so has enabled novel analyses from already collected molecular and clinical data (Garber et al. 2017; Sweeney et al. 2018a, b). Such annotation represents a large part of the work required to address the 'F' in FAIR data, making data 'findable.'

The utility of consistent annotation of research datasets is now widely accepted. As a result, there are several initiatives to build tools for consistent meta-data assignment, as well as indices of available datasets corresponding to specific terms of interest. CeDAR, the Center for Expanded Data Annotation and Retrieval, was formed with a goal of developing information technology to facilitate authoring and adoption of metadata (Musen, et al. 2015). BioCADDIE (Biomedical and healthCare Data Discovery Index Ecosystem) is an international effort to promote biomedical data discovery through the creation of a data discovery index called DataMed (Chen et al. 2018).

26.8 Network Analysis

Biology lends itself in various ways to modeling through networks or **graphs**. The term "graph" simply refers to a set of *nodes* or circles connected by a set of *edges* or lines. In a molecular context, a node represents a molecular entity, and an edge represents some form of relationship between those molecular entities. This relationship may be a physical interaction (e.g., binds to), an influence (e.g., activates), or a similarity (e.g., is co-expressed with), among other possibilities. One frequently sees graphical models of gene regulatory networks, protein-protein interactions, and signaling cascades. The set of all of these sorts of physical interactions has been referred to as the **interactome** (Barabasi et al. 2011). Studying this interactome and its properties from a graph theory perspective enables useful insights regarding gene modules and pathways, and how these are disrupted in disease.

A number of researchers have attempted to develop gene association networks using gene expression data either alone or together with other sources of network data such as protein-protein interactions. The general idea is that



26

Fig. 26.7 Three possible causal relationships between two co-expressed genes. **a** Gene X affects Y. **b** Gene Y affects X. **c** Both X and Yare affected by a third causal gene Z

co-expressed genes are likely to interact with each other or participate in the same pathway. But of course, correlation does not equal causality, and to be useful from a translational perspective, it is important to know the directionality of the influence between two molecules. Consider two genes, X and Y, whose expression is correlated (see Fig. 26.7). One can conclude that the genes interact in some way, whether directly or indirectly (i.e., through another molecule). However, without additional knowledge of any sort, we cannot know whether X influences Y (Fig. 26.7a), or Y influences X (Fig. 26.7b) or they share a third causal gene, Z (Fig. 26.7c). Which model represents the true underlying relationship is important to know because if Y is involved in poor outcome, then targeting X will help to alleviate this condition in the first model, but not in the second or third.

One way to determine the actual underlying relationship, used frequently in model systems, is to actively perturb a specific variable in the system. If the other molecule changes accordingly, then we know that the perturbed variable was causal. This is the approach frequently used in a **systems biology** approach (see ▶ Chap. 9). Unfortunately, this is much harder to do in human beings than in *E. coli* or yeast. One clever approach to determination of causality in human biological networks is to integrate gene expression with genotypic information, in which case DNA sequence can be assumed to be the independent variable. If differential gene expression is correlated with differential genotype, one can conclude that the genotype caused the gene expression pattern and not the other way around. This is the basis for the approach taken by Eric Schadt et al. to develop **probabilistic causal networks** which can then be used to identify key drivers of disease (Zhu et al. 2008).

Network analysis in translational research need not be confined to concrete objects such as molecules. The Human Disease Network is a graphical model where nodes represent both known disease genes and disorders, linked by known associations between a given gene and disease (Goh et al. 2007). Figure 26.8 shows the "diseaseome" bipartite network, as well as the Human Disease Network, which connects diseases based on common genes, and the Disease Gene Network, connecting genes based on diseases in common. Combining these disparate data types enables a graph theoretic approach to study the genetic basis for disease. Using this framework, one can analyze similarity between genes based not on co-expression or GO term annotation but based on the pathologies in which a gene is known to be involved. Such similarities could easily go undetected through gene expression analysis if, for example, the different diseases are caused by over-activation or inhibition of the gene respectively. A disease-gene network also enables the comparison of diseases not traditionally studied together, based on common underlying molecular mechanisms. Identifying disease similarities based on gene expression requires that one analyze expression data from those two diseases together in the first place, making it more difficult to discover novel, previously unsuspected relationships.

Building upon the Human Disease Network, Yildirim et al. created a network of drug-gene target interactions, thus enabling an additional layer of analysis regarding similarity between different drugs based on targeted genes, and between target molecules based on the drugs that target them (Yildirim et al. 2007). This type of network can be used as the basis for a number of different observations, including trends in drug development over time. For example, analysis of the structure of the graph revealed significant clustering of drug-gene interactions, suggesting a significant "me too" pattern to drug

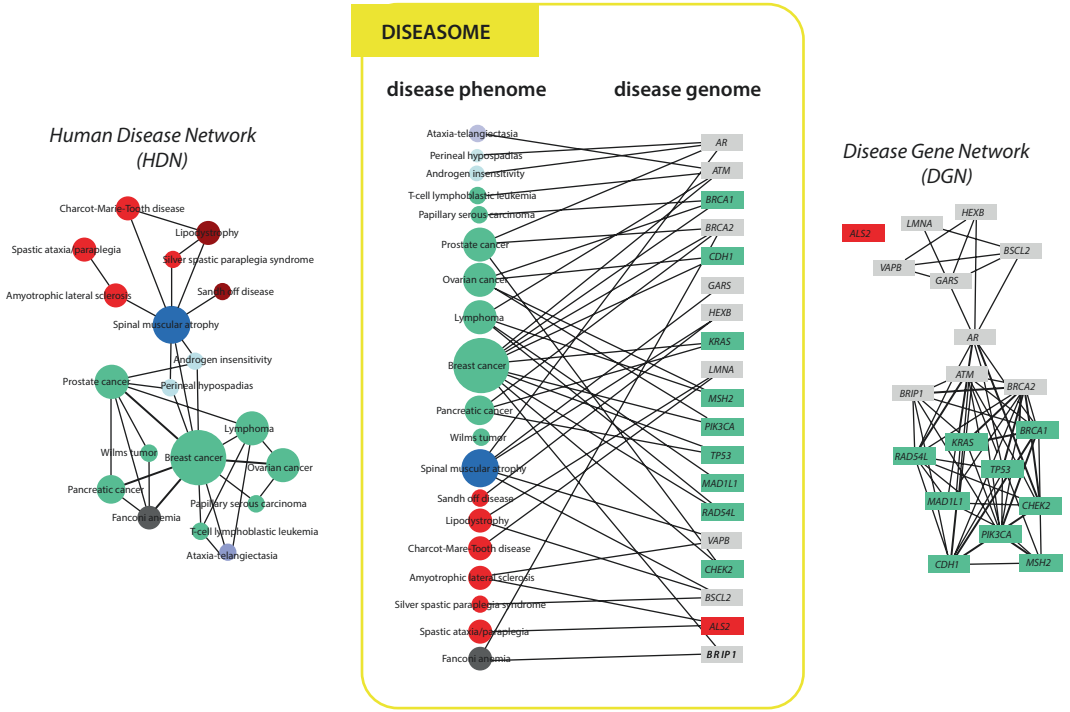


Fig. 26.8 The Human Disease Network. The *middle panel* shows a small subset of the bi-partite gene-disease network based on OMIM (Online Mendelian Inheritance in Man) gene-disease relationships. The Human Disease Network on the *left* shows diseases as nodes, with connections representing common related genes. The Disease Gene Network on the *right* depicts genes as node, with connections indicating that they have been implicated in one or more of the same disorders. (From Goh et al. (2007), ©2007 National Academy of Sciences, U.S.A. with permission)

development (see Fig. 26.9). Inclusion of drugs still under investigation, i.e., not yet FDA approved at the time of analysis, demonstrated that the breadth of drug targets is expanding, suggesting a trend toward target diversity. Incorporating the cellular component of target proteins showed that the distribution of cellular location for target proteins, previously nearly two-thirds membrane-associated, is becoming more diverse, better matching the known distribution for disease proteins. Finally, this group incorporated protein-protein interaction data to facilitate the study of network properties of drug target gene products. They looked for the shortest path between drug target genes and known disease genes for the disorder that drug was intended to treat and found that this number appears to be decreasing over time, suggesting that drugs are moving from a palliative approach (i.e. treating the symptoms and not the cause) to rational drug design (Yildirim et al. 2007). More recently, Hu et al. used a network-based approach to look at comorbidities and the effects of environmental perturbations (Hu et al. 2016). Using this approach they were able to generate hypotheses for molecular mechanisms of comorbidity which in turn can facilitate drug repurposing and the development of targeted therapeutics.

26.9 Basepairs to Bedside

Although the sequenced human genome has not been a panacea for human disease, it has enabled the beginnings of a new approach to human health and to the practice of **precision medicine** (Collins and Varmus 2015). As the price of genomic sequencing falls and as our knowledge regarding the meaning of genomic variation increases, genotypic data

Fig. 26.9 Yildirim et al.'s Drug-Target network. Circles represent FDA-approved drugs and rectangles represent target proteins. Diseases are color-coded by anatomical system and protein targets according to their cellular location. Clusters of drugs associated with one target reflect the pharmaceutical industry's tendency to develop 'follow-on' drugs. (Courtesy of Albert-Laszlo Barabasi, MD, with permission)

is poised to become a standard component of a person's medical record. In this section we describe the translational path of genomics, from sequencing in the lab to clinical relevance for individuals.

26.9.1 Whole Genome Sequencing

Technologic Advances

The DNA-probe approach to genotyping, described in ▶ Chap. 9, may be compared to looking for one's car keys under the proverbial street lamp. That is, the technology shines the light on a certain portion of the genetic landscape, and that is where we look. Which SNPs are included on a chip is determined in large part by which SNPs have been detected in the past, for example through the HapMap project (Kang et al. 2006). A number of new associations have been found in this way, whether because the SNPs themselves were of interest, or due to genetic linkage—the tendency for alleles located close to one another on a chromosome to be inherited together. However, more recent findings have demonstrated that the concept of "common disease-common variant" is flawed (Zhu et al. 2011). Indeed, there has been some disappointment in the extent to which GWAS has been able to explain common diseases with known genetic components (Manolio et al. 2009). Whole genome or, in some cases, whole exome sequencing allows researchers to identify rare variants (i.e., those with a minor allele frequency of <1%) that account for genetic disease. Advances in sequencing technologies (e.g., "**nextgen**" **sequencing**, and **third-generation sequencing**, see ▶ Chap. 9) and the corresponding decrease in cost, make genome-scale sequencing increasingly feasible in translational research and even in clinical care.

Whole Genome Versus Exome

Even with recent advances in genome sequencing technology, the cost to sequence a full genome at a rate of coverage that enables the identification of novel SNPs is still significant, on the order of one thousand dollars. However, recall that only about 1% of the genome actually codes for proteins and 85% of known disease-causing mutations with large effects occur in proteins (Choi et al. 2009). One way to further decrease the time and cost of sequencing is to look at only those stretches that code for actual proteins. This can be justified because most variants known to underlie Mendelian disorders disrupt protein-coding sequences. Of course, this approach will miss causal variations if they exist in the other 99% of the genome. Moreover, a recent cluster of publications from the ENCODE (**Enc**yclopedia **o**f **D**NA **E**lements) Consortium asserts assignment of biochemical function for 80% of the genome (Dunham et al. 2012). Some additional components, such as regulatory regions or splice acceptor and donor sites may be included as well to increase sensitivity without incurring significant additional cost. Exome sequencing in a small number of individuals has been used to identify the causal variant for rare diseases such as Miller's syndrome, a multiple malformation disorder (Ng et al. 2010), and Proteus syndrome, a disorder causing the overgrowth of tissues and organs, thought to have afflicted the nineteenth century Englishman known as *The Elephant Man* (Lindhurst et al. 2011).

Genomic Data Sharing

In 2008, researchers demonstrated that the presence of a single genome within a complex mixture of DNA samples could be ascertained (Homer et al. 2008). This caused both NIH and the Wellcome Trust[8] to limit access not only to individual genomes, but to aggregate genomic information as well. (Note that the ability to determine the presence or absence of an individual's DNA in a heterogeneous sample presupposes the availability of detailed genomic information about the individual in question.) These actions prompted responses that ranged from "too little, too late" to "a heavy-handed bureaucratic response to a practically minimal risk that will unnecessarily inhibit scientific research" (Church et al. 2009). Current NIH policy allows investigators to submit a data access request to be reviewed by an NIH Data Access Committee. Access to data is granted once a Data Use Certification is co-signed both by the investigator and the appropriate official[s] at the investigator's affiliated institution.[9]

The Global Alliance for Genomics & Health is a policy and standards oriented organization aimed at enabling responsible data sharing (Terry 2014). It was founded in 2013 when 50 colleagues from 8 countries met to discuss challenges and opportunities in genomic research and medicine, and now comprises more than 500 organizational members from 71 different countries. A number of Work Streams and Driver Projects guide development efforts and serve to pilot the organization's tools. In a similar vein, the RDA (Research Data Alliance) is an international organization meant to promote data sharing and data driven research, and to develop and promote the technical infrastructure to that end. The RDA's scope goes beyond genomic and biomedical data, but the mission is highly aligned with GA4GH. One of the RDA's many Working Groups is called "FAIRSharing Registry WG: connecting (meta) data standards, repositories, and policies.[10]"

Additional online genomic resources include TCGA (**T**he **C**ancer **G**enome **A**tlas) and WTCCC (**W**ellcome **T**rust **C**ase **C**ontrol **C**onsortium). TCGA is a joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) to accelerate understanding of the molecular basis for cancer through appli-

8 ► http://www.wellcome.ac.uk/ (Accessed 10/28/2018).

9 ► http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html (Accessed 10/29/2018).

10 ► https://rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html (Accessed 10/29/2018).

cation of genomic technologies, including genome sequencing.[11] The WTCCC, established in 2005, comprises 50 research groups across the UK who have performed a series of genome-wide association studies and made the data available through application to a Consortium Data Access Committee.[12]

26

26.9.2 Here Are Some Human Beings

■ **The Personal Genome Project (PGP)**

Promising though genomic medicine may be, much remains to be worked out technically, scientifically, and from the ELSI (ethical, legal, and social implications) perspective (see ► Chap. 12). Some notable pilot projects have been embarked upon in order to catalyze progress in all of these areas. Craig Venter was the first person to have his complete genome published in 2007. Since then, a number of human genomes have been sequenced, and some of those have been made available in the public domain. The question becomes: now what? What can any given individual learn from his or her complete genomic sequence? What does an individual *want* to learn, or not want to learn, as the case may be? The only reliable way to answer these questions is with empirical input.

George Church is a pioneer in genomic sequencing, inventor of the Polonator sequencer, and founder of personal genome sequencing company Knome. In 2005, Church started the Personal Genome Project (PGP), ultimately aiming to sequence 100,000 individuals in order to advance understanding of how genes contribute, along with environment, to human traits. The project "hopes to make personal genome sequencing more affordable, accessible, and useful for humankind."[13] A vanguard of ten volunteers—the PGP-10—were selected to have their genomes sequenced. This endeavor differs from other projects in one crucial way: in addition to making the sequence data publicly available, complete phenotypic data, including personal and health information, family history, and even name and photographs would be shared as well. This was a departure for the type of projects the NIH typically funds and supports. Generally, informed consent includes information on how the research team plans to secure privacy and confidentiality for the subject. In this case, sharing of personal data was part of the protocol itself. The first set of integrated data from this group was made available in October 2008.

Making this type of data both publicly available and personally identifiable was stepping out into socio-scientific *terra incognita*, generating some worry that it could affect health care, employment, insurance, and more. In 2008, the Genetic Information Nondiscrimination Act (GINA) was signed into law, but its scope is limited to employment and health care insurance. It does not address life, disability, or long term care insurance (Hudson et al. 2008; Tenenbaum and Goodman 2017). Though rare, there are a few notorious examples of lawsuits where employers performed genetic and health-related testing on employees without their consent (Angrist 2010), and though unlikely, the PGP warns prospective participants that their DNA could be artificially synthesized and planted at a crime scene (Lunshof et al. 2010).

As much as the PGP has pushed the boundaries and helped to advance the technology, data management, and clinical issues involved with personal genomes, and will continue to do so, it also serves as a weather balloon from the ELSI perspective, generating empirical data on sociological atmosphere, ethical pressures, and legal winds of change (see ► Chap. 12 for additional discussion on these points). Misha Angrist, a bioethicist at Duke University, is PGP Participant #4. As documented in his book *Here Is a Human Being*: *At the Dawn of Personal Genomics*, the early sequencing was slow going, the technology took time to work out the kinks, and the preliminary results were underwhelming even

11 ► http://cancergenome.nih.gov/ (Accessed 10/29/2018).

12 ► https://www.wtccc.org.uk/index.shtml (Accessed 10/29/2018).

13 ► http://www.personalgenomes.org/ (Accessed 10/29/2018).

to the individuals who had been sequenced (Angrist 2010). The infrastructure is not yet in place to empower someone with his complete genomic profile to do much with that information. Angrist describes his own attempts to make use of tools for genomic interpretation—SNPedia,[14] Sequence Variant Analyzer (Ge et al. 2011), and the Church lab's open source Trait-o-Matic[15]—which he compared to the dial-up days of the internet. Out of all of the variants carried by the PGP10, only one was deemed serious. Steven Pinker carried a mutation for MYL2, which had been shown in some cases to cause hypertrophic cardiomyopathy (Angrist 2010).

The resource created by the PGP enabled the Critical Assessment of Genome Interpretation (CAGI) to create a community challenge to assess the ability to predict traits from whole genomes in which researchers were asked to predict whether an individual had a particular trait or profile based on their whole genome. Overall, findings showed that predicting individual traits is difficult and that matching genomes to trait profiles depends strongly on a small number of common traits like ancestry, blood type, and eye color (Cai et al. 2017). Equally important, however, the project has created a publicly available, integrated resource for genomic, environmental, and trait (GET) data (Lunshof et al. 2010) and an empirical test bed for tackling the ELSI issues brought to bear by such a resource.

A personal genome for clinical assessment

As another proof of concept, collaborators at Stanford and Harvard did a complete sequencing, analysis, and genetic counseling for a 40-year-old male with family history of sudden death from cardiac arrest (Ashley et al. 2010). The goal was to determine how whole genome sequencing would translate to clinical application. The patient was found to have increased risk for myocardial infarction, type 2 diabetes, and some cancers. While most of the findings were not actionable, the patient had both increased risk for cardiovascular disease and genetic disposition to benefit from the use of statins and aspirin. Despite this, just over a year after publication, the patient maintained that he had "not been convinced that statins or aspirin would have enough beneficial effect relative to their risks," and had not therefore changed his pharmaceutical behavior (Quake, S, 2011, personal communication).

Just over a year after the Quake profile, the same group published their findings from performing whole exome sequencing on the first healthy nuclear family (Dewey et al. 2011). They generated an ethnically concordant reference sequence (i.e. a reference sequence based on a European population, reflecting the European background of the family in question), which enabled increased accuracy for rare mutations. Findings included high resolution inference of sites of recombination (i.e., where the parents' chromosomes "cross over" during meiosis), and a novel approach to HLA (Human Leukocyte Antigen) typing—important for risk in a number of diseases, particularly autoimmune disorders. For the family in question, they were able to determine that the father had passed down to his daughter a mutation for Factor V Leiden that poses increased risk for blood clotting. This is actionable information for women as the Factor V mutation is a contraindication for estrogen-based birth control pills (Singer 2011), and inherited thrombophilia is a known risk factor for pregnancy outcomes (Tenenbaum et al. 2012). Note that Factor V mutations are also included in chip-based genotyping services, so whole genome sequencing was not the key enabling technology in this case.

One key item reported in the paper by Ashley et al. was the fact that, in the absence of a centrally curated resource of all rare and disease-associated variants, the authors spent *hundreds of hours* reviewing databases. Moreover, the work was a collaborative effort among a number of highly trained experts in clinical genetics, genetic counseling, bioinformatics, internal medicine, pharmacogenomics, etc. (Ormond et al. 2010). Clearly new

14 ► http://www.snpedia.com/index.php/SNPedia (Accessed 12/6/2012).

15 ► https://github.com/xwu/trait-o-matic/wiki (Accessed 12/19/2012).

tools, automation, and infrastructure, as well as a whole new paradigm in genetic counseling, are required to make incorporation of genomic data into health care feasible for the population at large.

Ethical, Legal, and Social Issues (ELSI)

Pursuit of genomic medicine raises a number of ethical, legal, and social issues (see also ▶ Chap. 12). Some worry that people are ill-equipped to process the results of these tests. But it is not clear that a paternalistic approach is a better alternative; there was a time when it was considered acceptable for a doctor not to disclose a cancer diagnosis to the patient himself (Novack et al. 1979). In addition, new discoveries are being made all the time—what are the obligations to follow up if something new (and dire? and actionable?) is discovered about a given subject? Other questions include whether enough is known for the results to be of any practical use, whether the service should be provided outside of the context of a relationship with a clinical caregiver, and whether results could have detrimental effects on a person's ability to secure health insurance. Some states have banned the services, others have made stipulations requiring clinician involvement and **CLIA certification** for the labs that handle the samples and process the results.[16]

In a companion article to the Quake profile, it was asserted that consent for a process in which the risks of knowledge gained are not wholly understood is more complex than for simple genetic testing. People have trouble interpreting probabilities. Patients must be advised that they may find out things they did not want to know about. The eminent scientist James Watson made a point of requesting that his ApoE status be redacted from the release of his full genome because he did not want to know if he was at risk. His grandmother had died of Alzheimer's at 83, and he did not want to worry that every subsequent memory lapse marked the onset of dementia (Angrist 2010). Statistics predict that any given patient will find out he is a carrier for *some* lethal autosomal recessive disease. Illness aside, the average global non-paternity rate has been estimated to be as high as 10% (Olson 2007), though it is likely closer to 1% (Larmuseau et al. 2016). Genetic information could also have implications for the patient's children, present or future, and for other family members. Patients, this group concluded, must have access to trained professionals to provide answers to their questions, where answers exist. This will be difficult, lengthy, and expensive, but not to do it would undermine the consent process (Ormond et al. 2010).

Although knowing the "parts list" for the human genome is an important step, much remains to be understood about how genes factor into human health and disease. For most diseases, the environment plays as much, if not more, of a role as a person's DNA. Aside from some notable, deterministic exceptions such as Huntington's disease, most known risk alleles confer fairly low odds ratios unto themselves (see ▶ Chap. 3), making an individual, for example, approximately 1.1 times as likely as the average individual to develop a given condition. Even when ratios are as high as, say, twofold, it is of dubious actual utility to know that based on one's genotype, the odds of being diagnosed with Crohn's disease went from 0.5 in 100 to 1 in 100.

For certain disease markers, such as Alzheimer's or BRCA1 and BRCA2, it was, and largely still is, unknown what impact negative results might have on a customer's mental and emotion well-being. Some studies have shown that while a person experiences negative emotions immediately in the wake of learning the bad news, over a time period of months there is no significant difference in anxiety, depression, or test-related distress (Green et al. 2009). In any case, DTC genetics companies' websites must provide the ability to view sensitive results while protecting the customer from stumbling on these findings unintentionally. 23andMe, as an example, has spent considerable resources on the design of a user-friendly interface through which to present an individual's "health reports," or their individual genotype for markers that

16 ▶ http://www.genomeweb.com/dxpgx/will-other-states-follow-ny-calif-taking-dtc-genetic-testing-firms (Accessed 12/6/2012).

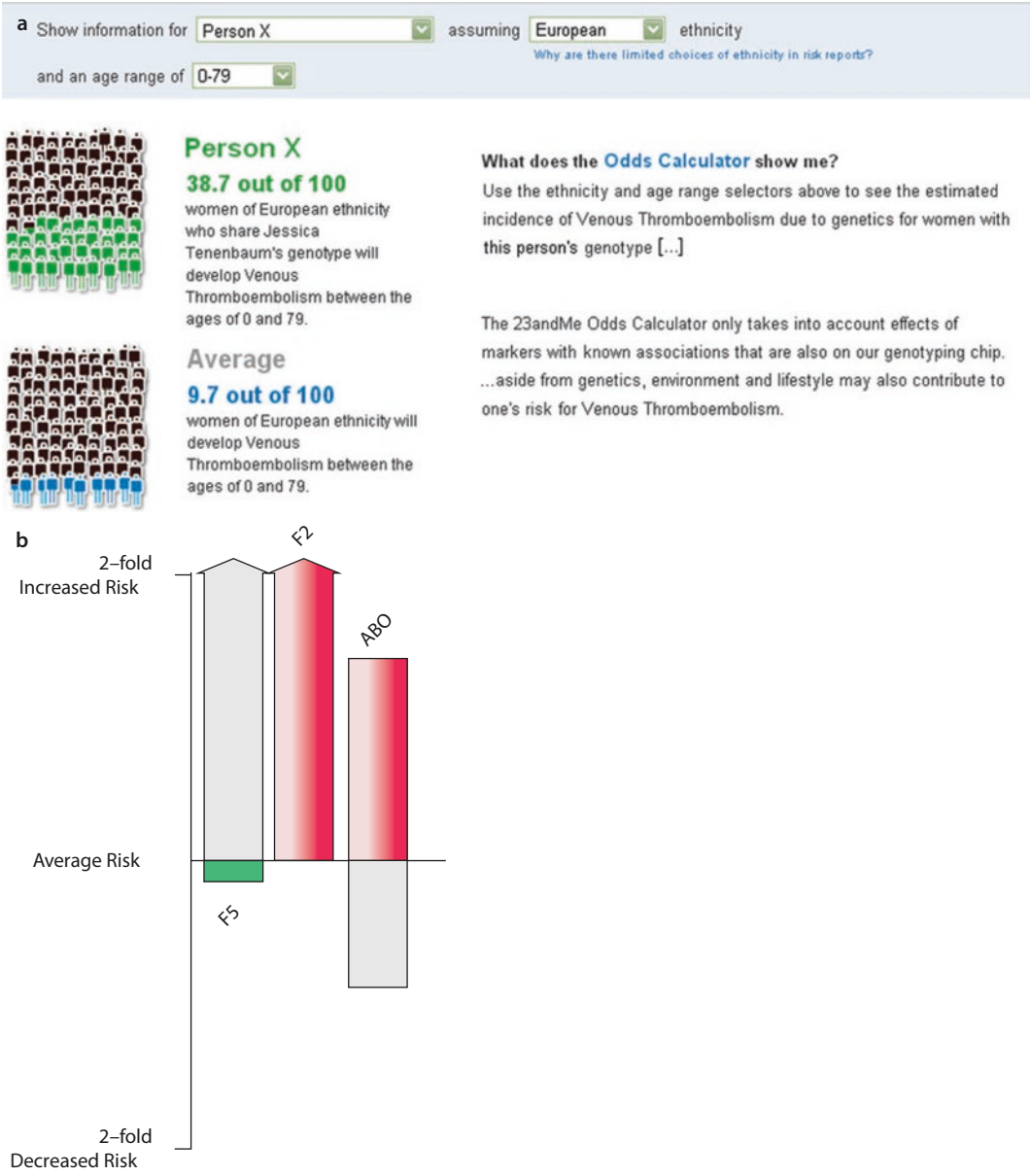


Fig. 26.10 23andMe's graphical representation of relative risk from their website before the FDA stepped in to regulate DTC genetic testing. **a** Colored figures represent the number of people on average out of 100 who are likely to develop venous thromboembolism over the course of a lifetime. Green figures represent the individual's personal reported risk; blue figures represent the average risk for females of European descent. (Accompanying text has been shortened for clarity.) **b** The individual's relative risk for each of three reported markers: factor 5, factor 2, and ABO. (Specific values were displayed on the website when the user hovered the mouse over the colored bars.) In a later version of the website, risk for hereditary thrombophilia was based only on factor 2 and factor 5; ABO was no longer included. (© 23andMe, Inc. 2007–2012. All rights reserved; distributed pursuant to a Limited License from 23andMe)

have been characterized through reliable, established research methods. Along with a text explanation, these health reports give a graphical depiction of a person's relative risk.

Figure 26.10 shows such a graphic for an individual's risk of venous thromboembolism from a circa 2012 version of the website, before the FDA began to regulate DTC genetic test-

ing. The graphical representation has been modified in more recent versions of the site to be less specific, more accurately reflecting the underlying uncertainty. For sensitive results such as BRCA1 and 2, and markers for Alzheimer's and Parkinson's disease, the information is initially "locked." Users must explicitly click through an additional screen to confirm that they truly want to know genotype and relative risk for that trait.

Rulings and Regulations

From a regulatory perspective, it was not initially clear whether these services qualify as medical devices as defined by the FDA, and are therefore subject to regulation by the Agency. In fall of 2013, the FDA sent a letter to 23andMe ordering it to "immediately discontinue marketing the PGS [Saliva Collection Kit and Personal Genome Service] until such time as it receives FDA marketing authorization for the device." (Annas and Elias 2014) A few weeks later, 23andMe announced that it was complying with the FDA's demands and suspending their health-related genetic tests. 23andMe continued to offer (and market) their ancestry-related testing, and less than 2 years later announced FDA approval for a carrier screening test for Bloom syndrome, with wording that left the option open for additional carrier screening tests without premarket review (Annas and Elias 2014, 2015). In October, 2017, the company announced that it would offer genetic risk for ten medical conditions including Parkinson's disease and late-onset Alzheimer's (Check Hayden 2017). The DTC testing landscape is still rapidly evolving, and will continue to do so for the foreseeable future.

Logistically, a prospective customer typically registers on the DTC company's website and a sample collection kit is sent in the mail, though 23andMe's kit is also available for purchase through ► Amazon.com and even on the shelves of local brick-and-mortar pharmacies. In May 2010, Pathway Genomics and Walgreens announced a plan to sell these kits in Walgreens drugstores, but the FDA sent a letter to Pathway Genomics indicating their belief that the company's genomic report qualified as medical device (Bradley et al. 2011) and as such required FDA approval. Plans to sell the saliva collection container in brick-and-mortar stores were put on hold until the regulatory issues could be resolved, or at least addressed.

Another high profile legal issue is the case of *Assoc. for Molecular Pathology v. Myriad Genetics, Inc.*, et al., regarding Myriad's patent on the BRCA1 and BRCA2 genes, which were included in 23andMe's offerings,[17] and more generally whether genes should be patentable at all. In 2011, a federal appeals court overturned a lower court in the case of and found that genes can, in fact, be patented (Pollack 2011). This ruling was upheld in a court of appeals in 2012, however, in 2013, the Supreme Court partially overturned that ruling and found that isolated genomic DNA (gDNA) is not patent-eligible, but cDNA is. Disappointingly, this ruling did not do much to reduce ambiguity around these issues.

26.10 Challenges and Future Directions

TBI as a discipline continues to evolve in an exciting and dynamic phase. Though challenges remain, the field is poised to become an increasingly crucial element of biomedical research and clinical practice in the era of precision medicine (Tenenbaum et al. 2016). We conclude this chapter with a discussion of future directions and key challenges for this burgeoning discipline.

26.10.1 Expansion of Data Types

Genomic data are already being used to guide clinical care. Genomic data themselves are relatively straightforward in that an individual's genome is relatively static, and through the intrinsic physical properties of ribonucleic acids and the transcriptional process, DNA and RNA are relatively easy to capture, observe, and quantify. Proteins and metabolites are more challenging in this regard. Proteomic

17 ► https://www.23andme.com/health/BRCA-Cancer/ (Accessed 12/19/12).

and metabolomic methodologies have primarily centered around isotopic labeling, but more recent approaches enable unbiased label-free identification and even quantification (Du et al. 2008; Wishart 2011). Identification of metabolites associated with disease has already enabled enzymatic drug targeting in diabetes, obesity, cardiovascular disease, and cancer, among other conditions (Chan and Ginsburg 2011). We expect that as proteomics and metabolomics standards and technologies continue to mature, they will play an increasingly significant role in translational research and practice.

The role of epigenetics needs to be understood more fully. It is clear that the environment can induce changes in the packaging and labeling of DNA. These environmental cues can include lifetime exposures to toxins, viruses, bacteria and nutritional compounds as well as drug exposures. Understanding the ways in which these epigenetic modifications affect phenotype is in its infancy, and so we must understand how to measure these effects, and then compute with them. The human microbiome is also an active area for translational research. Though one might expect associations between the gut microbiome and various gastro-intestinal conditions, surprising correlations and even causal relationships have been discovered with cancer, neurological, and even psychiatric disorders (Mayer et al. 2014; Zitvogel et al. 2015; Zheng et al. 2016; Clapp et al. 2017).

Finally, as standards are developed and clinicians and researchers see the value to be gained from structured data collection through studies such as The National Children's Study (Landrigan et al. 2006), structured environmental data is likely to be increasingly available to complete the picture for gene-environment interactions (Schwartz and Collins 2007).

26.10.2 Changes for Medical Training, Practice, and Support

Clinicians will need enhanced training in genetics and other areas described above. Curricular components relating to genetics, pharmacogenomics, statistics, and data standards will be increasingly important. Expertise in these fields will also need to be supplemented by an expanded workforce of genetic counselors. Increasingly, therapies will require accompanying diagnostic tests. As the opportunities for use of genomic data in clinical care continue to advance, it will become increasingly important to incorporate this information into both the electronic health record and into machine readable clinical care guidelines for clinical decision support. This in turn will require new standards to capture genomic findings, and new decision support tools to enable clinicians to incorporate this ever-increasing amount of information into their therapeutic decision making processes (Hoffman 2007). A number of standards exist in this space; the key will be in educating prospective users and enforcing adoption. This applies to the full translational spectrum, from annotation of experimentally generated datasets to a common format for the exchange of clinically relevant omic information between EHR systems. Most doctors have only a basic level of training in genetics, and are ill-equipped to answer in-depth questions from patients who bring to an appointment printouts of their results from these services (Frueh and Gurwitz 2004). More knowledge is required, in addition to training and tools, before family care providers, internists, and even specialists, are prepared to incorporate genomic information into their clinical practice (Ormond et al. 2010; Chan and Ginsburg 2011).

26.11 Conclusions

As the cost of data generation and storage continues to decrease, and the methods for data analysis and interpretation continue to advance, TBI is poised to be a key enabler of precision medicine (Tenenbaum et al. 2016). One can imagine a day when every newborn has his or her genome sequenced and this information becomes a part of the medical record, much as blood type is recorded today. The biggest challenges to achieving this vision are likely not to be technical ones, but rather

ethical, legal, and economic in nature (Schadt 2012). Society must strike a balance between privacy protection and facilitating progress in biomedical research. Legal issues will need to be worked out around direct-to-consumer genetic testing, gene patenting, preventing genetic discrimination, and many other such issues. Return on investment will need to be established through economic analysis combined with comparative effectiveness research (see ▶ Chaps. 11 and 26). Ultimately, someone will have to pay for these accompanying diagnostic tests. Major change is unlikely until an organization like the Center for Medicare and Medicaid Services (CMS) changes its policies. For example, CMS coverage for the genetic test to guide warfarin dosing is currently conditional upon it being ordered as part of a research protocol (Meckley and Neumann 2010). TBI will continue to play a key role in transforming these types of scientific discoveries into improvements in human health.

Suggested Readings

Altman, R. B., & Miller, K. S. (2011). 2010 translational bioinformatics year in review. Journal of the American Medical Informatics Association, 18, 358–366. This article summarizes Dr. Altman's third annual "year in review" presentation delivered at the 2010 AMIA Joint Summits on Translational Science in San Francisco.

Angrist, M. (2010). Here is a human being: At the dawn of personal genomics. New York: Harper. This text is written by one of the Personal Genome Project's first subjects, describing the project, the cohort, and the experience. It also gives a good overview of the background of the project and a number of ethical, legal, and social issues that it raises.

Capriotti, E., Nehrt, N. L., Kann, M. G., & Bromberg, Y. (2012, July). Bioinformatics for personal genome interpretation. Briefings in Bioinformatics, 13(4), 495–512. The authors of this review summarize key databases and bioinformatics tools that have been developed in recent years to aid in the interpretation of genomic variance. Resources covered include databases of variants, genotype/phenotype annotation databases, tools for gene prioritization and tools for interpretation of single nucleotide variants.

Davies, K. (2010). This text, written by the editor of BioIT World magazine, documents the characters, events, and issues in the race to achieve the $1000 Genome. In The $1000 genome: The revolution in DNA sequencing and the new era of personalized medicine. New York: Free Press.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning : Data mining, inference, and prediction. New York: Springer. A useful primer on the statistical concepts underlying machine learning approaches to biomarker discovery.

Kann, M. G, & Lewitter, F. (Eds.). (2012). Translational bioinformatics. PLOS Computational Biology Collections eBook. This eBook represents both the first "textbook" devoted entirely to TBI, and the first online, open access textbook from PLOS. In addition to many of the topics covered in this chapter, the collection includes chapters on related topics such as cancer genome analysis, micribiome analysis, structural variation, and protein interactions in disease.

Masys, D. R., Jarvik, G. P., Abernethy, N. F., Anderson, N. R., Papanicolaou, G. J., Paltoo, D. N., Hoffman, M. A., Kohane, I. S., & Levy, H. P. (2012). Technical desiderata for the integration of genomic data into electronic health records. Journal of Biomedical Informatics, 45(3), 419–422. The authors describe the characteristics of biomolecular data that differentiate it from other EHR data, enumerate a set of technical desiderata for management of biomolecular data in clinical settings (e.g., separation of molecular data observations from clinical interpretation, lossless data compression, support for readability by both humans and machines), and propose a technical approach to its representation.

Sarkar, I. N., & Payne, P. R. O. (2011, December). The joint summits on translational science: crossing the translational chasm. Journal of Biomedical Informatics, 44(Suppl 1), S1–S2. This editorial discusses the spectrum of biomedical informatics, from biology to medicine, in the context of the NIH Roadmap and the Clinical and Translational Science Award program. It gives the history of the AMIA Joint

Summits on Translational Science, and explains the emergence of TBI and CRI as disciplines unto themselves, intended to address the same issues that motivated those initiatives- namely translating scientific discoveries into meaningful changes in health care delivery.

Sarkar, I. N., Butte, A. J., Lussier, Y. A., Tarczy-Hornoch, P., & Ohno-Machado, L. (2011). Translational bioinformatics: Linking knowledge across biological and clinical realms. Journal of the American Medical Informatics Association, 18, 354–357. The authors present the field of TBI in the context of successes from bioinformatics and health informatics.

Questions for Discussion

1. Should DTC genetic testing for health-related traits be regulated by the FDA?
2. Should genes be patentable?
3. Are there sufficient legal protections in place to prevent discrimination based on genomic information? If not, what regulations are needed?
4. Are we headed toward full disclosure of genomic information?
5. What are some reasons a researcher might not want to share research data? Should they be required to share? If so, under what circumstances (e.g., 6 months after first publication)?
6. For novel analyses applied to complex, high-dimensional datasets, should there be new guidelines in place to prevent reporting erroneous results through user error or data fraud? Why or why not?
7. What are the major barriers to incorporating the benefits of personalized medicine fully into standard practice?

References

Alecu, I., Bousquet, C., Mougin, F., & Jaulent, M. C. (2006). Mapping of the WHO-ART terminology on Snomed CT to improve grouping of related adverse drug reactions. *Studies in Health Technology and Informatics, 124*, 833–838.

Alterovitz, G., Warner, J., Zhang, P., Chen, Y., Ullman-Cullere, M., Kreda, D., & Kohane, I. S. (2015). SMART on FHIR genomics: Facilitating standardized clinico-genomic apps. *Journal of the American Medical Informatics Association, 22*(6), 1173–1178.

Altman, R. B. (2007). PharmGKB: A logical home for knowledge relating genotype to drug response phenotype. *Nature Genetics, 39*(4), 426.

Altshuler, D. M., Durbin, R. M., & 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature, 467*(7319), 1061–1073.

Angrist, M. (2010). *Here is a human being : At the dawn of personal genomics*. New York: Harper.

Annas, G. J., & Elias, S. (2014). 23andMe and the FDA. *The New England Journal of Medicine, 370*(11), 985–988.

Aronson, A. R. (2001). Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. In *Proceedings of the AMIA symposium* (pp. 17–21).

Aronson, A. R., & Lang, F. M. (2010). An overview of MetaMap: Historical perspective and recent advances. *Journal of the American Medical Informatics Association, 17*(3), 229–236.

Aronson, A. R., Mork, J. G., Neveol, A., Shooshan, S. E., & Demner-Fushman, D. (2008). Methodology for creating UMLS content views appropriate for biomedical natural language processing. In *AMIA annual symposium proceedings* (pp. 21–25).

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics, 25*(1), 25–29.

Ashley, E. A., Butte, A. J., Wheeler, M. T., Chen, R., Klein, T. E., Dewey, F. E., Dudley, J. T., Ormond, K. E., Pavlovic, A., Morgan, A. A., Pushkarev, D., Neff, N. F., Hudgins, L., Gong, L., Hodges, L. M., Berlin, D. S., Thorn, C. F., Sangkuhl, K., Hebert, J. M., Woon, M., Sagreiya, H., Whaley, R., Knowles, J. W., Chou, M. F., Thakuria, J. V., Rosenbaum, A. M., Zaranek, A. W., Church, G. M., Greely, H. T., Quake, S. R., & Altman, R. B. (2010). Clinical assessment incorporating a personal genome. *Lancet, 375*(9725), 1525–1535.

Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., Spilker, B. A., Woodcock, J., & Zeger, S. L. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics, 69*(3), 89–95.

Auer, P. L., Reiner, A. P., Wang, G., Kang, H. M., Abecasis, G. R., Altshuler, D., Bamshad, M. J., Nickerson, D. A., Tracy, R. P., Rich, S. S., NHLBI GO Exome Sequencing Project, & Leal, S. M. (2016). Guidelines for large-scale sequence-based complex trait association studies: Lessons learned from the NHLBI exome sequencing project.

American Journal of Human Genetics, 99(4), 791–801.

Bahr, N. J., Nelson, S. D., Winnenburg, R., & Bodenreider, O. (2017). Eliciting the intension of drug value sets – principles and quality assurance applications. *Studies in Health Technology and Informatics, 245*, 843–847.

Barabasi, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nature Reviews. Genetics, 12*(1), 56–68.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets--update. *Nucleic Acids Research, 41*(Database issue), D991–D995.

Becker, K. G., Barnes, K. C., Bright, T. J., & Wang, S. A. (2004). The genetic association database. *Nature Genetics, 36*(5), 431–432.

Bodenreider, O. (2004). The unified medical language system (UMLS): Integrating biomedical terminology. *Nucleic Acids Research, 32*(Database issue), D267–D270.

Bradley, W. G., Golding, S. G., Herold, C. J., Hricak, H., Krestin, G. P., Lewin, J. S., Miller, J. C., Ringertz, H. G., & Thrall, J. H. (2011). Globalization of P4 medicine: Predictive, personalized, preemptive, and participatory--summary of the proceedings of the eighth international symposium of the International Society for Strategic Studies in Radiology, August 27–29, 2009. *Radiology, 258*(2), 571–582.

Brinkman, R. R., Courtot, M., Derom, D., Fostel, J. M., He, Y., Lord, P., Malone, J., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S. A., Soldatova, L. N., Stoeckert, C. J., Jr., Turner, J. A., & Zheng, J. (2010). Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics, 1*(Suppl 1), S7.

Brown, E. G., Wood, L., & Wood, S. (1999). The medical dictionary for regulatory activities (MedDRA). *Drug Safety, 20*(2), 109–117.

Bush, W. S., Crosslin, D. R., Owusu-Obeng, A., Wallace, J., Almoguera, B., Basford, M. A., Bielinski, S. J., Carrell, D. S., Connolly, J. J., Crawford, D., Doheny, K. F., Gallego, C. J., Gordon, A. S., Keating, B., Kirby, J., Kitchner, T., Manzi, S., Mejia, A. R., Pan, V., Perry, C. L., Peterson, J. F., Prows, C. A., Ralston, J., Scott, S. A., Scrol, A., Smith, M., Stallings, S. C., Veldhuizen, T., Wolf, W., Volpi, S., Wiley, K., Li, R., Manolio, T., Bottinger, E., Brilliant, M. H., Carey, D., Chisholm, R. L., Chute, C. G., Haines, J. L., Hakonarson, H., Harley, J. B., Holm, I. A., Kullo, I. J., Jarvik, G. P., Larson, E. B., McCarty, C. A., Williams, M. S., Denny, J. C., Rasmussen-Torvik, L. J., Roden, D. M., & Ritchie, M. D. (2016). Genetic variation among 82 pharmacogenes: The PGRNseq data from the eMERGE network. *Clinical Pharmacology and Therapeutics, 100*(2), 160–169.

Butte, A. J. (2008). Translational bioinformatics: Coming of age. *Journal of the American Medical Informatics Association, 15*(6), 709–714.

Butte, A. J., & Chen, R. (2006). Finding disease-related genomic experiments within an international repository: First steps in translational bioinformatics. In *AMIA annual symposium proceedings* (pp. 106–110).

Cai, B., Li, B., Kiga, N., Thusberg, J., Bergquist, T., Chen, Y. C., Niknafs, N., Carter, H., Tokheim, C., Beleva-Guthrie, V., Douville, C., Bhattacharya, R., Yeo, H. T. G., Fan, J., Sengupta, S., Kim, D., Cline, M., Turner, T., Diekhans, M., Zaucha, J., Pal, L. R., Cao, C., Yu, C. H., Yin, Y., Carraro, M., Giollo, M., Ferrari, C., Leonardi, E., Tosatto, S. C. E., Bobe, J., Ball, M., Hoskins, R. A., Repo, S., Church, G., Brenner, S. E., Moult, J., Gough, J., Stanke, M., Karchin, R., & Mooney, S. D. (2017). Matching phenotypes to whole genomes: Lessons learned from four iterations of the personal genome project community challenges. *Human Mutation, 38*(9), 1266–1276.

Camon, E., Magrane, M., Barrell, D., Binns, D., Fleischmann, W., Kersey, P., Mulder, N., Oinn, T., Maslen, J., Cox, A., & Apweiler, R. (2003). The gene ontology annotation (GOA) project: Implementation of GO in SWISS-PROT, TrEMBL, and InterPro. *Genome Research, 13*(4), 662–672.

Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J., & Bork, P. (2008). Drug target identification using side-effect similarity. *Science, 321*(5886), 263–266.

Cedar, H. (1988). DNA methylation and gene activity. *Cell, 53*(1), 3–4.

Chan, I. S., & Ginsburg, G. S. (2011). Personalized medicine: Progress and promise. *Annual Review of Genomics and Human Genetics, 12*, 217–244.

Check Hayden, E. (2017). The rise and fall and rise again of 23andMe. *Nature, 550*(7675), 174–177.

Chen, X., Gururaj, A. E., Ozyurt, B., Liu, R., Soysal, E., Cohen, T., Tiryaki, F., Li, Y., Zong, N., Jiang, M., Rogith, D., Salimi, M., Kim, H. E., Rocca-Serra, P., Gonzalez-Beltran, A., Farcas, C., Johnson, T., Margolis, R., Alter, G., Sansone, S. A., Fore, I. M., Ohno-Machado, L., Grethe, J. S., & Xu, H. (2018). DataMed – an open source discovery index for finding biomedical datasets. *Journal of the American Medical Informatics Association, 25*(3), 300–308.

Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., Nayir, A., Bakkaloglu, A., Ozen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S., & Lifton, R. P. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America, 106*(45), 19096–19101.

Church, G., Heeney, C., Hawkins, N., de Vries, J., Boddington, P., Kaye, J., Bobrow, M., & Weir, B. (2009). Public access to genome-wide data: Five

views on balancing research with privacy and protection. *PLoS Genetics, 5*(10), e1000665.

Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews. Genetics, 11*(6), 415–425.

Clapp, M., Aurora, N., Herrera, L., Bhatia, M., Wilen, E., & Wakefield, S. (2017). Gut microbiota's effect on mental health: The gut-brain axis. *Clinics and Practice, 7*(4), 987.

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *The New England Journal of Medicine, 372*(9), 793–795.

Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation, 115*(7), 928–935.

Coulet, A., Shah, N. H., Garten, Y., Musen, M., & Altman, R. B. (2010). Using text to build semantic networks for pharmacogenomics. *Journal of Biomedical Informatics, 43*(6), 1009–1019.

Crick, F. (1970). Central dogma of molecular biology. *Nature, 227*(5258), 561–563.

Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D. R., Roden, D. M., & Crawford, D. C. (2010). PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics, 26*(9), 1205–1210.

Denny, J. C., Bastarache, L., & Roden, D. M. (2016). Phenome-wide association studies as a tool to advance precision medicine. *Annual Review of Genomics and Human Genetics, 17*, 353–373.

Dewey, F. E., Chen, R., Cordero, S. P., Ormond, K. E., Caleshu, C., Karczewski, K. J., Whirl-Carrillo, M., Wheeler, M. T., Dudley, J. T., Byrnes, J. K., Cornejo, O. E., Knowles, J. W., Woon, M., Sangkuhl, K., Gong, L., Thorn, C. F., Hebert, J. M., Capriotti, E., David, S. P., Pavlovic, A., West, A., Thakuria, J. V., Ball, M. P., Zaranek, A. W., Rehm, H. L., Church, G. M., West, J. S., Bustamante, C. D., Snyder, M., Altman, R. B., Klein, T. E., Butte, A. J., & Ashley, E. A. (2011). Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. *PLoS Genetics, 7*(9), e1002280.

Doan, S., Lin, K. W., Conway, M., Ohno-Machado, L., Hsieh, A., Feupe, S. F., Garland, A., Ross, M. K., Jiang, X., Farzaneh, S., Walker, R., Alipanah, N., Zhang, J., Xu, H., & Kim, H. E. (2014). PhenDisco: Phenotype discovery system for the database of genotypes and phenotypes. *Journal of the American Medical Informatics Association, 21*(1), 31–36.

Du, X., Callister, S. J., Manes, N. P., Adkins, J. N., Alexandridis, R. A., Zeng, X., Roh, J. H., Smith, W. E., Donohue, T. J., Kaplan, S., Smith, R. D., & Lipton, M. S. (2008). A computational strategy to analyze label-free temporal bottom-up proteomics data. *Journal of Proteome Research, 7*(7), 2595–2604.

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., Shoresh, N., Simon, J. M., Song, L., Trinklein, N. D., Altshuler, R. C., Birney, E., Brown, J. B., Cheng, C., Djebali, S., Dong, X., Ernst, J., Furey, T. S., Gerstein, M., Giardine, B., Greven, M., Hardison, R. C., Harris, R. S., Herrero, J., Hoffman, M. M., Iyer, S., Kelllis, M., Kheradpour, P., Lassman, T., Li, Q., Lin, X., Marinov, G. K., Merkel, A., Mortazavi, A., Parker, S. C., Reddy, T. E., Rozowsky, J., Schlesinger, F., Thurman, R. E., Wang, J., Ward, L. D., Whitfield, T. W., Wilder, S. P., Wu, W., Xi, H. S., Yip, K. Y., Zhuang, J., Bernstein, B. E., Green, E. D., Gunter, C., Snyder, M., Pazin, M. J., Lowdon, R. F., Dillon, L. A., Adams, L. B., Kelly, C. J., Zhang, J., Wexler, J. R., Good, P. J., Feingold, E. A., Crawford, G. E., Dekker, J., Elinitski, L., Farnham, P. J., Giddings, M. C., Gingeras, T. R., Guigo, R., Hubbard, T. J., Kellis, M., Kent, W. J., Lieb, J. D., Margulies, E. H., Myers, R. M., Starnatoyannopoulos, J. A., Tennebaum, S. A., Weng, Z., White, K. P., Wold, B., Yu, Y., Wrobel, J., Risk, B. A., Gunawardena, H. P., Kuiper, H. C., Maier, C. W., Xie, L., Chen, X., Mikkelsen, T. S., Gillespie, S., Goren, A., Ram, O., Zhang, X., Wang, L., Issner, R., Coyne, M. J., Durham, T., Ku, M., Truong, T., Eaton, M. L., Dobin, A., Lassmann, T., Tanzer, A., Lagarde, J., Lin, W., Xue, C., Williams, B. A., Zaleski, C., Roder, M., Kokocinski, F., Abdelhamid, R. F., Alioto, T., Antoshechkin, I., Baer, M. T., Batut, P., Bell, I., Bell, K., Chakrabortty, S., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M. J., Gao, H., Gonzalez, D., Gordon, A., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Li, G., Luo, O. J., Park, E., Preall, J. B., Presaud, K., Ribeca, P., Robyr, D., Ruan, X., Sammeth, M., Sandu, K. S., Schaeffer, L., See, L. H., Shahab, A., Skancke, J., Suzuki, A. M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Hayashizaki, Y., Reymond, A., Antonarakis, S. E., Hannon, G. J., Ruan, Y., Carninci, P., Sloan, C. A., Learned, K., Malladi, V. S., Wong, M. C., Barber, G. P., Cline, M. S., Dreszer, T. R., Heitner, S. G., Karolchik, D., Kirkup, V. M., Meyer, L. R., Long, J. C., Maddren, M., Raney, B. J., Grasfeder, L. L., Giresi, P. G., Battenhouse, A., Sheffield, N. C., Showers, K. A., London, D., Bhinge, A. A., Shestak, C., Schaner, M. R., Kim, S. K., Zhang, Z. Z., Mieczkowski, P. A., Mieczkowska, J. O., Liu, Z., McDaniell, R. M., Ni, Y., Rashid, N. U., Kim, M. J., Adar, S., Zhang, Z., Wang, T., Winter, D., Keefe, D., Iyer, V. R., Sandhu, K. S., Zheng, M., Wang, P., Gertz,

J., Vielmetter, J., Partridge, E. C., Varley, K. E., Gasper, C., Bansal, A., Pepke, S., Jain, P., Amrhein, H., Bowling, K. M., Anaya, M., Cross, M. K., Muratet, M. A., Newberry, K. M., McCue, K., Nesmith, A. S., Fisher-Aylor, K. I., Pusey, B., DeSalvo, G., Parker, S. L., Balasubramanian, S., Davis, N. S., Meadows, S. K., Eggleston, T., Newberry, J. S., Levy, S. E., Absher, D. M., Wong, W. H., Blow, M. J., Visel, A., Pennachio, L. A., Elnitski, L., Petrykowska, H. M., Abyzov, A., Aken, B., Barrell, D., Barson, G., Berry, A., Bignell, A., Boychenko, V., Bussotti, G., Davidson, C., Despacio-Reyes, G., Diekhans, M., Ezkurdia, I., Frankish, A., Gilbert, J., Gonzalez, J. M., Griffiths, E., Harte, R., Hendrix, D. A., Hunt, T., Jungreis, I., Kay, M., Khurana, E., Leng, J., Lin, M. F., Loveland, J., Lu, Z., Manthravadi, D., Mariotti, M., Mudge, J., Mukherjee, G., Notredame, C., Pei, B., Rodriguez, J. M., Saunders, G., Sboner, A., Searle, S., Sisu, C., Snow, C., Steward, C., Tapanan, E., Tress, M. L., van Baren, M. J., Washieti, S., Wilming, L., Zadissa, A., Zhengdong, Z., Brent, M., Haussler, D., Valencia, A., Raymond, A., Addleman, N., Alexander, R. P., Auerbach, R. K., Bettinger, K., Bhardwaj, N., Boyle, A. P., Cao, A. R., Cayting, P., Charos, A., Cheng, Y., Eastman, C., Euskirchen, G., Fleming, J. D., Grubert, F., Habegger, L., Hariharan, M., Harmanci, A., Iyenger, S., Jin, V. X., Karczewski, K. J., Kasowski, M., Lacroute, P., Lam, H., Larnarre-Vincent, N., Lian, J., Lindahl-Allen, M., Min, R., Miotto, B., Monahan, H., Moqtaderi, Z., Mu, X. J., O'Geen, H., Ouyang, Z., Patacsil, D., Raha, D., Ramirez, L., Reed, B., Shi, M., Slifer, T., Witt, H., Wu, L., Xu, X., Yan, K. K., Yang, X., Struhl, K., Weissman, S. M., Tenebaum, S. A., Penalva, L. O., Karmakar, S., Bhanvadia, R. R., Choudhury, A., Domanus, M., Ma, L., Moran, J., Victorsen, A., Auer, T., Centarin, L., Eichenlaub, M., Gruhl, F., Heerman, S., Hoeckendorf, B., Inoue, D., Kellner, T., Kirchmaier, S., Mueller, C., Reinhardt, R., Schertel, L., Schneider, S., Sinn, R., Wittbrodt, B., Wittbrodt, J., Jain, G., Balasundaram, G., Bates, D. L., Byron, R., Canfield, T. K., Diegel, M. J., Dunn, D., Ebersol, A. K., Frum, T., Garg, K., Gist, E., Hansen, R. S., Boatman, L., Haugen, E., Humbert, R., Johnson, A. K., Johnson, E. M., Kutyavin, T. M., Lee, K., Lotakis, D., Maurano, M. T., Neph, S. J., Neri, F. V., Nguyen, E. D., Qu, H., Reynolds, A. P., Roach, V., Rynes, E., Sanchez, M. E., Sandstrom, R. S., Shafer, A. O., Stergachis, A. B., Thomas, S., Vernot, B., Vierstra, J., Vong, S., Weaver, M. A., Yan, Y., Zhang, M., Akey, J. A., Bender, M., Dorschner, M. O., Groudine, M., MacCoss, M. J., Navas, P., Stamatoyannopoulos, G., Stamatoyannopoulos, J. A., Beal, K., Brazma, A., Flicek, P., Johnson, N., Lukk, M., Luscombe, N. M., Sobral, D., Vaquerizas, J. M., Batzoglou, S., Sidow, A., Hussami, N., Kyriazopoulou-Panagiotopoulou, S., Libbrecht, M. W., Schaub, M. A., Miller, W., Bickel, P. J., Banfai, B., Boley, N. P., Huang, H., Li, J. J., Noble, W. S., Bilmes, J. A., Buske, O. J., Sahu, A. O., Kharchenko, P. V., Park, P. J., Baker, D., Taylor, J., & Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57–74.

Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research, 30*(1), 207–210.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C. D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., & D'Eustachio, P. (2018). The reactome pathway knowledgebase. *Nucleic Acids Research, 46*(D1), D649–D655.

Frueh, F. W., & Gurwitz, D. (2004). From pharmacogenetics to personalized medicine: A vital need for educating health professionals and the community. *Pharmacogenomics, 5*(5), 571–579.

Gamazon, E. R., Huang, R. S., Cox, N. J., & Dolan, M. E. (2010). Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. *Proceedings of the National Academy of Sciences of the United States of America, 107*(20), 9287–9292.

Garber, M. E., Saldanha, A., Parker, J. S., Jones, W. D., Kaukinen, K., Laurila, K., Lahdeaho, M. L., Khatri, P., Khosla, C., Adelman, D. C., & Maki, M. (2017). A B-cell gene signature correlates with the extent of gluten-induced intestinal injury in celiac disease. *Cellular and Molecular Gastroenterology and Hepatology, 4*(1), 1–17.

Garten, Y., & Altman, R. B. (2009). Pharmspresso: A text mining tool for extraction of pharmacogenomic concepts and relationships from full text. *BMC Bioinformatics, 10*(Suppl 2), S6.

Garten, Y., Coulet, A., & Altman, R. B. (2010). Recent progress in automatically extracting information from the pharmacogenomic literature. *Pharmacogenomics, 11*(10), 1467–1489.

Ge, D., Ruzzo, E. K., Shianna, K. V., He, M., Pelak, K., Heinzen, E. L., Need, A. C., Cirulli, E. T., Maia, J. M., Dickson, S. P., Zhu, M., Singh, A., Allen, A. S., & Goldstein, D. B. (2011). SVA: Software for annotating and visualizing sequenced human genomes. *Bioinformatics, 27*(14), 1998–2000.

Ghazvinian, A., Noy, N. F., & Musen, M. A. (2009). Creating mappings for ontologies in biomedicine: Simple methods work. In *AMIA annual symposium proceedings 2009* (pp. 198–202).

Ghazvinian, A., Noy, N. F., & Musen, M. A. (2011). How orthogonal are the OBO Foundry ontologies? *Journal of Biomedical Semantics, 2*(Suppl 2), S2.

Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabasi, A. L. (2007). The human disease network. *Proceedings of the National Academy of Sciences of the United States of America, 104*(21), 8685–8690.

Grando, A., & Schwab, R. (2013). Building and evaluating an ontology-based tool for reasoning about consent permission. In *AMIA annual symposium proceedings 2013* (pp. 514–523).

Green, R. C., Roberts, J. S., Cupples, L. A., Relkin, N. R., Whitehouse, P. J., Brown, T., Eckert, S. L., Butson, M., Sadovnick, A. D., Quaid, K. A., Chen, C., Cook-Deegan, R., & Farrer, L. A. (2009). Disclosure of APOE genotype for risk of Alzheimer's disease. *The New England Journal of Medicine, 361*(3), 245–254.

Gregory, S. (2011).

Hadley, D., Pan, J., El-Sayed, O., Aljabban, J., Aljabban, I., Azad, T. D., Hadied, M. O., Raza, S., Rayikanti, B. A., Chen, B., Paik, H., Aran, D., Spatz, J., Himmelstein, D., Panahiazar, M., Bhattacharya, S., Sirota, M., Musen, M. A., & Butte, A. J. (2017). Precision annotation of digital samples in NCBI's gene expression omnibus. *Scientific Data, 4*, 170125.

Hansen, N. T., Brunak, S., & Altman, R. B. (2009). Generating genome-scale candidate gene lists for pharmacogenomics. *Clinical Pharmacology and Therapeutics, 86*(2), 183–189.

Heikamp, K., & Bajorath, J. (2011). Large-scale similarity search profiling of ChEMBL compound data sets. *Journal of Chemical Information and Modeling, 51*(8), 1831–1839.

Hoffman, M. A. (2007). The genome-enabled electronic medical record. *Journal of Biomedical Informatics, 40*(1), 44–46.

Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, Pearson JV, Stephan DA, Nelson SF, Craig DW. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. *29*;4(8):e1000167. https://doi.org/10.1371/journal.pgen.1000167. PMID: 18769715; PMCID: PMC2516199.

Homerova, D., Sprusansky, O., Kutejova, E., & Kormanec, J. (2002). Some features of DNA-binding proteins involved in the regulation of the Streptomyces aureofaciens gap gene, encoding glyceraldehyde-3-phosphate dehydrogenase. *Folia Microbiologia (Praha), 47*(4), 311–317.

Hsi-Yang Fritz, M., Leinonen, R., Cochrane, G., & Birney, E. (2011). Efficient storage of high throughput DNA sequencing data using reference-based compression. *Genome Research, 21*(5), 734–740.

Hu, J. X., Thomas, C. E., & Brunak, S. (2016). Network biology concepts in complex disease comorbidities. *Nature Reviews. Genetics, 17*(10), 615–629.

Hudson KL, Holohan MK, Collins FS. (2008). pace with the times--the Genetic Information Nondiscrimination Act of 2008. N Engl J Med. *19*;358(25):2661-3. https://doi.org/10.1056/NEJMp0803964. PMID: 18565857.

Irwin, J. J., & Shoichet, B. K. (2005). ZINC--a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling, 45*(1), 177–182.

Jonquet, C., Musen, M. A., & Shah, N. H. (2010). Building a biomedical ontology recommender web service. *Journal of Biomedical Semantics, 1*(Suppl 1), S1.

Jonquet, C., LePendu, P., Falconer, S., Coulet, A., Noy, N. F., Musen, M. A., & Shah, N. H. (2011). NCBO Resource Index: Ontology-based search and mining of biomedical resources. *Web Semantics: Science, Services and Agents on the World Wide Web, 9*(3), 316–324.

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G. R., Wu, G. R., Matthews, L., Lewis, S., Birney, E., & Stein, L. (2005). Reactome: A knowledgebase of biological pathways. *Nucleic Acids Research, 33*(Database issue), D428–D432.

Kang, M. J., Lee, C. G., Cho, S. J., Homer, R. J., & Elias, J. A. (2006). IFN-gamma-dependent DNA injury and/or apoptosis are critical in cigarette smoke-induced murine emphysema. *Proceedings of the American Thoracic Society, 3*(6), 517–518.

Kaye, J., Heeney, C., Hawkins, N., de Vries, J., & Boddington, P. (2009). Data sharing in genomics--re-shaping scientific practice. *Nature Reviews. Genetics, 10*(5), 331–335.

Kaye J, Hawkins N. (2014). Data sharing policy design for consortia: challenges for sustainability. *Genome Med. 29*;6(1):4. https://doi.org/10.1186/gm523. PMID: 24475754; PMCID: PMC3978924.

Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A. I., Hufeisen, S. J., Jensen, N. H., Kuijer, M. B., Matos, R. C., Tran, T. B., Whaley, R., Glennon, R. A., Hert, J., Thomas, K. L., Edwards, D. D., Shoichet, B. K., & Roth, B. L. (2009). Predicting new molecular targets for known drugs. *Nature, 462*(7270), 175–181.

Khatri, P., & Draghici, S. (2005). Ontological analysis of gene expression data: Current tools, limitations, and open problems. *Bioinformatics, 21*(18), 3587–3595.

Kho, A. N., Pacheco, J. A., Peissig, P. L., Rasmussen, L., Newton, K. M., Weston, N., Crane, P. K., Pathak, J., Chute, C. G., Bielinski, S. J., Kullo, I. J., Li, R., Manolio, T. A., Chisholm, R. L., & Denny, J. C. (2011). Electronic medical records for genetic research: results of the eMERGE consortium. *Science Translational Medicine, 3*(79), 79re71.

Kim, H., Bell, E., Kim, J., Sitapati, A., Ramsdell, J., Farcas, C., Friedman, D., Feupe, S. F., & Ohno-Machado, L. (2017). iCONCUR: Informed consent for clinical data and bio-sample use for research. *Journal of the American Medical Informatics Association, 24*(2), 380–387.

Kinnings, S. L., Liu, N., Buchmeier, N., Tonge, P. J., Xie, L., & Bourne, P. E. (2009). Drug discovery using chemical systems biology: Repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Computational Biology, 5*(7), e1000423.

Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V.,

Djoumbou, Y., Eisner, R., Guo, A. C., & Wishart, D. S. (2011). DrugBank 3.0: A comprehensive resource for 'omics' research on drugs. *Nucleic Acids Research, 39*(Suppl 1), D1035–D1041.

Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y. A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., Megy, K., Pilicheva, E., Rustici, G., Tikhonov, A., Parkinson, H., Petryszak, R., Sarkans, U., & Brazma, A. (2015). ArrayExpress update--simplifying data submissions. *Nucleic Acids Research, 43*(Database issue), D1113–D1116.

Kraus, V. B., Burnett, B., Coindreau, J., Cottrell, S., Eyre, D., Gendreau, M., Gardiner, J., Garnero, P., Hardin, J., Henrotin, Y., Heinegard, D., Ko, A., Lohmander, L. S., Matthews, G., Menetski, J., Moskowitz, R., Persiani, S., Poole, A. R., Rousseau, J. C., & Todman, M. (2011). Application of biomarkers in the development of drugs intended for the treatment of osteoarthritis. *Osteoarthritis Cartilage, 19*(5), 515–542.

Kuhn, M., Campillos, M., Letunic, I., Jensen, L. J., & Bork, P. (2010). A side effect resource to capture phenotypic effects of drugs. *Molecular Systems Biology, 6*, 343.

Kweon, J., & Kim, Y. (2018). High-throughput genetic screens using CRISPR-Cas9 system. *Archives of Pharmacal Research, 41*(9), 875–884.

Lamb, J. (2007). The Connectivity Map: A new tool for biomedical research. *Nature Reviews. Cancer, 7*(1), 54–60.

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J. P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., & Golub, T. R. (2006). The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science, 313*(5795), 1929–1935.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., & Chen, Y. J. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860–921.

Landrigan, P. J., Trasande, L., Thorpe, L. E., Gwynn, C., Lioy, P. J., D'Alton, M. E., Lipkind, H. S., Swanson, J., Wadhwa, P. D., Clark, E. B., Rauh, V. A., Perera, F. P., & Susser, E. (2006). The National Children's Study: A 21-year prospective study of 100,000 American children. *Pediatrics, 118*(5), 2173–2186.

Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon,

R., Rubinstein, W., & Maglott, D. R. (2016). ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Research, 44*(D1), D862–D868.

Larmuseau, M. H. D., Matthijs, K., & Wenseleers, T. (2016). Cuckolded fathers rare in human populations. *Trends in Ecology & Evolution, 31*(5), 327–329.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D. N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M. I., Moonshine, A. L., Natarajan, P., Orozco, L., Peloso, G. M., Poplin, R., Rivas, M. A., Ruano-Rubio, V., Rose, S. A., Ruderfer, D. M., Shakir, K., Stenson, P. D., Stevens, C., Thomas, B. P., Tiao, G., Tusie-Luna, M. T., Weisburd, B., Won, H. H., Yu, D., Altshuler, D. M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J. C., Gabriel, S. B., Getz, G., Glatt, S. J., Hultman, C. M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M. I., McGovern, D., McPherson, R., Neale, B. M., Palotie, A., Purcell, S. M., Saleheen, D., Scharf, J. M., Sklar, P., Sullivan, P. F., Tuomilehto, J., Tsuang, M. T., Watkins, H. C., Wilson, J. G., Daly, M. J., MacArthur, D. G., & Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature, 536*(7616), 285–291.

Lenfant, C. (2003). Shattuck lecture--clinical research to clinical practice--lost in translation? *The New England Journal of Medicine, 349*(9), 868–874.

LePendu, P., Musen, M. A., & Shah, N. H. (2011a). Enabling enrichment analysis with the human disease ontology. *Journal of Biomedical Informatics, 44*(Suppl 1), S31–S38.

LePendu, P., Racunas, S. A., Iyer, S., Liu, Y., Fairon, C., & Shah, N. (2011b). Annotation analysis for testing drug safety signals. In *The 14th bio-ontologies SIG meeting at ISMB 2011*. Vienna, Austria.

LePendu, P., Iyer, S. V., Bauer-Mehren, A., Harpaz, R., Mortensen, J. M., Podchiyska, T., Ferris, T. A., & Shah, N. H. (2013). Pharmacovigilance using clinical notes. *Clinical Pharmacology and Therapeutics, 93*(6), 547–555.

Lindhurst, M. J., Sapp, J. C., Teer, J. K., Johnston, J. J., Finn, E. M., Peters, K., Turner, J., Cannons, J. L., Bick, D., Blakemore, L., Blumhorst, C., Brockmann, K., Calder, P., Cherman, N., Deardorff, M. A., Everman, D. B., Golas, G., Greenstein, R. M., Kato, B. M., Keppler-Noreuil, K. M., Kuznetsov, S. A., Miyamoto, R. T., Newman, K., Ng, D., O'Brien, K., Rothenberg, S., Schwartzentruber, D. J., Singhal, V., Tirabosco, R., Upton, J., Wientroub, S., Zackai, E. H., Hoag, K., Whitewood-Neal, T., Robey, P. G., Schwartzberg, P. L., Darling, T. N., Tosi, L. L., Mullikin, J. C., & Biesecker, L. G. (2011). A mosaic activating mutation in AKT1 associated with the Proteus syndrome. *The New England Journal of Medicine, 365*(7), 611–619.

Liu, T., & Altman, R. B. (2015). Relating essential proteins to drug side-effects using canonical component analysis: A structure-based approach. *Journal of Chemical Information and Modeling, 55*(7), 1483–1494.

Liu, T., Lin, Y., Wen, X., Jorissen, R. N., & Gilson, M. K. (2007). BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Research, 35*(Database issue), D198–D201.

Liu, M., Shah, A., Jiang, M., Peterson, N. B., Dai, Q., Aldrich, M. C., Chen, Q., Bowton, E. A., Liu, H., Denny, J. C., & Xu, H. (2012). A study of transportability of an existing smoking status detection module across institutions. In *AMIA annual symposium proceedings* (pp. 577–586).

Lunshof, J. E., Bobe, J., Aach, J., Angrist, M., Thakuria, J. V., Vorhaus, D. B., Hoehe, M. R., & Church, G. M. (2010). Personal genomes in progress: From the human genome project to the personal genome project. *Dialogues in Clinical Neuroscience, 12*(1), 47–60.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z. M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., & Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research, 45*(D1), D896–D901.

Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., Popova, N., Pretel, S., Ziyabari, L., Lee, M., Shao, Y., Wang, Z. Y., Sirotkin, K., Ward, M., Kholodov, M., Zbicz, K., Beck, J., Kimelman, M., Shevelev, S., Preuss, D., Yaschenko, E., Graeff, A., Ostell, J., & Sherry, S. T. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nature Genetics, 39*(10), 1181–1186.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. (2009) Finding the missing heritability of complex diseases. *Nature. 8*;461(7265):747–53. https://doi.org/10.1038/nature08494. PMID: 19812666; PMCID: PMC2831613.

Mayer, E. A., Knight, R., Mazmanian, S. K., Cryan, J. F., & Tillisch, K. (2014). Gut microbes and the brain: Paradigm shift in neuroscience. *The Journal of Neuroscience, 34*(46), 15490–15496.

Meckley, L. M., & Neumann, P. J. (2010). Personalized medicine: Factors influencing reimbursement. *Health Policy, 94*(2), 91–100.

Miller, G. C., & Britt, H. (1995). A new drug classification for computer systems: The ATC extension code. *International Journal of Bio-Medical Computing, 40*(2), 121–124.

Mork, J. G., Bodenreider, O., Demner-Fushman, D., Dogan, R. I., Lang, F. M., Lu, Z., Neveol, A., Peters, L., Shooshan, S. E., & Aronson, A. R. (2010). Extracting Rx information from clinical narrative. *Journal of the American Medical Informatics Association, 17*(5), 536–539.

Mort, M., Evani, U. S., Krishnan, V. G., Kamati, K. K., Baenziger, P. H., Bagchi, A., Peters, B. J., Sathyesh, R., Li, B., Sun, Y., Xue, B., Shah, N. H., Kann, M. G., Cooper, D. N., Radivojac, P., & Mooney, S. D. (2010). In silico functional profiling of human disease-associated and polymorphic amino acid substitutions. *Human Mutation, 31*(3), 335–346.

Musen, M. A., Noy, N. F., Shah, N. H., Chute, C. G., Store, M.-A., Smith, B., & NCBO team. (2011). The National Center for Biomedical Ontology. *Journal of the American Medical Informatics Association, 19*(2), 190–195.

Musen, M. A., Bean, C. A., Cheung, K. H., Dumontier, M., Durante, K. A., Gevaert, O., Gonzalez-Beltran, A., Khatri, P., Kleinstein, S. H., O'Connor, M. J., Pouliot, Y., Rocca-Serra, P., Sansone, S. A., Wiser, J. A., & CEDAR team. (2015). The center for expanded data annotation and retrieval. *Journal of the American Medical Informatics Association, 22*(6), 1148–1152.

Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., Shendure, J., & Bamshad, M. J. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics, 42*(1), 30–35.

Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., & Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genetics, 6*(4), e1000888.

NIH broadens genomic data-sharing policy. (2014). Cancer Discovery, 4(12): OF4. https://pubmed.ncbi.nlm.nih.gov/25477121/

Nikolsky, Y., Ekins, S., Nikolskaya, T., & Bugrim, A. (2005). A novel method for generation of signature networks as biomarkers from complex high throughput data. *Toxicology Letters, 158*(1), 20–29.

Novack, D. H., Plumer, R., Smith, R. L., Ochitill, H., Morrow, G. R., & Bennett, J. M. (1979). Changes in physicians' attitudes toward telling the cancer patient. *JAMA, 241*(9), 897–900.

Obeid, J., Gabriel, D., & Sanderson, I. (2010). A biomedical research permissions ontology: cognitive and knowledge representation considerations. In *Proceedings of the 2010 workshop on governance of technology, information and policies*. Austin, Texas: ACM.

Olson, S. (2007). Who's your daddy? *The Atlantic*. July/August.

Ormond, K. E., Wheeler, M. T., Hudgins, L., Klein, T. E., Butte, A. J., Altman, R. B., Ashley, E. A., & Greely, H. T. (2010). Challenges in the clinical application of whole-genome sequencing. *Lancet, 375*(9727), 1749–1751.

Overington, J. (2009). ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). Interview by Wendy A. Warr. *Journal of Computer-Aided Molecular Design, 23*(4), 195–198.

Parrish, F., Do, N., Bouhaddou, O., & Warnekar, P. (2006). Implementation of RxNorm as a terminology mediation standard for exchanging pharmacy medication between federal agencies. In *AMIA annual symposium proceedings* (p. 1057).

Pencina, M. J., D'Agostino, R. B., Sr., D'Agostino, R. B., Jr., & Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine, 27*(2), 157–172; discussion 207–112.

Percha, B., & Altman, R. B. (2015). Learning the structure of biomedical relationships from unstructured text. *PLoS Computational Biology, 11*(7), e1004216.

Percha, B., & Altman, R. B. (2018). A global network of biomedical relationships derived from text. *Bioinformatics, 34*(15), 2614–2624.

Pollack, A. (2011). Ruling Upholds gene patent in cancer test. *The New York Times*. New York City, The New York Times Company.

Rehm, H. L., Berg, J. S., Brooks, L. D., Bustamante, C. D., Evans, J. P., Landrum, M. J., Ledbetter, D. H., Maglott, D. R., Martin, C. L., Nussbaum, R. L., Plon, S. E., Ramos, E. M., Sherry, S. T., Watson, M. S., & ClinGen. (2015). ClinGen--the clinical genome resource. *The New England Journal of Medicine, 372*(23), 2235–2242.

Rhee, S. Y., Wood, V., Dolinski, K., & Draghici, S. (2008). Use and misuse of the gene ontology annotations. *Nature Reviews Genetics, 9*(7), 509–515.

Ritchie, M. D., Denny, J. C., Crawford, D. C., Ramirez, A. H., Weiner, J. B., Pulley, J. M., Basford, M. A., Brown-Gentry, K., Balser, J. R., Masys, D. R., Haines, J. L., & Roden, D. M. (2010). Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *American Journal of Human Genetics, 86*(4), 560–572.

Rocca-Serra, P., Brandizi, M., Maguire, E., Sklyar, N., Taylor, C., Begley, K., Field, D., Harris, S., Hide, W., Hofmann, O., Neumann, S., Sterk, P., Tong, W., & Sansone, S. A. (2010). ISA software suite: Supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics, 26*(18), 2354–2356.

Rohrer Vitek, C. R., Abul-Husn, N. S., Connolly, J. J., Hartzler, A. L., Kitchner, T., Peterson, J. F., Rasmussen, L. V., Smith, M. E., Stallings, S., Williams, M. S., Wolf, W. A., & Prows, C. A. (2017). Healthcare provider education to support integration of pharmacogenomics in practice: The

eMERGE Network experience. *Pharmacogenomics, 18*(10), 1013–1025.

Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S. S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J. C., Lashkari, D., Shalon, D., Myers, T. G., Weinstein, J. N., Botstein, D., & Brown, P. O. (2000). Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics, 24*(3), 227–235.

Rusk, N. (2010). Expanding HapMap. *Nature Methods, 7*(10), 780–781.

Sarkar, I. N., Butte, A. J., Lussier, Y. A., Tarczy-Hornoch, P., & Ohno-Machado, L. (2011). Translational bioinformatics: Linking knowledge across biological and clinical realms. *Journal of the American Medical Informatics Association, 18*(4), 354–357.

Schadt, E. E. (2012). The changing privacy landscape in the era of big data. *Molecular Systems Biology, 8*, 612.

Schaefer, C. F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., & Buetow, K. H. (2009). PID: The pathway interaction database. *Nucleic Acids Research, 37*(Database issue), D674–D679.

Schwartz, D., & Collins, F. (2007). Medicine. Environmental biology and human disease. *Science, 316*(5825), 695–696.

Seal, R. L., Gordon, S. M., Lush, M. J., Wright, M. W., & Bruford, E. A. (2011). genenames.org: The HGNC resources in 2011. *Nucleic Acids Research, 39*(Database issue), D514–D519.

Shah, N. H., & Musen, M. A. (2008). UMLS-Query: a perl module for querying the UMLS. In *AMIA annual symposium proceedings* (pp. 652–656).

Shah, N. H., Jonquet, C., Chiang, A. P., Butte, A. J., Chen, R., & Musen, M. A. (2009a). Ontology-driven indexing of public datasets for translational bioinformatics. *BMC Bioinformatics, 10*(Suppl 2), S1.

Shah, S. H., Crosslin, D. R., Nelson, S., Haynes, C., Boling, C. L., Bain, J., Muehlbauer, M., Stevens, R. D., Brantley, P. J., Stevens, V., Hollis, J. F., Appel, L. J., Lien, L., Newgard, C. B., & Svetkey, L. P. (2009b). Metabolomic profiles are associated with baseline insulin resistance and improvement in insulin resistance with weight loss in the WLM trial. *Ciculation, 1*(20), S467.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research, 29*(1), 308–311.

Shortliffe, E. H., Califano, A., & Hunter, L. (2009). New JBI emphasis on translational bioinformatics. *Journal of Biomedical Informatics, 42*(2), 199–200.

Shuldiner, A. R., O'Connell, J. R., Bliden, K. P., Gandhi, A., Ryan, K., Horenstein, R. B., Damcott, C. M., Pakyz, R., Tantry, U. S., Gibson, Q., Pollin, T. I., Post, W., Parsa, A., Mitchell, B. D., Faraday, N., Herzog, W., & Gurbel, P. A. (2009). Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA, 302*(8), 849–857.

Singer, E. (2011). A family learns the secrets of its genomes. *MIT Technology Review* (Sept).

Sirota, M., Dudley, J. T., Kim, J., Chiang, A. P., Morgan, A. A., Sweet-Cordero, A., Sage, J., & Butte, A. J. (2011). Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Science Translational Medicine, 3*(96), 96ra77.

Spackman, K. A., Campbell, K. E., & Cote, R. A. (1997). SNOMED RT: a reference terminology for health care. In *Proceedings of the AMIA annual fall symposium* (pp. 640–644).

Stein, L. D. (2010). The case for cloud computing in genome informatics. *Genome Biology, 11*(5), 207.

Stenson, P. D., Mort, M., Ball, E. V., Howells, K., Phillips, A. D., Thomas, N. S., & Cooper, D. N. (2009). The human gene mutation database: 2008 update. *Genome Medicine, 1*(1), 13.

Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., Hussain, M., Phillips, A. D., & Cooper, D. N. (2017). The Human Gene Mutation Database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human Genetics, 136*(6), 665–677.

Steyerberg, E. W., Pencina, M. J., Lingsma, H. F., Kattan, M. W., Vickers, A. J., & Van Calster, B. (2011). Assessing the incremental value of diagnostic and prognostic markers: A review and illustration. *European Journal of Clinical Investigation, 42*(2), 216–228.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America, 102*(43), 15545–15550.

Sung, N. S., Crowley, W. F., Jr., Genel, M., Salber, P., Sandy, L., Sherwood, L. M., Johnson, S. B., Catanese, V., Tilson, H., Getz, K., Larson, E. L., Scheinberg, D., Reece, E. A., Slavkin, H., Dobs, A., Grebb, J., Martinez, R. A., Korn, A., & Rimoin, D. (2003). Central challenges facing the national clinical research enterprise. *JAMA, 289*(10), 1278–1287.

Sweeney, T. E., Azad, T. D., Donato, M., Haynes, W. A., Perumal, T. M., Henao, R., Bermejo-Martin, J. F., Almansa, R., Tamayo, E., Howrylak, J. A., Choi, A., Parnell, G. P., Tang, B., Nichols, M., Woods, C. W., Ginsburg, G. S., Kingsmore, S. F., Omberg, L., Mangravite, L. M., Wong, H. R., Tsalik, E. L., Langley, R. J., & Khatri, P. (2018a). Unsupervised analysis of transcriptomics in bacterial sepsis across multiple datasets reveals three robust clusters. *Critical Care Medicine, 46*(6), 915–925.

Sweeney, T. E., Thomas, N. J., Howrylak, J. A., Wong, H. R., Rogers, A. J., & Khatri, P. (2018b). Multicohort analysis of whole-blood gene expres-

sion data does not form a robust diagnostic for acute respiratory distress syndrome. *Critical Care Medicine, 46*(2), 244–251.

Tarczy-Hornoch, P. (2007). Bio*medical informatics and genomic medicine: Research and training. *Journal of Biomedical Informatics, 40*(1), 1–4.

Tatonetti, N. P., Denny, J. C., Murphy, S. N., Fernald, G. H., Krishnan, G., Castro, V., Yue, P., Tsau, P. S., Kohane, I., Roden, D. M., & Altman, R. B. (2011a). Detecting drug interactions from adverse-event reports: Interaction between paroxetine and pravastatin increases blood glucose levels. *Clinical Pharmacology and Therapeutics, 90*(1), 133–142.

Tatonetti, N. P., Fernald, G. H., & Altman, R. B. (2011b). A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. *Journal of the American Medical Informatics Association, 19*(1), 79–85.

Taylor, C. F., Field, D., Sansone, S. A., Aerts, J., Apweiler, R., Ashburner, M., Ball, C. A., Binz, P. A., Bogue, M., Booth, T., Brazma, A., Brinkman, R. R., Michael Clark, A., Deutsch, E. W., Fiehn, O., Fostel, J., Ghazal, P., Gibson, F., Gray, T., Grimes, G., Hancock, J. M., Hardy, N. W., Hermjakob, H., Julian, R. K., Jr., Kane, M., Kettner, C., Kinsinger, C., Kolker, E., Kuiper, M., Le Novere, N., Leebens-Mack, J., Lewis, S. E., Lord, P., Mallon, A. M., Marthandan, N., Masuya, H., McNally, R., Mehrle, A., Morrison, N., Orchard, S., Quackenbush, J., Reecy, J. M., Robertson, D. G., Rocca-Serra, P., Rodriguez, H., Rosenfelder, H., Santoyo-Lopez, J., Scheuermann, R. H., Schober, D., Smith, B., Snape, J., Stoeckert, C. J., Jr., Tipton, K., Sterk, P., Untergasser, A., Vandesompele, J., & Wiemann, S. (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. *Nature Biotechnology, 26*(8), 889–896.

Tenenbaum, J. D., & Goodman, K. W. (2017). Beyond the Genetic Information Nondiscrimination Act: Ethical and economic implications of the exclusion of disability, long-term care and life insurance. *Personalized Medicine, 14*(2), 153–157.

Tenenbaum, J., James, A., & Paulyson-Nuñez, K. (2012). An altered treatment plan based on direct to consumer (DTC) genetic testing: Personalized medicine from the patient/pin-cushion perspective. *Journal of Personalized Medicine, 2*(4), 192–200.

Tenenbaum, J. D., Avillach, P., Benham-Hutchins, M., Breitenstein, M. K., Crowgey, E. L., Hoffman, M. A., Jiang, X., Madhavan, S., Mattison, J. E., Nagarajan, R., Ray, B., Shin, D., Visweswaran, S., Zhao, Z., & Freimuth, R. R. (2016). An informatics research agenda to support precision medicine: Seven key areas. *Journal of the American Medical Informatics Association, 23*(4), 791–795.

Terry, S. F. (2014). The global alliance for genomics & health. *Genetic Testing and Molecular Biomarkers, 18*(6), 375–376.

Thorisson, G. A., Smith, A. V., Krishnan, L., & Stein, L. D. (2005). The international HapMap project web site. *Genome Research, 15*(11), 1592–1593.

Tomczak, A., Mortensen, J. M., Winnenburg, R., Liu, C., Alessi, D. T., Swamy, V., Vallania, F., Lofgren, S., Haynes, W., Shah, N. H., Musen, M. A., & Khatri, P. (2018). Interpretation of biological experiments changes with evolution of the gene ontology and its annotations. *Scientific Reports, 8*(1), 5115.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Miklos, G. L. G., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K.,

Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., & Zhu, X. (2001). The sequence of the human genome. *Science, 291*(5507), 1304–1351.

Via, M., Gignoux, C., & Burchard, E. G. (2010). The 1000 Genomes Project: New opportunities for research and social challenges. *Genome Medicine, 2*(1), 3.

Waldman, S. A., & Terzic, A. (2010). Clinical and translational science: From bench-bedside to global village. *Clinical and Translational Science, 3*(5), 254–257.

Wan, Z., Vorobeychik, Y., Xia, W., Clayton, E. W., Kantarcioglu, M., & Malin, B. (2017). Expanding access to large-scale genomic data while promoting privacy: A game theoretic approach. *American Journal of Human Genetics, 100*(2), 316–322.

Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., & Bryant, S. H. (2009). PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research, 37*(Web Server issue), W623–W633.

Wang, Y., Bolton, E., Dracheva, S., Karapetyan, K., Shoemaker, B. A., Suzek, T. O., Wang, J., Xiao, J., Zhang, J., & Bryant, S. H. (2010). An overview of the PubChem BioAssay resource. *Nucleic Acids Research, 38*(Database issue), D255–D266.

Weston, A. D., & Hood, L. (2004). Systems biology, proteomics, and the future of health care: Toward predictive, preventative, and personalized medicine. *Journal of Proteome Research, 3*(2), 179–196.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, 160018.

Wishart, D. S. (2011). Advances in metabolite identification. *Bioanalysis, 3*(15), 1769–1782.

WTCCC. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature, 447*(7145), 661–678.

Yildirim, M. A., Goh, K. I., Cusick, M. E., Barabasi, A. L., & Vidal, M. (2007). Drug-target network. *Nature Biotechnology, 25*(10), 1119–1126.

Zerhouni, E. A. (2006). Clinical research at a crossroads: The NIH roadmap. *Journal of Investigative Medicine, 54*(4), 171–173.

Zheng, P., Zeng, B., Zhou, C., Liu, M., Fang, Z., Xu, X., Zeng, L., Chen, J., Fan, S., & Du, X. (2016). Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. *Molecular Psychiatry, 21*(6), 786.

Zhu, J., Zhang, B., & Schadt, E. E. (2008). A systems biology approach to drug discovery. *Advances in Genetics, 60*, 603–635.

Zhu, Q., Ge, D., Maia, J. M., Zhu, M., Petrovski, S., Dickson, S. P., Heinzen, E. L., Shianna, K. V., & Goldstein, D. B. (2011). A genome-wide comparison of the functional properties of rare and common genetic variants in humans. *American Journal of Human Genetics, 88*(4), 458–468.

Zitvogel, L., Galluzzi, L., Viaud, S., Vetizou, M., Daillere, R., Merad, M., & Kroemer, G. (2015). Cancer and the gut microbiota: An unexpected link. *Science Translational Medicine, 7*(271), 271ps271.

Clinical Research Informatics

Philip R. O. Payne, Peter J. Embi, and James J. Cimino

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_27

27

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What is clinical research and what factors influence the design of clinical studies?
- What are the types of information needs inherent to clinical research and how can those information needs be stratified by research project phase or activity?
- What types of information systems can be used to address or satisfy the information needs of clinical research teams?
- How can multi-purpose platforms, such as electronic health record (EHR) systems (see ► Chap. 14), be leveraged to enable clinical research?
- What is the role of a clinical trial or research management system (CTMS/CRMS) for supporting and enabling clinical research, and what types of functionality are common to such systems?
- What is the role of standards in supporting interoperability across and between actors and entities involved in clinical research?
- What are current and future clinical research informatics (CRI) research and development questions and how will they optimize or otherwise alter the conduct of clinical research?

27.1 Introduction

The conduct of clinical research is fundamental to the generation of evidence that can in turn facilitate improvements in human health. However, the design, execution, and analysis of clinical research is an inherently complex information- and resource-intensive endeavor, involving a broad variety of stakeholders, workflows, processes, data types, and computational resources. At the intersection point between biomedical informatics and clinical research, a robust and growing sub-discipline of informatics has emerged, which for the remainder of this chapter we will refer to as **Clinical Research Informatics (CRI)** (P. J. Embi and Payne 2009). Numerous reports have shown that innovations and best practices generated by the CRI community have contributed to improvements in the quality, efficiency, and expediency of clinical research (P. Embi 2013; P. J. Embi and Payne 2009, 2013; Johnson et al. 2016; Payne et al. 2005; Weng and Kahn 2016). Such benefits can be situated in a full spectrum of contexts that extends from the activities of individual clinical investigators to the operations of multi-center research consortia that involve geographically and temporally distributed participants.

Given the recognition of CRI as a distinct and increasingly important sub-discipline of biomedical informatics, it is imperative that a common basis for defining and understanding CRI science and practice be established. Such a foundation must, by necessity, include explicit linkages to the major challenges and opportunities associated with the planning, conduct, and evaluation of clinical research programs. To provide a common frame of reference for the remainder of this chapter, we will use the National Cancer Institute's (NCI) definition of clinical research, as follows:

» *Research in which people, or data or samples of tissue from people, are studied to understand health and disease. Clinical research helps find new and better ways to detect, diagnose, treat, and prevent disease. Types of clinical research include clinical trials, which test new treatments for a disease, and natural history studies, which collect health information to understand how a disease develops and progresses over time.*[1]

A lack of sufficient information technology (IT) and biomedical informatics tools and platforms, as well as relevant expertise and methodological frameworks, account for significant impediments to the rapid, effective, and resource-efficient conduct of clinical research projects (Payne et al. 2010; Payne

1 ► https://www.cancer.gov/publications/dictionaries/cancer-terms/def/clinical-research (Accessed January 1, 2019).

et al. 2005; Payne et al. 2013). Compounding these challenges is the rapid pace of advancement in biomedical research and the resulting need for advances in diagnostics and therapeutics that can be validated and disseminated quickly and cost effectively (Brightling 2017; Saad et al. 2017; Tenenbaum et al. 2016; Weng and Kahn 2016). The confluence of these factors has led to a number of major challenges and opportunities related to current and future CRI research and practice. For example, the importance of making clinical phenotype data available for the secondary use in support of clinical research has become a competitive requirement for research enterprises of all sizes. Similarly, the increasing complexity of clinical research programs and the difficulty of recruiting sufficiently large patient cohorts, when combined with the regulatory overhead of conducting studies in large academic institutions, has led to an increase in the conduct of clinical studies in community practice settings. Such community-based research paradigms introduce new levels of complexity to the technical and policy aspects of data capture, management, and sharing plans. This rapid evolution and the realities of an increasingly expansive clinical research landscape have led investigators and other decision makers in the health care and life sciences communities to call for increased investments in and delivery of innovative solutions to such information needs (P. J. Embi and Payne 2014; Pencina and Peterson 2016; R. Richesson et al. 2014; Saad et al. 2017; Tenenbaum et al. 2016). At the highest level, clinical research is a domain that has substantial information management needs, representing both a challenge and opportunity for biomedical informatics researchers and practitioners. Simultaneously, clinical research is an area of scientific endeavor that is at the forefront of attention for the governmental, academic, and private sectors, all of which have significant scientific and financial interests in the conduct and outcomes of such efforts. When viewed collectively, many have called, and continue to call, for the development and validation of innovative biomedical informatics methods and tools specifically designed to address clinical research information needs (P. J. Embi and Payne 2014; Pencina and Peterson 2016; R. Richesson et al. 2014; Saad et al. 2017; Tenenbaum et al. 2016). It is this overall context that has motivated an increasing focus on both basic and applied Clinical Research Informatics (CRI), which can be defined broadly as follows (P. J. Embi and Payne 2009):

Clinical Research Informatics* (*CRI*) *is the sub-domain of biomedical informatics concerned with the development*, *evaluation and application of informatics theory*, *methods and systems to improve the design and conduct of clinical research and to disseminate the knowledge gained.

Examples of focus areas in which CRI researchers and practitioners apply biomedical informatics theories and methods can include the following:

- Evaluating and modeling of clinical research workflow
- Social and behavioral studies involving clinical research professionals and participants
- Designing optimal human-computer interaction models for clinical research applications
- Improving information capture and data flow in clinical research
- Leveraging data collected in EHRs
- Optimizing site selection, investigator and patient recruitment
- Improving reporting to regulatory agencies
- Enhancing clinical and research data mining, integration, and analysis
- Phenomic characterization of patients for cohort discovery and analytical purposes
- Integrating research findings into individual and population level health care
- Defining and promoting ethical standards in CRI practice
- Educating researchers, informaticians, and organizational leaders about CRI
- Driving public policy around clinical and translational research informatics

Building upon the preceding definitions and state of knowledge and practice relevant to CRI, in the remainder of this chapter we will provide an overview of the types of activities commonly undertaken as part of a variety

of representative clinical research use cases, introduce the role of major classes and types of information system that enable or facilitate such activities, and conclude with a set analyses regarding the future directions of the field. The overall objective of this chapter is to provide the reader with the ability to evaluate critically the current and anticipated roles of biomedical informatics knowledge and practice as applied to clinical research.

27.2 A Primer on Clinical Research

In the following section, we will briefly introduce the characteristics of the modern clinical research environment, including the design and execution of an exemplary class of clinical studies that are known as randomized controlled trials (RCTs). This primer on clinical research will serve as the context for the remainder of the chapter, in which we will introduce major information needs and their relationships to a variety of basic and applied biomedical informatics practice areas and IT tools/platforms.

27.2.1 The Modern Clinical Research Environment

Clinical research comes in many forms and may include a variety of specific activities. All forms, however, share a common set of requirements related to the comprehensive management of study data – specifically, the collection of data on human research subjects – and analysis of those data. As clinical research designs span the spectrum from passive or observational studies to interventional trials, the acuity of activities and associated data-management needs increase commensurately. For example, in a retrospective study subjects are selected based on the presence or absence of a particular condition and retrospective or pre-existing data are obtained from historical records (such as EHRs, disease registries, and research-specific databases), whereas in natural history studies, subjects are recruited and followed in prospective manner, with additional collection of data performed solely for the purposes of research, rather than the normal process of patient care.

Further along the spectrum are clinical trials, in which research subjects participate in some additional activity, or *intervention*, that is intended either to induce a change in the subject or to prevent the occurrence of some change that would otherwise be expected. The intervention might be as simple as administering a substance already found in the human body (such as a vitamin) to measuring a change in that substance (such as the amount of the vitamin found in the blood or urine). More complex studies involve interventions that have an impact on human disease, such as the administration of a preventive vaccine, the administration of a curative drug, or a surgical procedure to remove, insert, repair or replace a structure or device in the subject's body. As with passive studies, data collection is critical to the proper performance of research and may become intense, with the collection of clinical information occurring more frequently and involving data describing the intervention materials (such as the purity of a drug or the performance of a device) in addition to data related to the human subject and their response to the intervention under study.

Although not an intrinsic requirement of clinical research, the inclusion of comparison groups is usually considered an important part of rigorous and reproducible clinical research method. In some cases, **historical controls** can be used for comparison with a group of subjects under study. For example, if a disease is known to have a particular fatality rate, subjects could be given a potentially life-saving treatment, and their fatality rate can be measured and compared to past experience. In **quasi-experiments**, comparison subject groups can also be selected based on some known characteristic that distinguishes the two groups, such as gender or race, or their willingness to undergo a particular intervention.

A more rigorous method of establishing comparison groups is through randomization, in which prospective subjects are assigned to

different groups (often referred to as **study arms**) and undergo different interventions as a result of the arm to which they are assigned. Typically, randomization might account for observable characteristics (such as gender, ethnicity, and race) to create balanced groups, especially where the characteristics are known to have some influence on the effect of the intended intervention. Randomization also serves to distribute subjects based on unobserved characteristics, for example, unknown genetic traits, in order to reduce differences in the groups that might bias the results of the study. In a randomized controlled trial (RCT), one subject group will often receive a **control intervention** (for example, the usual treatment or treatments for a condition, or even no treatment) while one or more other groups receive an **experimental intervention**.

Although intended to reduce bias, the randomization process itself must be carefully executed such that it does not introduce new sources of bias. For example, randomization can include **blinding**, in which the subject, the investigator, or both (as in **double-blinded** studies), are kept unaware of group assignment until after all assessments have been made. This might include the use of a **placebo** for a group receiving no treatment, in order to avoid the possibility that subjective improvement in a prior condition or the occurrence of random events (such as normally occurring illnesses), or are not ascribed to the intervention. This also may prevent subjects from deciding not to participate after randomization in a way that might unbalance the study groups (for example, if subjects prefer not to participate if they know they are not getting the experimental intervention) or even bias the assignments (for example, people less prone to take care of themselves might drop out if they find they are assigned to an intervention that requires a great deal of effort on their part).

The gold standard of clinical studies is the double-blinded, randomized, placebo-controlled trial (Hulley et al. 2013). However, such studies may not always be practical. For example, the use of a placebo when an effective therapy is known may be unethical, the blinding of a surgical repair may not be practical, or the condition under study may be so rare that only historical controls are available.

While different study designs have unique and differentiated data, information, and knowledge management needs, they usually involve some form of systematic data management, as noted previously. Such data management activities usually include initial data collection, aggregation, analysis, and results dissemination, to name a few of many such tasks. As shown in Fig. 27.1, different study methods introduce new issues as successively more complex interventions and study design patterns are employed. For the remainder of this chapter, we will focus our discussion on RCTs as our prototypical study design, as they tend to involve most if not all of the informatics issues and information needs encountered in other study designs. Further information on the design characteristics, data management needs, and associated best practices related to various types of clinical trials can be found in

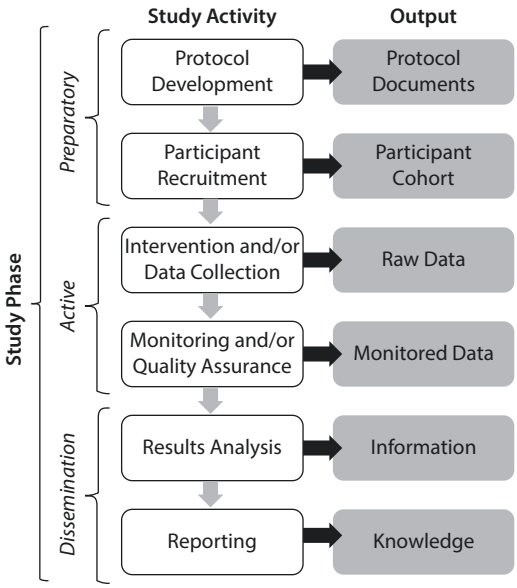


Fig. 27.1 Overview of clinical study phases and associated information and data management needs. Underlying such design patterns are a common thread of systematic data management, leveraging resources such as health records, research-specific laboratory data, as well as broader knowledge collections such as the published biomedical literature

a number of excellent works concerning this subject (Bhatt and Mehta 2016; Hulley et al. 2013; Prokscha 2011), and further discussion is beyond the scope of this chapter.

27.2.1.1 Phased Randomized Controlled Trials

Most clinical studies begin with the identification of a set of driving or motivating hypotheses. The research questions that serve to define such hypotheses might be raised through an analysis of gaps in knowledge as found in the published biomedical literature or be informed by the results of a previous study. It is important to note that clinical research endeavors exist on a spectrum of scientific activity that is often referred to as **clinical and translational research**. A particular type of translational research, often referred to as T1-type translation, is a process by which basic science discoveries are used to design novel therapies (Sung et al. 2003). Such discoveries are then evaluated during clinical research studies, first pre-clinical and subsequent clinical trial phases (Payne et al. 2005). A second type of translational research, often referred to as T2 translation, involves methods such as those borrowed from **implementation science** and clinical informatics, and focus on translating the findings of such clinical research studies into common practice (Sung et al. 2003). A common colloquialism for this process of translating a novel basic science discovery through clinical research and into clinical practice is "bench to bedside" science.

Individual and distinct RCTs are often conducted for different purposes, most often motivated by the need to fill fundamental knowledge gaps about a particular intervention under study. By combining such knowledge gaps with the underlying biomedical mechanisms of physiology and disease, a motivating hypothesis or collections of hypotheses are established as to why a given intervention might lead to a given result or finding. Such hypotheses result in a natural sequence of research questions that can be asked relative to a novel intervention. Usually, an individual research study is designed to address one specific research question and hypothesis. In the case of the development and evaluation of a new therapeutic intervention, like a new drug, an individual research study is designed to address each **phase** in a line of research inquiry that will determine the efficacy and effectiveness of such a therapy (Spilker 1984). In most cases, this adheres to the following model:

- **Phase I**: Investigators evaluate the novel therapy in a small group of participants in order to assess overall safety. This safety assessment includes dosing levels in the case of non-interventional therapeutic trials, and potential side effects or adverse effects of the therapy. Often, Phase I trials of non-interventional therapies involve the use of normal volunteers who do not have the disease state targeted by the novel therapy.
- **Phase II**: Investigators evaluate the novel therapy in a larger group of participants in order to assess the efficacy of the treatment in the targeted disease state. During this phase, assessment of overall safety is continued.
- **Phase III**: Investigators evaluate the novel therapy in an even larger group of participants and compare its performance to a reference standard which is usually the current standard of care for the targeted disease state. This phase typically employs an RCT design, and often a multi-center RCT given the numbers of variation of subjects that must be recruited to test the hypothesis. In general, this is the final study phase to be performed before seeking regulatory approval for the novel therapy and broader use in standard-of-care environments.
- **Phase IV**: Investigators study the performance and safety of the novel therapy after it has been approved and marketed. This type of study is performed in order to detect long-term outcomes and effects of the therapy. It is often called "post-market surveillance" and is, in fact, not an RCT at all, but a less formal, observational study.

The phase of an RCT has implications for the kinds of questions being asked and the kinds of processes carried out to answer them. From an informatics perspective, however, the

tasks are usually very similar. At a high level, the conduct of a Phase I, II or III clinical trial can be thought of in an operational sense as consisting of three major stages: preparatory, active, and dissemination. During these three stages, a specific temporal series of processes is executed. First, during the **preparatory phase**, a protocol document is generated as part of the project development process. The protocol document usually contains background information, scientific goals, aims, hypotheses and research questions to be addressed by the trial. In addition, the protocol describes policies, procedures, and data collection or analysis requirements. A critical aspect of the protocol document is the definition of a protocol schema, which defines at a highly granular level the temporal sequence of tasks and events required to both deliver the intervention under study and to ensure that data are collected and managed in a systematic manner commensurate with the study hypotheses and aims.

Once a protocol is deemed ready for execution, the feasibility of the study design (e.g., addressing questions such as "are there enough participants available in the targeted population to satisfy the study design defined in the protocol document?" is assessed either quantitatively (e.g., using historical data) and/or heuristically). Throughout the preparatory phase, a concurrent process of seeking regulatory approval from local and national bodies (e.g., local Institutional Review Boards, the Food and Drug Administration) occurs. Once a protocol plan is complete, deemed feasible, and regulatory approval has been received, potential participants are recruited and screened to determine if they meet the inclusion and exclusion criteria for the study (e.g., specific demographic and/or clinical parameters required for subjects to be eligible for the study). Once a potential participant has been deemed eligible for the study, they are provided with an informed consent document, which must be signed prior to proceeding with the enrollment process. **Enrollment** in the context of clinical trials means officially registering as a study subject, and the subsequent assignment of a study-specific identifier. Once a person agrees to become a participant, they are enrolled, and in the case of studies with multiple study groups or arms, randomized into one of those arms.

The preceding activities lead to the initiation of the next step in the research process, which we refer to as the **active phase**. During the active phase, the participant receives the therapeutic intervention indicated by their study arm and is actively monitored to enable the collection of study-specific data. This therapeutic intervention and active monitoring process is often iterative, involving multiple cycles of interventions and active monitoring. Follow-up activities begin once a participant has completed the interventional stage of a study. During this stage, subjects are contacted on a specified temporal basis in order to collect additional data of interest, such as long-term treatment effects, disease status or survival status.

Finally, during the **dissemination phase**, the results of the study are evaluated and formalized in publications or other knowledge dissemination media, for translation into the next phase of an RCT or into clinical practice. In some cases, such as is adaptive study designs (Bhatt and Mehta 2016), this dissemination phase feeds back into the planning and active phases. Such feedback cycles enable rapid revision of a study design, iterative participant enrollment, and dynamic data collection in support of such revised hypotheses and designs. Of note, these types of adaptive trial designs are particularly helpful when conducting studies of conditions where large numbers of patients may not be present for recruitment purposes, so that patients are assigned to an intervention and/or monitored in a manner that maximizes data collection that is most likely to demonstrate the safety, efficacy, and comparative effectiveness of the diagnostic or therapeutic approach being evaluated (Bhatt and Mehta 2016).

The quality of data produced by a clinical trial is assessed using multi-dimensional metrics that account for the design, execution, analysis and dissemination of the study results. The quality of a clinical trial is also judged with respect to the significance or relevance of the reported study results within a clinical context (Hulley et al. 2013; Prokscha

2011; Spilker 1984). One key metric used to assess clinical trial quality is validity, which can be defined both internally and externally. **Internal validity** is defined as the minimization of potential biases during the design and execution of the trial, while **external validity** is the ability to generalize study results into clinical care. It is important to note in a discussion of the role of biomedical informatics relative to clinical research that such platforms, interventions, and methods can play a major role in reducing or mitigating such sources of bias, thus enhancing the validity and generalizability of study results.

27.2.2 Information Needs and Systems in the Clinical Research Environment

As can be inferred by the preceding introduction to the definitional aspects of clinical research, such activities regularly involve a variety of data, information, and knowledge sources, as well a complicated set of complementary and overlapping workflows. At the highest level, these characteristics of the clinical research environment can be related to a number of critical information needs, as summarized in Table 27.1. This representation of the information needs inherent to clinical research is presented using the specific context of a prototypical RCT, but the basic types of needs and example solutions provided can be extended to apply to the broader spectrum of research designs and patterns introduced earlier.

Building upon this broad definition of the information needs inherent to clinical research, in the following sub-sections we: (1) review the types of information systems that can support the phases that comprise a clinical study, (2) explore the functional components that make up a clinical trials management system, (3) identify current consortia that share clinical and research data, and (4) discuss the role of standards in enabling interoperability between such information systems.

27.2.2.1 Information Systems Supporting Clinical Research Programs

It is helpful to conceptualize the conduct of clinical research studies as a multiple-stage sequential model, as was introduced previously and is expanded upon in this section (Payne et al. 2005). At each stage in such a model, a combination of research-specific and general technologies can be employed to support or address related information needs (Fig. 27.2).

There are numerous examples of general-purpose and clinical systems that are able to support the conduct of clinical research:

- **Publication (or bibliographic) databases** and **information retrieval (IR) tools** such as PubMed, Google Scholar, and OVID can be used to assist in conducting the background research necessary for the preparation of protocol documents.
- **Electronic health records (EHRs)** can be used to collect clinical data on research participants in a structured form that can reduce redundant data entry.
- **Data warehouses and associated data or text mining tools** can be used in multiple capacities, including: (1) determining if participant cohorts who meet the study inclusion or exclusion criteria can be practically recruited given historical trends, and (2) identifying specific participants and related data within existing databases.
- **Clinical decision-support systems (CDSS)** can be used to alert providers at the point-of-care that an individual may be eligible for a clinical trial

In addition to the preceding general technologies, a number of research-specific technologies have been developed:

- **Feasibility analysis applications and data simulation and visualization tools** can streamline the pre-clinical research process (e.g., disease models) and assist in the analysis of complex data sets in order to assess the feasibility of a given study design.

Table 27.1 Overview of definitional information needs in the clinical research environment, contextualized using the design of RCTs

Information needs	Major sub-components	Description
Support for research planning and conduct	Collaborative document and knowledge management	*Study teams often involve geographically and temporally distributed participants, who need to engage in iterative protocol development and approval processes. Such activities by necessity incorporate document versioning, annotation, and associated metadata management tasks. Once a protocol has been developed, access to data sets for the purposes of assessing the feasibility of a given study design is critical, and often involves the use of de-identified data sets drawn from a data warehouse or research registry. Finally, the submission, tracking, and documentation of regulatory approvals often necessitate the coordination and management of complex, document-oriented workflows and record keeping tasks.*
	Data sources and tools for feasibility analyses	
	Regulatory approval workflows	
Facilitation of data management, access, and integration	Secondary-use of EHR-derived data for research purposes	*The ability to use primary clinical data from EHR or equivalent platforms to support secondary use in a research program has the potential to reduce redundancy and potential errors while increasing data quality. However, using such data in a secondary capacity also requires that appropriately structured data be captured and codified in clinical systems, and then be made available to research teams and research data management systems in a timely and resource efficient manner. In addition to such secondary use of clinical data, most clinical studies require the regular capture and management of study-specific data elements, a task that is usually accomplished via the use of Electronic Data Capture (EDC) or Clinical Trial Management Systems (CTMS). Finally, given the propensity to conduct studies that span traditional organizational boundaries in order to realize economies of scale and/or access sufficiently large patient populations, it is often necessary to query, integrate, and manage distributed data sets, and ensure their syntactic and semantic interoperability. Such a need is usually addressed through the use of Service Oriented Architectures, Cloud Computing, Data Warehousing, and Metadata Management technologies.*
	Research project specific data capture, management, and reporting	
	Distributed data management (spanning traditional organizational boundaries)	
	Syntactic and semantic interoperability	
Workforce training and support	Dissemination of study, methodological, and technical training materials	*A central need when conducting clinical studies is the ability to ensure that individuals involved in the execution of a protocol share common methods, data management practices, and workflows (thus reducing potential sources of study bias). Ensuring such shared knowledge and practices, particularly in distributed or multi-site settings, requires the use of distance education and team-science tools and platforms to enable knowledge sharing and distance learning paradigms.*
	Support for team collaboration and knowledge sharing	

(continued)

Table 27.1 (continued)

Information needs	Major sub-components	Description
Management information capture and reporting	Support for research billing Operational instrumentation and reporting Regulatory monitoring Data quality assurance	*The business and management aspects of the conduct of clinical studies is complex, often requiring the disambiguation of standard-of-care and research specific charges as part of billing operations, as well as the tracking of key performance and data quality metrics that may be required to satisfy contractual commitments to the entities funding such studies. Furthermore, the monitoring of study data for critical or sentinel events that should or must be reported for regulatory purposes is both necessary and of extreme importance. All of the aforementioned activities require the application of a variety of management information system, business intelligence, and reporting tools, leveraging a broad variety of enterprise, administrative, and study-specific data sources.*
Participant recruitment tools and methods	Cohort discovery Eligibility determination and alerting Participant registration, consent, and enrollment execution and tracking	*The identification of participant cohorts that satisfy key study design criteria, such as inclusion and exclusion criteria, is frequently a major barrier to the timely and efficient execution of clinical studies. A variety of information needs, related to the identification and engagement of such cohorts, to point-of-care alerting regarding potential study eligibility, to the management of registration, consent, and enrollment records is inherent to this information need. Such requirements are usually satisfied through a multi-modal approach, leveraging both clinical and research-specific information systems.*
Data standards	Standards for interoperability between research systems Standards for interoperability between research, enterprise (e.g. EHR), and administrative systems	*As has been noted relative to several of the preceding information needs, there is a frequent and reoccurring requirement for both syntactic and semantic interoperability between research-specific information systems, as well as between research-specific and clinical or administrative systems. Such a need necessitates the design, selection, and application of a variety of data standards, as well as the ability to map and harmonize between shared information models to support interactions between systems using a variety of standards.*
Workflow support	Integration of tools for combined standard-of-care and research visits Data, information, and knowledge transfer between stakeholders, project phases, activities, and associated information systems or tools.	*Much as was the case related to data standards, a closely aligned information need exists relative to the ability to support complex workflows between information systems and actors involved in the conduct of clinical research. Such workflow support requires both computational and application-level workflow orchestration, as well as the ability to define and apply reusable data analytic "pipelines."*

Table 27.1 (continued)

Information needs	Major sub-components	Description
Data, information and knowledge dissemination	Knowledge management for clinical evidence generated during trials	*The ultimate objective of clinical research is to generate and apply new evidence in support of improvements in clinical care and human health. In order to do so, it is necessary to disseminate the findings generated during such studies in a variety of formats, including reusable/actionable knowledge resources, clinical guidelines, decision support rules, and/or publications and reports. In addition, increasing emphasis is being placed on the transparency and reproducibility of study designs, which is often accomplished through the creation of public registries via which study data sets can be shared and made available to the broader biomedical community.*
	Guidelines and CDSS delivery mechanisms	
	Publication mechanisms	
	Data registries	

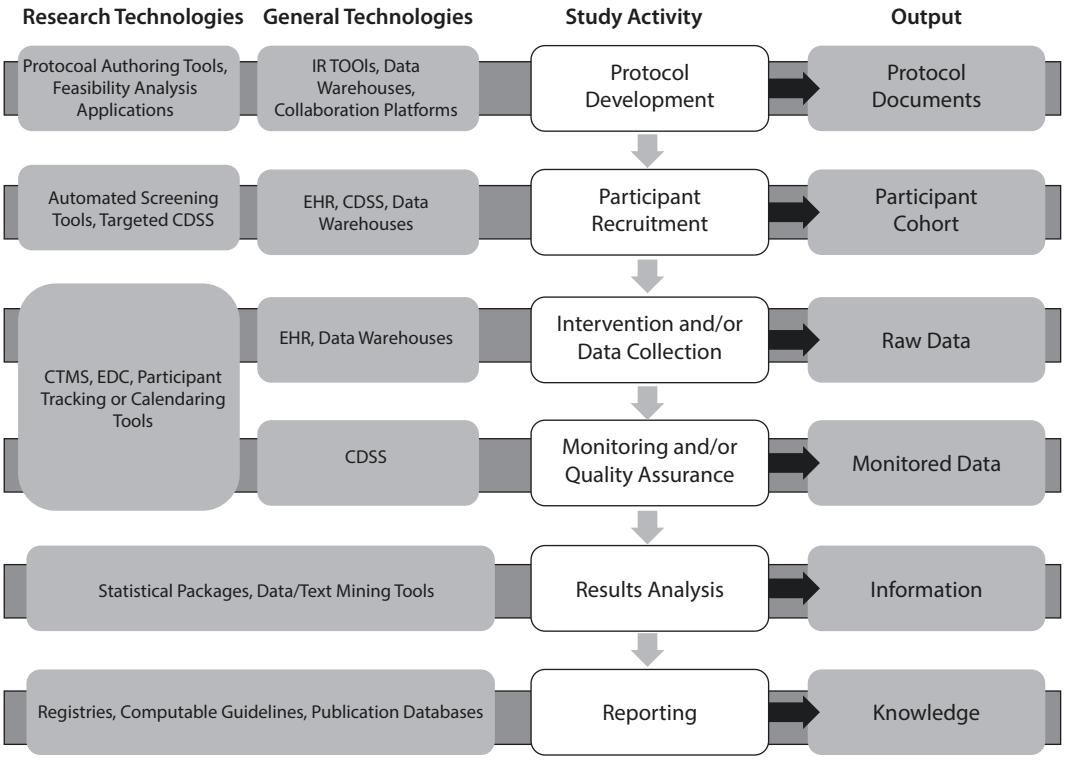


Fig. 27.2 Overview of study activities, and related research-specific and general information technologies, as well as targeted products or outputs associated with the sequential clinical research workflow paradigm

- **Protocol authoring tools** can allow geographically distributed authors to collaborate on complex protocol documents.
- **Automated screening tools and targeted alerts** can assist in the identification and registration of research participants.
- **Electronic data capture (EDC) and Clinical Trial and/or Research Management Systems (CTMS/CRMS)** can be used to collect research-specific data in a structured form and reduce the need for redundant and potentially error-prone paper-based

data collection techniques. More detail on these types of systems is provided in the following section of this chapter.

- **Research-specific decision support systems such as participant tracking or calendaring tools** provide protocol-specific guidelines and alerts to researchers, for example tracking the status of participants to ensure protocol compliance.

27.2.2.2 Clinical Research Management Systems

One of the most widely used technology platforms in the clinical research domain is the **clinical trial or research management system** (**CTMS/CRMS**). Such platforms were historically referred to as clinical *trials* management systems (CTMS), but the term CRMS is gaining popularity as such systems are increasingly used to manage the conduct of studies including but not limited to trials. CRMS platforms are usually architected as composite systems that incorporate a number of task and role-specific modules intended to address core research-related information needs (P. J. Embi and Payne 2009; Johnson et al. 2016; Payne et al. 2005). Exemplary instances of such modules include the following:

- **Protocol Management** components that support document management functionality to enable the submission, version control, and dissemination of protocol related artifacts and associated metadata annotations.
- **Participant Screening and Registration** tools that allow for the application of electronic eligibility "check lists" to individual patients or cohorts in order to assess study eligibility, and when appropriate, record the registration and associated "baseline" data that are required per the study protocol.
- **Participant Calendaring** functionality allows for the instantiation of general protocol schemas (e.g., a definition of a protocol's temporal series of tasks, events, and associated data collection tasks) in a participant specific manner, accounting for complex reasoning tasks including the dynamic recalculation of temporal intervals between events based on actual completion dates/times, as well as the "windowing" of events in which a given task or event is allowed to fall within a range of dates rather a specific, atomic temporal specification.
- **Electronic Data Capture** (**EDC**) components allow for the definition, instantiation, and use of electronic case report forms (e.g., forms that define study and task/event specific data elements to be collected in support of a given trial or research program). Such **electronic case report forms** (**eCRFs**) are the basic instrument by which the majority of study-specific data are collected and are usually populated via a combination of: (1) manual data entry (including abstraction from source documentation such as medical records); (2) the importation of secondary use data from clinical systems; or (3) a hybrid of the two preceding approaches.
- **Monitoring tools** enable the application of logical rules and conditions (e.g., range-checking, enforcement of data completion, etc.) using a rules engine or equivalent technology, in order to ensure the completeness and quality of research related data. Such tools may also be used to monitor patient compliance with study schemas, as reflected in the previously described patient calendar functionality.
- **Query and Reporting Tools** support the planned and ad-hoc extraction and aggregation of data sets from multiple eCRFs or equivalent data capture instruments as used with the CTMS. These types of tools are often used by biostatisticians and other quantitative scientists to perform interim and final analyses of study results, outcomes, and to enable higher-order safety analyses. In addition, such tools may be employed to comply with a broad variety of data submission and reporting standard set by both public- and private-sector entities.
- **Security and Auditing** functionality enables site, role, and study-specific access controls and end-user authentication/authorization relative to all of the preceding functionality, as well as the ability to track and report upon end-user interaction with and modifications to data con-

tained in the CRMS. Such functionality is critical to enabling compliance with a broad variety of regulatory and privacy/confidentiality frameworks that apply to the use of protected health information (PHI) for research purposes.

In most CRMS platforms, the aforementioned functional modules share one or more common research databases or in the case of service-oriented architectures (SOA), common data services. In more advanced platforms, these common data structures are populated with research-specific and/or clinical data from enterprise systems and sources (such as EHRs, personal health records, and data warehousing platforms) via either **API-level integration** (e.g., data service publication and consumption) or an **extract**, **transform**, **and load** (**ETL**) **based approach** (Raths 2013).

27.3 Data Sharing Resources and Networks for Clinical Research

In the following section, we provide an overview of the various data sharing resources and networks that are commonly encountered in the clinical research domain. These environments are used both for the design and execution of observational or pragmatic studies, as well as the conduct of retrospective analyses or de-novo re-analysis of data collected via clinical trials. Further, they also serve as a basis for disseminating the results and ensuring the reproducibility and rigor of a broad spectrum of clinical studies.

27.3.1 Publicly Deposited Clinical Research Metadata and Data Resources

For those seeking to share data, and to avail themselves of data shared by others, the National Center for Biotechnology Information (NCBI) at the NIH's National Library of Medicine has created a public repository of individual-level data, including exposure history, signs, symptoms, diagnostic test results, and genetic data. Called the **Database of Genome and Phenome (dbGAP)**, this project provides stable data sets that allow multiple researchers to reference the same samples in their publications of secondary analyses of the data (Mailman et al. 2007). Additional data from clinical trials, currently limited to summary results, are also being made available by the NLM through the ► ClinicalTrials.gov resource, which is a repository of descriptive metadata related to historical and actively recruiting clinical trials (Tse et al. 2009).

27.3.2 Clinical and Translational Science Award (CTSA) Network

The National Center for Advancing Clinical and Translational Science (NCATS) has – and continues to – fund a national-consortium of academic health centers (AHC) that are engaged in clinical and translational research under the auspices of the Clinical and Translational Science Award (CTSA) program (Zhang and Patel 2006). Each member of this network, known as a "hub", is responsible for creating a professional home that supports and enables the conduct of clinical and translational research. Such support includes the provision of Biomedical Informatics infrastructure and expertise is needed to facilitate the capture, storage, management, and analysis of data resulting from such research efforts. As such, the CTSA Network provides an important basis for the conduct of large-scale programs that involve the sharing of such data across and between "hubs." To coordinate and harmonize data sharing across this CTSA Network, NCATS has created a number of centers and sub-networks, including the Accrual to Clinical Trials Network (ACT, described in more detail below),[2] the Trial

2 ► http://www.actnetwork.us (Accessed January 27, 2019).

Innovation Network (TIN),[3] the Recruitment Innovation Center (RIC),[4] and the Center for Data 2 Health (CD2H)[5] which is charged with the coordination of network-wide informatics activities. These centers and sub-networks within the broader CTSA network engage in a range of activities, such as: (1) leveraging local data repositories and federated data query tools to enable the rapid assessment of feasibility when designing multi-site clinical trials (ACT); (2) providing expertise and shared best practices as they relate to study designs and regulatory frameworks for such multi-site clinical trials (TIN); (3) delivering novel tools and methods to accelerate the recruitment of participants into those trials (RIC); and (4) helping the Biomedical Informatics components at CTSA hubs to collaborate, share knowledge and tools, and harmonize data assets (CD2H).

27.3.3 i2b2 and SHRINE

One popular warehouse software platform that is frequently used in the context of clinical research activities is **Informatics for Integrating Biology and the Bedside (i2b2),** developed under a National Center for Biomedical Computing grant from NIH to Partners HealthCare System and Harvard University. i2b2 provides an information system framework to allow clinical researchers to use existing clinical data for discovery research (Murphy et al. 2010). Many of the over 60 institutions receiving CTSA grants have adopted i2b2 technologies to support research and collaboration.

A companion to i2b2 is the **Shared Health Research Information Network (SHRINE)**, which is a version of i2b2 that can pass data queries entered by a local user off to other i2b2 instances to provide patient counts and demographic information that is summarized for all remote sites (Weber et al. 2009). SHRINE networks have been established by many research institutions to support, among other functions, the estimation of available cohort sizes for multi-institutional clinical trials.

27.3.4 Accrual to Clinical Trials (ACT) Network

Recently, NCATS has funded the **Accrual to Clinical Trials (ACT)** network to bring together CTSA sites into a single SHINE network. Its intent is to allow clinical researchers to query the network in real time and to obtain aggregate counts of patients who meet clinical trial inclusion and exclusion criteria from sites across the United States (Visweswaran et al. 2018). Currently, 31 CTSA sites are fully operational in the ACT network, with an additional 15 being "staged" for integration.[6] Although the ACT SHRINE network currently returns only summary statistics for each site, future plans include the ability to obtain detailed patient data sets and, working with researchers are member sites, being able to contact individual patients for potential recruitment into clinical studies.

27.3.5 PCORNet

Another clinical research data sharing consortium is **PCORNet**, which is funded by the **Patient-Centered Outcomes Research Institute (PCORI)**, a non-profit corporation established by the **Patient Protection and Affordable Care Act** (see ▶ Chap. 12) to support a national clinical research agenda (Fleurence et al. 2014). PCORNet is made of clinical data research networks (CDRNs) that each have access to EHR data on a million or more patients, patient-powered research networks (PPRNs) that are led by patients and patient advocates with an interest in a particular disease (common or rare), and Health Plan

3 ▶ https://trialinnovationnetwork.org/ (Accessed January 27, 2019).

4 ▶ https://trialinnovationnetwork.org/recruitment-innovation-center/ (Accessed January 27, 2019).

5 ▶ https://ctsa.ncats.nih.gov/cd2h/ (Accessed January 27, 2019).

6 ▶ http://www.actnetwork.us/national/get-to-know-46EU-1128WI.html (Accessed January 1, 2019).

Research Networks (HPRNs) that link EHR data with insurance claims data (Fleurence et al. 2014). There are currently 13 CDRNs, 20 PPRNs, and 2 HPRNS nation-wide.[7] The intent of these networks is to provide data that can be used directly to answer clinical research questions based on the health data of large patient cohorts. Clinical researchers can also use the networks to help identify and contact patients who might be suitable for clinical studies.

27.3.6 Observational Health Data Sciences and Informatics (OHDSI)

The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is an international network of researchers and health databases (primarily EHRs) that work together to develop and share data query and analytic tools that can operate on members' databases. OHDSI grew out of the Observational Medical Outcomes Partnership (OMOP), which was a public-private partnership established in the US study how patient care databases could be used to study the effects (both beneficial and adverse) of medical products (Hripcsak et al. 2015). OHDSI currently involves 106 collaborators from 23 countries on six continents.[8]

Current tools include a browser-based data visualization tool called ACHILLES (Automated Characterization of Health Information at Large-scale Longitudinal Exploration System), a vocabulary browsing tool called HERMES (Health Entity Relationship and Metadata Exploration System), a predictive analytics tool called PLATO (Patient-Level Assessment of Treatment Outcomes), a cohort development tool called HERACLES (Health Enterprise Resource and Care Learning Exploration System) that includes analytic tools performing clinical quality metrics, and HOMER (Health Outcomes and Medical Effectiveness Research), a tool for risk identification and comparative effectiveness studies.

The OHDSI Research Network allows researchers at member sites to query data repositories to obtain high-quality observational data that can be used for study design, execution, and data analysis. An OHDSI research defines a "project" which is actualized as a query for specific patient data. Data owners are invited to participate in projects by querying their own databases and conducting data analysis locally, with results sent back to the initial research team for compilation and analysis.

27.3.7 Commercial and Health Care Information Technology Vendor Networks

In addition to the public and non-profit networks, commercial entities have begun to appear that bring private investment resources to bear on some of the challenges of access and integration of data. For example, the TriNetX network (Topaloglu and Palchuk 2018) includes over 20 academic and private health systems that agree to share data over the network in order to enable data queries that can be used to identify cohorts for studies, either led by institutional investigators or biopharmaceutical companies.

Another entity is Flatiron Health, a company recently acquired by Roche (Petrone 2018), which focuses on data from cancer patients. Rather than creating distributed queries against disparate data sets, Flatiron obtains data from major academic medical centers (currently 7[9]), and processes the data centrally, enhancing its utility through natural language processing machine learning to obtain a better understanding of the course of patients' conditions and their response to therapy.

7 ► https://pcornet.org/participating-networks (Accessed January 1, 2019).

8 ► https://www.ohdsi.org/who-we-are/collaborators (Accessed January 1, 2019).

9 ► https://flatiron.com/about-us (Accessed January 1, 2019).

Electronic Health Record (EHR) vendors are also creating networks to support or enable research data sharing among their clients. For example, Epic Corporation began working with a group of academic health center customers in 2013 on such a data sharing network and created an advisory group of CRI experts from such sites to inform governance, infrastructure, and research data sharing processes. Similarly, Cerner corporation has created a centralized data sharing platform that enables their clients to combine and access large volumes of de-identified patient-level data for trial design and population health management purposes.

Further, major computing and cloud computing vendors, such as Amazon, Apple, Facebook, Google, and Microsoft, have launched initiatives to provide patient-centered data sharing capabilities, which empower individuals to aggregate and share their own health care data from a variety of sources ("Large technology companies continue to ramp up healthcare forays," 2018). These capabilities introduce new data sharing scenarios in which individual patients will have the ability to "donate their data" to research projects without engaging their health care providers as intermediaries in such transactions. The impact of all of these activities remains to be understood, given their relative immaturity at the time this chapter is being written.

27.3.8 All of Us

The All of Us Research Program is the principle component of the US government's Precision Medicine Initiative (PMI), established in 2016 to establish a cohort of one million Americans who provide health status data, blood, and urine specimens for clinical and genetic analysis, and access to their complete EHR data (Collins and Varmus 2015). Recruitment is currently being carried out primarily by 11 consortia across the country, involving 51 health care organizations.[10]

Unlike other consortia described above, data and specimens are being consolidated centrally at a Data and Research Center (DRC) and a biobank, respectively. Plans include the collection of data from all of the participants' EHRs, not just those at recruitment centers, and the sequencing of participants full genomes.

The All of Us Research Program also differs from other consortia in its patient-centered focus. All of Us asks participants for full access to their fully identified data for use by researchers without prior approval of specific studies. In return, patients are repaid through full transparency and partnership. Participants have access to their information and play a role in helping to identify important research priorities for the program.

27.4 Data Standards in Clinical Research

The use of standards to represent clinical research information provides the same challenges and benefits found in other informatics application areas. Data may be captured with standard terminologies or translated into standards to support data reporting and sharing which, in turn, require agreed-upon standard frameworks to support such exchanges. Standards are even being developed for the representation of clinical trial protocols themselves. ◘ Figure 27.3 depicts how the various kinds of standards fit into the overall schema of clinical research, ranging from data models that define how data are to be represented, through standards for terminologies to actu-

10 ► https://allofus.nih.gov/about/program-partners/health-care-provider-organizations (Accessed January 1, 2019).

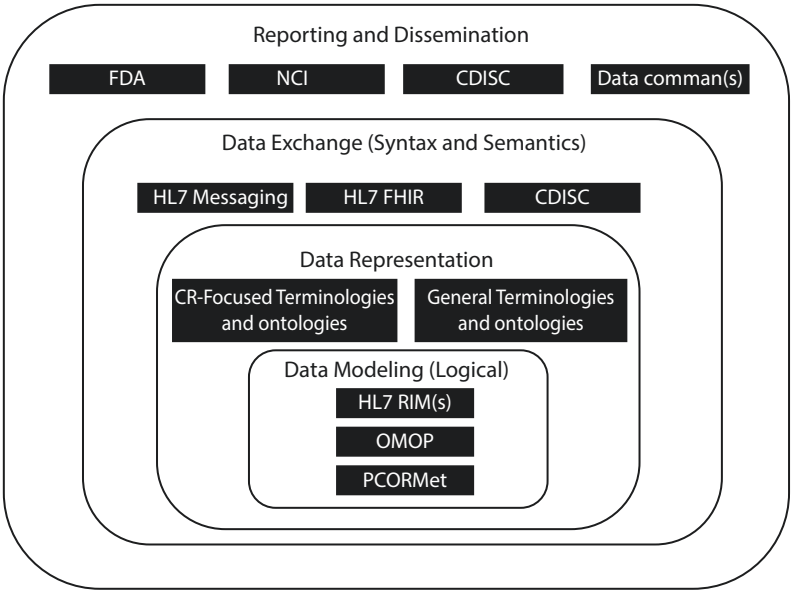


Fig. 27.3 Relationships among various general purpose and CRI-specific standards that are relevant to the design, conduct, and dissemination of clinical research studies. Data modeling determines how terms from terminologies and ontologies will be recorded in clinical research databases. Exchange standards determine how data will map from the model to the messages used for interchanging the data. The use of messages is determined by the requirements of regulatory agencies and collaborating research groups. See text for an explanation of acronyms used in this figure

ally represent the data and structures for exchanging them, out to standards for reporting and sharing. The standards described here are some of the current and most prevalent ones, but they continue to evolve and new standards relevant to the CRI domain are constantly emerging.

27.4.1 Data Modeling Standards to Support Clinical Research

Formats for data sharing typically include a data model for the information to be shared, leaving to individual contributors the later task of mapping local data into the exchange model. For example, i2b2 uses a standard data model internally that is based on a generic **entity-attribute-value model** (EAV) that essentially allows any data to be stored using any desired controlled terminology (Klann et al. 2016). This allows queries to be conducted across multiple i2b2 databases, as with SHRINE, but consolidation of the data may still require mapping of data to a common terminology (although the ACT network requires using a shared ontology for representing data). Note that the data model used for storing the data in a database (i2b2, for example) may not be acceptable for data exchange.

An alternative approach is the model-driven architecture, in which an underlying data model is created for the express purpose of representing all aspects of an information design, including data representation. Previously, the models used for clinical research management systems have been those required to support system functionality. New efforts are underway to create standards for modeling the actual research protocols, to enable a logical representation that includes the semantic aspects of the protocol (for example, the relationships between specific interventions and observations intended to measure their effects). While use of such mod-

els may make the research process somewhat more complicated, the mapping to standards used for exchanging data becomes greatly simplified.

For example, Health-Level 7 (HL7; see ▶ Chap. 7) is an open standards development organization that develops consensus standards for all manner of clinical and administrative data, and is also working on clinical research-specific standards, such as the **Regulated Clinical Research Information Management** (**RCRIM**) model in order to define messages, document structures, terminology, and semantics related to the collection, storage, distribution, integration and analysis of research information (R. L. Richesson and Krischer 2007). The main focus of the work is on data related to studies involving US Food and Drug Administration (FDA) regulated products (drugs and devices).

The **Biomedical Research Integrated Domain Group (BRIDG)** Model[11] is designed to harmonize models from the HL7 RCRIM, the **Clinical Data Interchange Standards Consortium (CDISC)** a standards group motivated by the needs of the pharmaceutical and bio-technology industry entities that sponsor or otherwise support many clinical studies. CDISC provides a standard for submitting regulatory information to the FDA (Fridsma et al. 2008).

In a similar manner, HL7 has created a standard **Clinical Document Architecture** (**CDA**; see ▶ Chap. 7) that specifies the structure and semantics of "clinical documents" for the purpose of data exchange. A CDA can contain any type of clinical content, including clinical notes typically found in EHRs but may also include case report forms from research studies. **HL7 Fast Healthcare Interoperability Resources** (**FHIR**; see ▶ Chap. 8) supports the exchange of CDA documents (Bender and Sartipi 2013).

The consortia that share clinical data various consortia have each developed their own data models to cover the common elements found in EHRs: patient demographics, encounters, diagnoses, laboratory results, procedures, and medications. PCORI uses the **PCORNet Common Data Model** (Belenkaya et al. 2015) based on the data model used in the FDA's Sentinel Initiative,[12] while OHDSI uses the **OMOP Common Data Model** (Ryan et al. 2009). The All of Us program has adopted the OMOP model as well.

27.4.2 Terminology Standards to Support Clinical Research

As described previously, the design of clinical protocols includes rigorous attention to the types of data to be collected and the format of those data. This often involves the use of controlled terminologies to capture categorical data. The terminology may be as small as "yes/no" or a ten-point pain scale for capturing subjects' symptoms, or it may be as vast as a list of all possible drugs or diseases in a subject's medical history. In many cases, researchers will simply compose sets of terms that meet their immediate needs and then require all investigators participating in the study to apply them consistently.

Because the terms used in clinical research are often identical to those used in clinical care, standard multi-use terminologies (such as those described in ▶ Chap. 7) are often appropriate for use in capturing clinical research data. However, there are some aspects of clinical research that are not well represented in mainstream terminologies; and in these cases, terminologies and their richer forms, ontologies, that are more focused on clinical research, are required. In particular, clinical research data and workflow models require controlled terminologies and ontologies that define domain-specific concepts and standard **common data elements** (**CDEs**). Collections of standard terms for CDEs can be found in the NIH's Common Data Element server (Rubinstein and McInnes 2015).[13] In a

11 ▶ https://www.cdisc.org/standards/domain-information-module/bridg (Accessed January 1, 2019).

12 ▶ https://www.sentinelinitiative.org (Accessed January 1, 2019).

13 ▶ https://www.nlm.nih.gov/cde (Accessed January 1, 2019).

similar manner, the Ontology for Biomedical Investigations (OBI) (Bandrowski et al. 2016) has been developed by a consortium of representatives from across the spectrum of biomedical research, and includes terms to represent the design of protocols and data collection methods, as well as the types of data obtained and the analyses performed on them.

There are several reasons for considering the use of *standard* controlled terminologies in the capture of clinical research data. One reason is to take advantage of clinical data that are already being collected on research subjects for other purposes. A common example is the use of data on morbidity and mortality that are collected using one of the various versions and derivatives of the **International Classification of Diseases** (ICD; see ▶ Chap. 7). In the US, for example, patient diagnoses are reported for billing purposes using the *Clinical Modifications* of the tent edition of ICD (ICD-10-CM). While such coded information is readily available, researchers repeatedly find that ICD-10-CM codes assigned to patient records have an undesired level of reliability or granularity, especially when compared to with the actual content of the records (Topaz et al. 2013). Thus, the convenience of using such standard codes may be outweighed by the imprecision, which can adversely affect study design and analytical results.

A second reason for adopting a standard controlled terminology is simply to avoid "reinventing the wheel." As is described in ▶ Chap. 7, a great deal of effort has been expended in the creation of domain-specific terminologies that are comprehensive, unambiguous, and maintained over time. Designating such terminologies for use in a protocol design can relieve researchers of having to worry about the quality of the terminology. For example, a researcher is unlikely to encounter novel concepts when recording subjects' demographic data, such as gender, marital status, religion, and race. Specifying, for example, that ISO standards should be used for these data elements greatly simplifies the protocol-design process.

A third reason for choosing standard terminologies relates to the ability to compare data collected in one study with those collected in others. For example, the use of a standard scale for recording a subject's pain will allow comparison of results from a study of one treatment with those from a second study of another treatment. The selection of an appropriate standard for a particular purpose is not straightforward (for example, the NIH Pain Consortium lists six different scales[14]). The choice may be determined simply based on the emerging popularity of one terminology over another in a wide community of those investigating similar problems. PCORNet, OHDSI, and All of Us each specify the use of terminologies such as ICD, SNOMED, CPT, and LOINC (See ▶ Chap. 7).

A fourth use of standard terminologies relates to reporting requirements. Government agencies sometimes require the reporting of clinical research data and, when they do, often require certain data to be reported using a particular standard. For example, the FDA requires the use of the **Medical Dictionary for Regulatory Activities (MedDRA)** for reporting all adverse events occurring in drug trials (Brown et al. 1999), while the **Cancer Therapy Evaluation Program (CTEP)** at the National Cancer Institute (NCI) requires the use of **Common Terminology Criteria for Adverse Events (CTCAE)** (Colevas and Setser 2004). In an analogous manner, at the international level, the World Health Organization requires the use of the Adverse Reactions Terminology (WHO-ART).[15] Faced with such reporting requirements, researchers sometimes choose to record data in these terminologies as they are being captured. In those cases where the clinical questions being answered require more detailed data, however, researchers must resort to recording data with some other standard (such as SNOMED; see ▶ Chap. 7), or a controlled terminology of their own creation, and then translating them to the terminology or terminologies required for reporting purposes.

14 ▶ http://nationalpainreport.com/how-to-measure-chronic-pain-8812496.html (Accessed January 1, 2019).

15 ▶ https://www.who-umc.org/vigibase/services/learn-more-about-who-art (Accessed January 1, 2019).

27.4.3 Clinical Research Reporting Requirements

Requirements for reporting research data, particularly those related to outcomes and adverse events, are generally accompanied by specifications for the format of the data being reported. For example, the FDA's Center for Drug Evaluation and Research (CDER) accepts reports using the HL7 Individual Case Safety Report, while the NCI's CTEP allows submission of adverse event information to its Cancer Therapy Evaluation Program Adverse Event Reporting System (CTEP-AERS[16]) either manually, using a Web-based application, or electronically via a web-services API. As mentioned earlier in this section, these agencies require that data be coded with standard terminologies, such as MedDRA and CTCAE, respectively.

Several reporting requirements have emerged for the purpose of making clinical trial results publicly available, both to support reuse of the data by researchers and as information sources for patients and their families. In 2000, the US National Library of Medicine launched ▶ ClinicalTrials.gov to provide a mechanism for researchers to voluntarily register their trials so that those interested in participating as research subjects can identify, via the World Wide Web, studies relevant to their condition. ▶ ClinicalTrials.gov currently includes information from over 300,000 trials from over 207 countries. In 2004, the European Union initiated a similar effort, called the European Union Drug Regulating Authorities Clinical Trials (EudraCT). ▶ ClinicalTrials.gov and EudraCT also support the reporting of the clinical trials results. While the submissions are nominally voluntary, federal agencies often mandate the reporting as a requirement for obtaining research funds or to obtain approval for regulated drugs and devices. In the US, for example, the Food and Drug Administration Amendments Act of 2007 (FDAAA) strongly reinforced these requirements. In addition, the over 4829 peer-reviewed biomedical journals that participate in the International Consortium of Medical Journal Editors (ICJME) now require public, prospective registration in ▶ ClincialTrials.gov or similar databases of clinical trials of all interventions (including devices) in order for resultant manuscripts to be considered for publication.[17]

Each repository has defined its own mechanisms for transmitting protocol data. ▶ ClinicalTrials.gov, for example, allows investigators to enter their data through an interactive Web site or to upload data in a defined XML (**eXtensible Markup Language**) format (see ▶ Chap. 7). Clinical research data management systems that can export their study in this format can save the researcher much manual effort and assure accurate data entry (Zarin et al. 2011).

27.5 CRI and the COVID-19 Pandemic

The emergence in 2020 of the COVID-19 pandemic has raised many biomedical and health issues on which informatics can have a major impact. Novel challenges include alteration of data collection functions of EHRs (▶ Chap. 14) and telemedicine (▶ Chap. 20) to support the needs of patient care and research functions. New data to be captured with these technologies will need to be met with advancements in data sharing and research analytics.

EHRs will need to be easily modifiable to capture new kinds of data when needed for patient care and for research. COVID-19 and other recent epidemics have demonstrated that patient travel and contact information need to be incorporated into the record for risk assessment (e.g., intensity of social contact; Meinert et al. 2020). Such data, in turn, can be used to study epidemiologic patterns across patient populations (e.g., developing predictive risk scores for disease exposure; Liu et al. 2020).

16 ▶ https://ctep.cancer.gov/protocolDevelopment/electronic_applications/adverse_events.htm (Accessed January 1, 2019).

17 ▶ http://www.icmje.org/journals-following-the-icmje-recommendations (Accessed January 1, 2019).

The value of telemedicine for bringing medical expertise to patients located in an area where expertise is lacking is well demonstrated. In a pandemic, such as the COVID-19 outbreak, the patients are by definition located almost everywhere, while expertise in this new condition is limited to a relatively small number of academic and government institutions. Perhaps as never before, telemedicine is also needed to protect the clinicians and other caregivers who must minimize contact with contagious patients. The technologies of telemedicine are also being exploited for "tele-research" where, again, potential research subjects are widely dispersed and hard to reach and at the same time pose a potential danger to the researchers. COVID-19 researchers are drawing lessons on research at a distance, including effective recruitment and consent, from previous work in HIV (Mgbako et al. 2020) and geriatric research (Nicol et al. 2020).

The rapid expansion of data collection in response to new domains and data types often outstrips the ability of standards organizations to keep up. This poses challenges for the sharing, aggregation, and analysis of data for surveillance, understanding natural history of the disease, and patient recruitment. In COVID-19, new data types and domains have outstripped the standards development organization's ability to keep up. As a result, consortia such as the the ACT network (see Sect. 29.3.4) and CD2H consortium in the CTSA network (see Sect. 29.3.2), both sponsored by NCATS, as well as the NIH-sponsored All of Us network (sees Sect. 29.3.8) have had to develop their own criteria for inferring which patients in EHR databases have COVID-19.[18] For ACT, the criteria support cross-institutional queries to obtain summary results that can be used for patient enrollment.[19] All of Us is ramping up data submissions from healthcare organizations to obtain more timely data on their participants who are afflicted with COVID-19. CD2H has established the National Clinical COVID Collaboratory (N3C) to create a centrally sponsored "data enclave", composed of de-identified COVID-19 patient records from CTSA-sponsored institutions that can be used directly for data analysis (see Fig. 27.4).[20]

COVID-19 research has placed new demands for analytic support. In part, this is due to the perceived urgency to find answers to epidemiologic, preventive, diagnostic, prognostic and therapeutic questions. Another issue is diversity of data sharing and harmonization efforts that exist at the regional and national levels (described above). Researchers are overwhelmed with trying to understand how to navigate various resources, their contents, and associated regulatory constraints, to find the best "fit" given a driving problem or hypothesis (see Fig. 27.5).

27.6 Future Directions for CRI

As the preceding sections illustrate, significant progress has been made to advance the state of the CRI domain, and such advances have already begun to enable significant improvements in the quality and efficiency of clinical research. These advances can be viewed as having been achieved at the individual investigator level (e.g., improvements in protocol development, study design, participant recruitment, etc.), through approaches and resources developed and implemented at the institutional level (e.g., development of methods and resources in data warehousing that enable storage and retrieval of clinical data for research, development of novel clinical trials management systems, etc.), and through mechanisms that have enabled and facilitated the endeavors multi-center research consortia to drive team science (e.g., innovations that enable data management and interchange for multi-center studies) (Bourne et al. 2015; P. J. Embi et al. 2019; Payne et al. 2018; Sanchez-Pinto et al. 2017; Smoyer et al. 2016).

18 ▶ https://allofus.nih.gov/news-events-and-media/announcements/coronavirus-update-all-us (Last accessed June 2, 2020).

19 ▶ https://ncats.nih.gov/pubs/features/ctsa-act (Last accessed June 2, 2020).

20 ▶ https://ncats.nih.gov/n3c (Last accessed June 2, 2020).

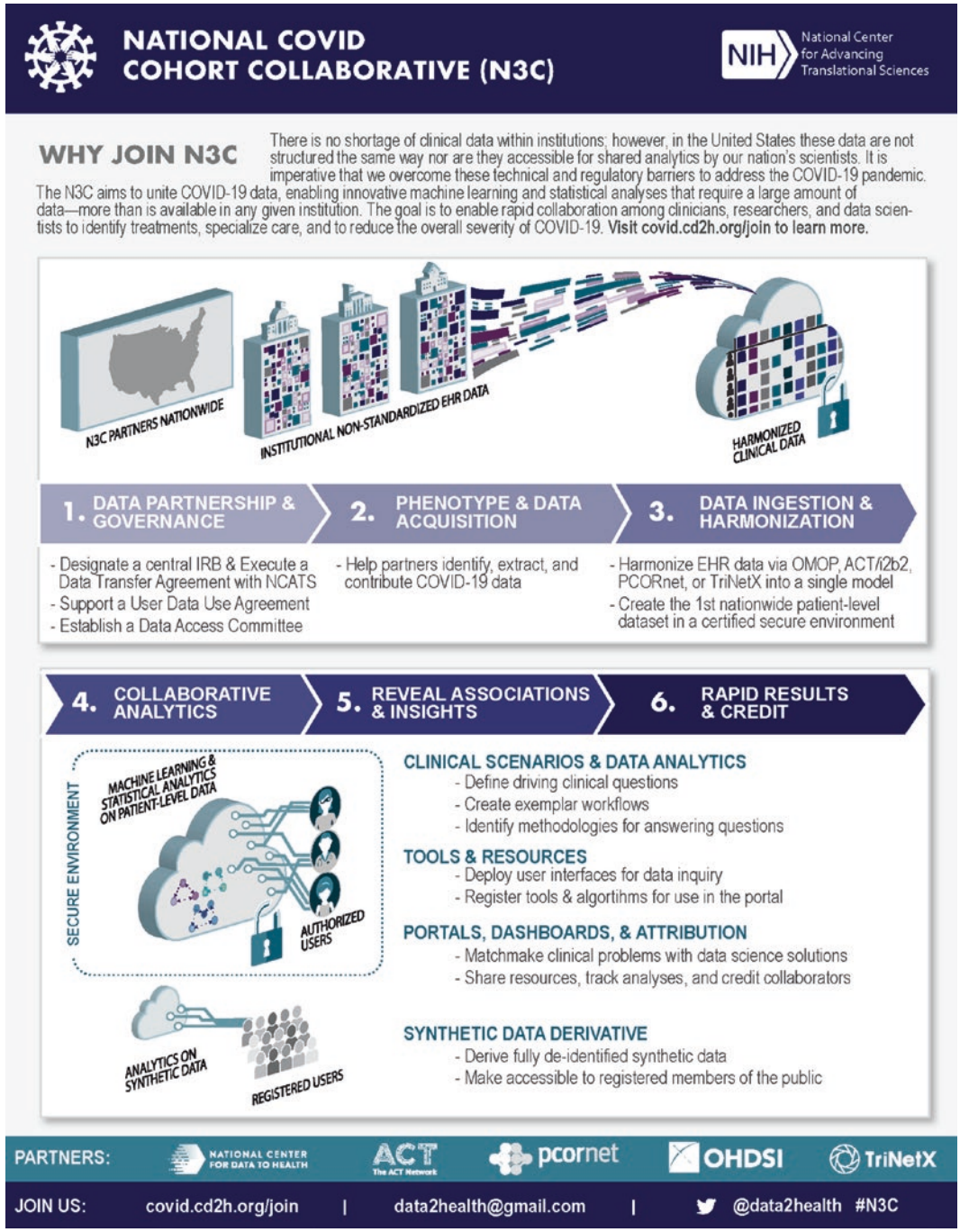


Fig. 27.4 The National COVID Cohort Collaboratory is a project at the National Institute for the Advancement of Translational Science (NCATS) which is pooling electronic data on COVID-19 patients in order to provide researchers with access to data on patient sets that are larger than those available in a single institution. Efforts include data acquisition and harmonization, creation of a centralized data enclave to support collaborative analytics, and development of a synthetic data set based on actual patient data, that can be downloaded for analysis but poses no risk of reidentification. (See ► https://covid.cd2h.org. Courtesy of the U.S. National Library of Medicine)

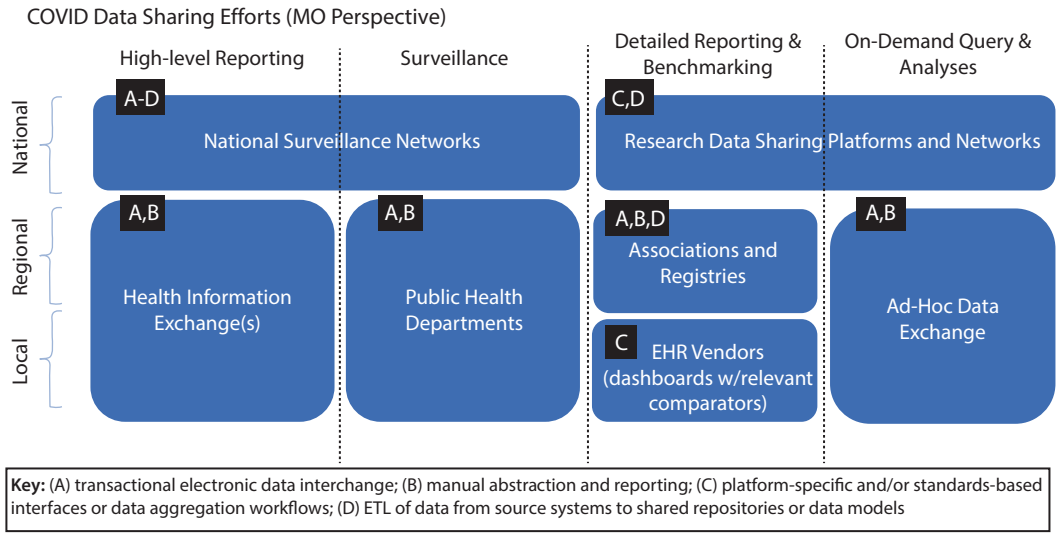


Fig. 27.5 The emergence of the COVID-19 pandemic has engendered many efforts to share clinical and epidemiologic data in order to rapidly learn patterns of natural history, predictions of risk for exposure and severe disease, and therapeutic outcomes. Efforts include local sharing in patient care networks as well as national coalitions established to support both patient care and research

Looking towards the future of the field of CRI, a number of national- and international-level trends are introducing new or evolved challenges and opportunities, including:

- Nearly ubiquitous adoption of EHRs at the national level, as well as the availability of lightweight and scalable data-level and platform-independent interoperability standards are making it possible to **build and leverage learning health care systems at-scale**, in which every patient encounter is an opportunity to learn and improve that patients care as well as the care delivered to broader populations (Friedman et al. 2010). However, while such systems do enable the pragmatic collection and integration of increasingly large volumes of clinical phenotype data, these "real world" data also exhibit new types of challenges in terms of scope, completeness, quality, and domain coverage. As such, new methods are needed to enable the characterization and analysis of such data in a rigorous and reproducible manner.
- Mobile, wearable, and other sensor technologies, in conjunction with new mechanisms for collecting patient reported outcome (PRO) data, are allowing for the conduct of **clinical research studies that incorporate data from beyond the clinic and hospital settings**. However, the temporality, granularity, and reliability of such data is vastly different from that collected via more "traditional" mechanisms as have been introduced in this chapter. As a result, important questions are now being raised concerning how such emergent data sources can/should be integrated with those "traditional" data types, and further, the methods that are appropriate when seeking to identify meaningful "signals" within ensuing multi-scale and complex data sets.
- The ready availability of **artificial intelligence (AI)** platforms and methods (e.g., knowledge-based systems, cognitive computing, high-throughput machine learning, deep learning, etc.) are making it possible to perform **analyses across and between data scales**, for example, identifying meaningful patterns that incorporate genomic, clinical, demographic, social, and environmental measures. Such multi-scale reasoning is essential to the generation of evidence in support of precision medicine and/or health paradigms, but also introduce critical questions concerning how to design and execute such

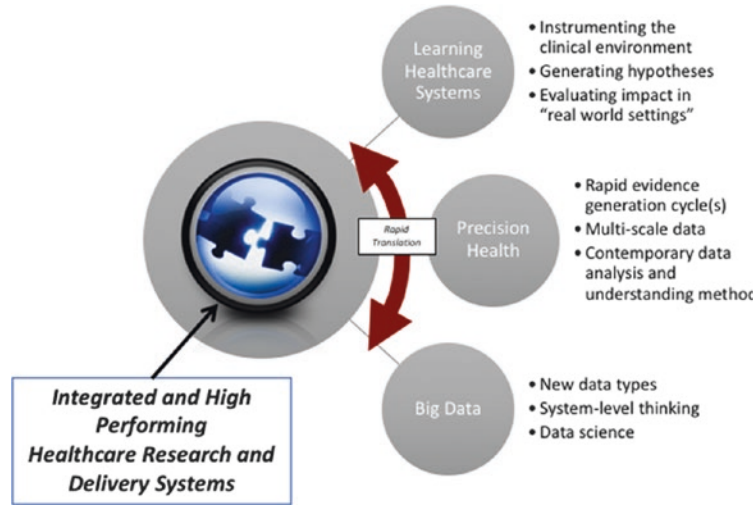


Fig. 27.6 Aspirational integrated and high performing healthcare research and delivery system model, enabled via emergent CRI frameworks, methods, and technologies

data-intensive studies, again emphasizing reproducibility and rigor as was the case in the two preceding trends.

As CRI capabilities and systems continue to evolve, an ability to assess the maturity of such systems and environments will become increasingly important. As with maturity assessments of EHR deployments and their use across health systems, so too will maturity and deployment models emerge for measuring the deployment and use of CRI systems to enable robust research enterprises (Knosp et al. 2017; Pettit 2013).

One key aspect of system maturity will be a mature workforce of CRI professionals and leaders. Such a workforce will become increasingly important to the successful deployment, management, and optimal use of CRI systems. Examples of such leaders, like the Chief Research Information Officer (CRIO) role first established in 2010, will continue to become common and as essential to the functioning of mature clinical and translational research enterprises as are their health IT (e.g., CIO) and clinical informatics (e.g., CMIO) leader counterparts (Sanchez-Pinto et al. 2017).

When viewed collectively, these emergent challenges and opportunities help to define a rapidly evolving CRI environment, in which we have the ability to create integrated and high performing health care research and delivery systems (Fig. 27.6), driven by rapid translation between and across:

- Functional **learning health care systems** in which we instrument the clinical environment to generate large-scale and pragmatic data, generate hypothesis to be testing in view of such data, and then evaluate the impact of ensuing data-driven intervention in such "real word settings."
- **Precision health** frameworks wherein the translation between research-generated evidence and practice is both rapid and cyclical, acting upon multi-scale data, and using contemporary analysis and understanding methods, such as those being made available through advances in artificial intelligence (AI).
- **Big data** resources, generated via the preceding areas and incorporating new or emergent data types, leveraging systems-level thinking (e.g., across scales) and employing data science methodologies to identify important signals or motifs in such high volume, velocity, and variability data.

27.7 Conclusion

This chapter has sought to introduce the following major themes: (1) design characteristics that serve to define contemporary clinical studies; (2) foundational information

needs inherent to clinical research programs and the types of information systems can be used to address or satisfy such requirements; (3) the role of multi-purpose platforms, such as Electronic Health Record (EHR) systems, that can be leveraged to enable clinical research programs; (4) the role of standards in supporting interoperability across and between actors and entities involved in clinical research activities; and (5) future directions for the CRI domain and how such endeavors may alter or optimize the conduct of clinical research. As we have explained, the clinical research environment is data, information, and knowledge intensive, thus calling for the application of biomedical informatics theories and methods. This set of features explains and justifies CRI's emergence as a distinct and highly valued sub-discipline of the broader field of biomedical informatics. Part of the evolution of CRI can be attributed to the extraordinary increase in the scope and pace of clinical and translational science research and development that has been catalyzed by a variety of funding and policy initiatives that seek to re-engineer the way in which governmental, public, and private entities advance basic science discoveries into practical therapies. Such evolution is further bolstered by technical and environmental changes, such as the advent of learning health care systems, the availability of new and novel data capture/generation mechanisms, and advances in our ability to analyze and understand "big data." As such, CRI has accordingly become a dynamic and relevant sub-domain of biomedical informatics knowledge and practice, providing a broad spectrum of research and development opportunities in context of both basic and applied informatics science.

Suggested Readings

Embi, P. J., & Payne, P. R. (2009). Clinical research informatics: Challenges, opportunities and definition for an emerging domain. Journal of the American Medical Informatics Association, 16(3), 316–327. This report defines the field of biomedical informatics knowledge and practice that applies to the design and conduct of clinical studies. Further, it presents a framework for the alignment of general purpose and research-specific technology.

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., et al. (2019). The REDCap consortium: Building an international community of software platform partners. Journal of Biomedical Informatics, 95, 103208. REDCap is one of the most common and widely adopted electronic data capture platforms used in the clinical research domain. This report describes the structure and function of the REDCap consortium, which has led the development and dissemination of this ubiquitous clinical research data management tool.

Hersh, W. R., Weiner, M. G., Embi, P. J., Logan, J. R., Payne, P. R., Bernstam, E. V., et al. (2013). Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical Care, 51(8 0 3), S30. This report introduces practical issues to consider when utilizing data from electronic health records to support and enable clinical research. It also presents a series of critical questions to be asked and answered when designing and executing such studies.

Hripcsak, G., Shang, N., Peissig, P. L., Rasmussen, L. V., Liu, C., Benoit, B., Carroll, R. J., Carrell, D. S., Denny, J. C., Dikilitas, O., & Gainer, V. S. (2019). Facilitating phenotype transfer using a common data model. Journal of Biomedical Informatics, 96, 103253. This report outlines the role of common data models in facilitating the systematic and reproducible phenotyping of individual patients as well as populations. In addition, the report provides a comparative assessment of existing data models and their utility for computational phenotyping and resultant data analyses.

Payne, P. R., Johnson, S. B., Starren, J. B., Tilson, H. H., & Dowdy, D. (2005). Breaking the translational barriers: The value of integrating biomedical informatics and translational research. Journal of Investigative Medicine, 53(4), 192–201. This report describes the critical role of biomedical informatics theories and methods in overcoming the T1 and T2 clinical and translational barriers. It also provides a conceptual model for the alignment of such capabilities with the spectrum of activi-

ties that make up the clinical and translational research "lifecycle.".

Tenenbaum, J. D., Avillach, P., Benham-Hutchins, M., Breitenstein, M. K., Crowgey, E. L., Hoffman, M. A., et al. (2016). An informatics research agenda to support precision medicine: Seven key areas. Journal of the American Medical Informatics Association, 23(4), 791–795. This perspective introduces an agenda for informatics research and practice in the context of precision medicine. The authors describe multiple axes of how supporting theoretical frameworks and applied methods can generate and deliver evidence in support of precision risk management, diagnosis, and treatment planning.

Weng, C., Shah, N., & Hripcsak, G. (2020). Deep phenotyping: Embracing complexity and temporality—Towards scalability, portability, and interoperability. Journal of Biomedical Informatics. As clinical and translational research becomes increasingly multi-institutional and involves the sharing of deep phenotypes across and between traditional organizational boundaries, there is a need for common tools and methods to derive and represent such constructs. This report outlines the current state-of-the-art in terms of computational phenotyping methods and interchange standards.

27

Questions for Discussion

1. How do the foundational information needs of clinical research differ depending on the type and phase of study being undertaken? Do study phases have an impact on the primacy of such information needs?
2. What is the role of biomedical informatics with regard to decreasing bias in RCTs and thus enhancing the internal validity, external validity, and generalizability of study results?
3. How can clinical or general-purpose information systems and research-specific tools be employed synergistically to address clinical research-specific information needs, such as participant recruitment or the population of study-specific data capture instruments?
4. How do the core functional components of common clinical trial management systems (CTMS) overlap with or otherwise replicate the functionality of electronic health record (EHR) systems? To what extent does this similarity or difference inform the need for syntactic and/or semantic interoperability among such systems?
5. In what situations is the use of clinical research-specific terminologies or ontologies appropriate? In such situations, what challenges exists relative to the selection, use, and maintenance of appropriate standards?
6. What is the role of data standards in enabling the dissemination and reuse of study-generated data sets? How can the use of such standards enable the cross-linkage or integrative analysis of data sets derived from multiple but independent studies?
7. Compare and contrast the future directions of CRI with those of other BMI sub-disciplines and focus areas described in this book. To what extent are they similar and different, and what are the implications of such findings relative to the role of common informatics theories and methods and their applicability to the clinical research domain?

References

Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M. H., Bug, B., Chibucos, M. C., et al. (2016). The ontology for biomedical investigations. *PLoS One, 11*(4).

Belenkaya, R., Mirhaji, P., Khayter, M., Torok, D., Khare, R., Ong, T., & Schilling, L. (2015). *Establishing interoperability standards between OMOP CDM v4, v5, and PCORnet CDM v1.* Paper presented at the OHDSI workshop.

Bender, D., & Sartipi, K. (2013). *HL7 FHIR: An Agile and RESTful approach to healthcare information exchange.* Paper presented at the Proceedings of the 26th IEEE international symposium on computer-based medical systems.

Bhatt, D. L., & Mehta, C. (2016). Adaptive designs for clinical trials. *New England Journal of Medicine, 375*(1), 65–74.

Bourne, P. E., Bonazzi, V., Dunn, M., Green, E. D., Guyer, M., Komatsoulis, G., et al. (2015). The NIH big data to knowledge (BD2K) initiative. *Journal of the American Medical Informatics Association, 22*(6), 1114–1114.

Brightling, C. E. (2017). Clinical trial research in focus: Do trials prepare us to deliver precision medicine in those with severe asthma? *The Lancet Respiratory Medicine, 5*(2), 92–95.

Brown, E. G., Wood, L., & Wood, S. (1999). The medical dictionary for regulatory activities (MedDRA). *Drug Safety, 20*(2), 109–117.

Colevas, A., & Setser, A. (2004). The NCI Common Terminology Criteria for Adverse Events (CTCAE) v 3.0 is the new standard for oncology clinical trials. *Journal of Clinical Oncology, 22*(14_suppl), 6098–6098.

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793–795.

Embi, P. (2013). Clinical research informatics: Survey of recent advances and trends in a maturing field. *Yearbook of Medical Informatics, 8*(1), 178–184.

Embi, P. J., & Payne, P. R. (2009). Clinical research informatics: Challenges, opportunities and definition for an emerging domain. *Journal of the American Medical Informatics Association, 16*(3), 316–327.

Embi, P. J., & Payne, P. R. (2013). Evidence generating medicine: Redefining the research-practice relationship to complete the evidence cycle. *Medical Care, 51*, S87–S91.

Embi, P. J., & Payne, P. R. (2014). Advancing methodologies in Clinical Research Informatics (CRI). *Journal of Biomedical Informatics, 52*(C), 1–3.

Embi, P. J., Richesson, R., Tenenbaum, J., Kannry, J., Friedman, C., Sarkar, I. N., & Smith, J. (2019). Reimagining the research-practice relationship: Policy recommendations for informatics-enabled evidence-generation across the US health system. *JAMIA 2*(1), 2–9.

Fleurence, R. L., Beal, A. C., Sheridan, S. E., Johnson, L. B., & Selby, J. V. (2014). Patient-powered research networks aim to improve patient care and health research. *Health Affairs, 33*(7), 1212–1219.

Fleurence, R. L., Curtis, L. H., Califf, R. M., Platt, R., Selby, J. V., & Brown, J. S. (2014). Launching PCORnet, a national patient-centered clinical research network. *Journal of the American Medical Informatics Association, 21*(4), 578–582.

Fridsma, D. B., Evans, J., Hastak, S., & Mead, C. N. (2008). The BRIDG project: A technical report. *Journal of the American Medical Informatics Association, 15*(2), 130–137.

Friedman, C. P., Wong, A. K., & Blumenthal, D. (2010). Achieving a nationwide learning health system. *Science Translational Medicine, 2*(57), 57cm29–57cm29.

Hripcsak, G., Duke, J. D., Shah, N. H., Reich, C. G., Huser, V., Schuemie, M. J., et al. (2015). Observational Health Data Sciences and Informatics (OHDSI): Opportunities for observational researchers. *Studies in Health Technology and Informatics, 216*, 574.

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2013). *Designing clinical research*. Lippincott Williams & Wilkins.

Johnson, S. B., Farach, F. J., Pelphrey, K., & Rozenblit, L. (2016). Data management in clinical research: Synthesizing stakeholder perspectives. *Journal of Biomedical Informatics, 60*, 286–293.

Klann, J. G., Abend, A., Raghavan, V. A., Mandl, K. D., & Murphy, S. N. (2016). Data interchange using i2b2. *Journal of the American Medical Informatics Association, 23*(5), 909–915.

Knosp, B., Barnett, W. K., & Embi, P. J. (2017). *Maturity models for research IT and informatics – Reports from the field.* Paper presented at the AMIA Summits on Translational Science, San Francisco, CA.

Large technology companies continue to ramp up healthcare forays. (2018, September 17). *Health Data Management*. Retrieved from https://www.healthdatamanagement.com/articles/large-technology-companies-continue-to-ramp-up-healthcare-forays

Liu, Y., Gu, Z., Xia, S., Shi, B., Zhou, X. N., Shi, Y., & Liu, J. (2020). What are the underlying transmission patterns of COVID-19 outbreak? – An age-specific social contact characterization. *E Clinical Medicine, 18*, 100354.

Meinert, E., Milne-Ives, M., Surodina, S., & Lam, C. (2020). Agile Requirements Engineering and Software Planning for a Digital Health Platform to Engage the Effects of Isolation Caused by Social Distancing: Case Study. *JMIR Public Health Surveill, 6*(2), e19297.

Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., et al. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nature Genetics, 39*(10), 1181–1186.

Mgbako, O., Miller, E. H., Santoro, A. F., Remien, R. H., Shalev, N., Olender, S., Gordon, P., & Sobieszczyk, M. E. (2020). COVID-19, telemedicine, and patient empowerment in HIV care and research. *AIDS Behav, 21*, 1–4.

Murphy, S. N., Weber, G., Mendis, M., Gainer, V., Chueh, H. C., Churchill, S., & Kohane, I. (2010). Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *Journal of the American Medical Informatics Association, 17*(2), 124–130.

Nicol, G. E., Piccirillo, J. F., Mulsant, B. H., & Lenze, E. J. (2020). Action at a distance: Geriatric research during a pandemic. *Journal of the American Geriatrics Society, 68*, 922–925. https://doi.org/10.1111/jgs.16443.

Payne, P. R., Bernstam, E. V., & Starren, J. B. (2018). Biomedical informatics meets data science: Current state and future directions for interaction. *JAMIA open, 1*(2), 136–141.

Payne, P. R., Embi, P. J., & Niland, J. (2010). Foundational biomedical informatics research in

the clinical and translational science era: A call to action. *Journal of the American Medical Informatics Association, 17*(6), 615–616.

Payne, P. R., Johnson, S. B., Starren, J. B., Tilson, H. H., & Dowdy, D. (2005). Breaking the translational barriers: The value of integrating biomedical informatics and translational research. *Journal of Investigative Medicine, 53*(4), 192–201.

Payne, P. R., Pressler, T. R., Sarkar, I. N., & Lussier, Y. (2013). People, organizational, and leadership factors impacting informatics support for clinical and translational research. *BMC Medical Informatics and Decision Making, 13*(1), 20.

Pencina, M. J., & Peterson, E. D. (2016). Moving from clinical trials to precision medicine: The role for predictive modeling. *Journal of the American Medical Association, 315*(16), 1713–1714.

Petrone, J. (2018). *Roche pays $1.9 billion for Flatiron's army of electronic health record curators*. Nature Publishing Group.

Pettit, L. (2013). Understanding EMRAM and how it can be used by policy-makers, hospital CIOs and their IT teams. *World Hospitals and Health Services: the Official Journal of the International Hospital Federation, 49*(3), 7–9.

Prokscha, S. (2011). *Practical guide to clinical data management*. CRC Press.

Raths, D. (2013, July 15). *Yale benefits from integration of epic, clinical trial management system*. Healthcare Informatics.

Richesson, R., Horvath, M., & Rusincovitch, S. (2014). Clinical research informatics and electronic health record data. *Yearbook of Medical Informatics, 9*(1), 215.

Richesson, R. L., & Krischer, J. (2007). Data standards in clinical research: Gaps, overlaps, challenges and future directions. *Journal of the American Medical Informatics Association, 14*(6), 687–696.

Rubinstein, Y. R., & McInnes, P. (2015). NIH/NCATS/GRDR® common data elements: A leading force for standardized data collection. *Contemporary Clinical Trials, 42*, 78–80.

Ryan, P. B., Griffin, D., & Reich, C. (2009). OMOP common data model (CDM) specifications. Accessed at: https://ohdsi.github.io/CommonDataModel/.

Saad, E. D., Paoletti, X., Burzykowski, T., & Buyse, M. (2017). Precision medicine needs randomized clinical trials. *Nature Reviews Clinical Oncology, 14*(5), 317–323.

Sanchez-Pinto, L. N., Mosa, A. S., Fultz-Hollis, K., Tachinardi, U., Barnett, W. K., & Embi, P. J. (2017). The emerging role of the Chief Research Informatics Officer in academic health centers. *Applied Clinical Informatics, 8*(03), 845–853.

Smoyer, W. E., Embi, P. J., & Moffatt-Bruce, S. (2016). Creating local learning health systems: Think globally, act locally. *Journal of the American Medical Association, 316*(23), 2481–2482.

Spilker, B. (1984). *Guide to clinical studies and developing protocols*. Raven Press.

Sung, N. S., Crowley, W. F., Jr., Genel, M., Salber, P., Sandy, L., Sherwood, L. M., et al. (2003). Central challenges facing the national clinical research enterprise. *Journal of the American Medical Association, 289*(10), 1278–1287.

Tenenbaum, J. D., Avillach, P., Benham-Hutchins, M., Breitenstein, M. K., Crowgey, E. L., Hoffman, M. A., et al. (2016). An informatics research agenda to support precision medicine: Seven key areas. *Journal of the American Medical Informatics Association, 23*(4), 791–795.

Topaloglu, U., & Palchuk, M. B. (2018). Using a federated network of real-world data to optimize clinical trials operations. *JCO clinical cancer informatics*, (2), 1–10.

Topaz, M., Shafran-Topaz, L., & Bowles, K. H. (2013). ICD-9 to ICD-10: Evolution, revolution, and current debates in the United States. *Perspectives in Health Information Management/AHIMA, American Health Information Management Association, 10*(Spring).

Tse, T., Williams, R. J., & Zarin, D. A. (2009). Reporting "basic results" in ClinicalTrials. Gov. *Chest, 136*(1), 295–303.

Visweswaran, S., Becich, M. J., D'Itri, V. S., Sendro, E. R., MacFadden, D., Anderson, N. R., et al. (2018). Accrual to clinical trials (ACT): A clinical and translational science award consortium network. *JAMIA open, 1*(2), 147–152.

Weber, G. M., Murphy, S. N., McMurry, A. J., MacFadden, D., Nigrin, D. J., Churchill, S., & Kohane, I. S. (2009). The Shared Health Research Information Network (SHRINE): A prototype federated query tool for clinical data repositories. *Journal of the American Medical Informatics Association, 16*(5), 624–630.

Weng, C., & Kahn, M. (2016). Clinical research informatics for big data and precision medicine. *IMIA Yearbook*, 211–218.

Zarin, D. A., Tse, T., Williams, R. J., Califf, R. M., & Ide, N. C. (2011). The ClinicalTrials.gov results database—Update and key issues. *New England Journal of Medicine, 364*(9), 852–860.

Zhang, J., & Patel, V. L. (2006). Distributed cognition, representation, and affordance. *Pragmatics & Cognition, 14*(2), 333–341.

Precision Medicine and Informatics

Joshua C. Denny, Jessica D. Tenenbaum, and Matt Might

Contents

Joshua C. Denny completed this work while working at Vanderbilt University School of Medicine.

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_28

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What is precision medicine, and how does it differ from traditional medical practice?
- How can Electronic Health Records (EHRs) be used to advance precision medicine discovery?
- How to identify and evaluate phenotypes algorithms from EHRs?
- How are EHRs aiding in implementation of precision medicine?
- How are genomic data being used today in research, clinical care, and consumer health?
- What is Mendelian randomization?
- What are some of the large cohorts being used to advance precision medicine, and what is the importance of diversity in these cohorts to accelerate discovery?

28.1 What Is Precision Medicine?

Precision medicine is a field focused on understanding the role molecular, environmental, and phenomic variation play in healthcare with the goals of more rational therapeutics, improved understandings of prognosis, and more optimal healthcare delivery. Fundamentally, precision medicine focuses on data-driven optimization of health care. This includes more precise understanding of disease through the amassing of both large research studies and amalgamations of truly massive amounts of routinely collected healthcare information from electronic health records. In addition, precision medicine is leveraging new technologies such as sensor-based measurements and omic technologies. The most commonly used technology is genomic assays, such as genotyping (which uses probes to assay large numbers of specific variants) or genomic sequencing (which assess each base pair present within a region or across the genome). However, epigenomics, transcriptomics, proteomics, microbiomes, and metabolomics are also being used and hold promise for further research and clinical care in precision medicine. The rapid growth and availability of other technologies such as sensors and imaging data will also contribute. The ultimate goals are to use all of these data to better guide diagnosis, improve understandings of prognosis, optimize existing treatments, develop new therapies, and design novel prevention schemes (Fig. 28.1).

Precision medicine as a field is closely related to personalized medicine, P4 medicine, individualized medicine, genomic medicine,

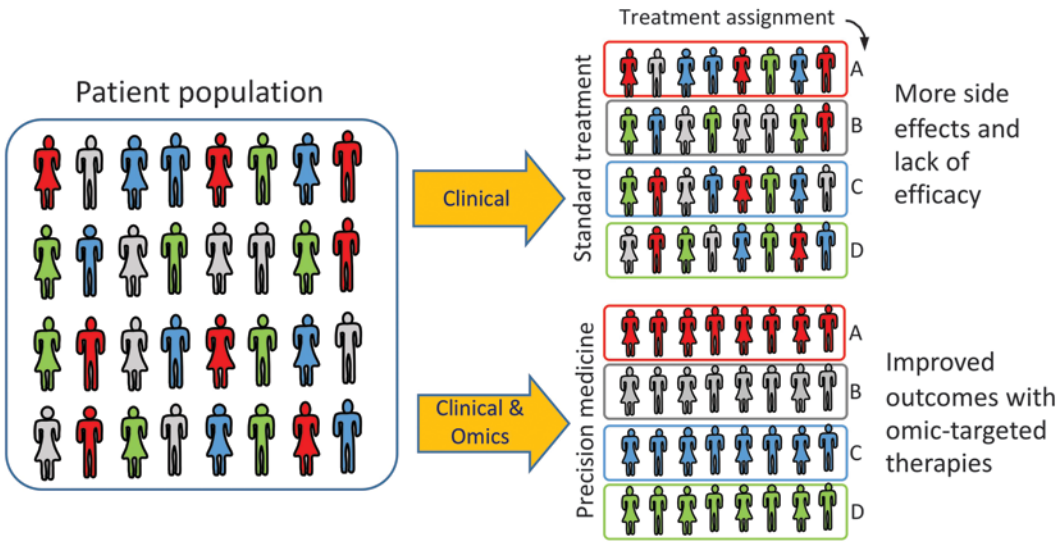


Fig. 28.1 Overview of the goals of Precision Medicine

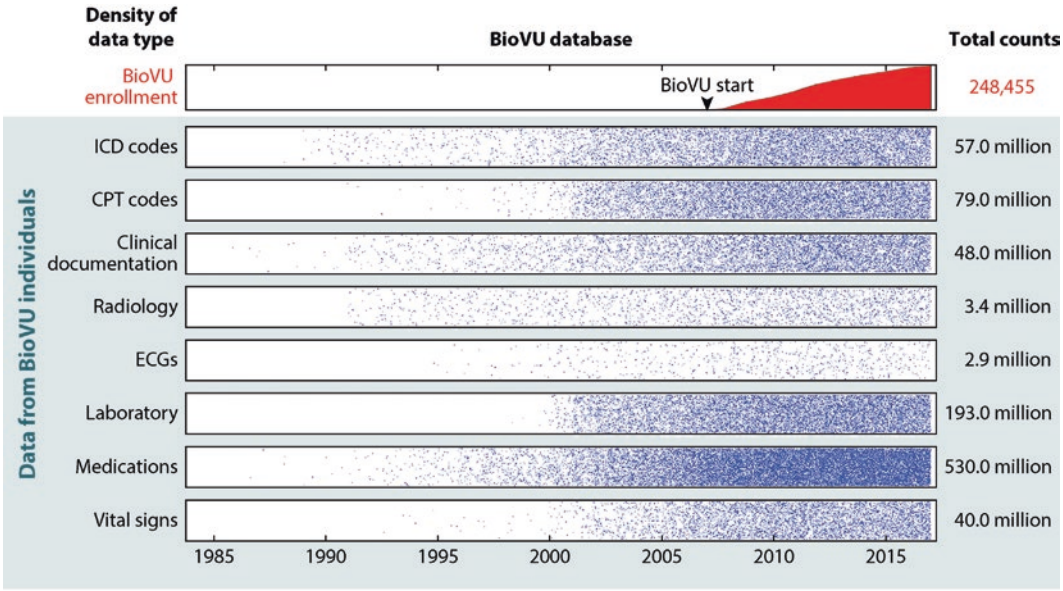
and other similar recent fields to all share as a goal improvement of health for both individual patients and populations through use of data. Indeed, in practice, most of these terms are used fairly interchangeably amongst most institutions focusing on discovery and implementation of precision medicine approaches. A key differentiator is that precision medicine focuses primarily on its goals through enhanced understandings and new taxonomies that lead to redefinition of the diseases themselves. Precision medicine shares with these other terms the centuries-old goal of individualizing care for individual patients, following Sir William Osler's maxim: "The good physician treats the disease; the great physician treats the patient who has the disease." The primary difference between now and then is the deluge of new modalities and quantities of information truly allowing us to leverage big data to solve problems of individual health in new ways. Such scale of huge data sets is needed, for instance, to untangle the contribution of rare genomic variation to the clinical impact on individual patients. For example, out of 6.4 billion base pairs in the (diploid) human genome, an average patient genome may have 4.1 million to 5 million genetic variants (1000 Genomes Project Consortium et al., 2015). Distilling the impact of these variants in such a way that clinicians can make use of this information is a substantial challenge for both rare and common disorders.

In this chapter, we will focus on the informatics resources for and implications of precision medicine. We will also provide a basic overview of precision medicine. Informatics is needed for the capture of data, transformation of data into information and knowledge, and – perhaps most importantly – implementation of precision medicine – acting on that new knowledge to improve the health of real people. Electronic health record (EHRs) are an important part of precision medicine both by providing a real world data source as well as a key modality for its implementation. Relevant other chapters to this chapter include ▶ Chaps. 9 and 26.

28.2 Using EHRs for Genomic Discovery

Electronic health records (EHRs) have been an important part of clinical care for decades, but over the last decade have become an increasing part of discovery to advance precision medicine. While retrospective epidemiological research has been performed using claims data for decades, the more recent use of electronic health records for molecular research, especially genomics, over the last decade has become transformational. EHRs contain a wealth of dense patient data that are valuable for discovery, and that would cost significant sums to reproduce in a research cohort (◘ Fig. 28.2). As such, they have become a valuable source of information for retrospective research (Robinson, Wei, Roden, & Denny, 2018).

The first successful use of electronic health records for genetic discovery was in 2010. Four papers published that year used EHR data as the sole phenotypic information to replicate known genetic effects (Denny, Ritchie, Basford, et al., 2010; Denny, Ritchie, Crawford, et al., 2010; Kullo, Ding, Jouni, Smith, & Chute, 2010; Ritchie et al., 2010). Ritchie et al. tested for the associations between 5 diseases and 21 SNPs that were known to be associated from prior literature and replicated all associations for which the study was adequately powered (Ritchie et al., 2010). The other studies added replications with endophenotypes of cardiac conduction and red blood cell traits. These studies included the first EHR-based **genome-wide association studies (GWAS;** ◘ Fig. 28.3a**)** – studies which test for association between a phenotype and hundreds of thousands to millions of single nucleotide polymorphisms (SNPs). GWAS is discussed more in ▶ Sect. 28.4.1. Importantly, these studies suggested a process for developing phenotype algorithms that identified both cases and controls as two separate groups. Each algorithm used a variety of types of EHR data to identify their populations with high positive predictive values (analytics for phenotype algorithms are



Robinson JR, et al. 2018.
Annu. Rev. Biomed. Data Sci. 1:69–92

Fig. 28.2 Density of phenotypic data in an EHR linked biobank. All rows represent data from patients that enrolled in BioVU, the Vanderbilt DNA biobank. The first patiented enroll ed in 2007. This figure demonstrates that individuals can have decades of extant EHR data prior to their enrollment, allowing both cross-sectional and *in silico* longitudinal studies. Each point represents real data transformed by square root divided by 20 of the actual count at that time period. (From Robinson et al., 2018)

discussed in more detail in ▶ Sect. 28.3 below). Another important early study validated known SNPs associated with rheumatoid arthritis using a validated EHR algorithm (Kurreeman et al., 2011). This study demonstrated the effect sizes from the EHR algorithms and from prior research studies were similar.

Another advance highlighted first in early EHR studies was the **phenome-wide association study (PheWAS; ◘ Fig. 28.3b),** which is discussed in ▶ Sect. 28.4.3 below. That the first systematic "PheWAS" was performed in EHRs was enabled by the broad collection of phenotypes found in EHRs that is essentially unrelated to an *a priori* research hypothesis (Denny, Ritchie, Basford, et al., 2010). PheWAS approaches have been applied in observational cohorts as well (Millard et al., 2015; Sarah A Pendergrass et al., 2013).

An important development in the use of EHRs for precision medicine was the National Human Genome Research Institute-funded **Electronic Medical Records and Genomics (eMERGE) network**, which began in 2007. eMERGE had the explicit goal of exploring the capabilities of EHRs for genomic discovery. The initial eMERGE network had five sites and has been renewed twice, now in its third iteration, with 10 sites. In subsequent iterations, eMERGE has grown from a primary focus of discovery to also include implementation of genomic medicine (Rasmussen-Torvik et al., 2014). The first novel discovery out of eMERGE was in 2011 by identifying variants in *FOXE1* associated with autoimmune hypothyroidism (some of the results from this study are shown **in** ◘ Fig. 28.3) (Denny et al., 2011). Since then, eMERGE sites have investigated nearly 100 different phenotypes, many with novel discoveries (Crawford et al., 2014).

28.3 Finding Research-Grade Phenotypes in EHRs

A characteristic of EHR phenotypes is the combination of multiple modalities of the

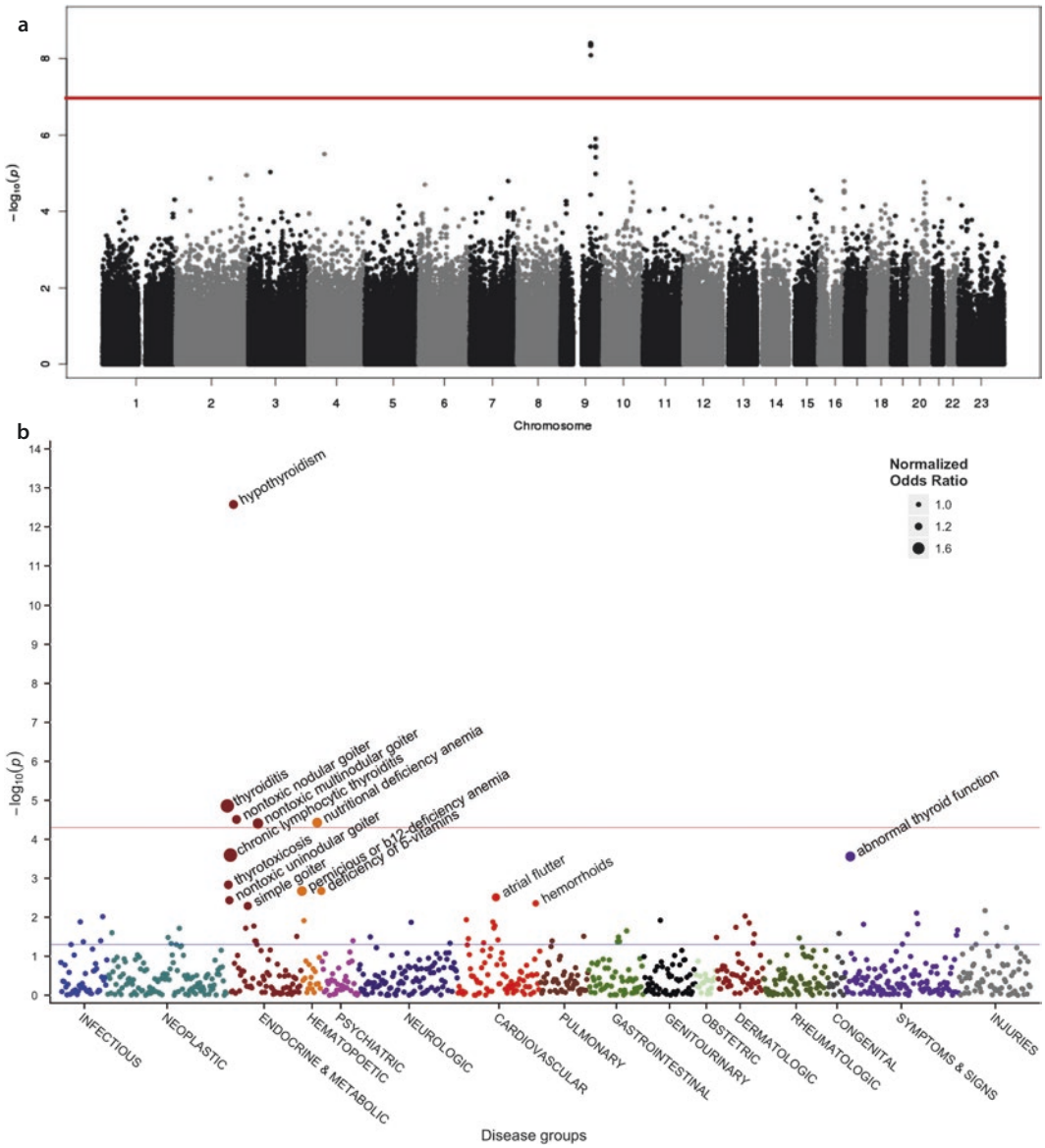


Fig. 28.3 Example GWAS and PheWAS using EHR data. **a** Manhattan plot from a GWAS of autoimmune hypothyroidism performed in the eMERGE Network. This GWAS identified variants in *FOXE1* as a risk factor for autoimmune hypothyroidisms, a novel finding at the time. **b** PheWAS of the variant identified in panel A. This PheWAS identified autoimmune thyroid disease associated with this variant and highlighted other conditions, like atrial flutter, that are inversely associated with hypothyroidism. (With permission from Denny et al. 2011 © Elsevier)

EHR information to identify high-quality research grade phenotypes. Most frequently, they include billing codes, medications, laboratory results, and some sort of text processing or natural language processing. Example phenotypes for autoimmune hypothyroidism and type 2 diabetes are shown in Fig. 28.4. Both of these were validated at more than one site and demonstrated successful genomic discovery or replication (Denny et al., 2011; Kho et al., 2012). Clearly, sufficient sample size (and consequently statistical power) is needed to identify associations. Thus, the eMERGE network has found it necessary to run pheno-

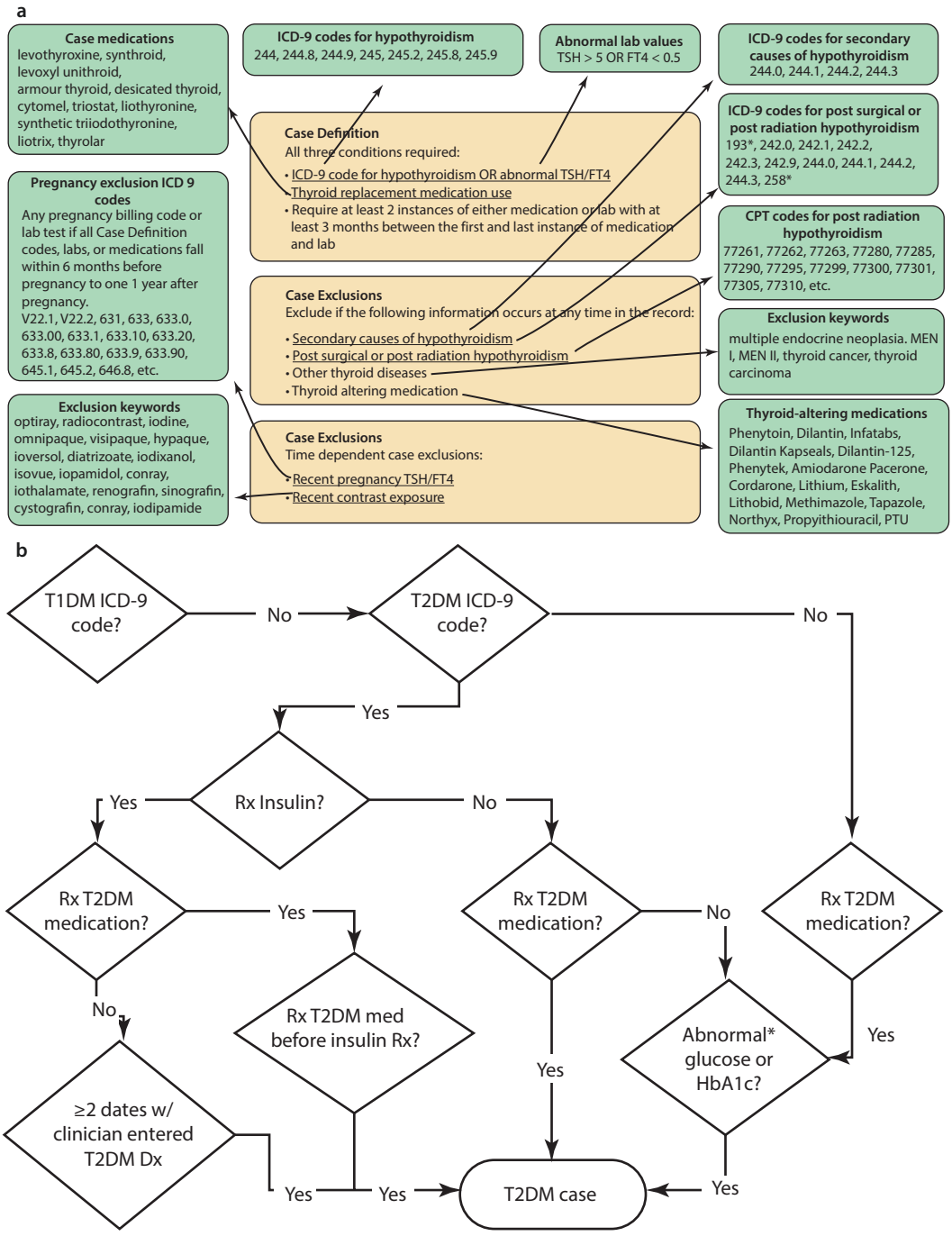


Fig. 28.4 EHR phenotype algorithms for Autoimmune Hypothyroidism and Type 2 Diabetes. Details for each of these algorithms are available on ► http://PheKB.org. Each algorithm was validated by manual chart review at multiple institutions. (Figures adapted from: **a** Conway et al., (2011) and **b** Kho et al., (2012))

type algorithms across different sites to increase sample size. In doing so, eMERGE found it helpful to collaboratively develop phenotypes across different sites given local variability in EHR systems, billing practices, and institutional practices. Algorithms from

eMERGE and other networks have been shared for use on ▶ PheKB.org (Kirby et al., 2016). Currently, PheKB houses more than 150 EHR-based algorithms.

There are many algorithmic approaches to creating high-quality phenotype algorithms. Perhaps the most common approach is through combinations of different elements via Boolean-logic, such as the algorithms depicted in Fig. 28.4. Other researchers have trained machine learning algorithms using a set of reviewed cases and controls. Among the first demonstration of this approach was using the Partners Biobank to find rheumatoid arthritis patients (Kurreeman et al., 2011; Liao et al., 2010). While algorithms using machine learning often can be overfit on a particular data set, research has shown that at least some machine learning approaches in EHR data can be portable across different EHR systems (Carroll et al., 2012). This machine learning approach has also been applied to Veteran Affairs EHRs from the Million Veteran Program to find cases of acute ischemic stroke (Imran et al., 2018).

A common theme among both rule-based and machine learning approaches is incorporation of different types of data (e.g., billing codes, laboratory data, note content, and medication data) from the EHR. Table 28.1 reviews the most common features used by algorithms posted in PheKB. In addition, rules requiring more than one instance of a given data type often improve algorithm performance as well. A study looking at 10 different diseases across ICD codes, clinical notes, and medications specific to diseases demonstrated that use of multiple modalities improves algorithm improves performance more than counting rules within a single data type (Wei et al., 2016). In this study, the average PPV of a single ICD code instance was only 0.37, but this increased to 0.84 when 2 or more instances of ICD codes were required. Requiring at least 2 different elements improved PPV to 0.91. Overall, notes were the most sensitive data type.

The overall process of defining and evaluating a phenotype algorithm is shown in Fig. 28.5. In evaluating a phenotype algorithm, it has become common to compare the algorithm to manual chart review as the "gold standard". Two typical approaches have been undertaken for this analysis (Newton et al., 2013). One is to use clinically-trained professionals to evaluate the patient records to determine whether or not the individual match the given clinical condition being assessed. The other approach is to develop a formal chart abstraction instrument that trained chart abstractors will review the chart to identify the elements to determine who it matches the case (or control) definition. This approach often proposes a set of rules for reviewers to validate in the patient record that determine if an individual meets a research-defined case definition. Regardless of the approach for chart review, the process is usually iterative: an investigator proposes an algorithm, executes it on a population, and then evaluates a set of charts. Especially if the chart review is using a trained professional instead of a chart abstraction instrument, it is best practice to include both individuals who match the algorithm

Table 28.1 Data modalities used in phenotyping algorithms available on PheKB

	Public (n = 44)	Nonpublic (n = 110)	Percent of total
ICD-9/-10 codes	39	73	73%
Medications	31	51	53%
CPT codes	23	44	44%
NLP	28	36	42%
Laboratory test results	21	37	38%
Vital signs	5	14	12%

Nonpublic algorithms include algorithms in development and those whose performance has not yet been validated. Data accessed Oct. 15, 2017

Abbreviations: *CPT* Current Procedural Terminology, *ICD-9/-10* International Classification of Diseases, Ninth Revision/Tenth Revision, *NLP* natural language processing, *PheKB* Phenotype Knowledgebase

From Robinson et al., (2018)

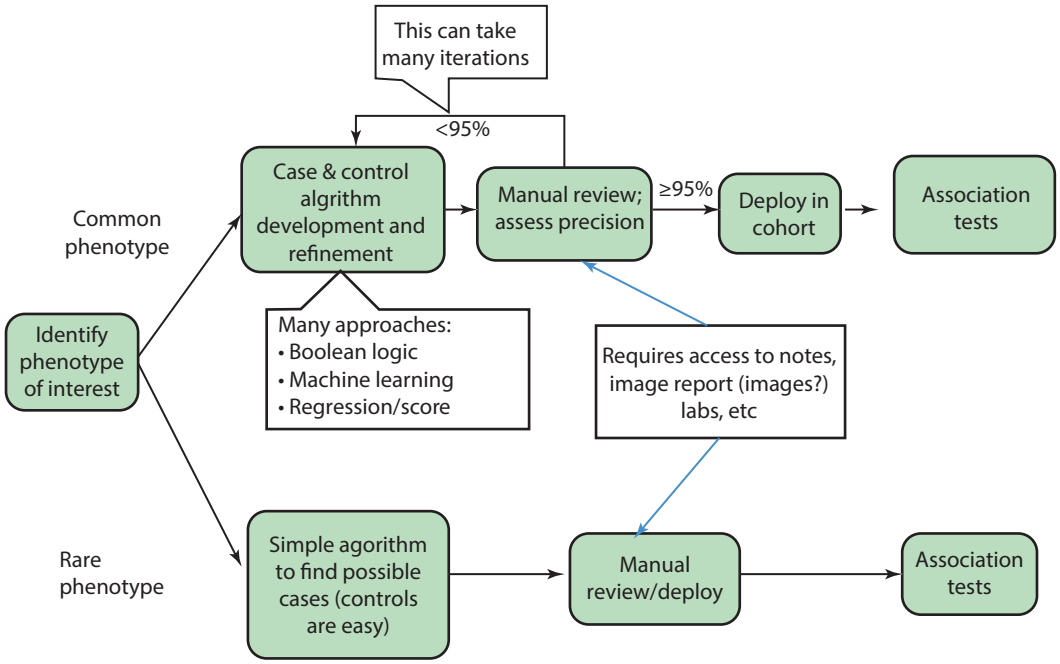


Fig. 28.5 Overview of the general approach to finding a phenotype in the EHR

and those who do not match the algorithm in the chart review to avoid anchoring bias. In this review, the order of the charts would be randomized and the reviewers blinded to the algorithm's determination of case (or control) status. After review, the investigators can calculate the **positive predictive value (PPV)** as the number of true positives divided by (the number of true positives plus the number of false positives). Additionally, researchers may want to calculate a **sensitivity** (or recall). Calculating recall for common diseases may be relatively straightforward but very challenging to assess for rare diseases. Thus, many reviewers often make simplifying assumptions about requirements to be a case. For example, many researchers make the assumption that a case must have at least one relevant ICD code to be a case (Carroll et al., 2012; Liao et al., 2010). This approach provides a reasonable estimate of sensitivity.

PPV is typically viewed as the most important metric performance for a case-control study, since it is usually more important to be sure to have high-quality cases and controls for a given analysis. For instance, review of phenotype algorithm implementations on ▶ PheKB.org demonstrated that out of 145 posted implementations (as of December 2018), 79 (54%) reported PPV and 33 (23%) reported recall. The condition in which sensitivity is more important is when the variable is being used as a covariate in an analysis.

To facility EHR based discovery, huge data sets are needed. While the growth through the early 2000's and this decade have often been focused on single EHR systems, there is an increasing need to combine data across sides for discovery. This is important both for the needs of amassing the necessary size of this size as well as representing diversity of geography, demographics, environmental exposures, and practice habits, which can vary between institutions and geography. In short, this effort has been facilitated through use of common data representations such as the Fast Health Interoperability Resource (FHIR) and common data models (CDMs). Most frequently used common data models have been the Informatics for Integrative Biology and the Bedside (i2b2) data model, PCORNet, and the OMOP data model. These are discussed in more detail in ▶ Chap. 25.

28.4 Omic Discovery Approaches

Over the last three decades, the explosion of efficient and relatively inexpensive dense molecular measures has led to the growth of more data-driven molecular investigations of traits and diseases. The most commonly investigated currently would be genomic investigations. Some of these are also finding translation into clinical practice (discussed in ▶ Sect. 28.7). The research observation of genetic pleiotropy (the condition in which one gene or genetic variant impacts multiple phenotypes) combined with more dense phenotypic assessments have led to similar hypothesis-free tests of association of the phenome. Specific technologies are discussed further in ▶ Chaps. (9 and 26).

28.4.1 Genome-Wide Association Studies (GWAS)

A GWAS systematically surveys polymorphisms across the genome to find variants associated with a trait or disease. Variants for a GWAS are typically weighted toward more common SNPs that can represent a broad range of genomic variation based on linkage disequilibrium, that is the nonrandom clustering of variation in the genome based on inheritance patterns. Thus, relatively small numbers of the >3 billion base pairs in the (haploid) human genome can represent a large fraction of inherited variation in the genome. Most GWAS assay >500,000 mostly SNPs common SNPs. More recent GWAS have incorporated rare variants, such as functional genomic variants known to be associated with disease and pharmacogenomic variants.

The first GWASs (◘ Fig. 28.3a) was conducted in 2005 and 2006 and discovered genetic variants associated with Age-related Macular Degeneration of ~100k SNPs (Dewan et al., 2006; Klein et al., 2005). The modern era of array-based GWAS approach with large case control populations identifying common variants influencing common disease was arguably introduced in a large scale by the 2007 by the Wellcome Trust Case Control Consortium, which successfully identified SNPs associated with 7 common diseases (Wellcome Trust Case Control Consortium, 2007). As mentioned above, the first GWAS using EHR information to define cases and controls was performed in 2010 (Denny, Ritchie, Crawford, et al., 2010; Kullo et al., 2010). Since then, the number of EHR-based GWAS and large consortia using EHR information has risen dramatically (Wei & Denny, 2015). The current inclusion of routine healthcare data in very large cohorts (see ▶ Sect. 28.6 below) has made reliance on healthcare data such as those from EHRs or administrative claims data now commonplace.

Since 2009, published GWAS have been curated and made available via the online GWAS Catalog, begun first by NHGRI and now hosted by EMBL-EBI (▶ https://www.ebi.ac.uk/gwas/) (MacArthur et al., 2017). By 2010, more than 500 GWASs had been performed on a wide variety of traits in common disease, and by the end of 2018, this had grown to 3675 publications reporting one of more GWASs, noting associations between 87,081 SNPs and phenotypes (Green, Guyer, & National Human Genome Research Institute, 2011; MacArthur et al., 2017).

28.4.2 Genomic Sequencing

The rapidly falling cost of genomic sequencing (to several hundred for a research whole exome and less than $1000 for a whole genome as of the end of 2018) is leading to a dramatic growth in use of genetic sequencing. The primary added benefit of genomic sequencing to precision medicine at the current time is a better and more detailed assessment of rare and very rare variants through a more comprehensive coverage of the genome. Sequencing approaches have enabled the discovery of novel variants for common disease and have been especially impactful for the uncovering of variants in rare disease. Sequencing is routinely used now clinically to aid cancer care or diagnose rare genetic diseases. In research, sequencing is rapidly expanding our ability to discover associations with rare conditions. The NIH's Undiagnosed Disease Network, for instance, routinely

employs whole exome sequencing (WES) or whole genome sequencing (WGS) to diagnose individuals. As a notable win for sequencing, the UDN has been able to diagnose 35% of individuals referred into their network, 74% of which were made with the addition of genomic sequencing to comprehensive clinical phenotyping (Splinter et al., 2018). In addition, they have defined 31 new syndromes through their comprehensive clinical and molecular assessments of undiagnosed patients.

28.4.3 Phenome-Wide Association Studies (PheWAS)

Growth of EHR based cohorts provided rich and diverse phenotype data to complement biologic data. Whereas GWASs provided a way to assess genomic accession associations in a hypothesis-free manner starting around 2005, GWAS usually assesses only one phenotype at a time. However, the growth of GWAS quickly highlighted the occurrence of genetic pleiotropy – the condition in which one gene influences multiple independent phenotypes. Thus, the rich collection of diverse phenotype information in EHRs and other growing cohorts provided the ability to simultaneously access phenotype associations in the same scanning hypothesis free manner as GWAS. The first PheWAS EHR-based aggregated billing codes into 744 PheWAS "cases" (Denny, Ritchie, Basford, et al., 2010). Each case was linked to a control group. After identification of case and control groups, a PheWAS (Fig. 28.3b) is essentially a pairwise test of all phenotypes against an independent variable, such as a genetic variant or laboratory value. For a genetic variant, PheWAS is analogous to genetic association tests performed in a GWAS, with a typical approach employing a logistic regression adjusted for demographic and genetic variables, such as genetic ancestry. The first PheWAS tested seven known SNP-disease associations, replicating four and suggested a couple of new associations. Newer approaches to PheWAS have leveraged increased density of phenotypes from the EHR, which current methods mapping ICD9 and ICD10 codes into >1800 phenotype case groups. A 2013 study shows that this approach was able to replicate 66% of adequately-powered SNP-phenotype pairs, and also identified several new associations that were replicated (Denny et al., 2013). A catalog of some of the PheWAS associations found to date is available at ▶ http://phewascatalog.org. A PheWAS of all phenotypes available in the UK Biobank has also been performed (▶ http://www.nealelab.is/uk-biobank).

PheWAS can essentially be performed on any broad collection of phenotypes. Researchers have used raw unaggregated ICD codes, other aggregation systems of ICD codes, or phenotypes collected from observational cohorts (Hebbring et al., 2013; Pathak, Kiefer, Bielinski, & Chute, 2012; S A Pendergrass et al., 2011). The disadvantage to using more granular ICD codes is the increased number of hypotheses being tested, which hinders the statistical power to detect a result. Lack of ICD code aggregation can also introduce variability in coding practices that decreases sample size for a given phenotype, such as the number of specific diagnostic codes available to represent common conditions and their complications, such as diabetes mellitus subtypes (e.g., with specific codes for controlled or uncontrolled glucose status and its resulting cardiovascular, renal, or neurological complications) or gout (e.g., chronic or acute, with or without tophi, etc).

PheWAS can quickly highlight potential pleiotropy of a given genetic variant or other independent variable by analyzing for associations with multiple phenotypes within a single population, one can test the independence of the potential pleiotropic findings with subsequent conditioned analyses. Other advantages of PheWAS is that they are quick to perform and easily implemented through existing R packages (▶ https://github.com/PheWAS/PheWAS) or iteration through common statistical packages. A disadvantage of PheWAS is that its phenotypes can be coarse and can have both lower sensitivity and PPV than custom phenotype algorithms as discussed in ▶ Sect. 28.3. Fortunately, these types of bias typically biased towards the null. Associations

found via PheWAS can require refinement and subsequent validation.

28.4.4 Other Omic Investigations

In addition to genomics, the growth of a number of other omic approaches are providing greater insight into an individual's environment, endophenotypes, and molecular measures. Some of these include the microbiome, proteome, metabolome, and other bioassays. Additional dense phenotypic and environmental assessments include dense measures of the environment and personal sensor-based technologies, such as consumer activity monitors. Publicly available datasets providing detailed measures of pollution, the built environment, weather patterns, availability of quality food or greenspace, and sociodemographic factors are available for linkage via geolocation, linked via smartphones and other devices that continuously track geolocation. These devices can also measure activity and heart rate to provide greater insight into a person's habits and physiological factors. Today, the clinical impact of many of these measures is not yet known. However, there growing ubiquity through both research and commercial interests are enabling deeper investigation into their clinical impact. They are also being included in large research cohorts (see ► Sect. 28.6).

28.5 Approaches to Using Dense Genomic and Phenomic Data for Discovery

28.5.1 Combining Genotypes and Phenotypes as Risk Scores

Most genetic variants discovered via GWAS have had relatively mild effect sizes for their phenotype of interest. However, the size of modern GWAS, now involving hundreds of thousands of individuals for more common traits, have allowed identification of many independent genetic loci, sometimes reaching into the hundreds of distinct loci (Locke et al., 2015; Okada et al., 2014; Wood et al., 2014). Collectively, these genetic variants can explain a much larger percentage of the variance in disease risk than the individual risk variants, even when the effect sizes of many of the individual variants may be rather small (e.g., having odds ratios of ~1.01). As a tool, researchers have aggregated genetic risk variants into a calculated score (called a **"genetic risk score", GRS, or "polygenic risk score", PRS**), typically as a sum of the presence of the variant multiplied by a weight, often taken from a regression analysis. These risk scores need to account linkage disequilibrium to find independent loci and may also produce a weighted model using penalized regression. A simple approach can be given as:

$$GRS = \sum_{i=1}^{k} w_i N_i$$

where w_i is the weight for the variant (e.g., the log odds ratio from a logistic regression) and N_i is the number of risk alleles for that variant (typically, 0,1, or 2). The clinical advantage of a GRS is that it provides a way to evaluate the aggregate risk of an individual having a given disease that takes into account many typically small risk factors.

For instance, consider breast cancer genetic testing. It has long been recognized that variants in *BRCA1* and *BRCA2* confer significant increased risk of breast cancer to carriers of these mutations. While pathogenic variants in these genes do confer a large risk of breast cancer (lifetime risk of 45–65%), the vast majority of breast cancer is not related to these variants, since they are present in <1% of the general population (Torkamani, Wineinger, & Topol, 2018). However, common SNPs from a large breast cancer GWAS published in 2017 represents about 41% of the familial risk of breast cancer (Michailidou et al., 2017). Studies with cardiovascular diseases have found similar results and potential clinical utility for PRS. In a study looking at 5 prospectively-followed cardiovascular cohorts with genetic testing found that polygenic risk scores and lifestyle factors were independently associated with incident cardiovascular events (Khera et al., 2016). Moreover, their study

identified that patients with high genetic risk of cardiovascular disease but healthy lifestyles were at similar risk to those with unhealthy lifestyles but low genetic risk. Importantly, healthy lifestyles decreased cardiovascular risk at any genetic risk threshold, suggesting the importance of potential preventative lifestyle modifications and therapies in those individuals at high genetic risk.

A more recently introduced approach is to do a similar process with phenotypes in a **phenotype risk score (PheRS)** (Bastarache et al., 2018). In the initial demonstration of PheRS, ICD codes were mapped to phecodes and summed weighted based on the inverse log of the frequency of the phecode in the EHR:

$$PheRS = \sum_{p=1}^{m} w_p x_{i,p}$$

where:

$x_{i,p}$ = 1 if $individual_i$ has $phenotype_p$ or 0 otherwise

$$w_p = \log \frac{N}{n_p}$$

where n_p is the number of individuals with phenotype p. By aggregating phenotypes in a similar way as genotypes, a combined score can increase the sensitivity to detect the phenotypic impact of a genetic variant. For example, the PheRS for cystic fibrosis includes component phenotypes such as bronchiectasis, pneumonia, infertility, and asthma. The disease code itself ("cystic fibrosis") is not part of the PheRS. Based on EHR weighting, bronchiectasis has a much higher weight than asthma since asthma is much more common. This approach was used in an initial demonstration exercise looking at ~1200 Mendelian diseases that could be tested in an EHR population. In this study, PheRS was able to identify diagnosed genetic diseases in the EHR using the component phenotypes of disease and was also able to be used to identify novel pathogenic variants for undiagnosed conditions.

28.5.2 Mendelian Randomization

Mendelian randomization (MR) is a technique used to provide evidence for the causality of a biomarker on a disease state in conditions in which randomized controlled trials are difficult or too expensive to pursue. For example, low density lipoprotein (LDL) and high-density lipoprotein (HDL) levels have long been associated with myocardial infarctions in observational cohorts, but it is unclear whether they are markers or causal: perhaps these levels are a marker of diet, activity level, or other unknown factors. It is essentially impossible (and probably unethical) to perform a randomized control trial that alters someone's LDL or HDL levels in isolation. However, a number of genetic variants have been found that alter LDL and HDL levels. Since alleles are randomly distributed to ova or sperm during meiosis, studying the impact of biomarker-influencing alleles provides a naturally occurring randomization of the risk factor. Genetic variants are generally not associated with behavioral, social, and some physiological factors – reducing confounding. Thus, by studying the impact on the clinical outcome of the variants associated with the biomarker, one can assert causality of the biomarker to the outcome of the variant. MR has proven a powerful tool in recent years (Fig. 28.6). MR studies have demonstrated clear associations between LDL and triglyceride levels and cardiovascular disease while casting doubt on the role of HDL in protecting against cardiovascular disease (Holmes et al., 2015; Voight et al., 2012). The latter is particularly interesting as cholesteryl ester transfer protein (CETP) inhibitors, medications targeted to raise HDL, have so far not been successful at reducing mortality (Mohammadpour & Akhlaghi, 2013).

By providing an approach to assess causality, MR can also provide an approach to investigate potential drug effects. An MR approach demonstrated the the lipid-lowering agent ezetimibe would reduce cardiovascular disease (by studying the clinical impact of

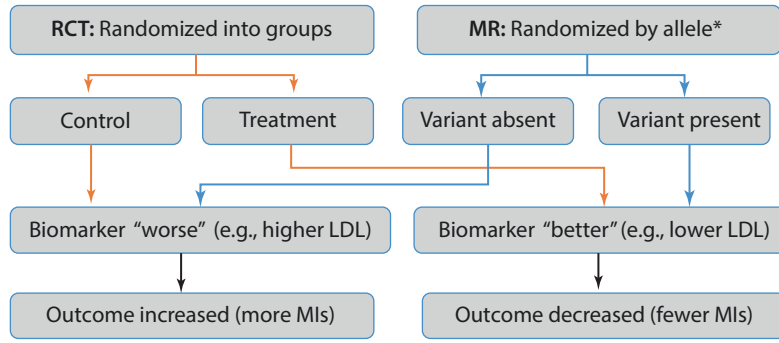


Fig. 28.6 Mendelian Randomization (MR) vs. Randomized Controlled Trials (RCT). MI myocardial infarction, *LDL* low density lipoprotein levels. *Allele could represent a single SNP or group of SNPs (e.g., combined via a genetic risk score)

genetic variants mimicking its effect) prior to a randomized controlled trial demonstrated this effect (Cannon et al., 2015; Myocardial Infarction Genetics Consortium Investigators et al., 2014). Similarly, MR has been used to show that diabetes is a potential concern for PCSK9 inhibitors and combined with PheWAS to highlight potential unanticipated side effects (Jerome et al., 2018; Schmidt et al., 2017).

28.5.3 Using Dense Data-Driven Measures to "Redefine" Disease

There is increasing enthusiasm that precision medicine will lead to new ways of defining disease and selecting treatments that will identify more rational therapeutic choices, improve our understanding of prognosis, and result in more effective disease screening. One example is cystic fibrosis pharmacotherapy, for which medications have been developed to target defects corresponding with specific genetic variants (O'Reilly & Elphick, 2013). These targeted therapies have dramatic influence on the disease course. There is a hope that similar approaches could be found for many common diseases aiding in drug selection and risk stratification for diseases such as depression, diabetes, hypertension, heart disease, and many other common diseases. Recent studies using clinical and molecular information have suggested subtypes of type 2 diabetes, heart failure, and autism (Ahmad et al., 2014; Doshi-Velez, Ge, & Kohane, 2014; Li et al., 2015); however, the clinical impact of such subtypes is unclear. Two examples of where targeted treatment for disease is currently being utilized in practice are pharmacogenomics and oncology, which are discussed in more detail in ▶ Sect. 28.7.

28.5.4 Use of Machine Learning and Artificial Intelligence to Advance Precision Medicine

The focus of this section is largely on the application of these methods to advance precision medicine. Other discussions on machine learning and artificial intelligence appear in ▶ Chap. 9.

Machine learning falls into two major classes of approaches: supervised and unsupervised, with the ability to apply many different algorithms. **Supervised machine learning** approaches tasks use a gold standard set as input to learn classifiers designed to optimally mimic the training set. **Unsupervised machine learning** learn patterns from the data without labeled training sets. Machine learning has the potential to augment any classification task, and has long been used with clinical data. Machine learning has been used for many tasks in EHRs (such as in natural language processing (Jiang et al., 2011; Y. Wu et al., 2017)) and in bioinformatics, such as aiding in interpretation of genetic variants (Kircher et al., 2014). Some of these use cases are referenced above, such as the learning of phenotype classifiers using EHR data and labeled cases or controls (Carroll, Eyler, & Denny, 2011; Liao et al., 2010; Lin et al., 2015; Peissig et al., 2014).

Recent areas of exploration in machine learning have seen a rapid rise in deep learning approaches. These take massive datasets and multi-layered neural networks to learn patterns in data that have proven superior to other machine learning techniques. They typically require very large data sets that have previously been unavailable in healthcare or biomedical research. However, the recent rapid growth of available EHR, genomic, and imaging data sources is enabling a new potential for machine learning to be applied to these data sets as well. Recent examples include training algorithms to identify malignant skin lesions and diabetic retinopathy from retinal scan (Esteva et al., 2017; Gulshan et al., 2016). These algorithms can present with performance that equals that of trained physicians at times. A challenge of these approaches is that they required huge data sets: the Gulshan et al. algorithm which used nearly 130,000 labeled retinal digital images to train its diabetic retinopathy algorithm (Gulshan et al., 2016). The growing availability of large-scale public biomedical data through cohorts such as those mentioned in ▶ Sect. 28.6 represent an important opportunity to accelerate such research. In addition, a number of for-profit companies, such as Alphabet, IBM, and many startups, have formed partnerships with diverse clinical entities from individual healthcare systems to the United Kingdom's National Health Service (Saria, Butte, & Sheikh, 2018).

28.6 Large Cohorts to Advance Precision Medicine Discovery

Clinical care since the 1960s has been dramatically influenced by observational cohort studies. Studies such as the Framingham Heart Study, the Nurses Health Study, or National Health and Nutrition Examination Survey (NHANES) have produced dramatic insights that have fundamentally changed our understanding of modifiable risk factors for many diseases. For instance, the Framingham Heart Study taught us that blood pressure, cholesterol, smoking, and activity contribute to cardiovascular disease risk. Its initial discoveries were derived from detailed longitudinal assessment of just under 5000 individuals. Most of these epidemiological cohorts have largely focused on answering exposure- or disease-focused questions. Two developments beginning in the early 2000s have brought new data resources to advanced discovery in more disease-neutral fashion. One has been the growth of large national-scale cohorts containing diverse phenotypic information connected with biosamples, and the other has been the growth of incorporation of routinely collected healthcare data linked to biological specimens. Example cohorts include the UK Biobank, the Million Veteran Program, the *All of Us* Research Program, and the China Kadoorie Biobank (◘ Table 28.2). Collectively, these cohorts will enroll millions of individuals across the world for longitudinal assessment of healthcare outcomes analyzed against molecular and environmental exposures. Each of these cohorts includes both participant-generated survey data and healthcare-derived data linked with biospecimens. Several of these cohorts also include the ability to recontact individuals. Collectively, these multiple complementary avenues of phenotype assessment augment passive phenotype collection (e.g., with EHR and claims-type data) with participant-provided information and the potential for reassessment with deeper phenotyping along topics of interest. Within *All of Us*, the participant provided survey information and the in-person research protocol physical measures are both being incorporated into the OMOP Common Data Model to simplify comparison of these data modalities. Digital engagement through email or websites or the collection of healthcare information enable cost-efficient follow-up for healthcare outcomes over long periods of time. Some of these larger resources are also pioneering newer models of researcher access that facilitate broad researcher communities to access the environments. An important aspect for these cohorts is the diversity of its participants, which is discussed in the next section.

Table 28.2 Selected biobanks and cohorts enabling precision medicine

Biobank	Region	Start year	Size	Website
eMERGE	U.S.	2007	143,896	► gwas.net
BioVU	U.S.	2007	~250,000	► victr.vanderbilt.edu
UK Biobank	U.K.	2006	512,000	► ukbiobank.ac.uk
Million Veteran Program	U.S.	2011	>600,000 Goal: 1 million	► www.research.va.gov/MVP/default.cfm
Kaiser Permanente Biobank	U.S.	2009	240,000	► www.rpgeh.kaiser.org
China Kadoorie Biobank	China	2004	510,000	► ckbiobank.org
All of Us Research Program	U.S.	2017	>80,000 Goal: 1 million or more	► joinallofus.org, researchallofus.org
Taiwan Biobank	Taiwan	2005	86,695 Goal: 200,000	► www.twbiobank.org.tw
Geisinger MyCode	U.S.	2007	>190,000	► www.geisinger.org/mycode

Limited to cohorts exceeding 100,000 individuals with biosamples. Sizes reported are as of 11/2018
eMERGE The Electronic Medical Records and Genomics Network

28.6.1 Need for Diversity, and Role of Precision Medicine in Health Disparities

Health disparities are abundant in health care. The same concerns can be said for precision medicine, for which variabilities in health insurance coverage, access to care, and financial situations may alter availability and accessibility for precision therapies (Bentley, Callier, & Rotimi, 2017). However, it is also true that precision medicine has the potential to identify and help alleviate some health disparities. Since genetic variants vary by ancestry, genetic testing has the opportunity to identify those most at risk for adverse events based not just on ancestry but on actual carriage of variants. Moreover, drugs traditionally have not been tested in all diverse populations and risk factors may not always be identified reach population. For instance, individuals of Asian ancestry are at much greater risk for severe skin reactions such as Stevens-Johnson syndrome from antiepileptics such as carbamazepine (Phillips et al., 2018). Similarly, it has been noted that carriage of *CYP2C19* loss of function of alleles is much more common in individuals of Pacific Island descent (Kaneko et al., 1999). Since diverse ancestries are often not tested in large numbers in clinical trials, the increased risks in diverse populations are not necessarily noticed. However, genomic testing would identify those at greater risk of adverse events thus identifying the opportunities to optimize care. The specific association of clopidogrel and reduced efficacy in individuals of Pacific Island descent was a subject of a lawsuit (A. H. Wu, White, Oh, & Burchard, 2015).

Unfortunately, the vast majority of individuals who have been genotyped or sequenced to date are of European ancestry. For instance, a 2016 study noted that 81 percent of all individuals who had undergone GWAS at that time were of European ancestry, and only ~4% represented African, Hispanic, or native

ancestries (Popejoy & Fullerton, 2016). Those latter populations represent about one-third of the current US population. A lack of diversity in genetic testing results in a lack of knowledge of the genetic architecture for diverse populations. For instance, variance in warfarin sensitivity vary by ancestry such that the variants needed to accurately guide prescribing for European and African ancestry are different (Perera et al., 2013; Ramirez et al., 2012). Moreover, it is known that individuals of African ancestry typically require higher doses of warfarin. However, most of the warfarin pharmacovariants that have been identified actually increase sensitivity to warfarin rather than reducing it.

The lack of diversity genotype populations affects not only our ability to adequately treat individuals with diverse ancestries, it also hinders discovery. For instance, the discovery of rare *PCSK9* loss-of-function variants as a drug target for cholesterol and cardiovascular disease was discovered in African Americans (Cohen, Boerwinkle, Mosley, & Hobbs, 2006). These loss of function variants led to production of monoclonal antibodies against PCSK9 that dramatically reduce cholesterol levels – and will treat individuals of essentially any ancestry (Sabatine et al., 2017).

28.7 Implementation of Precision Medicine in Clinical Practice

Currently, most efforts in precision medicine implementation focus on genomics. This comes in three main flavors: germline genomic changes to better tailor drug prescribing, diagnosing genetic disease, and identification of somatic variants to guide cancer therapy. A number of networks have been funded by the NIH to support these integration of genomic medicine into clinical care. They include the Implementing Genomics Into Practice (IGNITE) network, Electronic Medical Records and Genomics (eMERGE) Network, Clinical Sequencing Evidence-Generating Research (CSER) Network, the Pharmacogenomics Research Network (PGRN), and the Newborn Sequencing In Genomic medicine and public HealTh (NSIGHT) Network (■ Table 28.3).

Cancer genomic testing Perhaps the most widespread use of precision medicine currently is for somatic variation to target cancer therapies. Cancer therapies have long recognized the contribution of genetic variation to prognosis, starting with clinical karyotyping. One of the earliest applications of truly targeted therapy started with identification of the Philadelphia chromosome/translocation, which generates a fusion gene product BCR-ABL1. BCR-ABL1 results in the tyrosine kinase Abl being constitutively activated and is a marker for acute lymphoblastic leukemia and chronic myeloid leukemia. It's particular relevance to targeted therapy was noted in the 1990s when imatinib was identified through high throughput screening assays of tyrosine kinase inhibitors. Randomized controlled trials demonstrated a survival benefit on patients with chronic myelogenous leukemia (CML), thus leading to targeted therapies for individuals positive for this translocation.

The use of genetic changes to guide cancer therapy are proliferating rapidly. The growth of next generation sequencing of cancer patients has resulted in discovery of a number of mutations that have been successfully targeted for therapeutics. Examples include variants in *BRAF* for melanoma; *EGFR*, *ALK*, *ROS1*, and others for lung cancer; and many others. Hallmarks of genetically-focused therapies are applicability to smaller populations and a potential for fewer side effects compared to traditional chemotherapy. However, they also tend to be more expensive (Tannock & Hickman, 2016).

Given the focused care and workout for cancer patients, typical treatment for these individuals with cancers that have available genetically-targeted therapies is to clinically sequence tumor samples. These reports typically come in the form of PDFs; however, this is not a major impediment to accurate clinical care since it is a focus work up guided by professionals very knowledgeable in the field.

Table 28.3 Example projects exploring genetic medicine implementation

Program	Region	Website	Comments
eMERGE	U.S.	▶ gwas.net	Pharmacogenomics (PGx) and actionable Mendelian variants (AMV) for ~34 k
IGNITE	U.S.	▶ ignite-genomics.org/	Research demonstration projects exploring family medical history, PGx, *APOL1* variants
Alabama Genomic Health Initiative	U.S.	▶ www.uabmedicine.org/aghi	Community-based with GWAS-based AMV
Undiagnosed Disease Network	U.S.	▶ undiagnosed.hms.harvard.edu/	WGS, phenotyping for undiagnosed patients
Genomics England	U.K.	▶ www.genomicsengland.co.uk/	WGS for rare disease and cancer for 100 k
Thailand	SE Asia		Proactive genotyping for SJS/TEN risk alleles in carbamazepine-exposed patients
Sanford	U.S.	▶ imagenetics.sanfordhealth.org/	PGx and AMV among primary care population
All of Us Research Program	U.S.	▶ joinallofus.org	Stated goal of PGx and AMV for >1 million
Geisinger MyCode	U.S.	▶ www.geisinger.org/mycode	AMV; about 190 k enrolled

eMERGE The Electronic Medical Records and Genomics Network, *IGNITE* Implementing Genomics into Practice

28

Germline pharmacogenomics Medications have variable efficacy and potentials for adverse effects based on three major modes of action: altered metabolism, on-target side effects, or off-target side effects, each of which can result from a **drug-genome interaction** (See ▶ Chap. 26, ▶ Sect. 26.5 for more details.). A common scenario for altered metabolism resulting in lack of efficacy would be if a drug is a **prodrug**, meaning that the drug that is administered requires activation *in vivo* (typically by enzymes) into its active form. For example, clopidogrel is a prodrug that requires activation from CY2C19 to its active form 2-oxoclopidogrel (Scott et al., 2013). Thus, people with poor metabolizing variants of CYP2C19 are more likely to experience a lack of clopidogrel efficacy and be at higher risk of myocardial infarctions, need for revascularization, stroke, and death (Delaney et al., 2012). Similarly, decreased metabolism of thiopurines (e.g., azathioprine) due to *TPMT* polymorphisms can result in excessive bone marrow suppression (Relling et al., 2019). Second, drugs can produce adverse effects through off-target effects, such as an allergic reaction via an interaction with the immune system. Examples here include severe skin reactions from drugs such as carbamazepine and abacavir, which can be predicted by certain human leukocyte antigen variants (White et al., 2018). Third, drugs can have toxicity from on-target effects, such as increased sensitivity to warfarin resulting in an increased risk of bleeding with higher dose.

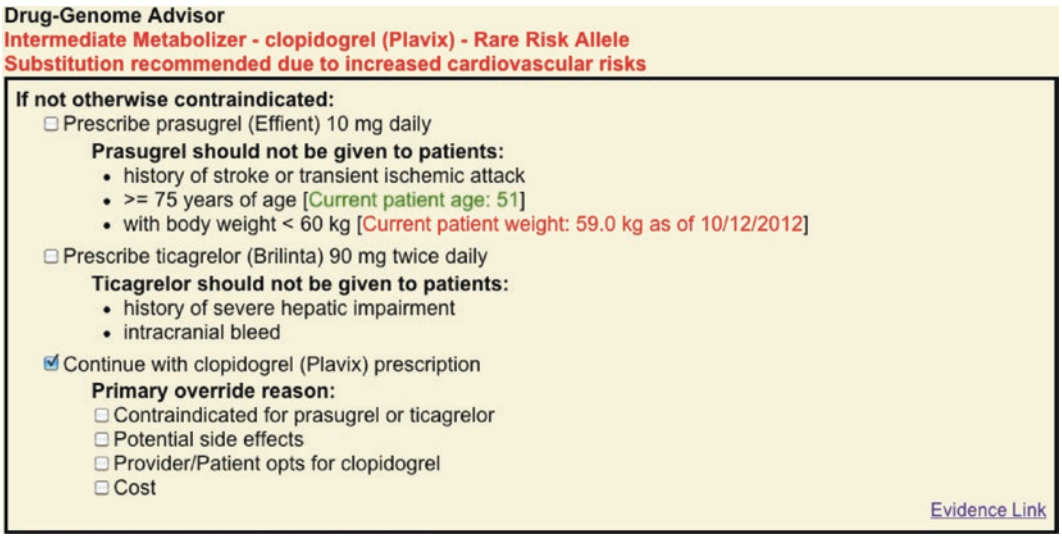
Germline pharmacogenomics holds the promise of tailoring medications to an individual's makeup to enable the "right drug for the right person" based on understanding of these effects. Unlike cancer genetic testing, pharmacogenetics requires a provider to potentially alter drug prescribing based on understanding of one's genotype. To allow for pharmacogenetics to work, the system must be able to intercept a drug order and provide

guidance. Drug-genome interactions could be accepted either by a computerized system of decision support (see ▶ Chap. 24) or via a human mechanism, e.g., via pharmacists. For decision support to work, the EHR requires a structured understanding of one's genotype, a clinical decision support system that can support action ability based on both a drug order and genotype.

Pharmacogenomic testing can be ordered in either a preemptive or reactive fashion. In a preemptive fashion, an individual has pharmacogenetic testing prior to drug prescribing. Then, when a medication would be prescribed that may be altered by one's genetic makeup, the system can intercept the order and recommend a genetically-tailored medication at the time of the prescribing event, such as the decision support alert in ◘ Fig. 28.7. This sort of genetic testing has been deployed at Vanderbilt, University of Chicago, and Indiana's INGENIOUS trial (Eadon et al., 2016; O'Donnell et al., 2012; Pulley et al., 2012). Further investigation of this approach is underway within the IGNITE Network.

Reactive pharmacogenetic testing is the more common approach to genetic testing and involves testing an individual when there is a specific indication for that test. Research has shown that having genetic testing available at the time of the prescribing event results in a higher frequency of genetically-tailored prescriptions (Peterson et al., 2016). Many genetic tests can take several days or more to receive results back for actionability, which may require a provider to recontact a patient to make a therapy change.

Genomics for disease diagnosis and risk assessment Clinical genetic testing has often occurred within the presence of specialized clinic visits with geneticists or genetic counselors, most commonly for prenatal screening or diagnosis of suspected genetic disease. These types of interactions typically require very little direct informatics support and results can be delivered effectively via send-out paper lab results. However, newer approaches underlying broaden understanding of individual disease risks based on genetics require greater intervention from informatics systems. While clinical use of genetic testing for common disease risk (such as through PRS as discussed above in ▶ Sect. 28.5.1) is uncommon in clinical care now, the explosion of genetic knowledge envisions a day in which people could clinically implement genetic risk to enhance their understanding of their degree of genetic risk for a disease. Understandings of genetic risk for disease is already implied through resource through direct-to-consumer genetic testing, discussed in the next section.



◘ **Fig. 28.7** Screenshot of clinical decision support advisor for Clopidogrel pharmacogenetic advice

28.8 Sequencing Early in Life

One crucial complication in the search for genomic explanations for any given disease or phenotype is the impact of environmental interactions. Over time, every person on earth is exposed to environmental factors that may differ based not only on a factory that disposes of industrial waste near a drinking water supply or the traffic on the street they grew up on, but also by the foods they eat, the climates in which they live, and the infections they have harbored. Those external variables, hard to control for and sometimes even to know, can have major effects on the downstream products and activities of one's genomic fingerprint. Early in life, however, those effects are less pronounced. Of course, the impact of the in-utero environment on the well-being of the developing fetus is well established. But a genetic defect is much more likely to be the cause in a newborn with an unidentified disease than in an adult patient who has undergone a lifetime of environmental insults. In this vein, a number of initiatives have been established across the US to offer clinical sequencing services for young patients, including programs at Children's Hospital of Philadelphia, Duke University, Partners Healthcare, the Baylor College of Medicine, and the Medical College of Wisconsin. More controversial on paper, and not yet being performed in practice, is prenatal genome sequencing. Ethicists are exploring the potential implications of this possible direction (Donley, Hull, & Berkman, 2012).

Addressing the time and resources needed to perform genome interpretation, one striking success story was achieved at Children's Mercy Hospitals and Clinics in Kansas City, MO (Saunders et al., 2012). Investigators used an Illumina HiSeq 2500 machine and an internally-developed automated analysis pipeline to perform whole-genome sequencing and make a differential diagnosis for genetic disorders in under 50 h. The diagnoses in question are among the ~3500 known monogenetic disorders that have been characterized. In this case, WGS is not being used to identify novel, previously unknown mutations. Rather, it is shortening the path to diagnosis to just over 2 days instead of the more traditional 4–6 weeks as a battery of tests were performed sequentially.

We offer one final example in which genome sequencing was used as a last resort in a medical odyssey to identify the cause of a mysterious bowel condition in a 4-year-old boy named Nicholas Volker (Worthey et al., 2011). Having ruled out every diagnosis they could conceive of, doctors resorted to exome sequencing, leading to the identification of 16,124 mutations, of which 1527 were novel. A causal mutation was discovered in the gene *XIAP*. This gene was already known to play a role in XLP, or X-linked lymphoproliferative syndrome and retrospective review showed that colitis had been observed in 2 XLP patients in the past. Based on these findings, a cord blood transplant was performed, and 2 years later, Nic's intestinal issues had not returned. News coverage of this story by the Milwaukee Journal Sentinel was awarded a Pulitzer Prize for explanatory reporting (Journal Sentinel wins Pulitzer Prize for "One in a Billion" DNA series, n.d.).

28.9 Direct to Consumer Genetics

In the wake of the human genome project and the commoditization of genotypic data, a number of companies were founded to provide consumers with their own genetic information directly. These direct-to-consumer (DTC) genomic companies began making the services broadly available when deCODE genetics launched the deCODEme service in November 2007, followed a few days later by 23andMe. Navigenics was launched the following spring. These companies offered consumers the opportunity to provide a saliva specimen or buccal swab through the mail, and in exchange to receive genotypic information for a range of known genetic markers. Different companies emphasized different aspects of genetic testing. Navigenics focused on known disease risk markers, while 23andMe was much broader, including disease markers but also ancestry information and "recreational" genetic information, for example earwax type and the ability to smell a distinct odor in urine after eating asparagus. Navigenics offered free genetic counseling as part of their service, while 23andMe and deCODEme provided referrals to genetic

counselors. A study of concordance between these three services found >99.6% agreement among them, but in some cases the predicted relative risks differed in magnitude or even direction (Imai, Kricka, & Fortina, 2011). This disagreement is likely due to differences in the specific SNPs and the reference population used to calculate risk.

From the companies' perspectives, their customers offer a rich resource of genomic data for potential research and data mining. 23andMe created a research initiative called 23andWe through which they enlist customers "to collaborate with us on cutting-edge genetic research."(23andWe: The First Annual Update – 23andMe Blog, n.d.) They invite users to fill out questionnaires and then use the phenotypic information to perform genome-wide analysis studies. This approach enabled researchers at the company to replicate a number of known associations, and to discover a number of novel associations, recreational though they may be, for curly hair, freckling, sunlight-induced sneezing, and the ability to smell a metabolite in urine after eating asparagus (Tung et al., 2011). deCODE, purchased by Amgen in 2012, boasts a large number of medically significant genetic discoveries to have come out of their volunteer registry of 160,000 Icelanders, more than half of the adult population of that country (SCIENCE | deCODE genetics, n.d.). Navigenics was purchased by Life Technologies (now part of Thermo Fisher Scientific Inc.) in 2012 and no longer offers their Health Compass genetic testing service.

28.10 Conclusion

Physicians have always sought to provide care personalized to the individual. The current era of large and deep data about individual patients is ushering in the promise of precision medicine that tailors care to the individual based on factors not previously observable by the clinician, such as genomic data, predictive patterns derived from mining clinical data, or dense sensors tracking activity and heart rate at density previously not possible. For precision medicine to become a reality, we will need informatics, to enable both its discovery and implementation. The irony of the ability to personalize care based on an individual's makeup is that it requires huge data sets of many individuals densley phenotyped to have statistical power to make predictions for rare variants, diseases, and outcomes. Thus, precision medicine requires that we have large data sets that are shareable and available for research. We will also need to effectively enroll diverse populations and ensure that the data includes both molecular data and social behavioral determinants of health. In addition, the ability to make accurate decisions for the individual patient requires implementation in the EHR, as the amount of data required to make decisions is vast and changing quickly.

Suggested Readings

Denny, J. C., Bastarache, L., & Roden, D. M. (2016). Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. Annual Review of Genomics and Human Genetics, 17(1), 353–373. https://doi.org/10.1146/annurev-genom-090314-024956. Provides an overview and history of phenome-wide association studies. Different approaches to PheWAS are described, along with the biases, advantages, and disadvantages of each.

Green, E. D., Guyer, M. S., & National Human Genome Research Institute. (2011). Charting a course for genomic medicine from base pairs to bedside. Nature, 470(7333), 204–213. https://doi.org/10.1038/nature09764. Provides an overview of the NHGRI strategic plan through 2020, including the plan moving discovery in large cohorts to implementation in clinical enterprises.

Kirby, J. C., Speltz, P., Rasmussen, L. V., Basford, M., Gottesman, O., Peissig, P. L., … Denny, J. C. (2016). PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. Journal of the American Medical Informatics Association, 23(6), 1046–1052. https://doi.org/10.1093/jamia/ocv202. Introduces the Phenotype KnowledgeBase website, which contains phenotype algorithms and related comments, plus implementation and validation data, for finding cases and controls for genomic analysis from EHR data. The paper includes some summary tables and experiences from the first several years of uploaded EHR phenotype algorithms.

Newton, K. M., Peissig, P. L., Kho, A. N., Bielinski, S. J., Berg, R. L., Choudhary, V., … Denny, J. C. (2013). Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. Journal of the American Medical Informatics Association, 20(e1), e147–54. https://doi.org/10.1136/amiajnl-2012-000896. This paper provides best practices and lessons learned from the Electronics Medical Records and Genomics (eMERGE) Network for how research-grade phenotypes are found from EHR data. This paper includes phenotype algorithm design, creation, and validation process, as well as some experiences regarding what worked well and what did not.

Pulley, J. M., Denny, J. C., Peterson, J. F., Bernard, G. R., Vnencak-Jones, C. L., Ramirez, A. H., … Roden, D. M. (2012). Operational implementation of prospective genotyping for personalized medicine: the design of the Vanderbilt PREDICT project. Clinical Pharmacology and Therapeutics, 92(1), 87–95. https://doi.org/10.1038/clpt.2011.371. This paper describes one of the first prospective implementations of pharmacogenomics. Patients were selected based on their risk for potentially needing a medication affected by pharmacogenes. They were tested on a multiplexed platform, and then medication recommendations were provided through computer-based provider order entry decision support. The first implementation was CYP2C19 and clopidogrel (an antiplatelet medication), but because the platform tested multiple pharmacovariants, drug-genome interactions could be added over time.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 447(7145), 661–678. https://doi.org/10.1038/nature05911. This was one of the first large scale genome-wide association studies, which found common genetic variants influencing seven common diseases. One interesting component, discovered loci for type 2 diabetes, was in FTO, whose effect on diabetes risk is largely mediated through adiposity. This shows the importance of considering phenotypes along the causal pathway when performing GWAS.

28

Questions for Discussion

1. Design a study to assess the genomic influences of a disease or drug response phenotype using EHR data. Who would be your cases and controls? What features would define each case and control, and how would you validate that the algorithms you picked for cases and controls were indeed finding the patients you wanted to find?
2. Research studies traditionally have not returned their research results to study subjects. However, genetic studies are on the forefront of changing paradigms in this space. What do you think about the implications of returning results to patients? How would you feel if you were a subject in a research study? Would you want results back or not?
3. What are the implications of returning results of actionable genetic variants (such as those causing breast and ovarian cancer) found incidentally during research studies or clinical testing purposes?
4. What are some ways in which precision medicine may improve health disparities between different populations? In what ways might precision medicine worsen them? How can researchers promote research that ameliorates this risk?
5. What are some requirements for a health system or a physician in the context of pharmacogenomic testing?
6. Given that genomics do not generally change over the lifetime, how can a patient take their genomic test results from one institution to another? What technological and non-technological solutions could be employed to allow a patient to take their genetic results with them?
7. Discuss the strengths and weaknesses of EHRs for precision medicine studies of diseases, drug responses, and exposures. What kinds of exposures and health outcomes does an EHR excel at capturing and where would traditional survey or in-person assessment measures perform better?

Bibliography

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68–74. https://doi.org/10.1038/nature15393.

23andWe: The First Annual Update – 23andMe Blog. (n.d.). Retrieved from https://blog.23andme.com/23andme-and-you/23andwe-the-first-annual-update/

Ahmad, T., Pencina, M. J., Schulte, P. J., O'Brien, E., Whellan, D. J., Piña, I. L., et al. (2014). Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *Journal of the American College of Cardiology, 64*(17), 1765–1774. https://doi.org/10.1016/j.jacc.2014.07.979.

Bastarache, L., Hughey, J. J., Hebbring, S., Marlo, J., Zhao, W., Ho, W. T., et al. (2018). Phenotype risk scores identify patients with unrecognized Mendelian disease patterns. *Science, 359*(6381), 1233–1239. https://doi.org/10.1126/science.aal4043.

Bentley, A. R., Callier, S., & Rotimi, C. N. (2017). Diversity and inclusion in genomic research: Why the uneven progress? *Journal of Community Genetics, 8*(4), 255–266. https://doi.org/10.1007/s12687-017-0316-6.

Cannon, C. P., Blazing, M. A., Giugliano, R. P., McCagg, A., White, J. A., Theroux, P., et al. (2015). Ezetimibe added to statin therapy after acute coronary syndromes. *The New England Journal of Medicine, 372*(25), 2387–2397. https://doi.org/10.1056/NEJMoa1410489.

Carroll, R. J., Eyler, A. E., & Denny, J. C. (2011). Naïve electronic health record phenotype identification for rheumatoid arthritis. *AMIA Annual Symposium Proceedings, 2011*, 189–196.

Carroll, R. J., Thompson, W. K., Eyler, A. E., Mandelin, A. M., Cai, T., Zink, R. M., et al. (2012). Portability of an algorithm to identify rheumatoid arthritis in electronic health records. *Journal of the American Medical Informatics Association, 19*(e1), e162–e169. https://doi.org/10.1136/amiajnl-2011-000583.

Cohen, J. C., Boerwinkle, E., Mosley, T. H., & Hobbs, H. H. (2006). Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *The New England Journal of Medicine, 354*(12), 1264–1272. https://doi.org/10.1056/NEJMoa054013.

Conway, M., Berg, R. L., Carrell, D., Denny, J. C., Kho, A. N., Kullo, I. J., et al. (2011). Analyzing the heterogeneity and complexity of Electronic Health Record oriented phenotyping algorithms. *AMIA Annual Symposium Proceedings, 2011*, 274–283.

Crawford, D. C., Crosslin, D. R., Tromp, G., Kullo, I. J., Kuivaniemi, H., Hayes, M. G., et al. (2014). eMERGEing progress in genomics-the first seven years. *Frontiers in Genetics, 5*, 184. https://doi.org/10.3389/fgene.2014.00184.

Delaney, J. T., Ramirez, A. H., Bowton, E., Pulley, J. M., Basford, M. A., Schildcrout, J. S., et al. (2012). Predicting clopidogrel response using DNA samples linked to an electronic health record. *Clinical Pharmacology and Therapeutics, 91*(2), 257–263. https://doi.org/10.1038/clpt.2011.221.

Denny, J. C., Bastarache, L., Ritchie, M. D., Carroll, R. J., Zink, R., Mosley, J. D., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature Biotechnology, 31*(12), 1102–1110. https://doi.org/10.1038/nbt.2749.

Denny, J. C., Crawford, D. C., Ritchie, M. D., Bielinski, S. J., Basford, M. A., Bradford, Y., et al. (2011). Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: Using electronic medical records for genome- and phenome-wide studies. *American Journal of Human Genetics, 89*(4), 529–542. https://doi.org/10.1016/j.ajhg.2011.09.008.

Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., et al. (2010). PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics, 26*(9), 1205–1210. https://doi.org/10.1093/bioinformatics/btq126.

Denny, J. C., Ritchie, M. D., Crawford, D. C., Schildcrout, J. S., Ramirez, A. H., Pulley, J. M., et al. (2010). Identification of genomic predictors of atrioventricular conduction: Using electronic medical records as a tool for genome science. *Circulation, 122*(20), 2016–2021. https://doi.org/10.1161/CIRCULATIONAHA.110.948828.

Dewan, A., Liu, M., Hartman, S., Zhang, S. S.-M., Liu, D. T. L., Zhao, C., et al. (2006). HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science, 314*(5801), 989–992. https://doi.org/10.1126/science.1133807.

Donley, G., Hull, S. C., & Berkman, B. E. (2012). Prenatal whole genome sequencing: Just because we can, should we? *The Hastings Center Report, 42*(4), 28–40. https://doi.org/10.1002/hast.50.

Doshi-Velez, F., Ge, Y., & Kohane, I. (2014). Comorbidity clusters in autism spectrum disorders: An electronic health record time-series analysis. *Pediatrics, 133*(1), e54–e63. https://doi.org/10.1542/peds.2013-0819.

Eadon, M. T., Desta, Z., Levy, K. D., Decker, B. S., Pierson, R. C., Pratt, V. M., et al. (2016). Implementation of a pharmacogenomics consult service to support the INGENIOUS trial. *Clinical Pharmacology and Therapeutics, 100*(1), 63–66. https://doi.org/10.1002/cpt.347.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118. https://doi.org/10.1038/nature21056.

Green, E. D., Guyer, M. S., & National Human Genome Research Institute. (2011). Charting a course for

genomic medicine from base pairs to bedside. *Nature, 470*(7333), 204–213. https://doi.org/10.1038/nature09764.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *The Journal of the American Medical Association, 316*(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216.

Hebbring, S. J., Schrodi, S. J., Ye, Z., Zhou, Z., Page, D., & Brilliant, M. H. (2013). A PheWAS approach in studying HLA-DRB1*1501. *Genes and Immunity, 14*(3), 187–191. https://doi.org/10.1038/gene.2013.2.

Holmes, M. V., Asselbergs, F. W., Palmer, T. M., Drenos, F., Lanktree, M. B., Nelson, C. P., et al. (2015). Mendelian randomization of blood lipids for coronary heart disease. *European Heart Journal, 36*(9), 539–550. https://doi.org/10.1093/eurheartj/eht571.

Imai, K., Kricka, L. J., & Fortina, P. (2011). Concordance study of 3 direct-to-consumer genetic-testing services. *Clinical Chemistry, 57*(3), 518–521. https://doi.org/10.1373/clinchem.2010.158220.

Imran, T. F., Posner, D., Honerlaw, J., Vassy, J. L., Song, R. J., Ho, Y.-L., et al. (2018). A phenotyping algorithm to identify acute ischemic stroke accurately from a national biobank: The million veteran program. *Clinical Epidemiology, 10*, 1509–1521. https://doi.org/10.2147/CLEP.S160764.

Jerome, R. N., Pulley, J. M., Roden, D. M., Shirey-Rice, J. K., Bastarache, L. A., Bernard, G., et al. (2018). Using human "experiments of nature" to predict drug safety issues: An example with PCSK9 inhibitors. *Drug Safety, 41*(3), 303–311. https://doi.org/10.1007/s40264-017-0616-0.

Jiang, M., Chen, Y., Liu, M., Rosenbloom, S. T., Mani, S., Denny, J. C., & Xu, H. (2011). A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. *Journal of the American Medical Informatics Association, 18*(5), 601–606. https://doi.org/10.1136/amiajnl-2011-000163.

Journal Sentinel wins Pulitzer Prize for "One in a Billion" DNA series. (n.d.). Retrieved from http://archive.jsonline.com/news/milwaukee/120091754.html/

Kaneko, A., Lum, J. K., Yaviong, L., Takahashi, N., Ishizaki, T., Bertilsson, L., et al. (1999). High and variable frequencies of CYP2C19 mutations: Medical consequences of poor drug metabolism in Vanuatu and other Pacific islands. *Pharmacogenetics, 9*(5), 581–590.

Khera, A. V., Emdin, C. A., Drake, I., Natarajan, P., Bick, A. G., Cook, N. R., et al. (2016). Genetic risk, adherence to a healthy lifestyle, and coronary disease. *The New England Journal of Medicine, 375*(24), 2349–2358. https://doi.org/10.1056/NEJMoa1605086.

Kho, A. N., Hayes, M. G., Rasmussen-Torvik, L., Pacheco, J. A., Thompson, W. K., Armstrong, L. L., et al. (2012). Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *Journal of the American Medical Informatics Association, 19*(2), 212–218. https://doi.org/10.1136/amiajnl-2011-000439.

Kirby, J. C., Speltz, P., Rasmussen, L. V., Basford, M., Gottesman, O., Peissig, P. L., et al. (2016). PheKB: A catalog and workflow for creating electronic phenotype algorithms for transportability. *Journal of the American Medical Informatics Association, 23*(6), 1046–1052. https://doi.org/10.1093/jamia/ocv202.

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics, 46*(3), 310–315. https://doi.org/10.1038/ng.2892.

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science (New York, N.Y.), 308*(5720), 385–389.

Kullo, I. J., Ding, K., Jouni, H., Smith, C. Y., & Chute, C. G. (2010). A genome-wide association study of red blood cell traits using the electronic medical record. *PLoS One, 5*(9). https://doi.org/10.1371/journal.pone.0013011.

Kurreeman, F., Liao, K., Chibnik, L., Hickey, B., Stahl, E., Gainer, V., et al. (2011). Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. *American Journal of Human Genetics, 88*(1), 57–69. https://doi.org/10.1016/j.ajhg.2010.12.007.

Li, L., Cheng, W.-Y., Glicksberg, B. S., Gottesman, O., Tamler, R., Chen, R., et al. (2015). Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Science Translational Medicine, 7*(311), 311ra174. https://doi.org/10.1126/scitranslmed.aaa9364.

Liao, K. P., Cai, T., Gainer, V., Goryachev, S., Zeng-treitler, Q., Raychaudhuri, S., et al. (2010). Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care & Research, 62*(8), 1120–1127. https://doi.org/10.1002/acr.20184.

Lin, C., Karlson, E. W., Dligach, D., Ramirez, M. P., Miller, T. A., Mo, H., et al. (2015). Automatic identification of methotrexate-induced liver toxicity in patients with rheumatoid arthritis from the electronic medical record. *Journal of the American Medical Informatics Association, 22*(e1), e151–e161. https://doi.org/10.1136/amiajnl-2014-002642.

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature, 518*(7538), 197–206. https://doi.org/10.1038/nature14177.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research,*

45(D1), D896–D901. https://doi.org/10.1093/nar/gkw1133.

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature, 551*(7678), 92–94. https://doi.org/10.1038/nature24284.

Millard, L. A. C., Davies, N. M., Timpson, N. J., Tilling, K., Flach, P. A., & Davey Smith, G. (2015). MR-PheWAS: Hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. *Scientific Reports, 5*, 16645. https://doi.org/10.1038/srep16645.

Mohammadpour, A. H., & Akhlaghi, F. (2013). Future of cholesteryl ester transfer protein (CETP) inhibitors: A pharmacological perspective. *Clinical Pharmacokinetics, 52*(8), 615–626. https://doi.org/10.1007/s40262-013-0071-8.

Myocardial Infarction Genetics Consortium Investigators, Stitziel, N. O., Won, H.-H., Morrison, A. C., Peloso, G. M., Do, R., et al. (2014). Inactivating mutations in NPC1L1 and protection from coronary heart disease. *The New England Journal of Medicine, 371*(22), 2072–2082. https://doi.org/10.1056/NEJMoa1405386.

Newton, K. M., Peissig, P. L., Kho, A. N., Bielinski, S. J., Berg, R. L., Choudhary, V., et al. (2013). Validation of electronic medical record-based phenotyping algorithms: Results and lessons learned from the eMERGE network. *Journal of the American Medical Informatics Association, 20*(e1), e147–e154. https://doi.org/10.1136/amiajnl-2012-000896.

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature, 506*(7488), 376–381. https://doi.org/10.1038/nature12873.

O'Donnell, P. H., Bush, A., Spitz, J., Danahey, K., Saner, D., Das, S., et al. (2012). The 1200 patients project: Creating a new medical model system for clinical implementation of pharmacogenomics. *Clinical Pharmacology and Therapeutics, 92*(4), 446–449. https://doi.org/10.1038/clpt.2012.117.

O'Reilly, R., & Elphick, H. E. (2013). Development, clinical utility, and place of ivacaftor in the treatment of cystic fibrosis. *Drug Design, Development and Therapy, 7*, 929–937. https://doi.org/10.2147/DDDT.S30345.

Pathak, J., Kiefer, R. C., Bielinski, S. J., & Chute, C. G. (2012). Applying semantic web technologies for phenome-wide scan using an electronic health record linked biobank. *Journal of Biomedical Semantics, 3*(1), 10. https://doi.org/10.1186/2041-1480-3-10.

Peissig, P. L., Santos Costa, V., Caldwell, M. D., Rottscheit, C., Berg, R. L., Mendonca, E. A., & Page, D. (2014). Relational machine learning for electronic health record-driven phenotyping. *Journal of Biomedical Informatics, 52*, 260–270. https://doi.org/10.1016/j.jbi.2014.07.007.

Pendergrass, S. A., Brown-Gentry, K., Dudek, S., Frase, A., Torstenson, E. S., Goodloe, R., et al. (2013). Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. *PLoS Genetics, 9*(1), e1003087. https://doi.org/10.1371/journal.pgen.1003087.

Pendergrass, S. A., Brown-Gentry, K., Dudek, S. M., Torstenson, E. S., Ambite, J. L., Avery, C. L., et al. (2011). The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genetic Epidemiology, 35*(5), 410–422. https://doi.org/10.1002/gepi.20589.

Perera, M. A., Cavallari, L. H., Limdi, N. A., Gamazon, E. R., Konkashbaev, A., Daneshjou, R., et al. (2013). Genetic variants associated with warfarin dose in African-American individuals: A genome-wide association study. *The Lancet, 382*(9894), 790–796. https://doi.org/10.1016/S0140-6736(13)60681-9.

Peterson, J. F., Field, J. R., Unertl, K. M., Schildcrout, J. S., Johnson, D. C., Shi, Y., et al. (2016). Physician response to implementation of genotype-tailored antiplatelet therapy. *Clinical Pharmacology and Therapeutics, 100*(1), 67–74. https://doi.org/10.1002/cpt.331.

Phillips, E. J., Sukasem, C., Whirl-Carrillo, M., Müller, D. J., Dunnenberger, H. M., Chantratita, W., et al. (2018). Clinical pharmacogenetics implementation consortium guideline for HLA genotype and use of carbamazepine and oxcarbazepine: 2017 update. *Clinical Pharmacology and Therapeutics, 103*(4), 574–581. https://doi.org/10.1002/cpt.1004.

Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature, 538*(7624), 161–164. https://doi.org/10.1038/538161a.

Pulley, J. M., Denny, J. C., Peterson, J. F., Bernard, G. R., Vnencak-Jones, C. L., Ramirez, A. H., et al. (2012). Operational implementation of prospective genotyping for personalized medicine: The design of the Vanderbilt PREDICT project. *Clinical Pharmacology and Therapeutics, 92*(1), 87–95. https://doi.org/10.1038/clpt.2011.371.

Ramirez, A. H., Shi, Y., Schildcrout, J. S., Delaney, J. T., Xu, H., Oetjens, M. T., et al. (2012). Predicting warfarin dosage in European-Americans and African-Americans using DNA samples linked to an electronic health record. *Pharmacogenomics, 13*(4), 407–418. https://doi.org/10.2217/pgs.11.164.

Rasmussen-Torvik, L. J., Stallings, S. C., Gordon, A. S., Almoguera, B., Basford, M. A., Bielinski, S. J., et al. (2014). Design and anticipated outcomes of the eMERGE-PGx project: A multicenter pilot for preemptive pharmacogenomics in electronic health record systems. *Clinical Pharmacology and Therapeutics, 96*(4), 482–489. https://doi.org/10.1038/clpt.2014.137.

Relling, M. V., Schwab, M., Whirl-Carrillo, M., Suarez-Kurtz, G., Pui, C.-H., Stein, C. M., et al. (2019). Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for thiopurine dosing

based on TPMT and NUDT15 genotypes: 2018 update. *Clinical Pharmacology and Therapeutics, 105*(5), 1095–1105. https://doi.org/10.1002/cpt.1304.

Ritchie, M. D., Denny, J. C., Crawford, D. C., Ramirez, A. H., Weiner, J. B., Pulley, J. M., et al. (2010). Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *American Journal of Human Genetics, 86*(4), 560–572. https://doi.org/10.1016/j.ajhg.2010.03.003.

Robinson, J. R., Wei, W.-Q., Roden, D. M., & Denny, J. C. (2018). Defining phenotypes from clinical data to drive genomic research. *Annual Review of Biomedical Data Science, 1*(1), 69–92. https://doi.org/10.1146/annurev-biodatasci-080917-013335.

Sabatine, M. S., Giugliano, R. P., Keech, A. C., Honarpour, N., Wiviott, S. D., Murphy, S. A., et al. (2017). Evolocumab and clinical outcomes in patients with cardiovascular disease. *The New England Journal of Medicine, 376*(18), 1713–1722. https://doi.org/10.1056/NEJMoa1615664.

Saria, S., Butte, A., & Sheikh, A. (2018). Better medicine through machine learning: What's real, and what's artificial? *PLoS Medicine, 15*(12), e1002721. https://doi.org/10.1371/journal.pmed.1002721.

Saunders, C. J., Miller, N. A., Soden, S. E., Dinwiddie, D. L., Noll, A., Alnadi, N. A., et al. (2012). Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Science Translational Medicine, 4*(154), 154ra135. https://doi.org/10.1126/scitranslmed.3004041.

Schmidt, A. F., Swerdlow, D. I., Holmes, M. V., Patel, R. S., Fairhurst-Hunter, Z., Lyall, D. M., et al. (2017). PCSK9 genetic variants and risk of type 2 diabetes: A mendelian randomisation study. The Lancet. *Diabetes & Endocrinology, 5*(2), 97–105. https://doi.org/10.1016/S2213-8587(16)30396-5.

SCIENCE | deCODE genetics. (n.d.). Retrieved from https://www.decode.com/research/

Scott, S. A., Sangkuhl, K., Stein, C. M., Hulot, J. S., Mega, J. L., Roden, D. M., et al. (2013). Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clinical Pharmacology and Therapeutics, 94*(3), 317–323. https://doi.org/10.1038/clpt.2013.105.

Splinter, K., Adams, D. R., Bacino, C. A., Bellen, H. J., Bernstein, J. A., Cheatle-Jarvela, A. M., et al. (2018). Effect of genetic diagnosis on patients with previously undiagnosed disease. *The New England Journal of Medicine, 379*(22), 2131–2139. https://doi.org/10.1056/NEJMoa1714458.

Tannock, I. F., & Hickman, J. A. (2016). Limits to personalized cancer medicine. *The New England Journal of Medicine, 375*(13), 1289–1294. https://doi.org/10.1056/NEJMsb1607705.

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews. Genetics, 19*(9), 581–590. https://doi.org/10.1038/s41576-018-0018-x.

Tung, J. Y., Do, C. B., Hinds, D. A., Kiefer, A. K., Macpherson, J. M., Chowdry, A. B., et al. (2011). Efficient replication of over 180 genetic associations with self-reported medical data. *PLoS One, 6*(8), e23473. https://doi.org/10.1371/journal.pone.0023473.

Voight, B. F., Peloso, G. M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen, M. K., et al. (2012). Plasma HDL cholesterol and risk of myocardial infarction: A mendelian randomisation study. *The Lancet, 380*(9841), 572–580. https://doi.org/10.1016/S0140-6736(12)60312-2.

Wei, W.-Q., & Denny, J. C. (2015). Extracting research-quality phenotypes from electronic health records to support precision medicine. *Genome Medicine, 7*(1), 41. https://doi.org/10.1186/s13073-015-0166-y.

Wei, W.-Q., Teixeira, P. L., Mo, H., Cronin, R. M., Warner, J. L., & Denny, J. C. (2016). Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. *Journal of the American Medical Informatics Association, 23*(e1), e20–e27. https://doi.org/10.1093/jamia/ocv130.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature, 447*(7145), 661–678. https://doi.org/10.1038/nature05911.

White, K. D., Abe, R., Ardern-Jones, M., Beachkofsky, T., Bouchard, C., Carleton, B., et al. (2018). SJS/TEN 2017: Building multidisciplinary networks to drive science and translation. *The Journal of Allergy and Clinical Immunology. In Practice, 6*(1), 38–69. https://doi.org/10.1016/j.jaip.2017.11.023.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics, 46*(11), 1173–1186. https://doi.org/10.1038/ng.3097.

Worthey, E. A., Mayer, A. N., Syverson, G. D., Helbling, D., Bonacci, B. B., Decker, B., et al. (2011). Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genetics in Medicine, 13*(3), 255–262. https://doi.org/10.1097/GIM.0b013e3182088158.

Wu, A. H., White, M. J., Oh, S., & Burchard, E. (2015). The Hawaii clopidogrel lawsuit: The possible effect on clinical laboratory testing. *Personalized Medicine, 12*(3), 179–181. https://doi.org/10.2217/pme.15.4.

Wu, Y., Denny, J. C., Trent Rosenbloom, S., Miller, R. A., Giuse, D. A., Wang, L., et al. (2017). A long journey to short abbreviations: Developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD). *Journal of the American Medical Informatics Association, 24*(e1), e79–e86. https://doi.org/10.1093/jamia/ocw109.

III

Biomedical Informatics in the Years Ahead

Contents

Health Information Technology Policy

Robert S. Rudin, Paul C. Tang, and David W. Bates

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_29

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- Why is the development and use of IT in healthcare so much slower than in other industries?
- How has public policy promoted the adoption and use of health IT?
- How does health IT support national agendas and priorities for health and health care?
- Why is it important to measure the value of health IT in terms of improvements in care quality and savings in costs?
- How can public policies safeguard patient privacy in an era of electronic health information?
- What are the main policy issues related to exchanging health information among health care organizations?
- What are the major tradeoffs for regulating electronic health records in the same way that other medical devices are regulated to ensure patient safety?
- What policies are needed to encourage clinicians to redesign their care practices to exploit better the capabilities of health IT?
- How does the U.S. approach to health IT policy compare with those of other countries?

29

29.1 Public Policy and Health Informatics

For decades after most industries had adopted IT as part of their core business and operational processes, clinical care in the U.S. remained largely in the paper world. Most developed countries adopted health IT sooner, especially in primary care. International leaders have included Denmark, Sweden and the Netherlands. However, health systems leaders in the U.S. have recognized that public policy played a role in the pace and nature of their health systems' adoption and use of IT, and that changes in policy had the potential to accelerate change.

The influence of policy can be found throughout a health care system. Policies shape the structure of health care delivery organizations and the markets for medical products. Directly or indirectly, policies influence the behaviors of all health care stakeholders including patients, providers, health plans, and researchers. Public policy changes can enhance or set back health care delivery through incentives, requirements, and restrictions.

In recent years, policy interventions have influenced health IT in major ways. In 2004, U.S. President George W. Bush established the Office of the National Coordinator for Health IT.[1] In 2009, during the Obama Administration, the U.S. Congress allocated approximately $30 billion to support providers' **meaningful use** of health IT. In 2015, the Medicare Access and CHIP Reauthorization Act (MACRA) absorbed the meaningful use program as part of a larger effort to harmonize how the federal government pays healthcare providers, called the Quality Payment Program (QPP). In 2016, the twenty-first Century Cures Act included provisions to improve patient access to their digital medical data and allow them to use the data in applications of their choice, which may accelerate innovation. Notably, healthcare information technology has been one of the few relatively non-partisan topics. Governments of many other countries have also spent significant public funds on health IT and are considering related policy issues.

In this chapter, we review some of the key policy goals relevant to informatics and discuss how researchers and policymakers are trying to address them. We discuss how health IT policy goals have changed substantially in recent years, from a focus on accelerating adoption to a greater emphasis on interoperability and fostering innovation. Protecting privacy of patients' health information, ensuring health IT products are safe for patients, and improving medical practice

1 ► http://www.healthit.gov/newsroom/about-onc (Accessed 12/9/2012).

workflows remain persistent challenges and of interest to policy.

While informatics research has been occurring for several decades, research in health IT policy is still relatively new. As stakeholders look to health IT to help address the major cost and quality problems in national health care systems, we expect the issues discussed in this chapter to become more important to policymakers and researchers in the fields of health policy and informatics.

29.2 How Health IT Supports National Health Goals: Promise and Evidence

Health IT is not an end in itself. Like all technology, it is simply a tool for achieving larger clinical, social and policy goals, such as improving health outcomes, improving the quality of care, and reducing costs. Health IT has the potential to have a tremendous impact on these goals.

Policymakers, however, are interested not only in the promise of health IT but also the reality. Like most software products, early versions of health IT products tend to have many problems, such as bugs, poor usability, and difficulties integrating with other products. Only after the technology matures is it possible to realize a larger portion of the promised benefits. Policymakers may be reluctant to invest public funds, which are raised primarily in the form of taxes, on technologies that have not been shown in empirical studies to produce benefits.

Many studies have demonstrated empirical benefits of health IT, especially CPOE (see ▶ Chap. 14) and some types of CDS (see ▶ Chap. 24) (Jones et al. 2014). Recent studies of HIE (see ▶ Chaps. 15 and 18) have also found some beneficial effects (Menachemi et al. 2018). However, substantial gaps in evidence exist. For example, many studies come from a small number of academic medical centers or geographical communities, and it is unknown if the benefits in terms of quality, safety and efficiency are being realized in other settings. Some have described this phenomenon as a health IT "productivity paradox" because of some observers' assessment that the benefits of IT have so far not justified the investment (Jones et al. 2012). Lessons from other industries suggest that the substantial benefits of IT will eventually be realized but will require more than just improvements in the technology itself. Care processes will likely also need to be redesigned so that users can take advantage of the technology's potential. New best practices may be needed for different care settings. And additional studies will likely also be needed to demonstrate benefits that may exist but are difficult to detect, especially in non-academic settings that do not have the expertise or incentives to conduct robust evaluation studies (see ▶ Chap. 13).

Despite the limits of the empirical evidence, policymakers have invested substantial sums in health IT hoping that the technology will realize its promised benefits and support national health goals. Further empirical studies will help to identify where health IT has been successful and what factors have made these investments effective, as well as to identify gaps that may benefit from further policy efforts. This section presents an overview of both the promise and the evidence of how health IT supports policy goals.

29.2.1 Improving Care Quality and Health Outcomes

As informatics professionals understand intuitively, health IT has enormous potential to improve care quality and health outcomes, which are, of course, central policy goals (■ Table 29.1). Just as computers have revolutionized many other industries, from banking to baseball, information technology is beginning to revolutionize health care through innovative applications. Policymakers in the U.S. appear to recognize this potential as demonstrated by the multiple pieces of state and federal legislation passed in recent years related to health IT. This activity began with a focus on encouraging adoption and has shifted to improving interoperability, patient

Table 29.1 The promise of health IT (selected functionality)

Health IT functionality	Expected effect on care quality	Expected effect on cost
Electronic health record (EHR) with clinical decisions support (CDS)	Improved clinical decisions, fewer medication and diagnostic errors, timelier follow up	Fewer unnecessary tests
Health information exchange (HIE)	Improved clinical decisions	Reduced burden of information gathering, reduced duplicate testing
Patient decision aids	More personalized treatment	Fewer procedures
Telehealth and personal health records (PHR)	More timely and accessible interactions with clinicians	Fewer office visits
E-prescribing	Fewer errors	Reduced costs from errors

access to records, and innovation. Many other countries also specifically encourage adoption and use of health IT to improve health care quality.

Electronic health records (EHRs; ▶ Chap. 14) probably represent the form of health IT that has been evaluated most extensively and are now widely adopted in hospitals and clinics. EHRs with CPOE and clinical decision support (CDS; ▶ Chap. 24) have been extensively studied and evaluated in terms of quality, safety, and efficiency benefits, with most studies finding positive results. For example, one study found EHRs with medication-related CDS can reduce the number of adverse drug events from 30% to 84% (Ammenwerth et al. 2008). A study that examined EHR use in several hospitals in Texas found that there are reduced rates of inpatient mortality, complications, and length of stay when EHRs are used (Amarasingham et al. 2009). Studies like these have supported the promotion of EHRs, medication-related CDS, and e-prescribing and are now widely, but not universally, adopted. Other functionalities, such as electronic patient decision aids, may have enormous potential to improve quality, safety and efficiency, but have not been evaluated as extensively and are not widely adopted (Friedberg et al. 2013).

Another component of health IT that may substantially improve quality of care is clinical data exchange, which is the ability to exchange health information among health care organizations and patients (see ▶ Chaps. 15 and 18). There is a great need for this kind of capability. In the U.S., the typical Medicare beneficiary visits seven different physicians in four different offices per year on average, and many patients with chronic conditions see more than 16 physicians per year (Pham et al. 2007). Not surprisingly, in such a fragmented system, information is often missing. One study shows that primary care doctors reported missing information in more than 13% of visits and other studies suggest much higher rates of missing data, affecting as much as 81% of visits (Smith et al. 2005; van Walraven et al. 2008; Tang et al. 1994). A study in one community found that there may be a need to exchange data among local medical groups in as many as 50% of patient visits (Rudin et al. 2011). Recent empirical studies have shown that real-world implementations of electronic clinical data exchange systems result in fewer duplicated procedures, reduced use of imaging, lower costs, and improved patient safety (Menachemi et al. 2018). However, these studies were concentrated in a small number of HIEs and some were restricted to a single vendor; it is not clear to what extent the results will generalize to other contexts.

Researchers and policymakers agree that improving the quality of health care must involve making it more patient-centric, and health IT will likely be crucial to achieving that goal on a large scale. For example, **personal health records** (**PHRs**) and patient portals were promoted by federal requirements in the US and are increasingly available – one

recent survey found that roughly half of older adults have accessed a PHR (Malani 2018). PHRs give patients access to their clinical data (see ▶ Chap. 11), facilitate communication between patients and providers, and provide relevant and customized educational materials so that patients can take a more active role in their care (Tang et al. 2006; Halamka et al. 2008; Wells et al. 2014). PHRs may also incorporate patient decision-aids to help them to make critical health care decisions, considering their personal preferences (Fowler et al. 2011; Friedberg et al. 2013). **Telehealth** technologies, which enable patients to interact with clinicians over the Internet (see ▶ Chap. 20), may make health care more patient-centric by allowing patients to receive some of their care without having to go physically to the doctor's office. Few empirical studies to date have shown that these technologies result in improvements in care quality or health outcomes (Milani et al. 2017).

A concern of policymakers is that there is an emerging "digital divide" in health IT, in which disadvantaged groups who might benefit most have less access to health IT than more affluent groups. One empiric study of this issue found that minority groups were less likely to access web-based PHRs and, in general, minorities and disadvantaged groups have less web access than other groups (Yamin et al. 2011). On the other hand, adoption rates of mobile platforms do not show as much of a divide and PHRs are increasingly accessible via these platforms. Still, policies may be necessary to ensure the technology is designed and implemented with minorities in mind to prevent disparities in health care from getting worse and to ensure that the improvements in care quality enabled by health IT are shared by all.

The digital divide has also been suggested to exist among hospitals. One study found that although EHRs are widely adopted among hospitals, critical access hospitals lagged in adoption of performance measurement and patient engagement functions, suggesting an "advanced use" digital divide (Adler-Milstein et al. 2017). However, even if critical access hospitals are slower to adopt advanced functionalities, that may not indicate a permanent divide but rather a typical technology diffusion curve in which some organizations adopt faster than others.

Unfortunately, health IT also has the potential to facilitate harmful unintended side effects (Bloomrosen et al. 2011). In one study involving a pediatric intensive care unit in Pittsburgh, patient mortality increased in patients transferred in after computerized physician order entry (CPOE) was installed (Han et al. 2005). The study found that certain aspects of the ordering system and some of the implementation decisions, restricted clinicians' ability to work efficiently, causing delays in treatment, which was especially deleterious because of the urgent nature of the children's conditions. Implementation decisions involving configuration of the system and changes in workflows appear to have been the major contributors to the increase in mortality– the same EHR product was installed in another hospital without such adverse impacts on mortality (Beccaro et al. 2006). Considering the volume of health IT studies, there are relatively few empirical assessments of adverse effects. Nonetheless, questions about the need to regulate the safety of EHRs are being debated. Balancing the need to protect patients from unintended harm is the concern, further discussed later in this chapter, that over-regulation may impede innovation. Most researchers tend to believe that if health IT systems are well-designed and implemented with close attention to the needs of the users, these kinds of unintended consequences can be avoided and health IT systems will result in tremendous improvements in quality of care (Berg 1999). Researchers have developed guides to help organizations implement health IT in a way that minimizes safety risks and improves patient safety (Sittig et al. 2014). In addition to unintended consequences on patients' health, IT has also been shown to be a source of physician professional dissatisfaction (Sinsky et al. 2017).

29.2.2 Reducing Costs

In addition to improving quality, health IT is expected to reduce costs of care substantially

(Table 29.1). Policies that promoted the use of health IT were informed by projections based on models showing large potential savings for many forms of health IT. One study by the RAND Corporation estimated that EHRs could save more than $81 billion per year (Hillestad et al. 2005). Another study estimated that electronic clinical data exchange has the potential to save $77.8 billion per year (Walker et al. 2005). Many of these savings were expected to come from reductions in redundant tests and use of generic drugs, as well as reductions in adverse drug events and other errors that EHRs might prevent (Bates et al. 1998; Wang et al. 2003). Telehealth and PHRs were also projected to result in billions of dollars in savings (Kaelber and Pan 2008; Cusack et al. 2008).

One weakness of these projections is that they relied on expert opinions for some point estimates because, other than several studies showing that EHRs reduce costs by reducing medical errors, few studies have tried to examine empirically the effect of health IT on costs (Tierney et al. 1987, 1993). Also, some of the projections have been criticized because they estimate potential savings rather than actual measured savings (Congressional Budget Office 2008). However, the projections do not include several types of savings that may result from providing better preventive care and care coordination, which would reduce the need for patients' use of high cost procedures in hospitals and emergency rooms. They also do not include potential reductions in costs that may result from decision aids for patients, which may, for example, reduce the number of unnecessary surgeries (O'Connor et al. 2009). And they do not include other innovations such as the impact of small changes in EHR displays. For example, one study found that when the fees associated with laboratory tests were shown to clinicians when they ordered the test, rates of test ordering decreased by more than 8% (Feldman et al. 2013). The actual savings, therefore, may be much greater than the projections suggest. As described above, realizing these savings will likely require more than simply adopting the technology – it will also require redesigning healthcare workflows to make greater use of the technology, and developing and spreading best practices.

29.2.3 Using Health IT to Measure Quality of Care

All health care stakeholders agree that a health care system should deliver high quality care. But how does one measure care quality? Current methods of quality measurement rely largely on administrative claims submitted by providers to insurers. These data may be useful for certain quality measurements such as for assessing a primary care physician's mammography screening rates, but they lack important clinical details, such as the results of laboratory tests. They also do not represent a comprehensive picture of the care that is delivered, assess the appropriateness of most medical procedures, or determine if a patient's quality of life has improved after treatment. Also, most patients in the U.S. switch insurance companies every few years, limiting the ability of any one insurer to measure quality improvements over longer periods of time, which is required to assess accurately the treatment of many medical conditions.

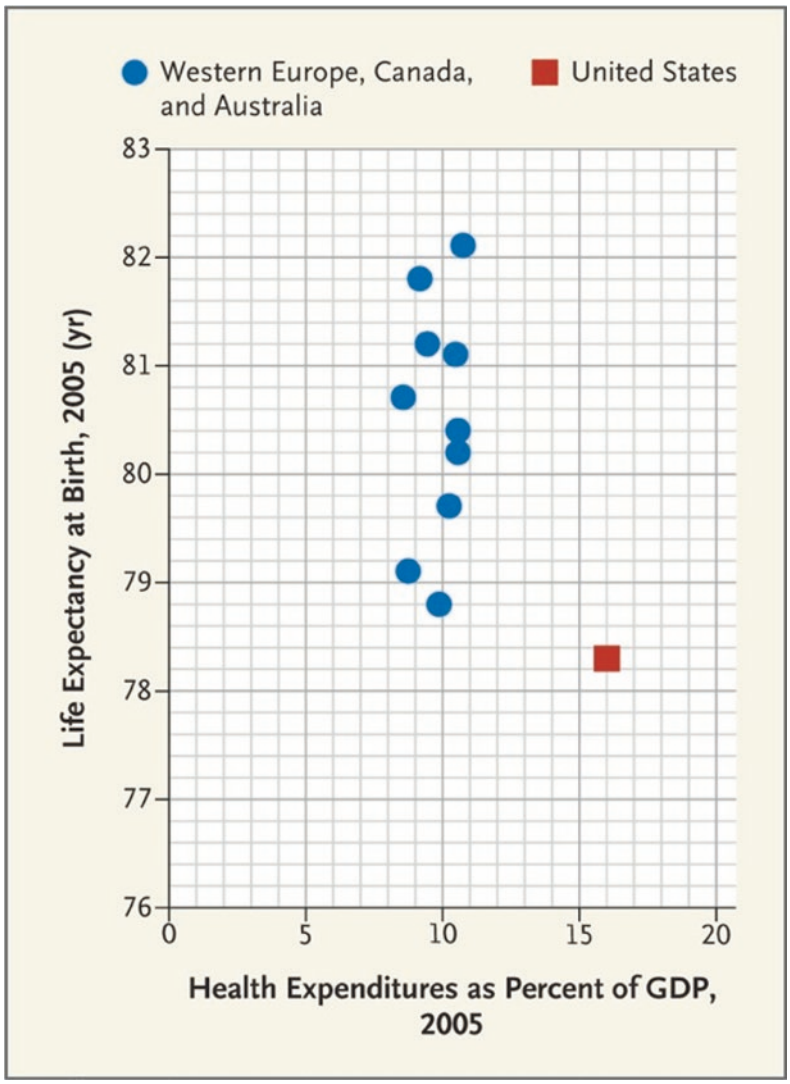
Increasingly, clinical data available through EHRs are used for quality measurement (Ancker et al. 2015). Clinical data are much more comprehensive than administrative claims, and methods for measuring clinical quality using these data are growing. In the U.S., there is growing policy interest in creating such measures as shown in the National Quality Strategy and other reports (AHRQ 2017). This approach has been used in the United Kingdom (U.K.) where nearly 200 quality measures have regularly been assessed, with up to 25% of payment for general practitioners based on performance on these measures (Roland and Olesen 2016). While initially popular, U.K. physicians have become increasingly disenchanted with the administrative requirements of the program. There is growing support for developing patient-reported **outcome measurements** which may be integrated in PHRs, or obtained

through other mechanisms and integrated with the patient's clinical data (Lavallee et al. 2016).

However, using electronic clinical data to generate quality measures is also associated with problems. Studies have found that clinical data in EHRs are often incomplete, inaccurate, and may not be comparable across different EHRs (Chan et al. 2010; Colin et al. 2018). Existing measures also tend to focus more on adherence to care processes rather than patient outcomes (Burstin et al. 2016). More research is needed to develop and standardize meaningful quality measures that would be worth the burden of reporting them.

29.2.4 Holding Providers Accountable for Cost and Quality

Currently, in the U.S., most care is delivered using a fee-for-service payment system, in which providers are paid for every procedure or patient visit. Under this payment method, providers have incentives to provide more care rather than less, which contributes to overtreatment (Lyu et al. 2017). It is therefore not surprising to find that in the U.S., costs are high and rising, nearly double those of many other industrial nations, and quality of care is mixed (Squires 2015). As ◘ Fig. 29.1 shows,



◘ Fig. 29.1 Health care expenditures and life expectancy in the United States and ten other developed countries. (From Fuchs and Milstein (2011), with permission © Massachusetts Medical Society)

the U.S. spends more money per capita on health care than any other country by a wide margin. Yet, many studies suggest that the U.S. is far from the world's leader in overall care quality (Squires 2015). A seminal study by McGlynn et al. in 2003 found that patients in the U.S. received recommended care only about half of the time across a broad array of quality measures (McGlynn et al. 2003). An updated version in 2016 found that those results had not changed much (Levine et al. 2016).

Policymakers are trying to replace the fee-for-service payment method with other methods that would hold providers accountable for the care they deliver. These policies create incentives for healthcare providers to constrain costs and may therefore motivate greater use of health IT tools to achieve this goal. In the U.S., one of the proposed mechanisms for accomplishing this is through Accountable Care Organizations (ACOs). As specified in the Affordable Care Act of 2010,[2] an ACO is a group of providers who are held accountable, to some extent, for both the cost and the quality of a designated group of patients (Berwick 2011; McClellan et al. 2010). ACOs are still a work in progress, but early indications suggest that they may reduce some costs (McWilliams et al. 2018). The concept of ACOs depends on having an electronic health information infrastructure in place, including widespread use of EHRs, because health IT would enable ACOs to improve quality, reduce costs, and measure their performance. Without prior federal incentives for health IT adoption, these policies to aim to change incentives may not have been feasible.

Many other countries have experimented with paying providers for quality and outcomes, or holding providers responsible for costs, although few have done both at the same time to a high degree. Health IT systems are critical for many of these efforts. Few policymakers or researchers believe providers can be held accountable to a substantial degree for the care they delivery without a robust health IT infrastructure.

29.2.5 Informatics Research

Although EHRs have become widespread, many health IT capabilities are still emerging, or standards have not yet been defined. New applications will still require additional research and development. For example, we are still in the early stages of understanding how to design applications for team care (▶ Chap. 17), remote patient monitoring (▶ Chaps. 20 and 21), online disease management (▶ Chaps. 11 and 19), clinical decision-making (▶ Chap. 24), alerts and reminders (▶ Chap. 24), public health and disease surveillance (▶ Chap. 18), clinical trial recruiting (▶ Chap. 27), and evaluations of the impact of technologies on care and costs (▶ Chap. 13). One concern is that most provider organizations, and increasingly even academic medical centers, are now using software applications made by private vendors, and innovating with them can be more challenging than with homegrown products. Private vendors may not be investing enough resources in research to produce transformational innovations (Shortliffe 2012). It will be essential to identify "sandboxes" in which new and innovative IT approaches can be developed and tested. More interactions between industry and academia may be a good way to accelerate progress (Rudin et al. 2016).

Federal funding plays a major role in supporting this kind of upstream informatics research to help to incubate these new technologies but is decreasing in recent years. Because the benefits of such research will accrue to everyone who uses the health care system, the investment of public funds is justified. Few private companies have taken the risk of doing this kind of experimental research to date in part because many health IT companies have been relatively small and were focused on adding the functionalities that are needed to meet federal certification requirements. More recently, some health IT

2 ▶ http://www.healthcare.gov/law/index.html (Accessed 12/9/2012).

companies have become larger but they have not sponsored much research. It is too early to know what impact private companies will have on health IT innovation, but historically, most of the innovation in health informatics has occurred at universities and other government-funded research organizations affiliated with academic medical centers.

29.3 Beyond Adoption: Policy for Optimizing and Innovating with Health IT

Many governments around the world have previously implemented policies to accelerate the adoption of health IT. The U.K. achieved near universal adoption of EHRs because it devoted substantial resources to the effort early on and has a national health care system which directly manages most of the health care providers in the country (Cresswell and Sheikh 2009; Ashworth and Millett 2008). Most other industrialized nations had achieved high levels of adoption in primary care by the early 2000s (Jha et al. 2008). Countries that achieved particularly high levels of adoption in non-hospital settings include Denmark, the Netherlands, Sweden, Hong Kong, Singapore, Australia, and New Zealand. Similar to the U.K., these countries devoted national resources for this effort. Levels of adoption in hospitals, however, lagged in many countries.

In the U.S., after years of slow adoption of health IT relative to other developed countries, the federal government began to address this issue in 2004 by establishing the Office of the National Coordinator for Health IT (ONC). This office is located within the U.S. Department of Health and Human Services and tasked with "promoting development of a nationwide Health IT infrastructure that allows for electronic use and exchange of information." The importance of this office grew considerably in 2009 when Congress passed legislation that is considered a major landmark in the history of health IT policy: the Health Information Technology for Economic and Clinical Health (HITECH) Act.[3] This legislation authorized $27 billion in stimulus funds to be paid to health care providers who demonstrate "meaningful use" of electronic health records as defined by specific criteria (Blumenthal 2010). Although there is debate as to the extent to which HITECH accelerated EHR adoption in ambulatory clinics, EHR adoption increased dramatically among hospitals and clinics in the U.S, after these incentives were put in place (Mennemeyer et al. 2016). Today, over 90% of hospitals and clinics have adopted some form of EHR, but there is large variation in adoption of specific EHR capabilities (HealthIT 2016).

Now that EHRs have become widely adopted, policymakers in many countries are shifting focus toward optimizing the technology and fostering innovation to achieve greater impact. In the U.S., policy efforts are now trying to improve interoperability and health information exchange among providers and patients and facilitate innovation by making health information accessible to third party applications using **application programming interfaces** (APIs). U.S. policy has also incorporated many health IT efforts into a larger program that affects how Center for Medicare and Medicare Services (CMS) pays health providers for services. This section describes some of these efforts.

29.3.1 Health Information Exchange

All countries have challenges sharing clinical data among providers (see ▶ Chap. 15). For many years, U.S. policy promoted data exchange through the formation of regional health information exchanges (HIEs). These organizations provided a variety of services including aggregating EHR data from local health care providers to create aggregate longitudinal patient records, automating the

3 ▶ https://www.healthit.gov/topic/laws-regulation-and-policy/health-it-legislation (Accessed 10/16/2018).

delivery of laboratory results, integrating with pharmacies to facilitate e-prescribing, and facilitating public health and quality reporting. Although some HIEs are well-established, the number of these organizations has been declining and many of the remaining ones may not be financially viable (Adler-Milstein et al. 2016).

Why is it so difficult to establish an HIE? Part of the problem is that EHR products did not always use the same technical data standards and are not interoperable. Recently developed technical and semantic standards have made considerable progress in making the standards robust (Health Level Seven International 2019). However, additional custom programming is still required to integrate EHRs with HIEs. HIEs face many other challenges including: recruiting providers who are reluctant to share data with competing medical groups, privacy and security concerns, legal issues, HIE-related fees, training clinicians to use the HIE, and the lack of a business case (▶ Chap. 15). The business case problem is perhaps the most pressing – for a business to thrive, key stakeholders must be willing to pay for the product or service. In HIE, the primary financial beneficiaries are employers and insurers, but they have been reluctant to pay for the exchange services (Walker et al. 2005). While regional HIEs have faltered, EHR vendor-based networks have emerged as an alternative, but the extent to which they will succeed in the long term is uncertain. These networks may be limited to one vendor or involve a consortium of vendors. Currently, the most prominent vendor-based networks are Epic CareEverywhere, CareQuality, and the CommonWell Health Alliance.

Policymakers have recognized that for data exchange to be comprehensive, these networks as well as regional HIEs will need to interact and share data. To address this concern, there are plans to establish a "trusted exchange framework" that facilitates this interaction (HealthIT 2018). Policymakers have also identified **information blocking** on the part of vendors and providers as a concern and plan to issue regulations to prevent it.

Some have proposed a different approach to data exchange in which patients can aggregate and control access to their complete health records (Szolovits et al. 1994). The history and details of this model are explained in ▶ Chap. 15. There has been an increase in interest in this approach recently. However, it is too early to tell if it will become widely adopted.

No country to date has completely solved the problem of clinical data exchange. In every country that attempts to foster data exchange, the hardest issues appear to be socio-political rather than technical, and there is clear agreement that health IT policy is particularly important to address these problems, especially in establishing standards. The U.K. has set up a "spine" which allows summary care documents to be widely exchanged (Greenhalgh et al. 2010). However, the overall program has encountered major political difficulties, and has been largely dismantled. Canada has established a program called Canada Health Infoway, which has emphasized setting up an infrastructure for data exchange (Rozenblum et al. 2011). While that effort has been somewhat successful, relatively little in the way of clinical data is being exchanged to date, in part because the adoption rate of electronic health records remains low. In Scandinavia, there has been substantial concern about the privacy aspects of data exchange, especially in Sweden, though data exchange is taking place in Denmark and its use is growing.

29.3.2 Patient Portals and Telehealth

Although EHRs have become widely adopted, other forms of health IT are still lagging. To encourage more patient-centric care, many countries are trying to foster the adoption of Patient Portals and Telehealth (see ▶ Chaps. 11 and 20). In the U.S., federal incentives promoted patient portals, and adoption rates are growing. To promote telehealth, policymakers are exploring the possibility of reimbursing for telehealth care, which would probably

improve adoption of this technology considerably (Mehrotra et al. 2016). Even though many states have passed "parity laws" that require commercial insurers to reimburse for telehealth visits, most healthcare encounters are still in-person.

29.3.3 Application Programming Interfaces

To accelerate innovation, policymakers in the U.S. have begun to promote Application Programming Interfaces (APIs) for EHR data. APIs are software mechanisms that allow different applications to connect to one another and share information. All modern software utilizes APIs for purposes ranging from communication with a computer's operating system to querying a website for the latest news stories. In healthcare, one use of APIs is to allow patients to more easily download their latest medical data into an application of their choosing, such as an application on their smartphone that helps them organize and understand their health data. Another use of APIs is to allow providers to install third party applications for use within their EHRs. If patients and providers can pick and choose applications, a new market of innovative applications may arise to take advantage of these data. Standardization of APIs across EHRs is critical because otherwise application developers will be required to spend effort customizing their product to integrate with every EHR vendor. In the U.S., new policies will require EHR vendors to support APIs as a condition for certification and for receiving certain payments (Leventhal 2018).

29.4 Policies to Ensure Safety of Health IT

As adoption of health IT accelerates and new innovations are developed, it is important to be vigilant about, and to reduce, the risk of unintended harmful side effects related to health IT use. Harm could arise from deficiencies in many areas when designing and deploying complex systems, including poor usability, inadequate testing and quality assurance, software flaws, poor implementation decisions, inattention to workflow design, or inadequate training. Policymakers have funded development of frameworks and guidelines to help implement and use health IT in a way that addresses safety concerns (Sittig and Singh 2012).

29.4.1 Should Health IT Be Regulated as Medical Devices?

One policy option for reducing the likelihood of health IT-related medical errors is to create regulations that require health IT products to adhere to strict principles of safe design and be tested and certified (see also ▶ Chap. 12) (Shuren et al. 2018). This is how many medical devices are regulated by the U.S. Food and Drug Administration.[4] While this approach may ensure some degree of patient safety, the regulatory burden will increase the price of health IT systems, raise barriers of entry for new companies, and could stifle innovation. Also, even with regulations, health IT products might still have safety issues because software products can be used in many ways, unlike other medical devices that have more limited utility.

There is an active debate about the appropriate types of regulation for medical apps for use by patients. Currently estimates have found more than 150,000 health apps available for download, but analysts have found that few have demonstrated clinical utility (Singh et al. 2016). The FDA does not regulate most apps but has recently begun a pilot "precertification" program for digital health which will provide information about vendors' software quality control processes but does not involve evaluations of outcomes (Bates et al. 2018; Lee and Kesselheim 2018). This is controversial, and some feel it does not go far enough (Bates et al. 2018).

4 ▶ http://www.fda.gov/ (Accessed 12/10/12).

29.4.2 Alternative Ways to Improve Patient Safety

There are many other policy options to support patient safety (Committee on Patient Safety and Health Information Technology; Institute of Medicine 2011). Policies may fund training programs to educate clinicians in how to use health IT safely and alert them to common mistakes. Policies might encourage providers to report problems with software, including usability issues and bugs, so that vendors can fix them quickly. Policies might also help to establish programs in which users can rate health IT products. Finally, funding research into the science of patient safety would improve our knowledge of how to design better products and identify risks of errors (Shekelle et al. 2011).

29.5 Policies to Ensure Privacy and Security of Electronic Health Information

It is almost impossible to have a conversation about digital medical records without discussing issues of privacy and security. Although the topic of privacy arose in the discussion of ethics in ► Chap. 12, it also has policy implications and warrants mention here. As healthcare has become digitized, there has been an increase in security events (Liu et al. 2015). Protecting privacy and security are clearly important policy goals.

29.5.1 Regulating Privacy

The Health Insurance Portability and Accountability Act (HIPAA) of 1996[5] and subsequent regulations created a legal category of "protected health information" which was defined to encompass most forms of clinical data. **Covered entities** which include providers and insurers are legally required under this law to safeguard electronic health information and would face fines if they did not.

Many states have additional privacy laws regarding data exchange (e.g., mental health and HIV status). The effectiveness of these privacy-protective laws has not been rigorously evaluated. They can inadvertently reduce privacy protection, particularly when exchanging data across state lines, and have been showed to slow the adoption of EHRs (Miller and Tucker 2009; Harmonizing State Privacy Law Collaborative 2009).

In other countries, privacy also has received a good deal of debate. Most recently, the European General Data Protection Regulation (GPDR) went into effect in 2018 and goes beyond healthcare in scope by encouraging "privacy by design" for all software products that store personal data (Haug 2018). Governments are still trying to find the best policies to protect privacy of medical records without slowing the adoption of health IT.

29.5.2 Security

Now that healthcare entities are mostly digital, they are increasingly targeted by cyberattacks, which may aim to steal patient data, demand money in return for unlocking a system, or make a political statement. HIPAA includes security policies that require health providers and other covered entities to implement various safeguards, and if data are breached, the federal government may charge a fine. The recent increase in cyberattacks on hospital and other healthcare stakeholders suggest that these regulations may not be adequate, and policymakers are considering additional moves. Security concerns exist in all countries. For example, the UK's National Health Service recently experienced a cyberattack that crippled many hospitals and required many clinics to close down completely (Clarke and Youngstein 2017).

5 ► http://www.hhs.gov/ocr/privacy/index.html (Accessed 12/9/2012).

29.5.3 Record Matching and Linking

For health IT to be effective, an essential prerequisite is that patients must be matched to their health data, and electronic records for the same patient must be linked together. If patients' identity attributes are used (e.g., name, address, date of birth), matching and linking errors often occur because many patients share attributes, attributes change over time, and clerical errors are common. Many countries have adopted a **unique health identifier** (**UHI**) to facilitate these processes. However, in response to concerns of privacy advocates, the U.S. congress prohibited use of HHS to expend federal dollars to support development of a UHI. There is little evidence that suggests UHIs pose an increased risk of privacy violations and, in fact, not having a UHI may be even more risky because many other kinds of personal data may be collected and used instead (Greenberg et al. 2009). But UHIs require substantial federal resources to implement and may not address all matching and linking issues. Currently, some estimates suggest that errors in linking records shared across providers in the U.S. can be as high as 50%.

Policymakers are therefore interested in alternative approaches, which include improving linking algorithms to better match identity attributes (e.g., name, address, date of birth), defining standards for the identity attributes, using biometrics-based methods, and allowing patients to participate more directly in the process, such as by verifying their phone number with their mobile phone or managing their data on their smartphone (Rudin et al. 2018). There are advantages and disadvantages to every approach, and it is likely that multiple approaches will be needed to substantially reduce matching and linking errors (Pew 2018). Policymakers may play a critical role in overseeing progress and supporting research to develop and more rigorously evaluate solutions.

29.6 The Growing Importance of Public Policy in Informatics

Public policy is becoming increasingly important to the field of informatics. Policies affect everything from what research projects receive funding to whether a physician in a solo practice allows her patients to access their medical records online. Many of the health IT policy issues we discuss in this chapter are just beginning to attract attention from policymakers, and further research is needed to understand the best role for policy. It is likely that new policy issues will emerge as technology capabilities become more advanced. For example, artificial intelligence may help with many clinical applications, but policies may be needed to ensure it is applied safely and to ensure accountability.

Traditionally, most informatics research has focused on the development of new technologies and how they integrate into clinical practice. Relatively few studies provide advice to policymakers on health IT policy issues, even though policies have enormous consequences for informatics research and practice. We hope that researchers and policymakers will recognize that technology and policy issues affect each other, and it is necessary to use both perspectives to understand how information technology can be used to improve health care.

Suggested Readings

Agency for Healthcare Research and Quality. (2013). A robust health data infrastructure. Retrieved from McLean, VA: https://www.healthit.gov/sites/default/files/ptp13-700hhs_white.pdf. This white paper makes the case for public policy to promote open APIs to improve interoperability and data exchange, and to promote innovation in healthcare.

Bloomfield, R. A., Jr., Polo-Wood, F., Mandel, J. C., & Mandl, K. D. (2017). Opening the Duke electronic health record to apps: Implementing SMART on FHIR. International Journal of Medical Informatics, 99, 1–10. https://doi.org/10.1016/j.ijmedinf.2016.12.005. This

research study discusses a successful early attempt to use APIs within a live EHR and emerging technical standards to implement patient- and provide-facing apps.

Clarke, R., & Youngstein, T. (2017). Cyberattack on Britain's National Health Service – a wake-up call for modern medicine. New England Journal of Medicine, 377(5), 409–411. https://doi.org/10.1056/NEJMp1706754. This brief perspective describes a harrowing cyber attack on the U.K.'s healthcare system and offers suggestion to help improve preparedness.

Jones, S. S., Heaton, P. S., Rudin, R. S., & Schneider, E. C. (2012). Unraveling the IT productivity paradox – lessons for health care. New England Journal of Medicine, 366(24), 2243–2245. This brief perspective addresses the contentious issue of why few studies have been able to show that health IT produces an improvement in economic productivity. It makes its case by pointing out that the IT industry had the same problem in the 1980s and 1990s but managed to overcome these difficulties through better measurement of productivity, improved management of technology, and better usability.

Sinsky, C., Colligan, L., Li, L., Prgomet, M., Reynolds, S., Goeders, L., et al. (2016). Allocation of physician time in ambulatory practice: A time and motion study in 4 specialties. Annals of Internal Medicine, 165(11), 753–760. https://doi.org/10.7326/M16-0961. This study reported direct observation of 57 U.S. physicians and found they spend almost 50% of their time on EHR and desk work, which was much more than time on direct clinical face time with patients. Other work by some of the same authors have identified EHRs as a source of professional dissatisfaction and burnout.

Sittig, D. F., & Singh, H. (2012). Electronic health records and national patient-safety goals. New England Journal of Medicine, 367(19), 1854–1860. https://doi.org/10.1056/NEJMsb1205420. This article proposes a 3-phased approach to implementing EHRs in a way that improves safety: address safety concerns unique to EHR technology, mitigate safety concerns arising from failure to use EHRs appropriately, and use EHRs to monitoring and improve patient safety.

Questions for Discussion

1. What are the key barriers to effective use of EHRs and exchange of health information? Which of these challenges are amenable to public policy decisions?
2. What are the key barriers to innovation in health IT? What can be done to accelerate innovation?
3. What might be some of the tradeoffs of using administrative claims data compared with using clinical data from health IT systems for care quality analysis?
4. What might be some of the tradeoffs of promoting health IT by paying for use compared with paying for quality?
5. Should health IT be regulated the same way as devices are regulated to protect patient safety? Why or why not?
6. If research finds strong evidence of a digital divide in health IT, what policy actions should be taken?
7. What kinds of health IT functionality are needed to support accountable care organizations and patient-centered medical homes?

References

Adler-Milstein, J., Lin, S. C., & Jha, A. K. (2016). The number of health information exchange efforts is declining, leaving the viability of broad clinical data exchange uncertain. *Health Affairs (Millwood), 35*(7), 1278–1285. https://doi.org/10.1377/hlthaff.2015.1439.

Adler-Milstein, J., Holmgren, A. J., Kralovec, P., Worzala, C., Searcy, T., & Patel, V. (2017). Electronic health record adoption in US hospitals: The emergence of a digital "advanced use" divide. *Journal of the American Medical Informatics Association, 24*(6), 1142–1148. https://doi.org/10.1093/jamia/ocx080.

Agency for Healthcare Research and Quality. (2017, March). *About the National Quality Strategy*. Retrieved from https://www.ahrq.gov/workingforquality/about/index.html.

Amarasingham, R., Plantinga, L., Diener-West, M., Gaskin, D. J., & Powe, N. R. (2009). Clinical information technologies and inpatient outcomes: A multiple hospital study. *Archives of Internal Medicine, 169*(2), 108–114.

Ammenwerth, E., Schnell-Inderst, P., Machan, C., & Siebert, U. (2008). The effect of electronic prescribing on medication errors and adverse drug events: A systematic review. *Journal of the American Medical Informatics Association: JAMIA, 15*(5), 585–600.

Ancker, J. S., Kern, L. M., Edwards, A., Nosal, S., Stein, D. M., Hauser, D., & Kaushal, R. (2015). Associations between healthcare quality and use of electronic health record functions in ambulatory care. *Journal of the American Medical Informatics Association, 22*(4), 864–871. https://doi.org/10.1093/jamia/ocv030.

Ashworth, M., & Millett, C. (2008). Quality improvement in UK primary care: The role of financial incentives. *The Journal of Ambulatory Care Management, 31*(3), 220–225.

Bates, D. W., Leape, L. L., Cullen, D. J., Laird, N., Petersen, L. A., Teich, J. M., et al. (1998). Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA, 280*(15), 1311–1316.

Bates, D. W., Landman, A., & Levine, D. M. (2018). Health apps and health policy: What is needed? *JAMA, 320*, 1975. https://doi.org/10.1001/jama.2018.14378.

Beccaro, M. A. D., Jeffries, H. E., Eisenberg, M. A., & Harry, E. D. (2006). Computerized provider order entry implementation: No association with increased mortality rates in an intensive care unit. *Pediatrics, 118*(1), 290–295.

Berg, M. (1999). Patient care information systems and health care work: A sociotechnical approach. *International Journal of Medical Informatics, 55*(2), 87–101.

Berwick, D. M. (2011). Launching accountable care organizations – the proposed rule for the Medicare Shared Savings Program. *The New England Journal of Medicine, 364*, e32.

Bloomrosen, M., Starren, J., Lorenzi, N. M., Ash, J. S., Patel, V. L., & Shortliffe, E. H. (2011). Anticipating and addressing the unintended consequences of health IT and policy: A report from the AMIA 2009 health policy meeting. *Journal of the American Medical Informatics Association: JAMIA, 18*(1), 82–90.

Blumenthal, D. (2010). Launching HITECH. *The New England Journal of Medicine, 362*(5), 382–385.

Burstin, H., Leatherman, S., & Goldmann, D. (2016). The evolution of healthcare quality measurement in the United States. *Journal of Internal Medicine, 279*(2), 154–159. https://doi.org/10.1111/joim.12471.

Chan, K. S., Fowles, J. B., & Weiner, J. P. (2010). Review: Electronic health records and the reliability and validity of quality measures: A review of the literature. *Medical Care Research and Review, 67*(5), 503–527.

Clarke, R., & Youngstein, T. (2017). Cyberattack on Britain's National Health Service – a wake-up call for modern medicine. *New England Journal of Medicine, 377*(5), 409–411. https://doi.org/10.1056/NEJMp1706754.

Colin, N. V., Cholan, R. A., Sachdeva, B., Nealy, B. E., Parchman, M. L., & Dorr, D. A. (2018). Understanding the impact of variations in measurement period reporting for electronic clinical quality measures. *EGEMS (Washington, DC), 6*(1), 17. https://doi.org/10.5334/egems.235.

Committee on Patient Safety and Health Information Technology; Institute of Medicine. (2011). *Health IT and patient safety: Building safer systems for better care*. Washington, DC: National Academies Press.

Congressional Budget Office. (2008). *Evidence on the costs and benefits of health information technology*. Washington, DC: CBO Paper.

Cresswell, K., & Sheikh, A. (2009). The NHS care record service (NHS CRS): Recommendations from the literature on successful implementation and adoption. *Informatics in Primary Care, 17*(3), 153–160.

Cusack, C. M., Pan, E., Hook, J. M., Vincent, A., Kaelber, D. C., & Middleton, B. (2008). The value proposition in the widespread use of telehealth. *Journal of Telemedicine and Telecare, 14*(4), 167–168.

Feldman, L. S., Shihab, H. M., Thiemann, D., Yeh, H.-C., Ardolino, M., Mandell, S., & Brotman, D. J. (2013). Impact of providing fee data on laboratory test ordering: A controlled clinical trial. *JAMA Internal Medicine, 173*(10), 903–908. https://doi.org/10.1001/jamainternmed.2013.232.

Fowler, F. J., Levin, C. A., & Sepucha, K. R. (2011). Informing and involving patients to improve the quality of medical decisions. *Health Affairs (Millwood), 30*(4), 699–706.

Friedberg, M. W., Van Busum, K., Wexler, R., Bowen, M., & Schneider, E. C. (2013). A demonstration of shared decision making in primary care highlights barriers to adoption and potential remedies. *Health Affairs (Millwood), 32*(2), 268–275. https://doi.org/10.1377/hlthaff.2012.1084.

Fuchs, V. R., & Milstein, A. (2011). The $640 billion question–why does cost-effective care diffuse so slowly? *The New England Journal of Medicine, 364*(21), 1985–1987.

Greenberg, M. D., Ridgely, M. S., & Hillestad, R. J. (2009). Crossed wires: How yesterday's privacy rules might undercut tomorrow's nationwide health information network. *Health Affairs (Millwood), 28*(2), 450–452.

Greenhalgh, T., Stramer, K., Bratan, T., Byrne, E., Russell, J., & Potts, H. W. W. (2010). Adoption and non-adoption of a shared electronic summary record in England: A mixed-method case study. *BMJ, 340*, c3111.

Halamka, J. D., Mandl, K. D., & Tang, P. C. (2008). Early experiences with personal health records. *Journal of the American Medical Informatics Association: JAMIA, 15*(1), 1–7.

Han, Y. Y., Carcillo, J. A., Venkataraman, S. T., Clark, R. S. B., Watson, R. S., Nguyen, T. C., et al. (2005). Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics, 116*(6), 1506–1512.

Harmonizing State Privacy Law Collaborative. (2009). *Harmonizing state privacy law*. Washington, DC: Office of the National Coordinator of Health Information Technology.

Haug, C. J. (2018). Turning the tables – the new European general data protection regulation. *New England Journal of Medicine, 379*(3), 207–209. https://doi.org/10.1056/NEJMp1806637.

Health Level Seven International. (2019). Retrieved from https://www.hl7.org/.

HealthIT. (2016). *Office-based physician electronic health record adoption*. Retrieved from https://dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php.

HealthIT. (2018). *Draft trusted exchange network*. Retrieved from https://www.healthit.gov/sites/default/files/draft-trusted-exchange-framework.pdf.

Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., & Taylor, R. (2005). Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs (Millwood), 24*, 1103–1117.

Jha, A. K., Doolan, D., Grandt, D., Scott, T., & Bates, D. W. (2008). The use of health information technology in seven nations. *International Journal of Medical Informatics, 77*(12), 848–854.

Jones, S. S., Heaton, P. S., Rudin, R. S., & Schneider, E. C. (2012). Unraveling the IT productivity paradox – lessons for health care. *The New England Journal of Medicine, 366*(24), 2243–2245.

Jones, S. S., Rudin, R. S., Perry, T., & Shekelle, P. G. (2014). Health information technology: An updated systematic review with a focus on meaningful use. *Annals of Internal Medicine, 160*(1), 48–54. https://doi.org/10.7326/M13-1531.

Kaelber, D., & Pan, E. C. (2008). The value of personal health record (PHR) systems. *AMIA Annual Symposium Proceedings, 2008*, 343–347.

Lavallee, D. C., Chenok, K. E., Love, R. M., Petersen, C., Holve, E., Segal, C. D., & Franklin, P. D. (2016). Incorporating patient-reported outcomes into health care to engage patients and enhance care. *Health Affairs (Millwood), 35*(4), 575–582. https://doi.org/10.1377/hlthaff.2015.1362.

Lee, T. T., & Kesselheim, A. S. (2018). U.S. Food and Drug Administration precertification pilot program for digital health software: Weighing the benefits and risks. *Annals of Internal Medicine, 168*(10), 730–732. https://doi.org/10.7326/m17-2715.

Leventhal, R. (2018). *BREAKING: CMS finalizes "promoting interoperability" rule*. Retrieved from https://www.healthcare-informatics.com/article/value-based-care/breaking-cms-finalizes-promoting-interoperability-rule.

Levine, D. M., Linder, J. A., & Landon, B. E. (2016). The quality of outpatient care delivered to adults in the United States, 2002 to 2013. *JAMA Internal Medicine, 176*(12), 1778–1790. https://doi.org/10.1001/jamainternmed.2016.6217.

Liu, V., Musen, M. A., & Chou, T. (2015). Data breaches of protected health information in the United States. *JAMA, 313*(14), 1471–1473. https://doi.org/10.1001/jama.2015.2252.

Lyu, H., Xu, T., Brotman, D., Mayer-Blackwell, B., Cooper, M., Daniel, M., et al. (2017). Overtreatment in the United States. *PLoS One, 12*(9), e0181970. https://doi.org/10.1371/journal.pone.0181970.

Malani, P. (2018). *Logging in: Using patient portals to access health information*. Retrieved from https://deepblue.lib.umich.edu/handle/2027.42/145683.

McClellan, M., McKethan, A. N., Lewis, J. L., Roski, J., & Fisher, E. S. (2010). A national strategy to put accountable care into practice. *Health Affairs (Millwood), 29*(5), 982–990.

McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., & Kerr, E. A. (2003). The quality of health care delivered to adults in the United States. *The New England Journal of Medicine, 348*, 2635–2645.

McWilliams, J. M., Hatfield, L. A., Landon, B. E., Hamed, P., & Chernew, M. E. (2018). Medicare spending after 3 years of the Medicare Shared Savings Program. *New England Journal of Medicine, 379*(12), 1139–1149. https://doi.org/10.1056/NEJMsa1803388.

Mehrotra, A., Jena, A. B., Busch, A. B., Souza, J., Uscher-Pines, L., & Landon, B. E. (2016). Utilization of telemedicine among rural medicare beneficiaries. *JAMA, 315*(18), 2015–2016. https://doi.org/10.1001/jama.2016.2186.

Menachemi, N., Rahurkar, S., Harle, C. A., & Vest, J. R. (2018). The benefits of health information exchange: An updated systematic review. *Journal of the American Medical Informatics Association, 25*(9), 1259–1265. https://doi.org/10.1093/jamia/ocy035.

Mennemeyer, S. T., Menachemi, N., Rahurkar, S., & Ford, E. W. (2016). Impact of the HITECH Act on physicians' adoption of electronic health records. *Journal of the American Medical Informatics Association, 23*(2), 375–379. https://doi.org/10.1093/jamia/ocv103.

Milani, R. V., Lavie, C. J., Bober, R. M., Milani, A. R., & Ventura, H. O. (2017). Improving hypertension control and patient engagement using digital tools. *The American Journal of Medicine, 130*(1), 14–20. https://doi.org/10.1016/j.amjmed.2016.07.029.

Miller, A. R., & Tucker, C. (2009). Privacy protection and technology diffusion: The case of electronic medical records. *Management Science, 55*, 1077–1093.

O'Connor, A. M., Bennett, C. L., Stacey, D., Barry, M., Col, N. F., Eden, K. B., et al. (2009). Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews*, (3), CD001431.

Pew. (2018). *Enhanced patient matching is critical to achieving full promise of digital health records*. Retrieved from https://www.pewtrusts.org/-/media/assets/2018/09/healthit_enhancedpatientmatching_report_final.pdf.

Pham, H. H., Schrag, D., O'Malley, A. S., Wu, B., & Bach, P. B. (2007). Care patterns in medicare and

their implications for pay for performance. *The New England Journal of Medicine, 356*(11), 1130–1139.

Roland, M., & Olesen, F. (2016). Can pay for performance improve the quality of primary care? *BMJ, 354*, i4058. https://doi.org/10.1136/bmj.i4058.

Rozenblum, R., Jang, Y., Zimlichman, E., Salzberg, C., Tamblyn, M., Buckeridge, D., et al. (2011). A qualitative study of Canada's experience with the implementation of electronic health information technology. *CMAJ: Canadian Medical Association Journal, 183*(5), E281–E288.

Rudin, R. S., Salzberg, C. A., Szolovits, P., Volk, L. A., Simon, S. R., & Bates, D. W. (2011). Care transitions as opportunities for clinicians to use data exchange services: How often do they occur? *Journal of the American Medical Informatics Association: JAMIA, 18*(6), 853–858.

Rudin, R. S., Bates, D. W., & MacRae, C. (2016). Accelerating innovation in Health IT. *New England Journal of Medicine, 375*(9), 815–817. https://doi.org/10.1056/NEJMp1606884.

Rudin, R. S., Hillestad, R., Ridgely, M. S., Qureshi, N. S., II, Davis, J. S., & Fischer, S. H. (2018). *Defining and evaluating patient-empowered approaches to improving record matching*. Santa Monica: The RAND Corporation.

Shekelle, P. G., Pronovost, P. J., Wachter, R. S., Taylor, S. L., Dy, S. M., Foy, R., et al. (2011). Advancing the science of patient safety. *Annals of Internal Medicine, 154*(10), 693–696.

Shortliffe, T. (2012). *The future of biomedical informatics: A perspective from academia*. Keynote Presentation, Medical Informatics Europe 2012, Pisa, Italy.

Shuren, J., Patel, B., & Gottlieb, S. (2018). FDA regulation of mobile medical apps. *JAMA, 320*(4), 337–338. https://doi.org/10.1001/jama.2018.8832.

Singh, K., Drouin, K., Newmark, L. P., Lee, J., Faxvaag, A., Rozenblum, R., et al. (2016). Many mobile health apps target high-need, high-cost populations, but gaps remain. *Health Affairs (Millwood), 35*(12), 2310–2318. https://doi.org/10.1377/hlthaff.2016.0578.

Sinsky, C. A., Dyrbye, L. N., West, C. P., Satele, D., Tutty, M., & Shanafelt, T. D. (2017). Professional satisfaction and the career plans of US physicians. *Mayo Clinic Proceedings, 92*(11), 1625–1635. https://doi.org/10.1016/j.mayocp.2017.08.017.

Sittig, D. F., & Singh, H. (2012). Electronic health records and national patient-safety goals. *New England Journal of Medicine, 367*(19), 1854–1860. https://doi.org/10.1056/NEJMsb1205420.

Sittig, D. F., Ash, J. S., & Singh, H. (2014). The SAFER guides: Empowering organizations to improve the safety and effectiveness of electronic health records. *The American Journal of Managed Care, 20*(5), 418–423.

Smith, P. C., Araya-Guerra, R., Bublitz, C., Parnes, B., Dickinson, L. M., Vorst, R. V., et al. (2005). Missing clinical information during primary care visits. *JAMA, 293*(5), 565–571.

Squires, D. (2015). *U.S. health care from a global perspective: Spending, use of services, prices, and health in 13 countries*. Retrieved from https://www.commonwealthfund.org/publications/issue-briefs/2015/oct/us-health-care-global-perspective.

Szolovits, P., Doyle, J., Long, W. J., Kohane, I., & Pauker, S. G. (1994). *Guardian angel: Patient-centered health information systems*. Retrieved from https://smarthealthit.org/1994/10/guardian-angel-patient-centered-health-information-systems/.

Tang, P. C., Fafchamps, D., & Shortliffe, E. H. (1994). *Traditional medical records as a source of clinical data in the outpatient setting*. Proceedings of the Annual Symposium Computer Applications in Medical Care, 575–579.

Tang, P. C., Ash, J. S., Bates, D. W., Overhage, J. M., & Sands, D. Z. (2006). Personal health records: Definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics Association: JAMIA, 13*(2), 121–126.

Tierney, W. M., McDonald, C. J., Martin, D. K., & Rogers, M. P. (1987). Computerized display of past test results. Effect on outpatient testing. *Annals of Internal Medicine, 107*(4), 569–574.

Tierney, W. M., Miller, M. E., Overhage, J. M., & McDonald, C. J. (1993). Physician inpatient order writing on microcomputer workstations. Effects on resource utilization. *JAMA, 269*(3), 379–383.

van Walraven, C., Taljaard, M., Bell, C. M., Etchells, E., Zarnke, K. B., Stiell, I. G., & Forster, A. J. (2008). Information exchange among physicians caring for the same patient in the community. *Canadian Medical Association Journal, 179*(10), 1013–1018.

Walker, J., Pan, E., Johnston, D., Adler-Milstein, J., Bates, D. W., & Middleton, B. (2005). The value of health care information exchange and interoperability. *Health Affairs (Millwood)*, (Suppl Web Exclusives), W5–10–W5–18.

Wang, S. J., Middleton, B., Prosser, L. A., Bardon, C. G., Spurr, C. D., Carchidi, P. J., et al. (2003). A cost-benefit analysis of electronic medical records in primary care. *American Journal of Medicine, 114*(5), 397–403.

Wells, S., Rozenblum, R., Park, A., Dunn, M., & Bates, D. W. (2014). Personal health records for patients with chronic disease: A major opportunity. *Applied Clinical Informatics, 5*(2), 416–429.

Yamin, C. K., Emani, S., Williams, D. H., Lipsitz, S. R., Karson, A. S., Wald, J. S., & Bates, D. W. (2011). The digital divide in adoption and use of a personal health record. *Archives of Internal Medicine, 171*(6), 568–574.

The Future of Informatics in Biomedicine

James J. Cimino, Edward H. Shortliffe, Michael F. Chiang, David Blumenthal, Patricia Flatley Brennan, Mark Frisse, Eric Horvitz, Judy Murphy, Peter Tarczy-Hornoch, and Robert M. Wachter

Contents

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5_30

Learning Objectives

After reading this chapter, you should know the answers to these questions:

- What does the past evolution of the field of biomedical informatics tell us about its future trajectory?
- How will data science methods influence biomedical informatics research
- What roles will electronic health records and artificial intelligence play in health care of the future?

30.1 The Present and Its Evolution from the Past

Every good look forward should start with a look back to provide perspective regarding the past and an assessment of the pace of change, thereby helping us to anticipate a trajectory for the future. This book first appeared in 1990 at a time when the field was much younger (the word *informatics* had come into common use only in the previous decade) and was still being defined. Thus that early edition, and the ones that followed (in 2000, 2006, and 2014) offer a glimpse of what topics appeared over time, which ones faded away, and how even the terminology evolved (as it will no doubt continue to do in the future). Consider, for example, the list of chapter titles from the 1990 edition (■ Table 30.1). The first edition was titled *Medical Informatics: Computer Applications in Medical Care*, reflecting the field's original roots in clinical medicine. In those days, the field was called medical informatics (see ▶ Chap. 1) and the first edition was focused largely on clinical application areas, such as electronic health records, nursing systems, laboratory systems, radiology system, and education systems.

■ **Table 30.1** Table of contents sections and chapters from all five editions of this book, aligned by subject matter

Medical Informatics: Computer Applications in Medical Care (1990)	*Medical Informatics: Computer Applications in Health Care and Biomedicine (2000)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2006)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2014)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2020)*
Recurrent themes in medical informatics	**Recurrent themes in medical informatics**	**Recurrent themes in biomedical informatics**	**Recurrent themes in biomedical informatics**	**Recurrent themes in biomedical informatics**
1. The computer meets medicine: Emergence of a discipline	1. The computer meets medicine and biology: Emergence of a discipline	1. The computer meets medicine and biology: Emergence of a discipline	1. Biomedical informatics: The science and the pragmatics	1. Biomedical informatics: The science and the pragmatics
2. Medical data: Their acquisition, storage, and use	2. Medical data: Their acquisition, storage, and use	2. Biomedical data: Their acquisition, storage, and use	2. Biomedical data: Their acquisition, storage, and use	2. Biomedical data: Their acquisition, storage, and use
3. Medical decision making: Probabilistic medical reasoning	3. Medical decision-making: Probabilistic medical reasoning	3. Biomedical decision making: Probabilistic clinical reasoning	3. Biomedical decision making: Probabilistic clinical reasoning	3. Biomedical decision making: Probabilistic clinical reasoning
4. Essential concepts for medical computing	4. Essential concepts for medical computing	5. Essential concepts for biomedical computing		

Table 30.1 (continued)

Medical Informatics: Computer Applications in Medical Care (1990)	*Medical Informatics: Computer Applications in Health Care and Biomedicine (2000)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2006)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2014)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2020)*
		4. Cognitive science and biomedical informatics	4. Cognitive science and biomedical informatics	4. Cognitive informatics
				5. Human-computer interaction, usability, and workflow
5. System design and evaluation	5. System design and engineering	6. System design and engineering in health care	5. Computer architectures for health care and biomedicine	
			6. Software engineering for health care and biomedicine	6. Software engineering for health care and biomedicine
	6. Standards in medical informatics	7. Standards in biomedical informatics	7. Standards in biomedical informatics	7. Standards in biomedical informatics
		8. Natural language and text processing in biomedicine	8. Natural language processing in health care and biomedicine	8. Natural language processing for health-related texts
		9. Imaging and structural informatics	9. Biomedical imaging informatics	10. Imaging and structural informatics
				9. Bioinformatics
				11. Personal health informatics
	7. Ethics and health informatics: Users, standards, and outcomes	10. Ethics and health informatics: Users, standards, and outcomes	10. Ethics in biomedical and health informatics: Users, standards, and outcomes	12. Ethics in biomedical and health informatics: Users, standards, and outcomes
	8. Evaluation and technology assessment	11. Evaluation and technology assessment	11. Evaluation of biomedical and health information resources	13. Evaluation of biomedical and health information resources

(continued)

Table 30.1 (continued)

Medical Informatics: Computer Applications in Medical Care (1990)	*Medical Informatics: Computer Applications in Health Care and Biomedicine (2000)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2006)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2014)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2020)*
Medical computing applications	**Medical computing applications**	**Biomedical informatics applications**	**Biomedical informatics applications**	**Biomedical informatics applications**
6. Medical-record systems	9. Computer-based patient record systems	12. Electronic health record systems	12. Electronic health record systems	14. Electronic health records
			13. Health information infrastructure	15. Health information infrastructure
7. Hospital information systems	10. Management of information in integrated delivery networks	13. Management of information in healthcare organizations	14. Management of information in health care organizations	16. Management of information in health care organizations
8. Nursing information systems	12. Patient care systems	16. Patient-care systems	15. Patient-centered care systems	17. Patient-centered care systems
9. Laboratory information systems				
10. Pharmacy systems				
	11. Public health and consumer uses of health information	15. Public health informatics and the health information infrastructure	16. Public health informatics	18. Population and public health informatics
		14. Consumer health informatics and telehealth	17. Consumer health informatics and personal health records	19. mHealth and applications
			18. Telehealth	20. Telemedicine and telehealth
11. Radiology systems	14. Imaging systems	18. Imaging systems in radiology	20. Imaging systems in radiology	22. Imaging systems in radiology
12. Patient-monitoring systems	13. Patient monitoring systems	17. Patient-monitoring systems	19. Patient monitoring systems	21. Patient monitoring systems
13. Information systems for office practice				

Table 30.1 (continued)

Medical Informatics: Computer Applications in Medical Care (1990)	*Medical Informatics: Computer Applications in Health Care and Biomedicine (2000)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2006)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2014)*	*Biomedical Informatics: Computer Applications in Health Care and Biomedicine (2020)*
14. Bibliographic-retrieval systems	15. Information retrieval systems	19. Information retrieval and digital libraries	21. Information retrieval and digital libraries	23. Information retrieval
15. Clinical decision-support systems	16. Clinical decision support systems	20. Clinical decision-support systems	22. Clinical decision-support systems	24. Clinical decision-support systems
16. Clinical research systems			26. Clinical research informatics	27. Clinical research informatics
17. Computers in medical education	17. Computers in medical education	21. Computers in medical education	23. Computers in health care education	25. Digital technology in health science education
18. Health-assessment systems				
	18. Bioinformatics	22. Bioinformatics	24. Bioinformatics	(see ► Chap. 9 under "recurrent themes in biomedical informatics", above)
			25. Translational bioinformatics	26. Translational bioinformatics
				28. Precision medicine and informatics
Medical informatics in the years ahead	**Medical informatics in the years ahead**	**Biomedical informatics in the years ahead**	**Biomedical informatics in the years ahead**	**Biomedical informatics in the years ahead**
19. Health-care financing and technology assessment	19. Health care and information technology: Growing up together	23. Health care financing and information technology: A historical perspective	27 health information technology policy	29. Health information technology policy
20. The future of computer applications in health care	20. The future of computer applications in health care	24. The future of computer applications in biomedicine	28. The future of informatics in biomedicine	30. The future of informatics in biomedicine

The next decade was revolutionary, however, and it had a profound effect on informatics. During the 1990s, the Human Genome Project made it clear that much of what needed to be accomplished in human biology and genetics could not be achieved with the use of the computational methods available or introduced at that time. Many of the informatics techniques that had been developed in the clinical world became relevant to genomics research, where investigators coined the term *bioinformatics* for their computational explorations. Thus, the field of informatics began to broaden to span both basic and applied clinical sciences. In an effort to acknowledge this evolution, the second edition of this book was renamed, with "medical care" giving way to "health care" (to acknowledge the field's growing role in prevention and public health) and the addition of "biomedicine" (to embrace the role of informatics in human biology research) (Table 30.1). In addition, a new chapter on bioinformatics was added to the edition when it appeared in 2000. Similarly, the second edition took a broader view to consider topics such as standards, ethics, integrated delivery networks, and public health. Bibliographic retrieval expanded to be information retrieval, and there were changes in emphasis in several other chapters as well.

In an attempt to acknowledge and emphasize the shared methods that applied in both the human life sciences and in clinical medicine and health, the academic discipline began to change its name from "medical informatics" to "biomedical informatics". Several departments were renamed or created with this new name for the field. Hence, when the third edition of this book appeared in 2006, it adopted the title *Biomedical Informatics*, discarding the more limited "medical informatics" focus. Although several chapters were simply updated and some were deleted, others were divided into two components (e.g., the Imaging Systems chapter from the second edition was divided into a methodologic chapter on imaging/structural informatics plus an application chapter on Imaging Systems in Radiology) (Table 30.1). Furthermore, totally new chapters were drawn from other fields, including cognitive science, natural language processing, and consumer-facing systems. In addition, chapter authorship evolved substantially as new topics were introduced and authors from earlier editions brought on coauthors whose expertise complemented their own.

The book title remained unchanged in the fourth edition in 2014 (and in this edition), but changes to the chapter titles provide more detail to what was evolving (Table 30.1). The fourth edition introduced several new topics, including telehealth, translational bioinformatics, and clinical research informatics. And the current edition has added new chapters in the areas of human-computer interaction, mHealth, and precision medicine.

Thus, a review of the titles and tables of contents of the five editions of this book, spanning 30 years with 20 chapters at the outset evolving to 30 chapters now, provides a thumbnail view of the evolution of the field as a whole. So, what evolution can we observe and what does it tell us about where we are headed? We see a field that started with a strong focus on computer programming for clinical medicine and ancillary services. The field grew to embrace research areas related to medicine and health, ranging from molecular to biologic systems to organisms, and beyond to populations. And, as with any emerging discipline, biomedical informatics began to differentiate its activities into practice (the lion's share), research (not just biomedical research informatics, but research on informatics in its own right), and education. Connections among these three types of activities and overlap across domains were often scant, as shown in Fig. 30.1. Four trends, described throughout this book, have blurred many of the distinctions (Fig. 30.2).

First, the broader field of biomedicine itself has begun to blur the distinctions among its traditional areas of research domains. This is evident in the emergence of the "translational science" philosophy, with clear recognition that each scientific endeavor builds on the discoveries made at some other, usually smaller scale. One culmination of this trend is *precision medicine* (see ▶ Chap. 28), in which discoveries at the genomic level are translated into knowledge that supports decisions for patients and populations.

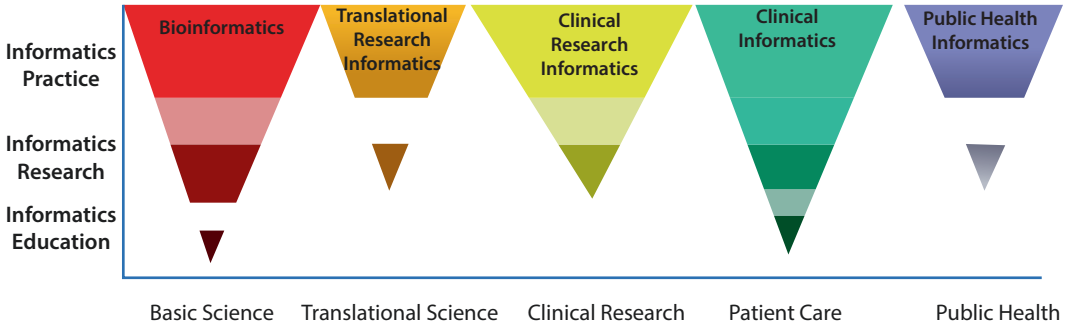


Fig. 30.1 Prior state of biomedical informatics domains and activities. Clinical informatics was the earliest and continues to be the largest domain, as reflected in this book, with other domains following in time. The advent of the Human Genome Project led to rapid expansion of the application of bioinformatics. Research in each domain followed practice, depicted here as more or less semi-transparent connections. Education in informatics, both for research and practice, began in the clinical domain, especially with nursing informatics, to be followed by nascent bioinformatics training programs. Connections between the education and research varied

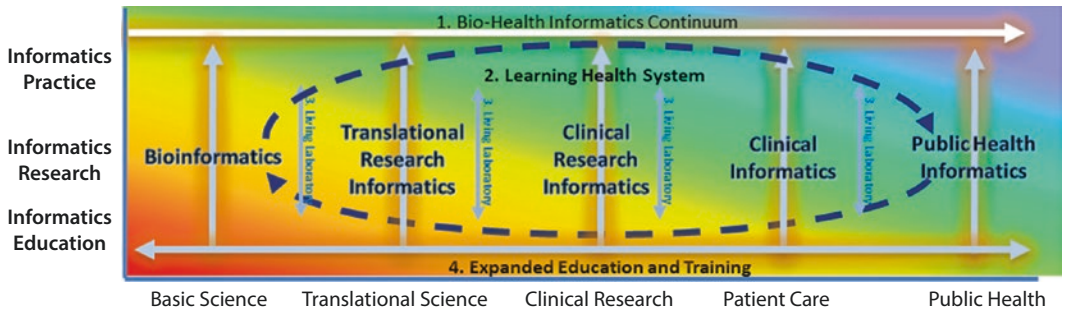


Fig. 30.2 Today, biomedical informatics is becoming a continuum, with fewer distinctions among domains and activities. This mirrors the continuum of biomedicine, with its recognition of the translational nature ranging from basic science public health (1), exemplified by precision medicine which draws on genomic knowledge to directly improve care of the individual patient. Increased clinical informatics activity has resulted in increased availability of clinical data, which informs research to produce better evidence to guide practice, resulting in a "learning health system" (2). Users of informatics applications in research and practice settings are increasingly seen as research subjects in "living laboratories" (3) for guiding the improvement of the tools they use and learning new ways to apply informatics methods. Finally, education and training in informatics is increasingly leaving the classroom and moving to practice sites for observation and learning and as settings for experimenting with new solutions (4). (► https://systems.jhu.edu/research/public-health/2019-ncov-map-faqs/. Copyright 2020, Johns Hopkins University. All rights reserved)

Second, the *learning health system* (► Chaps. 1 and 17) is making use of lare-scale data collected from patients and populations to frame research questions that are answered in the laboratory. Knowledge discovered there is then returned to the point of care to support evidence-based diagnostic, preventive and therapeutic decisions.

Third, informatics research is moving from the computer lab out to where potential users of informatics tools are actually working. Harnessing tools from cognitive science and mixed methods evaluation, the biological laboratory, the clinic, and the hospital are becoming *living informatics laboratories* for testing informatics ideas and observing their impact.

Fourth, the connections between informatics education and informatics practice are being strengthened. As electronic health records have become ubiquitous in clinical practice, computing devices and information technologies have become virtually the only tools used by every health care provider. Therefore, the importance of rigorous training in the use of these tools has increased. *Informatics training programs* are now able to train their students with the

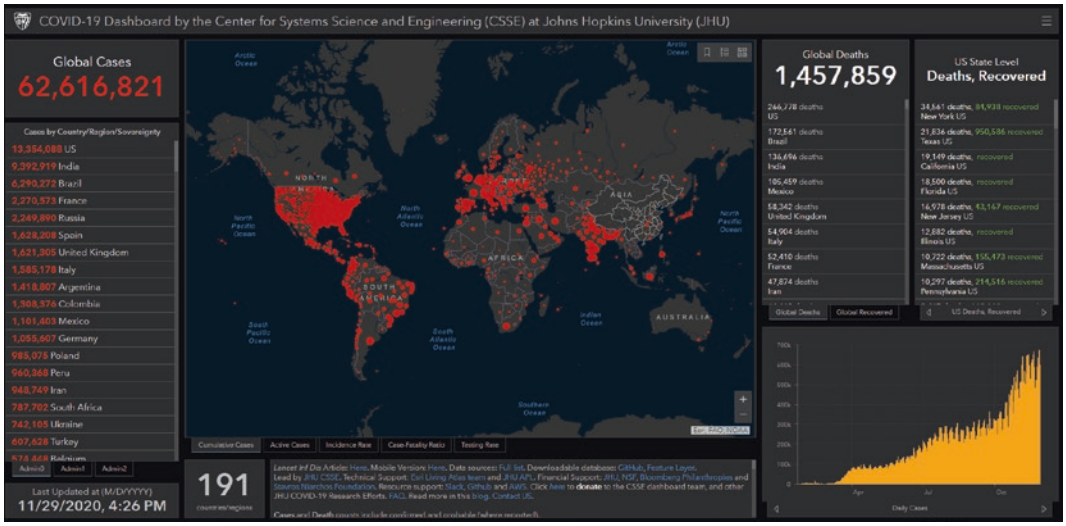


Fig. 30.3 An example of the application of informatics to increase availability of large data sets and to facilitate their processing for public consumption. Depicted is a COVID-19 dashboard developed at the Johns Hopkins University Center for Systems Science and Engineering for presenting COVID-19 data ingested from a variety of sources to allow lay people easy access to up-to-date information on the COVID-19 pandemic in their area (Dong et al. 2020). (► https://www.jhu.edu/. Copyright 2020, Johns Hopkins University. All rights reserved)

living laboratories by studying how research and patient care systems are being used and can be improved. The long tradition of formal informatics training in nursing programs is now being adopted in clinical medicine, which has recently added clinical informatics as a board-certified subspecialty through the American Board of Medical Specialties (ABMS).

30.2 Looking to the Future

Given the general trends we have outlined, what can we expect in terms of specific advances for the field? For that, we invited seven visionaries to share their predictions for the directions we will, or least should, be moving. We chose innovative thinkers who could provide insights on the future of biomedical informatics from a variety of perspectives: bioinformatics (Tarczy-Hornoch), industry (Horvitz), nursing (Murphy), health policy (Blumenthal), academic informatics (Frisse), clinical medicine (Wachter), and federal government (Brennan). Individually, they provide perspectives on government efforts, policy changes, research advances, and clinical practice. Together, they weave a rich tapestry that presages how biomedical informatics is likely to influence the twenty-first century. As it happens, shortly after these pieces were written, events have unfolded to put these predictions to the test.

Writing in early 2020, we are referring of course of the COVID-19 pandemic. The chapters in this book were largely written in the preceding year or two. Some have been updated to discuss the current situation (e.g., see ► Chap. 18), but no textbook can keep current with the rapidly unfolding events of this current natural disaster. The level of public interest in biomedical information, ranging from virology and immunology, to pharmacology and epidemiology, has risen to unprecedented levels. Up-to-the-minute data are being provided with great volume, variety and velocity (the hallmarks of "big data – see ► Chap. 13) through government, academic and news media sources for popular consumption (Fig. 30.3).[1] All of this requires rapid development and delivery of informatics solutions on an unprecedented scale, along with approaches for confirming the veracity of data.

1 ► https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 (accessed 2/13/2021).

The lessons and predictions discussed by our guest visionaries can be applied directly to the care of patients with suspected or confirmed COVID-19. For example, Tarczy-Hornoch (▶ Box 30.1) describes correlation of genomic and functional data with clinical outcomes data. Thanks to adoption and interoperability of electronic health records (▶ Chaps. 14 and 15), sufficient data are becoming available to provide an understanding of risk factors for disease severity as well as benefits risks of putative treatments much more rapidly than could be achieved with formal human subject studies (Xu et al. 2020).

Horvitz (▶ Box 30.2) provides an inventory of the methods drawn from the field of artificial intelligence (see ▶ Chaps. 1 and 24) that stand ready to use these data to infer answers to such pressing questions through machine learning. He also describes supplemental methods for "machine teaching" that can be brought to bear when some data remain sparse (Feijóo et al. 2020).

Murphy (▶ Box 30.3) describes the advantages of tele-visits for improving access to care; the social distancing required for helping to control the pandemic provides additional incentive for care at a distance, even for those who otherwise have sufficient access to in-person care. Fortunately, the technology of tele-visits (▶ Chap. 20) has progressed to the point where healthcare institutions have been able to make the necessary transition with an ease that would not have been possible 10 years earlier (Hong et al. 2020).

Blumenthal (▶ Box 30.4) enumerates issues related to safety of information systems and the government's role in developing policies that address privacy concerns (▶ Chaps. 12 and 29). The immediate need for such policies relates to the balance between individual rights and the protection of the public, as software developers race to create patient contact tracing applications (Abeler et al. 2020) and patients begin to collect their own intimate, detailed data through the wearables mentioned by Frisse (see also ▶ Chap. 19 and Ding et al. 2020).

Frisse (▶ Box 30.5) recognizes that with the success of informatics, and its growing impact on both science and health, the challenges and complexities involved with "doing it right" extend beyond the protection of data privacy. Other impacts on society may be both deep and unanticipated, with the possibility that unintended consequences may exacerbate differences among how different people, with difference education, cultures, and financial means, may experience health care and manage their own health. Unintended consequences of technology have been rampant in many fields (did anyone anticipate that television would engender generations of "couch potatoes"?).

Wachter, who is well known for his provocative book characterizing the "digital doctor" who is already somewhat upon us (Wachter 2017), focuses on the informatician (or *informaticist*) of the future and the impact that such individuals will have in clinical settings (▶ Box 30.6). He acknowledges the problems that are highlighted in his popular book, but envisions an ultimately positive future in which "the experience of being both patient and healthcare professional will be far more satisfying", due in part to the role that informatics, and those who practice this new specialty, will play in the clinical environment.

Tarczy-Hornoch, Horvitz, Murphy, Blumenthal, Frisse, and Wachter all describe ways in which clinicians' workflow can be influenced for the better through informatics, with particular attention to their quality of life, which the pandemic has demonstrated can require as much attention, for some individuals, as do the lives of the patients they serve (Dewey et al. 2020). And of course, all of these activities are supported through access to data and literature (including the works cited here), made possible by the National Institutes of Health, with the National Library of Medicine at the fore (Zayas-Cabán et al. 2020), as well as other governmental and nongovernmental organizations. As Brennan notes in her perspective (▶ Box 30.7), the federal government must provide the resources for the things that only it can do (such as gathering and consolidating epidemiologic data) and collaborate with leaders in the private sector that can provide additional breadth and depth of expertise. The COVID monitoring dashboard shown in ◘ Fig. 30.3 is just the tip of an iceberg of such cooperation (in this case between the Centers for Disease Control &

Prevention and Johns Hopkins University), when one considers all the work that went into obtaining the underlying data.

Today, biomedical informatics is moving out of the shadows. Instead of hearing "biomedical informatics? What's that?", we hear "How is biomedical informatics helping to solve this problem?" There are many answers and although they may seem specific to the current pandemic, they will remain applicable long after the current challenges are overcome. Public recognition of the importance of informatics will lead to increased resources for research, increased interest in education and training, and new opportunities for applications in biomedicine, in preparation for the inevitable challenges we know to anticipate. The intent of this book is to prepare those who wish to understand, support and lead these changes.

Box 30.1 A Perspective on the Future of Translational Bioinformatics and Precision Medicine

Peter Tarczy-Hornoch

30

The first part of the twenty-first century saw the establishment of the fields of translational bioinformatics (TBI) and precision medicine (PM), accompanied by the movement of this research into the early T-phases [1] of translational research (e.g. T0–T2 research focused on discovery and early application). The next decade will see the fields move into the later T-phases of broader adoption and diffusion (T3) and into evaluating the population impact in terms of health outcomes (T4). Due to shared core methodologies plus pressures on the health system, T3/T4 research will demonstrate a convergence between TBI/PM (predicting individual outcomes) and integration with more population-based approaches such as comparative effectiveness research and, more broadly, the concept of the learning healthcare system.

The earliest work in TBI and PM focused on the identification of opportunities and new approaches (T0), discovery to early health applications (T1), and assessment of value (T2). In the TBI area, T0 work focused on studies piloting the combination of both genomic data and electronic health record data for discovery (e.g. the early phase of the eMERGE project) and proof-of-concept T1 translational applications of genomic discoveries to clinical care (e.g. targeted pharmacogenomic decision-support systems). We also see an emerging body of T2 translational research that is beginning to assess the value of these new discoveries for health practice and for the development of evidence-based guidelines. The validation of genomic discovery and demonstration of its suitability for widespread adoption (T2/T3) is just beginning. This evolution can be illustrated on the clinical T2 front by the work of the American College of Medical Genetics, which monitors new genomic discoveries to identify what secondary findings in genome and exome sequencing meet criteria for reporting [2]. Thus far only selected mutations of around 60 genes (out of over 20,000 in the genome) meet the ACMG's rigorous criteria for clinical reporting. The clinical validation of new discoveries facilitated by TBI is a key step in the development of informatics tools that apply this knowledge to practice (e.g. decision-support tools). As an example of informatics T2 work, researchers have begun to assess the cost/benefit of genomic decision-support tools in the electronic health record [3]. The research and application of PM informatics approaches for T1 discovery and T2 application parallel those of TBI (oftentimes incorporating genomic elements as part of the input data for the development of predictive models).

In the coming decade the types and volume of data used for TBI and PM discovery and application will continue to expand and the distinctions between TBI and PM will blur even further. As the cost of genome sequencing continues to drop, increasing numbers of patients will have genotypic information available to correlate with clinical and other information, which will enable both larger scale discovery and application. In the cancer domain, for example, new single-cell sequencing approaches will provide additional granular data on a specific patient's

clonal mutational profiles. In the metabolomics and proteomics areas, the cost of gathering these data is dropping, both at the patient and more targeted (e.g. organ) levels. The ability to begin to correlate these functional data with clinical outcome data and with response-to-therapy data will provide powerful new biological and clinical insights. These new sources of biological process data will complement new sources of phenotypic and environmental data. Text mining will enable free-text notes describing phenotype and environment (e.g. social determinants of health) to be transformed into more discrete data suitable for machine learning. Increasing availability of geocoded environmental data (e.g. climate data, pollution data, air quality data, pollen counts, etc.) will enable cross-links to patient data. With patient engagement and permission (and substantial work on standards and security), specific environmental data from the Internet of Things (e.g. lighting and temperature data in a home) may also be linked with genomic, biological and clinical data for a patient. Similarly other patient data can be integrated, such as questionnaire and survey data (patient reported outcomes and mMeasures) and data from consumer wearables (counts of steps, heart rate monitoring, sleep monitoring) as well as consumer medical devices (home glucose and blood pressure monitoring, and, recently, more experimental transdermal monitoring of metabolic processes). This increase in data about individual patients, as well as the number of patients for which these rich data are available, will greatly accelerate the T0–T1 discovery and initial clinical application phases of TBI/PM. The volume of data and potential outcomes are such that the informatics tools will become ever more important for discovery. Already health care providers struggle with information overload and with the need to be current on new medical discoveries. The anticipated complexity and volume of new findings and correlations will be such that computer-based decision-support tools will be obligatory for application of these new findings. All of these approaches will fit into the paradigm of using predictions to provide early/preventive interventions that are tailored to the unique profile of the individual patient. The core methods and approaches used for analysis and discovery, and the ones used for decision support, will be fundamentally similar, whatever mix of input variables is used across the spectrum of genetic, biologic, clinical, patient provided, or environmental data. In light of this, the distinction between TBI and PM will likely vanish.

There are number of promising data analytics methods currently under development that are likely to be useful in the TBI/PM informatics area. One category is the creation of more automated model-selection and tuning methods. Without these it will be difficult to scale a number of the approaches currently being used since they are dependent on the involvement of human data scientists. Similarly, there is foundational work being done in academia and industry that is seeking better unsupervised learning approaches. These are needed because the ability to develop gold-standard training sets is now often constrained by the amount of human effort required.

Another broad category is methods that provide some explanatory power related to predictions. As one example, current machine learning approaches identify correlation but generally cannot provide insight into causation. New approaches show promise when they leverage large enough data sets to begin to infer causation. Another example is using automated tools both to develop predictive models and then to use artificial intelligence techniques to develop an explanatory model. Both these examples illustrate ways in which new methods may begin to address the concern raised by some overly opaque "black box" predictive models. A final broad category is methods that begin to leverage available data more effectively, including new AI-based image-analysis approaches, next-generation hybrid statistical and rules-based text-mining approaches, and new approaches to improve the use of temporal information in prediction algorithms (e.g. the slope and tempo of visits and laboratory values).

The rapidly rising costs of health care in the United States, without a corresponding improvement in quality, will influence the development of informatics tools for precision medicine.

Broadly this will mean that work in the TBI/PM informatics space will need to factor in the perspectives of the Quadruple Aim: (1) enhancing patient experience, (2) improving population health, (3) reducing costs, and (4) improving the work life of healthcare providers. Regarding the first of these, it will be important to ensure that predictive model-based decision-support tools are built in ways to ensure the pursuit of shared decision making involving the patient. Tools must also provide the appropriate support for ensuring that behavioral changes occur (e.g. if a model predicts the need for increased aerobic exercise, there must be methods to ensure that occurs). It will similarly be important to ensure, as data are shared and models are developed, that attention is paid to ethical, legal and social aspects of data sharing. This will help to maintain the trust of patients and to avoid unintended biases in the models (consider, for example, the recent issues with facial recognition software that works well on white males but not on women of color). The ethical lens will be particularly important to ensure that the privacy and trust of patients and public are preserved as these large scale data-intensive methods are developed and deployed. The recent academic and popular press discussions of the breeches of trust by large scale social media and other Internet companies should serve as a cautionary tale.

Regarding the next two elements in the Quadruple Aim, it will be critical that the work in TBI/PM be subject to the same kinds of assessments that we expect for other diagnostic and therapeutic interventions. Currently there are deployed tools that demonstrate a sensitivity and specificity that are far below the values that we would otherwise demand of diagnostic and screening tests. Informatics interventions have not been treated in quite the same way as laboratory tests or medications. Efforts to demonstrate value and real-world impact of TBI/PM tools will align with broader efforts to demonstrate effectiveness in the real world (e.g. Comparative Effectiveness Research). They will also form a key aspect of the Learning Healthcare System approach, since TBI/PM tools will help to assure that learning can occur from analysis of the data artifacts generated in the care-delivery process (e.g. the electronic health record and related data).

Finally, in order to address the fourth element in the Quadruple Aim, we will need to determine how best to deploy predictive analytics tools. It will be important to preserve provider decision-making autonomy, to provide sufficient explanatory ability and rigorous validation to ensure that providers trust the results, and to diminish the information and alert overload that providers face today.

In summary, we have just begun to see TBI and PM informatics discoveries and applications have an impact on achieving the broader goals of improving health and the more focused goals of the Quadruple Aim. Over the next decade, with advances in data analytics methods and increasing sources of data regarding an increasing number of patients, we are likely to see remarkable progress in the development of more easily developed and more accurate predictive models that will allow us to intervene at the patient level. These advances will be integrated into the broader trends in health care as encapsulated in the Quadruple Aim, which will require additional research and innovation to ensure that the full potential of TBI and PM are realized.

1. Khoury, M. J., Gwinn, M., Yoon, P. W., Dowling, N., Moore, C. A., & Bradley, L. (2007). The continuum of translation research in genomic medicine: How can we accelerate the appropriate integration of human genomic discoveries into health care and disease prevention? *Genetics in Medicine, 9*(10), 665–674.
2. Kalia, S. S., Adelman, K., Bale, S. J., Chung, W. K., Eng, C., Evans, J. P., et al. (2017). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): A policy statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine, 19*(2), 249–255.
3. Mathias, P. C., Tarczy-Hornoch, P., & Shirts, B. H. (2017). Modeling the costs of clinical decision support for genomic precision medicine. *Clinical Pharmacology & Therapeutics, 102*(2), 340–348.

Box 30.2 The Future of Biomedical Informatics: Bottlenecks and Opportunities

Eric Horvitz

I see the rich, interdisciplinary field of biomedical informatics as the gateway to the future of health care. The concepts, methods, rich history of contributions, and the aspirations of biomedical informatics define key opportunities ahead in biomedicine—and shine light on the path to achieving true evidence-based health care.

Progress with influences of biomedical informatics on health care over the last three decades has been slower than I had hoped. However, I remain optimistic about a forthcoming biomedical informatics revolution, made possible by a confluence of advances across industry and academia. Such a revolution will accelerate discovery in biomedicine, enhance the quality of health care, and reduce the costs of health care delivery.

From my perch as an investigator and director of a worldwide system of computer science research labs, I view key opportunities ahead as hinging on (1) addressing the often underappreciated bottleneck of *translation*—moving biomedical informatics principles and prototypes into real-world practice, and (2) making progress on persisting challenges in principles and applications of artificial intelligence (AI). I am optimistic that we will make progress on both fronts and that there will be synergies among these advances.

On challenges of translation, I believe that the difficulties of transitioning ideas and implementations from academic and industry research centers into the open world of medical practice have been widely underappreciated. Numerous factors are at play, including poor understanding of how computing solutions can assist with the tasks and day-to-day needs of health care practitioners and patients, inadequate appreciation of the needs and difficulties of developing site-specific solutions, poor compute infrastructure, and a constellation of challenges with human factors, including entrenched patterns of practice and difficulties of integrating new capabilities and services into existing clinical workflows.

Multiple advances coming with the march of computer science will help to address challenges of translating ideas and methods that have been nurtured by biomedical informaticians for decades. At the base level, such advances include ongoing leaps in computing power and in storage, but also key innovations with computing principles and methods in such subdisciplines as databases, programming languages, security and privacy, human-computer interaction, visualization, and sensing and ubiquitous computing.

Faster and more effective translation of ideas and methods from biomedical informatics will also be enabled by jumps in the quality of available computing tools and infrastructure. Increases in the power and ease-of-use of cloud computing platforms are being fueled by unprecedented investments in research and development by information technology companies—companies that are competing intensely with one another for contracts with enterprises that are hungry for digital transformation and the latest in modern computing tools. Cloud computing companies are packaging in their offerings sets of development tools and constellations of specialized services. Many of these offerings are relevant to biomedical informatics efforts, including machine learning toolkits, suites for analysis and visualization of data, and computer vision, speech recognition, and natural language analysis services made available via programmatic interfaces.

Beyond developing generic platform capabilities, cloud service providers are motivated to gain understandings in key vertical markets, such as health care, finance, and defense, and have been working to custom-tailor their general platforms with tools, designs, and services for use in specific sectors. For example, there is incentive to support rising standards on schemata (e.g., Fast Healthcare Interoperability Resources (FHIR)) for storing and transferring electronic health records and on methods to ensure the privacy of patient data. There is also pressure to develop special versions of computing services for medicine, such as language

models and, more generally, natural language capabilities specialized for medical terminology, enabling more accurate understanding and analysis of medical text and speech. Competitor cloud providers have also worked to identify and provide efficient methods and tools for important vertical needs, such as the rising importance of determining DNA sequences and interpreting protein expression data. Such special needs of researchers and clinicians have led to the availability of efficient and inexpensive cloud-computing services for genomic and proteomic analyses.

Moving on to the second realm of opportunities, around harnessing advances in the constellation of technologies that we call AI, I believe that our community can do more to leverage existing methods and also to closely follow, push, and contribute to advances in AI subfields. Beyond methods available today, key developments will be required in principles and applications to realize the long-term goals of biomedical informatics. I am seeing good progress and am optimistic that the advances coming over the next decade will be deeply enabling.

On existing technologies, and focusing on the example of developing effective decision support systems, we have been very slow to leverage the visionary ideas proposed by Robert Ledley and Lee Lusted in 1959 [1]. Ledley and Lusted provided a blueprint for constructing differential diagnoses and to use decision-theoretic analyses to generate recommendations for action. Biomedical informatics investigators have been top leaders with exploring prototypes for decision support systems, and systems constructed over 60 years of research have been shown to perform at expert levels. However, real-world impact has been limited to date. A key bottleneck has been the scarcity and cost of expertise and data. I believe that harnessing advances in machine learning will be particularly critical for delivering on the vision of evidence-based clinical decision making. Machine learning techniques available today can and should be playing a more central role in health care for assisting with pattern recognition, diagnosis, and prediction of outcomes. There are multiple opportunities to build and to integrate pipelines where data flow via machine learning to predictions and via automated decision analyses to recommendations about testing and treatment. Making key investments to build and refine effective *data-to-prediction-to-decision* pipelines will provide great value in multiple areas of medicine [2].

Opportunities ahead for biomedical informatics include leveraging recent advances in deep learning in medical applications, especially for image recognition and natural language tasks. These multilayered neural network architectures are celebrated for providing surprising boosts in classification accuracy in multiple application areas and for easing engineering overhead, as they do not require special feature engineering. The methods have been shown to perform well for recognition in the image-centric areas of pathology and radiology. Different variants of deep learning are also being explored for building predictive models from clinical data drawn from electronic health records. Beyond direct applications, deep learning methods have led to enhanced capabilities in multiple areas of AI with relevance to goals in biomedical informatics, including key advances in computer vision, speech recognition, text summarization, and language translation.

With all of the recent fanfare about deep learning, it is easy to overlook the applicability of other machine learning methods, including probabilistic graphical models, generalized additive models, and even logistic regression for serving as the heart of predictions in recommendation engines. While excitement about deep learning is appropriate, it is important to note that the methods typically require large amounts of data of the right form and that such datasets may not be available for medical applications of interest. Other approaches have proven to be as accurate for clinical applications and provide other benefits such as providing more intelligible, explainable inferences. Also, when sufficiently large corpora of data labeled with ground truth are not available, knowledge acquisition techniques, referred to broadly as *machine teaching*, can provide value. While work

is moving forward on machine teaching, existing methods and tools can be valuable in building models for prediction and classification.

I believe that it is important to note that having access to powerful machine learning procedures may be insufficient for addressing goals in biomedical informatics. Key challenges for moving ahead with developing and deploying effective decision support systems include identifying where and when such systems would provide value, collecting sufficient amounts of the right kind of data for applications, developing and integrating automated decision analyses to move from predictions to recommendations for action [2], maintaining systems over time, developing means to build and apply learned models at multiple sites, and addressing human-factors, including formulating means for achieving smooth integration of inferences and recommendations into clinical workflows, and providing explanations of inferences to clinicians [3]. Providing explanations of predictions generated by machine-learned models is a topic of rising interest [4]. I hope to see revitalized interest and similar enthusiasm extended to addressing challenges identified in biomedical informatics with the intelligibility and explanation of the advice provided by other forms of reasoning employed in decision support systems, including logical, probabilistic, and decision-theoretic inference [5].

Key opportunities in AI research for progress with developing and fielding effective decision support systems include efforts in principles and applications of *transfer learning, unsupervised learning, and causal inference*. Transfer learning refers to methods that allow for data or task competencies learned in one area to be applied to another [6]. Unsupervised and semi-supervised learning refers to methods that can be used to build models and perform tasks without having a complete set of labeled data, such as labels about the final diagnoses of patients when working with electronic health records data. Causal inference refers to methods that can be used to identify causal knowledge, versus statistical associations that are commonly inferred from data. Advances in these areas promise to provide new sources of biomedical knowledge, and to address the challenge of data scarcity and related difficulties with the generalizability of data resources for health care applications.

On data scarcity and generalizability, an important, often underappreciated challenge in biomedical informatics is that the accuracy of diagnosis and decision support may not transfer well across institutions. In our work at Microsoft Research, we found that accuracies of a system trained on data obtained from a site can plummet when used at another location. The poor generality of datasets is based on multiple factors, including differences in patient populations—with site-specific incidence rates, covariates, and presentations of illness, site-specific capture of evidence in the electronic health record, and site-specific definitions of signs, symptoms, and lab results. As an example, we found site-specificity when my team studied the task of building models to predict the likelihood that patients being discharged from a hospital would be readmitted within 30 days. The accuracy of prediction for a model learned from a massive dataset drawn from single large urban hospital dropped when the model was applied at other hospitals. This observation of poor generalizability was behind our decision to develop a capability for performing automated, recurrent machine learning separately at each site that would rely on local data for predictions. This local train-and-test capability served as the core engine of an advisory system for readmissions management, named Readmission Manager, that was commercialized by Microsoft.

Moving forward, research on a set of methods jointly referred to as *transfer learning* may help to address challenges of data scarcity and generalizability. Transfer learning algorithms for mapping the learnings from one hospital to another show promise in medicine [6]. Such methods include *multitask learning*. Also, obtaining spanning datasets, composed of large amounts of data drawn from multiple sites, may provide effective generalization. In support of this approach, methods called *multiparty computation* have been developed that

can enable learning from multiple, privately held databases, where there is no violation of privacy among the contributing organizations.

Beyond the daily practice of health care, and uses in such applications as diagnosis and treatment, methods for learning and reasoning from data can provide the foundations for new directions in the clinical sciences via tools and analyses that identify subtle but important signals in the fusing of clinical, behavioral, environmental, genetic, and epigenetic data. I see many directions springing from applications of machine learning, reasoning, planning, and causal inference for health care delivery as well as in supporting efforts in health care policy and in the discovery of new biomedical understandings.

I remain excited about advances in biomedical informatics and see a biomedical informatics revolution on the horizon. Such a revolution will build on the glowing embers of decades of contributions and the flames of late-breaking activities that address long-term challenges and bottlenecks.

1. Ledley, R. S., & Lusted, L. B. (1959). Reasoning foundations of medical diagnosis. *Science, 130*(3366), 9–21.
2. Bayati, M., Braverman, M., Gillam, M., Mack, K. M., Ruiz, G., Smith, M. S., & Horvitz, E. (2014). Data-driven decisions for reducing readmissions for heart failure: General methodology and case study. *PLOS One Medicine, 9*(10), e109264. https://doi.org/10.1371/journal.pone.0109264
3. Teach, R. L., & Shortliffe, E. H. (1981). An analysis of physician attitudes regarding computer-based clinical consultation systems. *Computers and Biomedical Research, 14*(6), 542–558. ► https://doi.org/10.1016/0010-4809(81)90012-4.
4. Caruana, R., Koch, P., Lou, Y., Sturm, M., Gehrke, J., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. *KDD*, August 10–13, 2015, Sydney, NSW, Australia.
5. Horvitz, E., Heckerman, D., Nathwani, B., & Fagan, L. M. (1986). *The use of a heuristic problem-solving hierarchy to facilitate the explanation of hypothesis-directed reasoning*. Proceedings of Medical Informatics, Washington, DC (October 1986), North Holland: New York, pp. 27–31.
6. Wiens, J., Guttag, J., & Horvitz, E. (2014). A study in transfer learning: Leveraging data from multiple hospitals to enhance hospital-specific predictions. *Journal of the American Medical Informatics Association, 21*(4), 699–706. ► https://doi.org/10.1136/amiajnl-2013-002162.

Box 30.3 The Future of Nursing Informatics

Judy Murphy

The focus of this commentary is on the future of biomedical informatics from a nursing perspective, but it is helpful to understand the background and history of nursing's role in the field. Starting there, the focus will move to looking at nursing informatics today and then looking to the future of the field from a nursing point of view.

Nurses have contributed to the purchase, design, and implementation of health information technology (IT) since the 1970s. The term "nursing informatics" (NI) first appeared in the literature in the 1980s [1–3]. The definition of NI has evolved ever since, molded by maturation of the field and influenced by health policy. In a classic article that described its domain, NI was defined as the combination of nursing, information, and computer sciences to manage and process data into information and knowledge for use in nursing practice [4]. Nurses who worked in NI during that time were pioneers who often got into informatics practice because they were good clinicians, were involved in IT projects as educators or project team members or were just technically curious and willing to try new things. Their roles, titles, and responsibilities varied greatly.

A solid foundation for the NI profession continued to be laid over the ensuing 40 years. Today, informatics has been built into undergraduate nursing education and there are over a hundred schools offering post-graduate NI education. NI is recognized as a specialty by the American Nursing Association (ANA) and has a specialty certification [5]. NI is now described as the specialty that integrates nursing science with multiple information and analytical sciences to identify, define, manage, and communicate data, information, knowledge, and wisdom in nursing practice. NI supports nurses, consumers, patients, the interprofessional healthcare team, and other stakeholders in their decision-making in all roles and in all settings to achieve desired health and healthcare outcomes. This support is accomplished using information structures, information processes, and information technology [6].

NI continues to grow. In the most recent Health Information and Management Systems Society (HIMSS) NI Workforce Survey, 57% of respondents held a post-graduate degree in nursing or nursing informatics and 44% were specialty certified by ANA in NI or other nursing specialty. Another 32% were currently pursuing NI certification, and over half have been working in an informatics role for more than 7 years [7].

Since the HITECH Act of 2009, nursing informatics specialists have played a pivotal role in influencing the adoption of electronic health records (EHR) for meaningful use. Having the breadth and depth of healthcare knowledge and understanding clinical practice workflows, nurses help all clinicians understand the application and value of the EHR. Nurses have a perspective of the many venues of care and working with all care team members, as well as working with patients at different points in their care continuum. Nurses help the patient utilize health IT to improve engagement in their own care, take control of their own health and become an integral part of the decision-making process and care team. As patient advocates, nurses understand the power of the patient in a participatory role and how this can improve outcomes.

The type and quality of care that nurses provide to their patients will benefit immensely from the continued advancement of technology and informatics in healthcare. Although there are many ways those advancements will impact nursing, here are two areas that hold the greatest promise for nursing's future.

Data and the Continuous Learning Health System: Nursing research has not been as prolific as medical research, so there is a lot less known about the true impact/outcomes of nursing interventions. But now that organizations are aggregating health data electronically in an EHR and other Health IT, nurses can more easily identify practices that measurably

impact individuals by mining the data and using prescriptive, predictive and cognitive analytics to correlate actions to improved outcomes. The collection, summarization and analysis of data can be from multiple venues and sources, including social determinants and patient-generated information for personalization. Then, it's not just about impacting traditional care, but about the impact across the continuum for the individual and including public health and population health management. The learnings can be iterated back into nursing practice in months instead of years, using protocols/guidelines, documentation templates, and clinical decision support – making it easier to do the right thing and 'hard-wiring' new best practices – thus, creating a continuous learning health system.

Care Coordination and Healthcare Anywhere: The advancement of technology has provided us the opportunity to provide care anytime/anywhere and there's little question that both patients and providers are increasingly drawn to the concept of healthcare services that are virtual. This includes "visits" using communication technologies such as email, phone and videoconference, as well as telehealth technologies for remote monitoring and management of conditions or chronic disease. Coupling this with engaged patients using portals and mobile apps creates a new ecosystem for nurses and their patients to interact. Care coordination between venues of care and across the continuum will be directly impacted in a positive way. As nurses have primary responsibility for coordinating care and helping patients navigate the complexities of the healthcare system, this will be a way for them to extend their reach to more patients and to improve the quality of the care provided to each patient. Nurses can more easily close care gaps for preventive and disease management services, monitor patients' conditions while they live their lives and not just when they visit a healthcare facility, and provide consulting and educational services.

The future of nursing informatics has no bounds; technologies of all kinds will continue to evolve, and informatics will help nurses both integrate new technologies into their practice as well as manage the impact of new technologies on that practice. Informatics will help invent the future of nursing care transformation.

1. Ball, M., & Hannah, K. (1984). *Using computers in nursing*. Reston, VA: Reston Publishers.
2. Grobe, S. (1988). Nursing informatics competencies for nurse educators and researchers. In H.E. Petersen & U. Gerdin-Jelger (Eds.), *Preparing nurses for using information systems: Recommended informatics competencies*. New York: National League for Nursing.
3. Hannah, K. J. (1985). Current trends in nursing informatics: Implications for curriculum planning. In K. J. Hannah, E. J. Guillemin, & D. N. Conklin (Eds.), *Nursing uses of computers and information science*. Amsterdam: Elsevier.
4. Graves, J., & Corcoran, S. (1989). The study of nursing informatics. *Image: Journal of Nursing Scholarship, 21*(4), 227–231.
5. ANCC. (2018). *Informatics nursing certification*. Retrieved from ► https://www.nursingworld.org/our-certifications/informatics-nurse/.
6. ANA. (2014). *Nursing informatics: Scope and standards of practice* (2nd ed.). Silver Spring: ANA.
7. HIMSS. (2017). *Nursing informatics workforce survey.* Retrieved from ► https://www.himss.org/ni-workforce-survey.

Box 30.4 Biomedical Informatics: The Future of the Field from a Health Policy Perspective

David Blumenthal

Policy issues and developments in the United States will be vital to the evolution and efficacy of health information technology (HIT) in the future. This is true because health policy has made HIT a mainstream feature of the U.S. health care system and a vital tool for improving it.

Two types of health policy issues will vitally affect the future of HIT, its uses and its benefits. The first type is generic to the U.S. health care system but will indirectly affect how HIT evolves. The second type of policy issue focuses particularly on HIT.

Generic policy issues include payment reform and the push toward consumer empowerment. There is an urgent need for payment reform to address issues such as the high costs associated with the U.S. health system. HIT has the potential to be a powerful tool in health system improvement but whether that potential is exploited will depend on the needs and priorities of its users, especially health care providers. In a fee-for-service environment, where volume and revenue maximization are prioritized, purchasers of HIT will demand that it serve these purposes. The requirement to capture detailed information for billing purposes will be paramount to the design and configuration of electronic health records (EHRs) and other IT. Information systems will be used to assure that providers capture every billable service in a way that maximizes revenue collected.

Payment approaches that prioritize value will favor different HIT configurations, especially if those payment methods hold providers accountable through risk-sharing for the cost and quality of services. HIT will have to facilitate the capture and reporting of quality and cost information for the purpose of demonstrating the value of services provided and to manage resource use continuously over a reporting period. Interoperability and exchange of health care data will become a business imperative to the extent that accountable providers must absorb the costs of services provided to their patients at other health care facilities in their communities.

HIT for value maximization will also put much greater emphasis on improving clinical decisions so as to enhance the value of services performed. In a value-oriented environment, usable and helpful decision support will achieve a priority it has never had in the current fee-for-service environment. Another priority will likely be the capability to assess the comparative performance of clinicians within organizations so as to evaluate reasons for variation in decision-making and health care outcomes.

A bipartisan interest in making health care markets more competitive and responsive to patients' needs is also motivating a push toward patient empowerment through sharing electronic data with patients and their families. This movement is reflected in legislation and regulations that encourage providers to share EHR data with individuals or their designated third parties. The Office of the National Coordinator for Health Information Technology (ONC) issued a rule in 2015 that requires certified EHRs to have standardized application programming interfaces (APIs) [1], which will facilitate access to EHR data by patients and their agents. A new ONC rule, proposed in the spring of 2018 [2], would also discourage so called information blocking.

The growing interest in data-sharing with patients is also apparent in Apple's decision to work with 13 prominent health systems [3] to accept their patients' EHR data. Large, innovative technology companies like Apple may be able to support patient empowerment by fashioning user-friendly applications that use patients' data to inform their decision-making.

The emergence of such applications will raise a host of policy issues. Finding ways to assure the safety of these consumer-facing applications will be a critical part of consumer empowerment, and constitutes a key policy agenda. To this end, the Food and Drug Administration (FDA) is making an effort to adjust its traditional regulatory approaches for the special circumstances of HIT applications.

One example of their efforts is the Accelerated Digital Clinical Ecosystem (ADviCE), a partnership between the University of California, San Francisco, several other universities and health systems, and the FDA to share best practices and data for using, integrating, and deploying health technology services and applications. ADviCE will make recommendations on the types of data needed, data sharing, transparency, and use.

Policymakers must also find ways to protect privacy of patients, either through enforceable voluntary standards or governmental regulation of emerging private organizations, like Apple, that play the role of data stewards.

Some HIT specific policy issues are also likely to influence the future development of health information technology. On this front, the increased use of EHRs has also given rise to safety challenges, as enumerated in a recent report from the Pew Charitable Trusts [4]. For example, patients may receive the incorrect dose of a medication or clinicians may select the wrong person when inputting an order. These safety issues are probably linked with the usability of EHRs, and suggest the need for improved user-centered design focused on the needs of both clinicians and patients. The EHR certification process will likely play a role in pursuing improved safety of patient data.

To address these and other health IT safety concerns, multiple experts have proposed the establishment of a safety collaborative composed of EHR developers, hospitals, government, health practitioners, and other key organizations to work together to resolve safety problems.

Finally, policy interventions may be required to improve equity of access to benefits of HIT in rural areas and for underserved populations. Lack of connectivity and sophisticated technical support can handicap rural providers in their efforts to use advanced HIT.

With the increasing power of HIT in health care will come increased reliance on its capabilities for responding to policy challenges, both general and HIT-specific. For the most part, these challenges will stimulate evolution in HIT design that makes it even more useful and important for the future of our health care system, and its patients and providers.

1. ► https://www.healthit.gov/sites/default/files/facas/HITSC_Onc_2015_edition_final_rule_presentation_2015-11-03.pdf
2. ► https://www.reginfo.gov/public/do/eAgendaViewRule?pubId=201804&RIN=0955-AA01
3. ► https://hbr.org/2018/03/apples-pact-with-13-health-care-systems-might-actually-disrupt-the-industry
4. ► https://www.pewtrusts.org/en/research-and-analysis/reports/2017/12/improving-patient-care-through-safe-health-it

Box 30.5 Future Perspective

Mark Frisse

As this textbook demonstrates, biomedical informatics paradigms are changing. Original paradigms of necessity were moored by an environment where data sets were small; data storage was limited; computation required massive and costly hardware; and high-bandwidth network connections were rare. Most major biomedical research was conducted in large laboratories and, with a few exceptions, computational needs were limited. Health care delivery and clinical research generally took place in hospitals and large clinics both affiliated with medical schools and endowed with talent, revenues, and capital necessary for their successful operation. Payment models and reimbursement for health care operations took place behind the scenes without excessive complexity and with few burdens on providers. Public health workers, health policy researchers and related groups had access only to selective, retrospective, and often manually-collected data and limited analytic capabilities. Informatics was a select, expensive, and time-consuming endeavor. Despite great challenges, remarkable feats were accomplished.

Recent paradigms are untethered from many early constraints. Today, data sets are massive and plentiful; data storage is inexpensive and seems unlimited; computation is ubiquitous and extends from minute sensor devices to massive cloud-based virtual machines; high-bandwidth network connections are pervasive and central to American life. The range of biomedical research activities is far broader and is constrained more by funding and talent limitations than by facilities; and computation is not only central to traditional research approaches but has extended the reach of scientific investigation dramatically through the analysis of data sets ranging from molecules to genomes. Largely because of Internet-based services, more providers and other care givers have access to information they need. Public health workers, health policy researchers and other interest groups can access large and broad data sets collected in near real-time. Patients and their families can also access much more information and are have become truly central to health care; patients are speaking up, and our health system is listening.

Other academic disciplines, once working on the periphery of biomedical informatics, are converging and taking center stage. Social scientists explore care complexity both in delivery settings in and in the home. For example, cognitive scientists seek more effective and efficient ways of managing care tasks. Operations research professionals seek to improve patient access, scheduling, workflow analysis, capacity management, throughput, and systems science. Behavioral psychologists are studying how mobile technologies can "nudge" patients and providers into better behaviors. As a result, informatics has become even more imaginative, extensive, rigorous, broad, accessible, and inexpensive.

The accomplishments have been many, the future seems bright, and the potential for societal good is promising. But, to paraphrase novelist William Gibson, this bright future is not now nor will it quickly become evenly distributed. Both in biomedicine and in society at large, new paradigms and technologies transformed commerce, interpersonal communication, social interactions, and behaviors have upended almost every aspect of society. By integrating and analyzing the multiple data streams emerging from our personal behavior, communication, reading habits, purchasing patterns, and social interactions, data and algorithms are capable of startlingly accurate predictions that in turn can profoundly influence behavior. The velocity of these changes carries biomedical informatics – and all of society – into an uncertain future full of promise and peril.

Consider the American healthcare system. The United States has the highest per capita expenditures for health care in the world, yet, by many measures, important health care quality measures lag far behind these of other countries [1]. Despite significant advances in technology and clinical informatics, this trend continues. This may be due in part because technologies are

not capable only of reducing complexity, they are also capable of introducing additional complexity whether such complexity is warranted or not. One cannot argue against effective informatics support for prescribing decisions; biology and the clinical condition warrant extreme detail to complexity. Similarly, knowledge of total and out-of-pocket drug costs would be helpful if patients were presented with choices, but it is difficult to rationalize the hundreds (if not thousands) of different formularies imposed by health plans. One can argue that effective measurement of outcomes and care metrics is essential for demonstrably increasing quality of care, but the value of many quality metrics is uncertain and the administrative burdens imposed on clinicians who must collect these data borders on the intolerable, often coming at the expense of patient interaction. As the economist Uwe Reinhardt wrote: "I have been at many conferences at which concerned clinicians explore so-called 'evidence-based medicine,' replete with 'evidence-based best-clinical-practice guidelines' and the associated 'clinical pathways.' I cannot recall a conference on the topic of 'evidence-based best administrative practices,' (although I may have missed it.)" [2].

Consider the future role of the traditional institution-centric electronic health record. Federal incentives greatly accelerated the introduction of EHRs into hospitals and clinics and made transactions like e-Prescribing routine. Data and communications standards allow communication across different clinical systems and expand capabilities for medication management, care coordination and other clinical activities outside of hospitals and clinics. Common EHR data elements and organizational data warehouses are simplifying secondary data use for quality reporting, administration, population health, research, and other uses. Web portals, mobile communications, and patient-accessible EHRs are engaging patients and their families to a greater degree. But this rapid introduction of EHRs has been a mixed blessing. Critics claim that EHRs focus on administrative and payment at the expense of providing the cognitive support patients and clinicians desperately need. EHRs cannot simply continue their current approach at the expense of providing the cognitive support patients and clinicians desperately need. To improve clinician morale and productivity, the urge to introduce even further unnecessary administrative burdens on care providers must be resisted.

Given the many turbulent transformations in care delivery methods, care delivery organizations, and patient-centered health technologies, many clinical informatics advances will be the realized through extension of traditional EHRs and still others will be the product of experimentation with clinical technologies address immediate consumer-directed needs and view EHR connectivity as secondary objective. Since both models will be introduced, evaluated, and adopted, one must understand how informatics can influence the evolution of many different types of clinical systems.

The ascendancy of data science has been a central theme of biomedical informatics. Broadly construed, these activities expand fundamental biomedical informatics activities through the introduction of new technologies and techniques. Findings emerging from increasingly interoperable clinical databases like i2b2, OMOP, and PCORNet further stimulate essential large-scale, collaborative data standardization and ontology development. These in turn will simplify the inclusion of a broader array of personal, environmental, and biologic computable knowledge structures. Machine learning and related disciplines arising from these activities foster discovery of previously unknown medication interactions, genetic propensities, behavioral risks, predictions, and actionable care interventions.

Social networks and other forms of informal communication are having similar impacts. In principle, these networks can gather isolated individuals sharing common concerns and can reinforce positive behaviors and combat impediments to health – social isolation, misinformation, and costs. Some forms of "digital group therapy" or "group telemedicine" may be particularly well-suited in these circumstances.

A dazzling array of new technologies must also be understood and when appropriate intro-

duced into clinical research and care delivery. The collection, integration, and analysis of new data streams produced by these devices are already being used to manage diet, weight, exercise, and even cardiac rhythm problems. Untethered from traditional EHRs, these products are producing new and valuable sources of ambiently-collected data at lower costs.

Speech and gesture recognition will simplify human-computer interaction. Ambient data collection methods simplify collection of routine data and provide additional context for documentation and interpretation. Clinician-computer interactions may be unobtrusive and allow greater focus on patients rather than computer screens. Ambient data collection – including video interpretation of clinician – patient interactions may be used to more completely summarize the clinical encounter. Image recognition technologies can diagnose skin disorders, radiographs, and some other medical images. Machine learning algorithms will reliably screen for abnormalities and complement human judgement.

We cannot can fully control how innovations will be adopted, nor can we predict their societal impact. Informatics – and innovation more broadly – is a two-edged sword.

For example, clinical systems have improved care, reduced costs, and contributed to new insights through translational informatics and data science. At the same time, they have added considerably to administrative burdens and cost, and in practice, may emphasize administrative tasks over the critical cognitive work that is the foundation of clinical medicine. At the clinical and policy level, efforts to simplify programs and processes become even more important.

Similarly, social networks and telemedicine allow previously isolated individuals to reinforce possibly socially objectionable attitudes or behaviors. But these same networks can rapidly distribute and reinforce exaggerated or false claims about the efficacy of vaccinations, treatments, and scientific evidence; these practices challenge society's very idea of a common truth.

Advances in data science and analytics, when combined with sensors and devices on the person, in the home or in public spaces raise fears that "someone/something is always watching." If data are aggregated and used by an unauthorized "data-industrial complex" working outside of socially acceptable norms, privacy rights are threatened. Better means of anonymizing data and more realistic privacy and data use policies will become even more important.

Although paradigms change, an emphasis on data, information, knowledge, and effective use remains foundational. A primary responsibility of biomedical information is to ensure that everything from data generation to knowledge generation is continually improving through greater consistency and efficiency. These improvements in turn should result in systems that more effectively address real needs and not merely automate flawed behaviors or practices. Our future depends on the extent to which we can introduce efficient means of presenting needed, reliable, and consistent information and the extent to which our efforts ensure better outcomes for individuals and society. To be effective, informatics professionals proceed based on their experience, knowledge, and values. They must, in other words, practice wisdom.

1. Schneider, E. C., Sarnak, D. O., Squires, D., Shah, A., & Doty, M. M. (2017). *Mirror, mirror 2017: International comparison reflects flaws and opportunities for better U.S. Health Care*. Commonwealth Fund. ► http://www.commonwealthfund.org/interactives/2017/july/mirror-mirror/.
2. Reinhardt, U. E. (2013, September 13). Waste vs value in American Health Care. *New York Times*. ► https://economix.blogs.nytimes.com/2013/09/13/waste-vs-value-in-american-health-care/.

Box 30.6 The Future of Health IT: A Clinical Perspective

Robert M. Wachter

About a decade ago, I hired a young clinical informaticist for a faculty position at UCSF. I told him he had an incredibly bright future, since we would soon implement a well-respected vendor-built electronic health record (EHR). I was confident that this would be exciting and important work, work that would keep him fully employed for years to come.

I didn't share with him my worry: what would his job be after the EHR was installed?

Needless to say, despite the fact that our EHR has been up and running for 6 years, he remains gainfully employed. In fact, he is busier than ever. His experience taught me something I did not understand at the time: the implementation of the EHR is merely the first step in the process of extracting value from healthcare digitization.

In fact, I have come to see the process of digitization as involving four steps:

1. Digitizing the record
2. Connecting all the digital parts ("interoperability")
3. Gaining insights from the digital data now being generated by and traversing the system
4. Taking advantage of digitization to build and/or implement new tools and approaches that deliver healthcare value (improving quality, safety, patient experience, access, and equity while also lowering costs and improving efficiency and productivity)

In the United States, the $30 billion of incentive payments distributed by the government under the HITECH Act from 2010–2014 succeeded in achieving the first step – nearly all hospitals and 90% of physician offices now use an EHR. While we see sporadic examples of activities under Steps 2, 3, and even 4, they are by far the exception.

As we look beyond the present, let's fantasize about a future in world in which we have substantially accomplished all four steps. What might our healthcare system look like? The answer is that the experience of being both patient and healthcare professional will be far more satisfying.

Let's turn first to the hospital. Much of the care that we currently think of as requiring hospitalization will undoubtedly be accomplished within less expensive settings (including the patient's home), aided by a variety of technologies ranging from clinical sensors to advanced audio and video capabilities. The hospital will mostly exist to care for very sick patients – the types we might today associate with being in the ICU. And the ICU will likely no longer be a walled off physical space. Rather, every hospital bed will be modular, capable of supporting ICU level care with the push of a few buttons.

Decision-making about who needs higher levels of care will not be left to the clinician's "eyeball test." Instead, clinicians' experience will be augmented by sophisticated AI-based prediction tools constantly humming in the background, alerting doctors and nurses that, say, a patient's probability of death just spiked up and thus she bears closer watching. Of course, taking advantage of all these AI-generated predictions will require cracking the tough nut of alert fatigue. This will be accomplished by markedly decreasing false positive rates, implementing advanced data visualization and other prioritization methods, and likely through the discovery of approaches that haven't yet been invented.

Patient rooms will have large video screens and sophisticated camera and audio equipment to allow for tele-visits. Patients and families will be able to review clinicians' notes, test results, and treatment recommendations, either on the big screen or on their hospital-issued tablet computer. Patients will not only have full access to their EHR but will also receive educational materials ("here's what to expect from your MRI tonight") and motivation ("Good job on your incentive spirometer today!") – provided by the technology. The dreaded nurse call button will be replaced by a voice-activated system in which a patient's request results in a nurse appearing on screen and even taking some actions (increasing the IV flow rate or adjusting

the bed, for example) remotely. If a new pill is needed, likely as not a robot will deliver it.

When the hospital doctor comes to visit the patient, the room's telemedicine capabilities will allow additional parties to participate. For example, a palliative care discussion can involve distant family members, the inpatient palliative care team, and a physician at an outside hospice. An infectious disease consult might involve a discussion between the patient, the hospitalist, and the ID consultant in real time, rather than the serial visits and imperfect communication through chart notes that marks current practice.

Speaking of notes, in both inpatient and outpatient settings physicians will no longer spend hours typing notes into the EHR. Rather, natural language processing technology will "listen" to the doctor-patient conversation via room-based microphones and create a useful note, improving itself over time as it learns each physicians' individual practice style and patient population ("digital scribes"). Documentation will increasingly become the byproduct of the doctor-patient encounter, not a central focus on the physician's attention.

On the other hand, clinicians will glean far more useful information and insights from their digital tools, including the EHR. As data are entered into the patient's chart, the EHR will suggest possible diagnoses and testing approaches, and guidelines and recommended treatment approaches will be a click or a voice command away. In essence, the EHR and the electronic textbook will merge into one integrated tool.

Turning to the outpatient arena, much of the care that currently requires in-person visits will be conducted via IT-enabled home care and televisits. The care of patients with chronic diseases will be utterly transformed, with a far greater emphasis on real-time, home-based, technology-enabled decision support and disease management. The heart failure patient will begin his day by weighing himself on a digital scale and answering a few questions on the computer ("How is your breathing? How did you sleep?"). It might even know how much salt he used in the past day (through the "Internet of Things"). The technology will integrate this information, along with streaming data on heart rate and blood pressure drawn from wearable or stick-on sensors, to offer recommendations for drug and activity adjustments. Ditto for patients with diabetes, emphysema, asthma, and the like.

Making all of this function will require a new workflow, and with it a new set of healthcare professionals. Sometimes called "care traffic controllers," they will be clinicians (likely nurses with advanced training in population health and some informatics) who will monitor, via advanced digital dashboards, the status of 100, or 1000 such patients, contacting and coaching the ones who seem to be having problems. The initial contacts may be generated by AI-driven algorithms and delivered by the technology, but the care traffic controller will intervene if the patient continues to have problems. For patients continuing to do poorly, a physician will become engaged. Even then, many of these encounters will be IT-enabled remote ones.

For patients with acute medical issues, much of the care will be delivered by apps, which will also offer AI-derived recommendations for simple diagnoses and interventions. Patients who require higher levels of care will see a clinician through telemedicine or community-based urgent care. Urgent care clinics will be conveniently placed in supermarkets and pharmacies, and our eventual success in achieving complete interoperability – the patient's record always available via the cloud – will enhance the ability to view the relevant parts of the EHR and to record data that then becomes available to all subsequent practitioners.

The promise of precision medicine will finally be realized. For example, the guidelines for treating a 50-year old woman with high blood pressure or elevated cholesterol will become far more complex and customized, considering a variety of patient- and population-based risk factors and large amounts of genetic information. This same complexity also means that the clinician will depend on the computer to "know" all of these variables and

suggest the best approach. Rather than remembering the correct approach to hypercholesteremia in middle-aged women (there will no longer be any one correct approach), the role of the clinician will become more about interpreting the computer's output (including intervening when it seems wrong), communicating the findings to the patient, and motivating the necessary behavioral change. Of course, this changed role will require a significant evolution in medical education.

In fact, the ability to analyze vast amounts of digital data will transform all clinical research. Rather than basing most of our treatment recommendations on small randomized clinical trials, many advances will come through analyses of actual clinical data, seeing which approaches are associated with better outcomes. Of course, this will require sophisticated adjustment for confounders, which should also be facilitated by the vast amounts of fully integrated digital clinical information. Individual healthcare systems will take advantage of these data as well, transforming themselves into so-called "learning healthcare systems" by mining their own data and experience to determine which approaches lead to the best outcomes.

The vision that I've described here is not around the corner – it is likely 10–15 years away. And achieving it will require not just the technology of a few large EHR vendors, but also the contribution of companies, large and small, some built specifically to solve specific healthcare problems, others digital giants (the Apples, Googles, and Amazons of the world) taking advantage of their capabilities in areas like app development, supply chain, and AI to attack healthcare problems.

Importantly, in such a multidimensional digital world, success cannot come simply by buying pieces of technology, peeling off the bubble wrap, and dropping them into healthcare systems and workflows. It will be up to clinical informaticists to deeply understand the needs of patients, clinicians, families, and administrators, the complexities of the technologies, and the economic, regulatory, privacy, and often ethical context. Informatics professionals will be the ones making the clinical and business case for change, and working with both vendors and clinicians to ensure that these new approaches and technologies actually achieve their aims.

This is why the job of the clinician informaticist will remain highly secure for the foreseeable future. While the job of the informaticist will no longer be to implement a core enterprise EHR, he or she will be doing something more complex and likely more valuable: reimagining the work and the workflow to take advantage of evolving digital capabilities to improve healthcare value.

Box 30.7 The Future of Biomedical Informatics from the Federal Government Perspective

Patricia Flatley Brennan

Advances in biomedical informatics, including computational bioinformatics, are essential to accelerating scientific discovery and assuring the health of society. Finally, after 40 years of promise, there is sufficient data and computing power to realize the visions of early biomedical informatics leaders that data-powered health could become a reality. Decades of slow but steady progress towards formalizing biomedical knowledge through effective use of language and messaging standards is now complemented by improvement in heuristics and algorithms that can translate those formalizations into actionable decisions. The attention of the field to key users has broadened to include basic science researchers and clinicians as well as patients and families.

As a major provider of health care services, as well as a key funder of health care services, supporter of biomedical and health related research, and guardian of key health quality initiatives, the United States federal government plays and will continue to play a significant role in advancing biomedical informatics over the next decade. Federal investment will lead to advances in data management and protection, new ways to draw knowledge out of health data, and delivery of better, accurate and complete health information at the point of need, anywhere. Perspectives of open science, ensuring economic advancement through research, and a recognition of the accountability of the government to the taxpayer are engendering a new commitment of openness and responsiveness to society.

The National Library of Medicine (NLM), one of the 27 institutes and centers at the National Institutes of Health, is key among the several federal agencies committed to ensuring the availability of high-quality data to characterize patient problems, account for health care resource expenditure and foster research driven by greater understanding of clinical phenomena. The NLM partners with other health related divisions and agencies, including the Center for Disease Control and Prevention, the Center for Medicare and Medicaid, and the Agency for Health Care Research and Quality to rapidly respond to public health threats, monitor health care expenditures and quality and foster systemic interoperability. Partnerships between the NIH and with other federal agencies outside of the health sector will allow the investments in biomedical informatics to benefit from generalized investment in data curation, large scale data management and storage, privacy and network platforms.

The NLM will do for data what it has done for the literature – making them findable, accessible, interoperable and re-usable (FAIR). These attributes, linked under the rubric of the FAIR principles, provide guidance for how a federal library makes its resources available to the public. Making data FAIR requires improved curation strategies, ones that balance automated approaches with human indexing and metadata developments in a way that takes advantage of the speed of automation while preserving human talent for the most-complex cases. The Library-of-the-Future will continue to see the NLM serving as the custodian of key collections, but also increasing its reach as a connector of important information and data resources that exist outside of its boundaries. Future developments may also lead to a discovery-on-demand approach to locating and obtaining information that has not be previously archived. The NLM will invest in research that advances use of these important collections and provides novel methodologies to interrogate them.

Some agencies within the Department of Health and Human Services, such as the Office of the National Coordinator of Health Information Technology (ONC), will continue to invest in broad, societal resources to maintain the health information infrastructure. Other agencies, such as the Center for Medicare and Medicaid Services, will continue to be both data consumers (payment schemes resting on claims for health services) as well as data contributors, making their information accessible to consumers for enhanced self-monitoring and to

researchers to foster discovery informed by care. Several trans-federal initiatives are on the horizon, designed to ensure efficient investment in scalable, re-suable information resources.

The recognition across NIH of the importance linking clinical information and biological data portends expanded investment in the methods for curating and integrating information across time within a person and across people to better understand the health of individuals and populations. Rapid growth of data from research taxes existing technical capabilities and demands additional policy development and financial investment to house important data resources. The federal government fosters policies that protect patient privacy and develop the incentive structures to accelerate the adoption of effective computer systems for health care. With the rapidly growing data generating initiatives, the federal government must take a critical role in determining how to best select and preserve the full range of information. The federal government will host public discussion and dialogs that ensure the clinical information is sufficiently broad to reflect the clinical experience of all persons. It is responsible for ensuring the cross-national arrangements needed to keep scientific exchange of health data open and free-flowing.

Interagency coordination is needed to ensure that technological advances benefit health care and that health dollars leverage investment made in other sectors. The primary point of coordination is through the Networking and Information Technology Research and Development (NITRD) Program. NITRD is a trans-agency initiative designed to provide the research and development foundations for advancing information technologies, and also to deploy those technologies in the service of the country. The NIH reports its technological research and development expenditures to the President through the NITRD program. The NIH broadly, and the NLM specifically, participate in the many workgroups that focus on broad ranging topics such as computing-enabled human interaction, communication and augmentation, cybersecurity and privacy, and high capability computing infrastructure and applications.

Federal resources should be spent on those things that only the federal government should do. These investments include short and long-term research and development programs that advance the health and well-being of society, educating the workforce of the future, and protecting key assets in perpetuity. Most of this investment is likely to occur through the NLM. In the biomedical informatics arena this means investing in research to develop method that are scalable, sustainable and reproducible, creating computational approaches to data management capable of curation at scale, developing the libraries of the future that not only encompass literature and data but also the interim product of research such as protocols, ethics and human subjects agreement, as well as novel methods of documenting research activities, such as the next generation of Jupyter notebooks. Development efforts should be applied to the ever-growing amount of text-based journal articles and reports, to devise new and creative ways to expose the literature to a variety of publics. Educational programs and efforts of the future will infuse data science and advanced biomedical informatics lessons not only in the training programs of specialists, but across the biomedical research and clinical training programs, and even extending into equipping patients and lay people with access to data and information and tools to make use of those resources.

It's worthy to note two very important trends that will shape the future of the federal engagement with health information technologies. First, there will be an increase in public private partnerships to leverage knowledge in the technical and information technology sector in support of health care. Such partnership should lead to a more robust and interoperative health information environment. Second, there will be certain roles that the federal government must preserve, such as maintaining accurate and freely-accessible information resources for the public good and overseeing the development of policies that foster data sharing while protecting individual and institutional rights.

Future federal efforts will be accompanied by collaborations with industry. These collaborations could take the form of joint investments in common problems, such as data quality or cura-

30

tion. Other forms of partnership may emerge that engage the federal investment for research and development with accelerated pathways for technology transfer. Including industrial members on Federal Advisory Committees will provide pathways for exchange of information.

The NLM will continue to play a leadership role in maintaining accurate and freely-accessible information resources. The NLM has taken a major step towards this by migrating all of our public facing information resources onto a common, sustainable technical platform. This migration will not only enhance efficiencies but also allow for increased interoperability across our resources. A common technical platform, coupled with enhancement of terminology and vocabulary systems, will make it more feasible for intended users to traverse the information resources housed here.

In the future there will be an increasing role of the federal government in protecting and preserving information in perpetuity. The enabling legislation of the NLM directs it to collect the medical knowledge of the time and store it permanently in ways that make it accessible for a wide range of users. As the largest funder of public health and heath care, the federal government indirectly shapes what constitutes health information and how it is used and valued. The federal government has two key levers for expanding the definition of what constitutes health: investing in research to demonstrate the consideration of health data, including social and behavioral predictors, on the impact of what constitutes health is a major contribution. Additionally, because of its role as a major funder of health care through the CMS, the federal government shapes what is considered of value in health care, such as research that finds ways to incorporate the social and behavioral predictors of health into routine data collection, and then to ensure the use of this information in the diagnostic, treatment and evaluation aspects of the health care process.

The future of biomedical informatics from the federal perspective is one characterized by openness, partnerships and perpetual storage of biomedical knowledge. A vibrant research program will be needed to develop and deploy the tools needed to accomplish this vision. Thoughtful deliberation is essential to protect the privacy rights of individuals while fostering the greatest degree of sharing of data and information needed to achieve the goals enabled by data driven discovery.

Suggested Readings

Cimino, J. J. (2019). Putting the "why" in "EHR": Capturing and coding clinical cognition. Journal of the American Medical Informatics Association, 26(11), 1379–1384. Cimino identifies fundamental changes that will be needed to correct the common criticisms of today's electronic health records to transform them from glorified billing diaries into true electronic assistants.

Mesko, B. The Medical Futurist. https://medicalfuturist.com/magazine (accessed June 12, 2020). Mesko's online magazine (and other postings on the Futurist's web site) provides a glimpse of technologies that are currently emerging or envisioned for the future, in many cases leveraging innovations in biomedical engineering or biomedical informatics.

Topol, E. (2016). The patient will see you now: The future of medicine is in your hands. New York: Basic Books. Topol envisions the future world that follows today's "Guttenberg moment." Much as the printing press took learning out of the hands of a special class that had access to manuscripts, the Internet and modern computing devices are doing the same for medicine, giving individuals control over their own health care.

Wachter, R. (2017). The digital doctor: Hope, hype, and harm at the dawn of medicine's computer age. New York: McGraw-Hill Education. Offers a thoughtful critique of today's modern application of digital technologies in health care, identifying today's limitations but emphasizing the promise for a greatly enhanced world for both patients and physicians.

Questions for Discussion

1. How are the advances in bioinformatics likely to affect clinical care and vice versa?
2. Identify one potential setting for an informatics "living laboratory". Who or what is the subject of evaluation? How would you "instrument" the setting to measure activity and performance?
3. Identify one area for informatics education and describe the living laboratory that would support training objectives.

References

Abeler, J., Bäcker, M., Buermeyer, U., & Zillessen, H. (2020). COVID-19 contact tracing and data protection can go together. *JMIR Mhealth and Uhealth, 8*(4), e19359.

Dewey, C., Hingle, S., Goelz, E., & Linzer, M. (2020). Supporting clinicians during the COVID-19 pandemic. *Annals of Internal Medicine, 172*, M20-1033. https://doi.org/10.7326/M20-1033.

Ding, X. R., Clifton, D., Ji, N., Lovell, N. H., Bonato, P., Chen, W., et al. (2020). Wearable sensing and telehealth technology with potential applications in the coronavirus pandemic. *IEEE Reviews in Biomedical Engineering*, 1. https://doi.org/10.1109/RBME.2020.2992838.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet. Infectious Diseases, 20*(5), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1.

Feijóo, C., Kwon, Y., Bauer, J. M., Bohlin, E., Howell, B., Jain, R., et al. (2020). Harnessing artificial intelligence (AI) to increase wellbeing for all: The case for a new technology diplomacy. *Telecomm Policy, 44*, 101988.

Hong, Y. R., Lawrence, J., Williams, D., Jr., & Mainous, I. I. I. A. (2020). Population-level interest and telehealth capacity of US hospitals in response to COVID-19: Cross-sectional analysis of Google Search and National Hospital Survey Data. *Journal of Medical Internet Research, 6*(2), e18961.

Wachter, R. (2017). *The digital doctor: Hope, hype, and harm at the dawn of medicine's computer age*. New York: McGraw-Hill Education.

Xu, B., Kraemer, M. U. G., & Open COVID-19 Data Curation Group. (2020). Open access epidemiological data from the COVID-19 outbreak. *The Lancet. Infectious Diseases, 20*(5), 534.

Zayas-Cabán, T., Abernethy, A. P., Brennan, P. F., Devaney, S., Kerlavage, A. R., Ramoni, R., & White, P. J. (2020). Leveraging the health information technology infrastructure to advance federal research priorities. *Journal of the American Medical Informatics Association, 27*(4), 647–651.

Supplementary Information

E. H. Shortliffe, J. J. Cimino (eds.), *Biomedical Informatics*, https://doi.org/10.1007/978-3-030-58721-5

Glossary

21st Century Cures Act A comprehensive bill that promotes and funds the acceleration of research into preventing and curing serious illnesses; accelerates drug and medical device development; attempts to address the opioid abuse crisis; and tries to improve mental health service delivery. It also includes a health IT-related provisions on interoperability, data sharing/exchange and electronic health records.

Abductive reasoning Can be characterized as a cyclical process of generating possible explanations or a set of hypotheses that are able to account for the available data and then each of these hypotheses is evaluated on the basis of its potential consequences. In this regard, abductive reasoning is a data-driven process that relies heavily on the domain expertise of the person.

Accountability Security function that ensures users are responsible for their access to and use of information based on a documented need and right to know.

Accountable care A descendant of managed care, accountable care is an approach to improving care and reducing costs. See: Accountable Care Organizations.

Accountable Care Organizations (ACOs) An organization of health care providers that agrees to be accountable for the quality, cost, and overall care of their patients. An ACO will be reimbursed on the basis of managing the care of a population of patients and are determined by quality scores and reductions in total costs of care.

ACO See: Accountable Care Organizations.

Active failures Errors that occur in an acute situation, the effects of which are immediately felt.

Active phase The phase of a clinical research study during which investigators collect data from participants receiving an intervention or interventions under study. It is also common to monitor study participants for adverse events during this phase.

Active storage In a hierarchical data-storage scheme, the devices used to store data that have long-term validity and that must be accessed rapidly.

Acute Physiology and Chronic Health Evaluation, Version III [APACHE III] A scoring system for rating the disease severity for particular use in intensive care units.

Adaptive learning Adapting the presentation of learning content in response to continuous assessment of the learner's performance.

Address An indicator of location; typically a number that refers to a specific position in a computer's memory or storage device; see also: Internet Address.

ADE See: Adverse Drug Events.

Admission-discharge-transfer (ADT) The core component of a hospital information system that maintains and updates the hospital census, including bed assignments of patients.

ADT See: Admission-discharge-transfer.

Advanced Cardiac Life Support A course to train providers on the procedure and set of clinical interventions for urgent treatment of cardiovascular emergencies.

Advanced Research Projects Agency Network (ARPANET) A large wide-area network created in the 1960s by the U.S. Department of Defense Advanced Research Projects Agency (DARPA) for the free exchange of information among universities and research organizations; the precursor to today's Internet.

Advanced Trauma Life Support A training program for medical providers for the management of acute trauma cases. ATLS is developed by the American College of Surgeons.

Adverse drug events (ADEs) Undesired patient events, whether expected or unexpected, that are attributed to administration of a drug.

Aggregations In the context of information retrieval, collections of content from a variety of content types, including bibliographic, full-text, and annotated material.

AHIMA See: American Health Information Management Association.

Alert message A computer-generated warning that is generated when a record meets pre- specified criteria, often referring to a potentially dangerous situation that may require action; e.g., receipt of a new laboratory test result with an abnormal value.

Algorithmic process An algorithm is a well-defined procedure or sequence of steps for solving a problem. A process that follows prescribed steps is accordingly an algorithmic process.

Alphanumeric Descriptor of data that are represented as a string of letters and numeric digits, without spaces or punctuation.

Amazon Mechanical Turk Amazon's crowdsourcing website for businesses or researchers (known as Requesters) that allows hiring of remotely located "crowdworkers" to perform discrete on-demand tasks that computers are currently unable to do.

Ambulatory medical record system (AMRS) A clinical information system designed to support all information requirements of an outpatient clinic, including registration, appointment scheduling, billing, order entry, results reporting, and clinical documentation.

American Health Information Management Association (AHIMA) Professional association devoted to the discipline of health information management (HIM).

American Heart Association A non-profit organization dedicated to improving heart health.

American Immunization Registry Association (AIRA) is a membership organization that exists to promote the development and implementation of immunization information systems (IIS) as an important tool in preventing and controlling vaccine-preventable diseases. ► https://www.immregistries.org/about-aira.

American Medical Informatics Association (AMIA) Professional association dedicated to biomedical and health informatics.

American National Standards Institute [ANSI] A private organization that oversees voluntary consensus standards.

American Public Health Association (APHA) Represents a broad array of health professionals and others who care about the health of all people and all communities. It is the leading not-for-profit public health organization in the U.S. and seeks to strengthens the impact of public health professionals and provides a science-based voice in policy debates. APHA seeks to advance prevention, reduce health disparities and promote wellness. ► http://www.apha.org/.

American Recovery and Reinvestment Act of 2009 Public Law 111–5, commonly referred to as the Stimulus or Recovery Act, this legislation was designed to create jobs quickly and to invest in the nation's infrastructure, education and healthcare capabilities.

American Standard Code for Information Interchange (ASCII) A 7-bit code for representing alphanumeric characters and other symbols.

AMIA See: American Medical Informatics Association.

AMRS See: Ambulatory medical record systems.

Analog signal A signal that takes on a continuous range of values.

Analog-to-digital conversion (ADC) Conversion of sampled values from a continuous-valued signal to a discrete-valued digital representation.

Anchoring and adjustment A heuristic used when estimating probability, in which a person first makes a rough approximation (the anchor), then adjusts this estimate to account for additional information.

Annotated content In the context of information retrieval, content that has been annotated to describe its type, subject matter, and other attributes.

Anonymize Applied to health data and information about a unique individual, the act of de-identifying or stripping away any and all data which could be used to identify that individual.

ANSI See: American National Standards Institute.

Antibiogram Pattern of sensitivity of a microorganism to various antibiotics.

APACHE III See Acute Physiology and Chronic Health Evaluation, Version III.

Apache Open source Web server software that was significant in facilitating the initial growth of the World Wide Web.

Applets Small computer programs that can be embedded in an HTML document and that will execute on the user's computer when referenced.

Application program A computer program that automates routine operations that store and organize data, perform analyses, facilitate the integration and communication of information, perform bookkeeping functions, monitor patient status, aid in education.

Application programming interface (API) A specification that enables distinct software modules or components to communicate with each other.

Applications (applied) research Systematic investigation or experimentation with the goal of applying knowledge to achieve practical ends.

Apps Software applications, especially ones downloaded to mobile devices.

Archival storage In a hierarchical data-storage scheme, the devices used to store data for long- term backup, documentary, or legal purposes.

Arden Syntax for Medical Logic Module A coding scheme or language that provides a canonical means for writing rules that relate specific patient situations to appropriate actions for practitioners to follow. The Arden Syntax standard is maintained by HL7.

Argument A word or phrase that helps complete the meaning of a predicate.

ARPANET See Advanced Research Projects Agency Network.

Artificial intelligence (AI) The branch of computer science concerned with endowing computers with the ability to simulate intelligent human behavior.

Artificial neural network A computer program that performs classification by taking as input a set of findings that describe a given situation, propagating calculated weights through a network of several layers of interconnected nodes, and generating as output a set of numbers, where each output corresponds to the likelihood of a particular classification that could explain the findings.

ASCII See: American Standard Code for Information Interchange.

Assembler A computer program that translates assembly-language programs into machine-language instructions.

Assembly language A low-level language for writing computer programs using symbolic names and addresses within the computer's memory.

Association of American Medical Colleges (AAMC) A non-profit organization that includes all US and Canadian medical colleges and many teaching hospitals, and supports them in their education and research mission.

Asynchronous Transfer Mode (ATM) A network protocol designed for sending streams of small, fixed length cells of information over very high-speed, dedicated connections, often digital optical circuits.

Audit trail A chronological record of all accesses and changes to data records, often used to promote accountability for use of, and access to, medical data.

Augmented reality Imposition of synthetic three-dimensional and text information on top of a view of the real world seen through specialized glasses worn by the learner.

Authenticated A process for positive and unique identification of users, implemented to control system access.

Authorized Within a system, a process for limiting user activities only to actions defined as appropriate based on the user's role.

Automated indexing The most common method of full-text indexing; words in a document are stripped of common suffixes, entered as items in the index, then assigned weights based on their ability to discriminate among documents (see vector-space model).

Availability In decision making, a heuristic method by which a person estimates the probability of an event based on the ease with which he can recall similar events. In security systems, a function that ensures delivery of accurate and up-to-date information to authorized users when needed.

Averaging out at chance nodes The process by which each chance node of a decision tree is replaced in the tree by the expected value of the event that it represents.

Backbone links Sections of high-capacity trunk (backbone) network that interconnect regional and local networks.

Backbone Network A high-speed communication network that carries major traffic between smaller networks.

Background question A question that asks for general information on a topic (see also: foreground question).

Backward chaining Also known as goal-directed reasoning. A form of inference used in rule-based systems in which the inference engine determines whether the premise (left-hand side) of a given rule is true by invoking other rules that can conclude the values of variables that currently are unknown and that are referenced in the premise of the given rule. The process continues recursively until all rules that can supply the required values have been considered.

Bag-of-words A language model where text is represented as a collection of words, independent of each other and disregarding word order.

Bandwidth The capacity for information transmission; the number of bits that can be transmitted per unit of time.

Baseline rate: population The prevalence of the condition under consideration in the population from which the subject was selected; individual: The frequency, rate, or degree of a condition before an intervention or other perturbation.

Basic Local Alignment Search Tool (BLAST) An algorithm for determining optimal genetic sequence alignments based on the observations that sections of proteins are often conserved without gaps and that there are statistical analyses of the occurrence of small subsequences within larger sequences that can be used to prune the search for matching sequences in a large database.

Basic research Systematic investigation or experimentation with the goal of discovering new knowledge, often by proposing new generalizations from the results of several experiments.

Basic science The enterprise of performing basic research.

Bayes' theorem An algebraic expression often used in clinical diagnosis for calculating post-test probability of a condition (a disease, for example) if the pretest probability. (prevalence) of the condition, as well as the sensitivity and specificity of the test, are known (also called Bayes' rule). Bayes' theorem also has broad applicability in other areas of biomedical informatics where probabilistic inference is pertinent, including the interpretation of data in bioinformatics.

Bayesian diagnosis program A computer-based system that uses Bayes' theorem to assist a user in developing and refining a differential diagnosis.

Before-after study (aka Historically controlled study) A study in which the evaluator attempts to draw conclusions by comparing measures made during a baseline period prior to the information resource being available and measures made after it has been implemented.

Behaviorism A social science framework for analyzing and modifying behavior.

Belief network A diagrammatic representation used to perform probabilistic inference; an influence diagram that has only chance nodes.

Best of breed An information technology strategy that favors the selection of individual applications based on their specific functionality rather than a single application that integrates a variety of functions.

Best of cluster Best of cluster became a variant of the "best of breed" strategy by selecting a single vendor for a group of similar departmental systems, such laboratory, pharmacy and radiology.

Bibliographic content In information retrieval, information abstracted from the original source.

Bibliographic database A collection of citations or pointers to the published literature.

Binary The condition of having only two values or alternatives.

Biobank A repository for biological materials that collects, processes, stores, and distributes biospecimens (usually human) for use in research.

Biocomputation The field encompassing the modeling and simulation of tissue, cell, and genetic behavior; see biomedical computing.

Bioinformatics The study of how information is represented and transmitted in biological systems, starting at the molecular level.

Biomarker A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

Biomed Central An independent publishing house specializing in the publication of electronic journals in biomedicine (see ► www.biomedcentral.com).

Biomedical computing The use of computers in biology or medicine.

Biomedical engineering An area of engineering concerned primarily with the research and

development of biomedical instrumentation and biomedical devices.

Biomedical informatics The interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, driven by efforts to improve human health.

Biomedical Information Science and Technology Initiative (BISTI) An initiative launched by the NIH in 2000 to make optimal use of computer science, mathematics, and technology to address problems in biology and medicine. It includes a consortium of senior-level representatives from each of the NIH institutes and centers plus representatives of other Federal agencies concerned with biocomputing. See: ► http://www.bisti.nih.gov.

Biomedical taxonomy A formal system for naming entities in biomedicine.

Biomolecular imaging A discipline at the intersection of molecular biology and *in vivo* imaging, it enables the visualisation of cellular function and the follow-up of the molecular processes in living organisms without perturbing them.

Biopsychosocial model A model of medical care that emphasizes not only an understanding of disease processes, but also the psychological and social conditions of the patient that affect both the disease and its therapy.

Biosample Biological source material used in experimental assays.

Biosurveillance A public health activity that monitors a population for occurrence of a rare disease of increased occurrence of a common one. Also see Public Health Surveillance and Surveillance.

Bit The logical atomic element for all digital computers.

Bit depth The number of bits that represent an individual pixel in an image; the more bits, the more intensities or colors can be represented.

Bit rate The rate of information transfer; a function of the rate at which signals can be transmitted and the efficacy with which digital information is encoded in the signal.

BLAST See: Basic Local Alignment Search Tool.

Blinding In the context of clinical research, blinding refers to the process of obfuscating from the participant and/or investigator what study intervention a given participant is receiving. This is commonly done to reduce study biases.

Blog A type of Web site that provides discussion or information on specific topics.

Blue Button A feature of the Veteran Administration's VistA system that exports an entire patient's record in electronic form.

BlueTooth A standard for the short-range wireless interconnection of mobile phones, computers, and other electronic devices.

Body The portion of a simple electronic mail message that contains the free-text content of the message.

Body of knowledge An information resource that encapsulates the knowledge of a field or discipline.

Boolean operators The mathematical operators and, or, and not, which are used to combine index terms in information retrieval searching.

Boolean searching A search method in which search criteria are logically combined using and, or, and not operators.

Bootstrap A small set of initial instructions that is stored in read-only memory and executed each time a computer is turned on.

Execution of the bootstrap is called booting the computer. By analogy, the process of starting larger computer systems.

Bottom-up An algorithm for analyzing small pieces of a problem and building them up into larger components.

Bound morpheme A morpheme that creates a different form of a word but must always occur with another morpheme (*e.g.*, *−ed*, *−s*).

B-pref A method for measuring retrieval performance in which documents without relevance judgments are excluded.

Bridge A device that links or routes signals from one network to another.

Broadband A data-transmission technique in which multiple signals may be transmitted simultaneously, each modulated within an assigned frequency range.

Browsing Scanning a database, a list of files, or the Internet, either for a particular item or for anything that seems to be of interest.

Bundled payments In the healthcare context, refers to the practice of reimbursing providers based on the total expected costs of a particular episode of care. Generally occupies a "middle ground" between fee-for-service and capitation mechanisms.

Business logic layer A conceptual level of system architecture that insulates the applications and processing components from the underlying data and the user interfaces that access the data.

Buttons Graphic elements within a dialog box or user-selectable areas within an HTML document that, when activated, perform a specified function (such as invoking other HTML documents and services).

C statistic The area under an receiver operating characteristic (ROC) curve.

CAD See: Computer-aided diagnosis.

Cadaver An embalmed human body used for teaching anatomy through the process of dissecting tissue.

Canonical form A preferred string or name for a term or collection of names; the canonical form may be determined by a set of rules (e.g., "all capital letters with words sorted in alphabetical order") or may be simply chosen arbitrarily.

Capitated payments System of health-care reimbursement in which providers are paid a fixed amount per patient to take care of all the health-needs of a population of patients.

Capitation Payments to providers, typically on an annual basis, in return for which the clinicians provide all necessary care for the patient and do not submit additional fee-for-service bills.

Cardiac output A measure of blood volume pumped out of the left or right ventricle of the heart, expressed as liters per minute.

Care coordinator See: Case Manager.

Care plan A document that provides direction for individualized patient care.

Cascading finite state automata (FSA) A tagging method in natural language processing in which as series of finite state automata are employed such that the output of one FSA becomes the input for another.

Case Refers to the capitalization of letters in a word.

Case manager A person in charge of coordinating all aspects of a patient's care.

CCD See: Continuity of Care Document.

CCOW See: Clinical Context Object Workgroup.

CDC See Centers for Disease Control and Prevention.

CDE See Common Data Element.

CDR See: Clinical data repository.

CDS Hooks A technical approach designed to invoke external CDS services from within the EHR workflow based upon a triggering event. Services may be in the form of (a) *information card*s – provide text for the user to read; (b) *suggestion card*s – provide a specific suggestion for which the EHR renders a button that the user can click to accept, with subsequent population of the change into the EHR user interface; and (c) *app link card*s – provide a link to an app.

CDSS See: Clinical decision-support system.

CDW See: Clinical data warehouse.

Cellular imaging Imaging methods that visualize cells.

Center for Medicare & Medicaid Services The Center for Medicare & Medicaid Services (CMS) is a federal agency within the United States Department of Health and Human Services that administers the Medicare program and works in partnership with state governments to administer Medicaid, the Children's Health Insurance Program, and health insurance portability standards. In addition to these programs, CMS has other responsibilities, including the administrative simplification standards from the Health Insurance Portability and Accountability Act of 1996 (HIPAA).

Centering theory A theory that attempts to explain what entities are indicated by referential expressions (such as pronouns) by noting how the center (focus of attention) of each sentence changes across the text.

Centers for Disease Control and Prevention (CDC) An agency within the US Department of Health and Human Services that provides the public with health information and promotes health through partnerships with state health departments and other organizations.

Central computer system A single system that handles all computer applications in an institution using a common set of databases and interfaces.

Central processing unit (CPU) The "brain" of the computer. The CPU executes a program stored in main memory by fetching and executing instructions in the program.

Central Test Node (CTN) DICOM software to foster cooperative demonstrations by the medical imaging vendors.

Certificate Coded authorization information that can be verified by a certification authority to grant system access.

Challenge evaluation An evaluation of information systems, often in the field of information retrieval or related areas, that provides a public test collection or gold standard data collection for various researchers to compare and analyze results.

Chance node A symbol that represents a chance event. By convention, a chance node is indicated in a decision tree by a circle.

Character sets and encodings Tables of numeric values that correspond to sets of printable or displayable characters. ASCII is one example of such an encoding.

Chart parsing A dynamic programming algorithm for structuring a sentence according to grammar by saving and reusing segments of the sentence that have been parsed.

Chat A synchronous mode of text-based communication.

Check tags In MeSH, terms that represent certain facets of medical studies, such as age, gender, human or nonhuman, and type of grant support; check tags provide additional indexing of bibliographic citations in databases such as Medline.

CHI See: Consumer health informatics.

CHIN See: Community Health Information Network.

Chunking A natural language processing method for determining non-recursive phrases where each phrase corresponds to a specific part of speech.

CINAHL (or CINHL) See: Cumulative Index to Nursing and Allied Health Literature.

CINAHL Subject Headings A set of terms based on MeSH, with additional domain-specific terms added, used for indexing the Cumulative Index to Nursing and Allied Health Literature (CINAHL).

CIS See: Clinical information system.

Citation database A database of citations found in scientific articles, showing the linkages among articles in the scientific literature.

Classification In image processing, the categorization of segmented regions of an image based on the values of measured parameters, such as area and intensity.

Classroom Technologies All technology used in a classroom setting including projection of two-dimensional slides or views of three-dimensional objects, electronic markup of a screen presentation, real time feedback systems such as class polling, and digital recording of a class session.

CLIA certification See: Clinical Laboratory Improvement Amendments of 1988 Certification.

Client–server Information processing interaction that distributes application processing between a local computer (the client) and a remote computer resource (the server).

Clinical and translational research A broad spectrum of research activities involving the translation of findings from initial laboratory- based studies into early-stage clinical studies, and subsequently, from the findings of those studies in clinical and/or population-level practice. This broad area incorporates multiple biomedical informatics sub-domains, including both translational bioinformatics and clinical research informatics.

Clinical Context Object Workgroup (CCOW) A common protocol for single sign-on implementations in health care. It allows multiple applications to be linked together, so the end user only logs in and selects a patient in one application, and those actions propagate to the other applications.

Clinical data repository (CDR) Clinical database optimized for storage and retrieval for individual patients and used to support patient care and daily operations.

Clinical data warehouse (CDW) A database of clinical data obtained from primary sources such as electron health records, organized for re-use for secondary purposes.

Clinical datum Replaces medical datum with same definition.

Clinical decision support Any process that provides health-care workers and patients with situation-specific knowledge that can inform their decisions regarding health and health care.

Clinical decision-support system (CDSS) A computer-based system that assists physicians in making decisions about patient care.

Clinical Document Architecture An HL7 standard for naming and structuring clinical documents, such as reports.

Clinical expert system A computer program designed to provide decision support for diagnosis or therapy planning at a level of sophistication that an expert physician might provide.

Clinical guidelines Systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.

Clinical informatics The application of biomedical informatics methods in the patient-care domain; a combination of computer science, information science, and clinical science designed to assist in the management and processing of clinical data, information, and knowledge to support clinical practice.

Clinical information system (CIS) The components of a health-care information system designed to support the delivery of patient care, including order communications, results reporting, care planning, and clinical documentation.

Clinical judgment Decision making by clinicians that incorporates professional experience and social, ethical, psychological, financial, and other factors in addition to the objective medical data.

Clinical Laboratory Improvement Amendments of 1988 certification Clinical Laboratory Improvement Amendments of 1988, establishing laboratory testing quality standards to ensure the accuracy, reliability and timeliness of patient test results, regardless of where the test was performed.

Clinical modifications A published set of changes to the International Classification of Diseases (ICD) that provides additional levels of detail necessary for statistical reporting in the United States.

Clinical pathway Disease-specific plan that identifies clinical goals, interventions, and expected outcomes by time period.

Clinical Quality Language An expression language standardized by HL7 that is used to characterize both quality measure logic and decision-support logic.

Clinical research The range of studies and trials in human subjects that fall into the three sub-categories: (1) Patient-oriented research: Research conducted with human subjects (or on material of human origin such as tissues, specimens and cognitive phenomena) for which an investigator (or colleague) directly interacts with human subjects. Patient-oriented research includes: (a) mechanisms of human disease; (b) therapeutic interventions; (c) clinical trial; and (d) development of new technologies. (2) Epidemiologic and behavioral studies. (3) Outcomes research and health services research.

Clinical research informatics (CRI) The application of biomedical informatics methods in the clinical research domain to support all aspects of clinical research, from hypothesis generation, through study design, study execution and data collection, data analysis, and dissemination of results.

Clinical Research Management System (CRMS) A clinical research management system is a technology platform that supports and enables the conduct of clinical research, including clinical trials, usually through a combination of functional modules targeting the preparatory, enrollment, active, and dissemination phases of such research programs. CRMS systems are often also referred to as Clinical Trials Management Systems (CTMS), particularly when they are used to manage only clinical trials rather than various types of clinical research.

Clinical subgroup A subset of a population in which the members have similar characteristics and symptoms, and therefore similar likelihood of disease.

Clinical trials Research projects that involve the direct management of patients and are generally aimed at determining optimal modes of therapy, evaluation, or other interventions.

Clinical-event monitors Systems that electronically and automatically record the occurrence or changes of specific clinical events, such as blood pressure, respiratory capability, or heart rhythms.

Clinically relevant population The population of patients that is seen in actual practice. In the context of estimating the sensitivity and specificity of a diagnostic test, that group of patients in whom the test actually will be used.

Closed loop Regulation of a physiological variable, such as blood pressure, by monitoring the value of the variable and altering therapy without human intervention.

Closed loop medication management system A workflow process (typically supported electronically) through which medications are ordered electronically by a physician, filled by the pharmacy, delivered to the patient, administered by a nurse, and subsequently monitored for effectiveness by the physician.

Cloud technology or computing Cloud computing is using computing resources located in a remote location. Typically, cloud computing is provided by a separate business, and the user pays for it on per usage basis. There are variations such as private clouds, where the "cloud" is provided by the same business, but leverages methods that permit easier virtualization and expandability than traditional methods. Private clouds are popular with healthcare because of security concerns with public cloud computing.

Clustering algorithms A method which assigns a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters.

CMS See: Center for Medicare and Medicaid Services.

Coaching system An intelligent tutoring system that monitors the session and intervenes only when the student requests help or makes serious mistakes.

Cocke-Younger-Kasami (CYK) A dynamic programming method that uses bottom-up rules for parsing grammar-free text; used only in conjunction with a grammar written in Chomsky normal form.

Code As a verb, to write a program. As a noun, the program itself.

Cognitive artifacts human-made materials, devices, and systems that extend people's abilities in perceiving objects, encoding and retrieving information from memory, and problem-solving.

Cognitive engineering An interdisciplinary approach to the development of principles, methods and tools to assess and guide the design of computerized systems to support human performance.

Cognitive heuristics Mental processes by which we learn, recall, or process information; rules of thumb.

Cognitive Informatics (CI) is an interdisciplinary field consisting of cognitive and information sciences, specifically focusing on human information processing, mechanisms and processes within the context of computing and computer applications. The focus of CI is on understanding work processes and activities within the context of human cognition and the design of interventional solutions (often engineering, computing and information technology solutions).

Cognitive load An excess of information that competes for few cognitive resources, creating a burden on working memory.

Cognitive science Area of research concerned with studying the processes by which people think and behave.

Cognitive task analysis The analysis of both the information-processing demands of a task and the kinds of domain-specific knowledge required performing it, used to study human performance.

Cognitive walkthrough (CW) An analytic method for characterizing the cognitive processes of users performing a task. The method is performed by an analyst or group of analysts "walking through" the sequence of actions necessary to achieve a goal, thereby seeking to identify potential usability problems that may impede the successful completion of a task or introduce complexity in a way that may frustrate users.

Collaborative workspace A virtual environment in which multiple participants can interact, synchronously or asynchronously, to perform a collaborative task.

Color resolution A measure of the ability to distinguish among different colors (indicated in a digital image by the number of bits per pixel). Three sets of multiple bits are required to specify the intensity of red, green, and blue components of each pixel color.

Commodity internet A general-purpose connection to the Internet, not configured for any particular purpose.

Common Data Elements (CDEs) Standards for data that stipulate the methods by which the data are collected and the controlled terminologies used to represent them. Many standard sets of CDEs have been developed, often overlapping in nature.

Communication Data transmission and information exchange between computers using accepted protocols via an exchange medium such as a telephone line or fiber optic cable.

Community Health Information Network (CHIN) A computer network developed for exchange of sharable health information among independent participant organizations in a geographic area (or community).

Comparative effectiveness research A form of clinical research that compares examines outcomes of two or more interventions to determine if one is statistically superior to another.

Compiler A program that translates a program written in a high-level programming language to a machine-language program, which can then be executed.

Comprehensibility and control Security function that ensures that data owners and data stewards have effective control over information confidentiality and access.

Computational biology The science of computer-based mathematical and statistical techniques to analyze biological systems. See also bioinformatics.

Computed check A procedure applied to entered data that detects errors based on whether values have the correct mathematical relationship; (e.g., white blood cell differential counts, reported as percentages, must sum to 100.

Computed tomography (CT) An imaging modality in which X rays are projected through the body from multiple angles and the resultant absorption values are analyzed by a computer to produce cross-sectional slices.

Computer architecture The basic structure of a computer, including memory organization, a scheme for encoding data and instructions, and control mechanisms for performing computing operations.

Computer memories Store programs and data that are being used actively by a CPU.

Computer program A set of instructions that tells a computer which mathematical and logical operations to perform.

Computer simulated patient See Virtual patient.

Computer-aided diagnosis (CAD) Any form of diagnosis in which a computer program helps suggest or rank diagnostic considerations.

Computer-based (or computerized) physician order entry (CPOE) A clinical information system that allows physicians and other clinicians to record patient-specific orders for communication to other patient care team members and to other information systems (such as test orders to laboratory systems or medication orders to pharmacy systems). Sometimes called provider order entry or practitioner order entry to emphasize such systems' uses by clinicians other than physicians.

Computer-based patient records (CPRs) An early name for electronic health records (EHRs) dating to the early 1990s.

Concept A unit of thought made explicit through the representation of properties of an object or a set of common objects. An abstract idea generalized from specific instances of objects that occur in the world.

Conceptual graph A formal notation in which knowledge is represented through explicit relationships between concepts. Graphs can be depicted with diagrams consisting of shapes and arrows, or in a text format.

Conceptual knowledge Knowledge about concepts.

Concordant test results Test results that reflect the true patient state (true-positive and true- negative results).

Conditional probability The probability of an event, contingent on the occurrence of another event.

Conditionally independent Two events, A and B, are conditionally independent if the occurrence of one does not influence the probability of the occurrence of the other, when both events are conditioned on a third event C. Thus, p[A | B,C] = p[A | C] and p[B | A,C] = p[B | C]. The conditional probability of two conditionally independent events both occur- ring is the product of the individual conditional probabilities: p[A,B | C] = p[A | C] × p[B | C]. For example, two tests for a disease are conditionally independent when the probability of the result of the second test does not depend on the result of the first test, given the disease state. For the case in which disease is present, p[second test positive | first test positive and disease present] = p[second test positive | first test negative and disease present] = p[second test positive | disease present]. More succinctly, the tests are conditionally independent if the sensitivity and specificity of one test do not depend on the result of the other test (See independent).

Conditioned event A chance event, the probability of which is affected by another chance event (the conditioning event).

Conditioning event A chance event that affects the probability of occurrence of another chance event (the conditioned event).

Confidentiality The ability of data owners and data stewards to control access to or release of private information.

Consistency check A procedure applied to entered data that detects errors based on internal inconsistencies; e.g., recognizing a problem with the recording of cancer of the prostate as the diagnosis for a female patient.

Constructivism Argues that humans generate knowledge and meaning from an interaction between their experiences and their ideas.

Constructivist Argues that humans generate knowledge and meaning from an interaction between their experiences and their ideas.

Consumer health informatics(CHI) Applications of medical informatics technologies that focus on patients or healthy individuals as the primary users.

Content In information retrieval, media developed to communicate information or knowledge.

Content based image retrieval Also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases.

Context free grammar A mathematical model of a set of strings whose members are defined as capable of being generated from a starting symbol, using rules in which a single symbol is expanded into one or more symbols.

Contingency table A 2 × 2 table that shows the relative frequencies of true-positive, true-negative, false-positive, and false-negative results.

Continuity of care The coordination of care received by a patient over time and across multiple healthcare providers.

Continuity of Care Document (CCD) An HL7 standard that enables specification of the patient data that relate to one or more encounters with the healthcare system. The CCD is used for interchange of patient information (e.g., within Health Information Exchanges). The format enables all the electronic information about a patient to be aggregated within a standardized data structure that can be parsed and interpreted by a variety of information systems.

Continuous glucose monitor (CGM) A device that automatically tracks a diabetic patient's blood glucose levels throughout the day and night using a tiny sensor inserted under the skin.

Continuum of care The full spectrum of health services provided to patients, including health maintenance, primary care, acute care, critical care, rehabilitation, home care, skilled nursing care, and hospice care.

Contract-management system A computer system used to support managed-care contracting by estimating the costs and payments associated with potential contract terms and by comparing actual with expected payments based on contract terms.

Contrast The difference in light intensity between dark and light areas of an image.

Contrast resolution A metric for how well an imaging modality can distinguish small differences in signal intensity in different regions of the image.

Control intervention In the context of clinical research, a control intervention represents the intervention (e.g. placebo, standard care, etc.) given to the group of study participants assigned to the control or comparator arm of a study. Depending on the study type, the goal is to generate data as the basis of comparison with the experimental intervention of interest in order to determine the safety, efficacy, or benefits of an experimental intervention.

Controlled terminology A finite, enumerated set of terms intended to convey information unambiguously.

Copyright law Protection of written materials and intellectual property from being copied verbatim.

Coreference chains Provide a compact representation for encoding the words and phrases in a text that all refer to the same entity.

Coreference resolution In natural language processing, the assignment of specific meaning to some indirect reference.

Correctional Telehealth The application of telehealth to the care of prison inmates, where physical delivery of the patient to the practitioner is impractical.

Covered entities Under the HIPAA Privacy Rule, a covered entity is an organization or individual that handles personal health information. Covered entities include providers, health plans, and clearinghouses.

COVID-19 A disease that was identified in late 2019 and was declared a global pandemic on March 11, 2020. COVID-19 became an international public health emergency, affecting essentially all countries on the planet. It is characterized by contagion before symptoms, high rate of transmission between human beings, variable severity among affected individuals, and relatively high mortality rate.

CPOE See: Computer-based (or Computerized) Physician (or Provider) Order Entry.

CPR (or CPRS) See: Computer-based patient records.

CPU See: Central processing unit.

CRI See: Clinical research informatics.

CRMS (or CRDMS) See: Clinical Research Management System.

Cryptographic encoding Scheme for protecting data through authentication and authorization controls based on use of keys for encrypting and decrypting information.

CT (or CAT) See: Computed tomography.

Cumulative Index to Nursing and Allied Health Literature (CINHL) A non-NLM bibliographic database the covers nursing and allied health literature, including physical therapy, occupational therapy, laboratory technology, health education, physician assistants, and medical records.

Curly Braces Problem The situation that arises in Arden Syntax where the code used to enumerate the variables required by a medical logic module (MLM) cannot describe how the variables actually derive their values from data in the EHR database. Each variable definition in an MLM has {curly braces} that enclose words in natural language that indicate the meaning of the corresponding variable. The particular database query required to supply a value for the variable must be specified by the local implementer, however. The curly braces problem makes it impossible for an MLM developed at one institution to operate at another without local modification.

Cursor A blinking region of a display monitor, or a symbol such as an arrow, that indicates the currently active position on the screen.

Cybersecurity Measures that seeks to protect against the criminal or unauthorized use of electronic data.

CYK See: Cocke-Younger-Kasami.

Dashboard A user-interface element that displays data produced by several computer programs simultaneously and that allows users to interact with those programs in standardized ways.

Data buses An electronic pathway for transferring data—for instance, between a CPU and memory.

Data capture The process of collecting data to be stored in an information system; it includes entry by a person using a keyboard and collection of data from sensors.

Data Encryption Standard (DES) A widely-used method of for securing encryption that uses a private (secret) key for encryption and requires the same key for decryption (see also, public key cryptography).

Data independence The insulation of applications programs from changes in data-storage structures and data-access strategies.

Data layer A conceptual level of system architecture that isolates the data collected and stored in the enterprise from the applications and user interfaces used to access those data.

Data Recording The documentation of information for archival or future use through mechanisms such as handwritten text, drawings, machine-generated traces, or photographic images.

Data science The field of study that uses analytic, quantitative, and domain expertise for knowledge discovery, typically using "big data" which could be structured and/or unstructured.

Database A collection of stored data—typically organized into fields, records, and files—and an associated description (schema).

Database management system (DBMS) An integrated set of programs that manages access to databases.

Data-interchange standards Adopted formats and protocols for exchange of data between independent computer systems.

Datum Any single observation of fact. A medical datum generally can be regarded as the value of a specific parameter (for example, *red-blood-cell count*) for a particular object (for example, a patient) at a given point in time.

DBMIS See: Database Management System.

DCMI See: Dublin Core Metadata Initiative.

Debugger A system program that provides traces, memory dumps, and other tools to assist programmers in locating and eliminating errors in their programs.

Decision analysis A methodology for making decisions by identifying alternatives and assessing them with regard to both the likelihood of possible outcomes and the costs and benefits of those outcomes.

Decision node A symbol that represents a choice among actions. By convention, a decision node is represented in a decision tree by a square.

Decision support The process of assisting humans in making decisions, such as interpreting clinical information or choosing a diagnostic or therapeutic action. See: Clinical Decision Support.

Decision tree A diagrammatic representation of the outcomes associated with chance events and voluntary actions.

Deductive reasoning is a process of reaching specific conclusions (e.g., a diagnosis) from a hypothesis or a set of hypotheses. Deductive logic helps in building up the consequences of each hypothesis, and this kind of reasonning is customarily regarded as a common way of evaluating diagnostic hypotheses.

De-duplicate/Deduplication The process that matches, links, and or merges data to eliminate redundancies.

De-identified aggregate data Data reports that are summarized or altered slightly in a way that makes the discernment of the identity of any of the individuals whose data was used for the report impossible or so difficult as to be extremely improbable. The process of de-identifying aggregate data is known as statistical disclosure control.

Delta check A procedure applied to entered data that detects large and unlikely differences between the values of a new result and of the previous observations; e.g., a recorded weight that changes by 100 lb in 2 weeks.

Demonstration study Study that establishes a relation—which may be associational or causal—between a set of measured variables.

Dental informatics The application of biomedical informatics methods and techniques to problems derived from the field of dentistry. Viewed as a subarea of clinical informatics.

Deoxyribonucleic acid (DNA) The genetic material that is the basis for heredity. DNA is a long polymer chemical made of four basic subunits. The sequence in which these subunits occur in the polymer distinguishes one DNA molecule from another and in turn directs a cell's production of proteins and all other basic cellular processes.

Department of Health and Human Services (DHSS) that provides the public with health information and promotes health through partnerships with state health departments and other organizations. It is the federal agency charged with protecting the health and safety of U.S. citizens, both at home and abroad. It also oversees the development and application of programs for disease prevention and control, environmental health, and health promotion and education. ► http://www.cdc.gov/.

Departmental system A system that focus on a specific niche area in the healthcare setting, such as a laboratory, pharmacy, radiology department, etc.

Dependency grammar A linguistic theory of syntax that is based on dependency relations between words, where one word in the sentence is independent and other words are dependent on that word. Generally, the verb of a sentence is independent and other words are directly or indirectly dependent on the verb.

Dependent variable (also called outcome variable) In a correlational or experimental study, the main variable of interest or outcome variable, which is thought to be affected by or associated with the independent variables (q.v.).

Derivational morphemes A morpheme that changes the meaning or part of the speech of a word (e.g., *–ful* as in painful, converting a noun to an adjective).

DES See: Data Encryption Standard.

Descriptive study One-group study that seeks to measure the value of a variable in a sample of subjects. Study with no independent variable.

Design validation A study conducted to inform the design of an information resource, e.g., a user survey.

DHHS See: Department of Health and Human Services.

Diagnosis The process of analyzing available data to determine the pathophysiologic explanation for a patient's symptoms.

Diagnosis-based reimbursement Payments to providers (typically hospitals) based on the diagnosis made by a physician at the time of admission.

Diagnosis-related group (DRG) One of almost 500 categories based on major diagnosis, length of stay, secondary diagnosis, surgical procedure, age, and types of services required. Used to determine the fixed payment per case that Medicare will reimburse hospitals for providing care to elderly patients.

Diagnostic decision-support system A computer- based system that assists physicians in rendering diagnoses; a subset of clinical decision-support systems. See clinical decision support system.

Diagnostic process The activity of deciding which questions to ask, which tests to order, or which procedures to perform, and determining the value of the results relative to associated risks or financial costs.

DICOM See: Digital Image Communications in Medicine.

Dictionary A set of terms representing the system of concepts of a particular subject field.

Differential diagnosis The set of active hypotheses (possible diagnoses) that a physician develops when determining the source of a patient's problem.

Digital computer A computer that processes discrete values based on the binary digit or bit. Essentially all modern computers are digital, but analog computers also existed in the past.

Digital divide Term referring to disparity in economic access to technology between "haves" and "have-nots".

Digital image An image that is stored as a grid of numbers, where each picture element (pixel) in the grid represents the intensity, and possibly color, of a small area.

Digital Image Communications in Medicine (DICOM) A standard for electronic exchanging digital health images, such as x-rays and CT scans.

Digital library Organized collections of electronic content, intended for specific communities or domains.

Digital object identifier (DOI) A system for providing unique identifiers for published digital objects, consisting of a prefix that is

assigned by the International DOI Foundation to the publishing entity and a suffix that is assigned and maintained by the entity.

Digital radiography (DR) The process of producing X-ray images that are stored in digital form in computer memory, rather than on film.

Digital signal A signal that takes on discrete values from a specified range of values.

Digital signal processing (DSP) An integrated circuit designed for high-speed data manipulation and used in audio communications, image manipulation, and other data acquisition and control applications.

Digital subscriber line (DSL) A digital telephone service that allows high-speed network communication using conventional (twisted pair) telephone wiring.

Digital subtraction angiography (DSA) A radiologic technique for imaging blood vessels in which a digital image acquired before injection of contrast material is subtracted pixel by pixel from an image acquired after injection. The resulting image shows only the differences in the two images, highlighting those areas where the contrast material has accumulated.

Direct entry The entry of data into a computer system by the individual who personally made the observations.

Discharge Plan A plan that supports the transition of a patient from one care facility to home or another care facility and includes evaluation of the patient by qualified personnel, discussion with the patient or his representative, planning for homecoming or transfer to another care facility, determining whether caregiver training or other support is needed, referrals to a home care agency and/or appropriate support organizations in the community, and arranging for follow-up appointments or tests.

Discourse Large portions of text forming a narrative, such as paragraphs and documents.

Discrete event simulation model A modeling approach that assesses interactions between people, typically composed of patients that have attributes and that experience events.

Discussion board An on-line environment for exchanging public messages among participants.

Discussion lists and messaging boards Online tools for asynchronous text conversation.

Disease Any condition in an organism that is other than the healthy state.

Dissemination phase During the dissemination phase of a clinical research study, investigators analyze and report upon the data generated during the active phase.

Distributed cognition A view of cognition that considers groups, material artifacts, and cultures and that emphasizes the inherently social and collaborative nature of cognition.

Distributed computer systems A collection of independent computers that share data, programs, and other resources.

DNA See: Deoxyribonucleic Acid.

DNS See: Domain name system.

Document structure The organization of text into sections.

DOI See: Digital object identifier.

Domain name system (DNS) A hierarchical name-management system used to translate computer names to Internet protocol (IP) addresses.

Doppler shift A perceived change in frequency of a signal as the signal source moves toward or away from a signal receiver.

Double blind A clinical study methodology in which neither the researchers nor the subjects know to which study group a subject has been assigned.

Double-blinded study In the context of clinical research, a double blinded study is a study in which both the investigator and participant are blinded from the assignment of an intervention. In this scenario, a trusted third party must maintain records of such study arm assignments to inform later data analyses.

Draft standard for trial use A proposal for a standard developed by HL7 that is sufficiently well defined that early adopters can use the specification in the development of HIT. Ultimately, the draft standard may be refined and put to a ballot for endorsement by the members of the organization, thus creating an official standard.

DRG See Diagnosis-Related Groups.

Drug repurposing Identifying existing drugs that may be useful for indications other than those for which they were initially approved.

Drug screening robots A scientific instrument that can perform assays with potential drugs in a highly parallel and high throughput manner.

Drug-genome interaction A relationship between a drug and a gene in which the gene product affects the activity of the drug or the drug influences the transcription of the gene.

DSA See: Digital subtraction angiography.

DSL See: Digital subscriber line.

DSP See: Digital signal processing.

Dublin Core Metadata Initiative (DCMI) A standard metadata model for indexing published documents.

Dynamic A simulation program that models changes in patient state over time and in response to students' therapeutic decisions.

Dynamic programming A computationally intensive computer-science technique used, for example, to determine optimal sequence alignments in many computational biology applications.

Dynamic transmission model A model that divides a population into compartments (for example, uninfected, infected, recovered, dead), and for transitions between compartments are governed by differential or difference equations.

Dynamical systems models Models that describe and predict the interactions over time between multiple components of a phenomenon that are viewed as a system. Dynamical systems models are often used to construct "controllers," algorithms that adjust functioning of the system (an airplane, artificial pancreas, etc.) to maximize a set of optimization criteria.

Earley parsing A dynamic programming method for parsing context-free grammar.

EBM See Evidence-Based Medicine.

EBM database For Evidence-Based Medicine database, a highly organized collection of clinical evidence to support medical decisions based on the results of controlled clinical trials.

Ecological momentary assessment (EMA) A range of methods for collecting ecologically-valid self-report by enabling study participants and patients to report their experiences in real-time, in real-world settings, over time and across contexts.

eCRF See: Electronic Case Report Form.

EDC See: Electronic Data Capture.

EDI See: Electronic data interchange.

EEG See: Electroencephalography.

EHR See: Electronic health record.

EIW See: Enterprise information warehouse.

Electroencephalography (EEG) A method for measuring the electromagnetic fields generated by the electrical activity of the neurons using a large arrays of scalp sensors, the output of which are processed to localize the source of the electrical activity inside the brain.

Electronic Case Report Form (eCRF) A computational representation of paper case report forms (CRFs) used to enable EDC.

Electronic Data Capture (EDC) EDC is the process of capturing study-related data elements via computational mechanisms.

Electronic Data interchange (EDI) Electronic exchange of standard data transactions, such as claims submission and electronic funds transfer.

Electronic Health Record (EHR) A repository of electronically maintained information about an individual's lifetime health status and health care, stored such that it can serve the multiple legitimate users of the record. See also EMR and CPR.

Electronic health record system An electronic health record and the tools used to manage the information; also referred to as a computer-based patient-record system and often shortened to electronic health record.

Electronic Medical Record (EMR) The electronic record documenting a patient's care in a provider organization such as a hospital or a physician's office. Often used interchangeably with Electronic Health Record (EHR), although EHRs refer more typically to an individual's lifetime health status and care rather than the set of particular organizationally- based experiences.

Electronic Medical Records and Genomics (eMERGE) network A network of academic institutions that is exploring the capabilities of EHRs for genomic discovery and implementation.

Electronic-long, paper-short (ELPS) A publication method which provides on the Web site supplemental material that did not appear in the print version of the journal.

ELPS See: Electronic-long, paper-short.

EMBASE A commercial biomedical and pharmacological database from ExcerptaMedica, which provides information about medical and drug-related subjects.

Emergent design Study where the design or plan of research can and does change as the study progresses. Characteristic of subjectivist studies.

Emotion detection A natural language technique for determining the mental state of the author of a text document.

EMPI See: Enterprise master patient index.

EMR (or EMRS) See: Electronic Medical Record.

EMTREE A hierarchically structured, controlled vocabulary used for subject indexing, used to index EMBASE.

Enabling technology Any technology that improves organizational processes through its use rather than on its own. Computers, for example, are useless unless "enabled" by operation systems and applications or implemented in support of work flows that might not otherwise be possible.

Encryption The process of transforming information such that its meaning is hidden, with the intent of keeping it secret, such that only those who know how to decrypt it can read it; see decryption.

Endophenotypes An observable characteristic that is tightly linked to underlying genetics and less dependent on environmental exposures or chance.

Enrichment analysis A statistical method to determine whether an a priori defined set of concepts shows statistically significant overrepresentation in descriptions of a set of items (such as genes) compared to what one would expect based on their frequency in a reference distribution.

Enrollment during enrollment of a clinical research study, potential participants are identified and research staff determine their eligibility for participation in a study, based upon the eligibility criteria described in the study protocol. If a participant is deemed eligible to participate, there are then officially "registered" for the study. It is during this phase that in some trial designs, a process of randomization and assignment to study arms occurs.

Enterprise information warehouse (EIW) A data base in which data from clinical, financial and other operational sources are collected in order to be compared and contrasted across the enterprise.

Enterprise master patient index (EMPI) An architectural component that serves as the name authority in a health-care information system composed of multiple independent systems; the EMPI provides an index of patient names and identification numbers used by the connected information systems.

Entrez A search engine from the National Center for Biotechnology Information (NCBI), at the National Library of Medicine; Entrez can be used to search a variety of life sciences databases, including PubMed.

Entry term A synonym form for a subject heading in the Medical Subject Headings (MeSH) controlled, hierarchical vocabulary.

Epidemiologic Related to the field of epidemiology.

Epidemiology The study of the patterns, causes, and effects of health and disease conditions in defined populations.

Epigenetics Heritable phenotypes that are not encoded in DNA sequence.

Epigenomics The study of heritable phenotypes that are not encoded in the organisms DNA.

e-prescribing The electronic process of generating, transmitting and filling a medical prescription.

Error analysis In natural language processing, a process for determining the reasons for false-positive and false-negative errors.

Escrow Use of a trusted third party to hold cryptographic keys, computer source code, or other valuable information to protect against loss or inappropriate access.

Ethernet A network standard that uses a bus or star topology and regulates communication traffic using the Carrier Sense Multiple Access with Collision Detection (CSMA/CD) approach.

Ethnography Set of research methodologies derived primarily from social anthropology. The basis of much of the subjectivist, qualitative evaluation approaches.

ETL See: Extract, Transform, and Load.

Evaluation contract A document describing the aims of a study, the methods to be used and resources made available, usually agreed between the evaluator and key stakeholders before the study begins.

Event-Condition-Action (ECA) rule A rule that requires some *event* (such as the availability of a new data value in the database) to cause the *condition* (premise, or left-hand side) of the rule to be evaluated. If the condition is determined to be true, then some *action* is performed. Such rules are commonly found in active database systems and form the basis of medical logic modules.

Evidence-based guidelines(EBM) An approach to medical practice whereby the best possible evidence from the medical literature is incorporated in decision making. Generally such evidence is derived from controlled clinical trials.

Exabyte 10^{18} bytes.

Exome The entire sequence of all genes within a genome, approximately 1–3% of the entire genome.

Expected value The value that is expected on average for a specified chance event or decision.

Experimental intervention In the context of clinical research, an experimental intervention represents the treatment or other intervention delivered to a participant assigned to the experimental arm of the study in order to determine the safety, efficacy, or benefits of that intervention.

Experimental science Systematic study characterized by posing hypotheses, designing experiments, performing analyses, and interpreting results to validate or disprove hypotheses and to suggest new hypotheses for study.

Extensible markup language (XML) A subset of the **Standard Generalized Markup Language** (**SGML**) from the World Wide Web Consortium (W3C), designed especially for Web documents. It allows designers to create their own custom-tailored tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.

External router A computer that resides on multiple networks and that can forward and translate message packets sent from a local or enterprise network to a regional network beyond the bounds of the organization.

External validity In the context of clinical research, external validity refers to the ability to generalize study results into clinical care.

Extract, Transform, and Load (ETL) ETL is the process by which source data is collected and manipulated so as to adhere to the structure and semantics of a receiving data construct, such as a data warehouse.

Extrinsic evaluation An evaluation of a component of a system based on an evaluation of the performance of the entire system.

F measure A measure of overall accuracy that is a combination of precision and recall.

Factual knowledge Knowledge of facts without necessarily having any in-depth understanding of their origin or implications.

False negative A negative result that occurs in a true situation. Examples include a desired entity that is missed by a search routine or a test result that appears normal when it should be abnormal.

False positive A positive result that occurs in a false situation. Examples include an inappropriate entity that is returned by a search routine or a test result that appears abnormal when it should be normal.

False-negative rate (FNR) The probability of a negative result, given that the condition under consideration is true—for example, the probability of a negative test result in a patient who has the disease under consideration.

False-negative result (FN) A negative result when the condition under consideration is true—for example, a negative test result in a patient who has the disease under consideration.

False-positive rate (FPR) The probability of a positive result, given that the condition under consideration is false—for example, the probability of a positive test result in a patient who does not have the disease under consideration.

False-positive result (FP) A positive result when the condition under consideration is false—for example, a positive test result in a patient who does not have the disease under consideration.

Fast Healthcare Interoperability Resource (FHIR) An HL7 standard for information exchange using a well-defined and limited set of resources.

FDDI See: Fiber Distributed Data Interface.

Feedback In a computer-based education program, system-generated responses, such as explanations, summaries, and references, provided to further a student's progress in learning.

Fee-for-service Unrestricted system of health care reimbursement in which payers pay provider for those services the provider has deemed necessary.

Fiber Distributed Data Interface [FDDI] A transmission standard for local area networks operating on fiberoptic cable, providing a transmission rate of 100 Mbit/s.

Fiberoptic cable A communication medium that uses light waves to transmit information signals.

Fiducial An object used in the field of view of an imaging system which appears in the image produced, for use as a point of reference or a measure.

Field In science, the setting, which may be multiple physical locations, where the work under study is carried out. In database design, the smallest named unit of data in a database. Fields are grouped together to form records.

Field function study Study of an information resource where the system is used in the context of ongoing health care. Study of a deployed system (cf. Laboratory study).

Field user effect study A study of the actual actions or decisions of the users of the resource.

File In a database, a collection of similar records.

File format Representation of data within a file; can refer to the method for individual characters and values (for example, ASCII or binary) or their organization within the file (for example, XML or text).

File server A computer that is dedicated to storing shared or private data files.

File system An organization of files within a database or on a mass storage device.

Filtering algorithms A defined procedure applied to input data to reduce the effect of noise.

Finite state automaton An abstract, computer-based representation of the state of some entity together with a set of actions that can transform the state. Collections of finite state automata can be used to model complex systems.

Fire-wall A security system intended to protect an organization's network against external threats by preventing computers in the organization's network from communicating directly with computers external to the network, and vice versa.

Flash memory card A portable electronic storage medium that uses a semiconductor chip with a standard physical interface; a convenient method for moving data between computers.

Flexnerian One of science-based acquisition of medically relevant knowledge, followed by on-the-job apprentice-style acquisition of experience, and accompanied by evolution and expansion of the curriculum to add new fields of knowledge.

Floppy disk An inexpensive magnetic disk that can be removed from the disk-drive unit and thereby used to transfer or archive files.

FM See: Frequency modulation.

fMRI See: Functional magnetic resonance imaging.

FN See: False-negative result.

Force feedback A user interface feature in which physical sensations are transmitted to the user to provide a tactile sensation as part of a simulated activity. See also Haptic feedback.

Foreground question Question that asks for general information related to a specific patient (see also background question).

Form factor Typically refers to the physical dimensions of a product. With computing devices, refers to the physical size of the device, often with specific reference to the display. For example, we would observe that the form factor of a desktop monitor is significantly larger than that of a tablet or smart phone, and therefore able to display more characters and larger graphics on the screen.

Formative evaluation An assessment of a system's behavior and capabilities conducted during the development process and used to guide future development of the system.

Forward chaining Also known as data-driven reasoning. A form of inference used in rule-based systems in which the inference engine uses newly acquired (or concluded) values of variables to invoke all rules that may reference one or more of those variables in their premises (left-hand side), thereby concluding new values for variables in the conclusions (right-hand side) of those rules. The process continues recursively until all rules whose premises may reference the variables whose values become known have been considered.

FP See: False-positive result.

FPR See: False-positive rate.

Frame An abstract representation of a concept or entity that consists of a set of attributes, called slots, each of which can have one or more values to represent knowledge about the entity or concept.

Frame Relay A high-speed network protocol designed for sending digital information over shared wide-area networks using variable length packets of information.

Free morpheme A morpheme that is a word and that does not contain another morpheme (*e.g.*, *arm*, *pain*).

Frequency modulation(FM) A signal representation in which signal values are represented as changes in frequency rather than amplitude.

Front-end application A computer program that interacts with a database-management system to retrieve and save data and to accomplish user-level tasks.

Full-text content The complete textual information contained in a bibliographic source.

Functional magnetic resonance imaging (fMRI) A magnetic resonance imaging method that reveals changes in blood oxygenation that occur following neural activity.

Functional mapping An imaging method that relates specific sites on images to particular physiologic functions.

Gateway A computer that resides on multiple networks and that can forward and translate message packets sent between nodes in networks running different protocols.

Gbps See: Gigabits per second.

GEM See: Guideline Element Model.

GenBank A centralized repository of protein, RNA, and DNA sequences in all species, cur-

rently maintained by the National Institutes of Health.

Gene expression microarray Study the expression of large numbers of genes with one another and create multiple variations on a genetic theme to explore the implications of changes in genome function on human disease.

Gene Expression Omnibus (GEO) A centralized database of gene expression microarray datasets.

Gene Ontology(GO) A structured controlled vocabulary used for annotating genes and proteins with molecular function. The vocabulary contains three distinct ontologies, Molecular Function, Biological Process and Cellular Component.

Genes Units encoded in DNA and they are transcribed into ribonucleic acid (RNA).

Genetic data An overarching term used to label various collections of facts about the genomes of individuals, groups or species.

Genetic risk score (GRS) A calculation of the likelihood of a particular phenotype being present based on a weighed score of one or more genetic variants; also referred to as a polygenic risk score (PRS).

Genome-Wide Association Studies (GWAS) An examination of many common genetic variants in different individuals to see if any variant is associated with a given trait, e.g., a disease.

Genomic medicine (also known as stratified-medicine) The management of groups of patients with shared biological characteristics, determined through molecular diagnostic testing, to select the best therapy in order to achieve the best possible outcome for a given group.

Genomics The study of all of the nucleotide sequences, including structural genes, regulatory sequences, and noncoding DNA segments, in the chromosomes of an organism.

Genomics database An organized collection of information from gene sequencing, protein characterization, and other genomic research.

Genotype The genetic makeup, as distinguished from the physical appearance, of an organism or a group of organisms.

Genotypic Refers to the genetic makeup of an organism.

GEO See: Gene Expression Omnibus.

Geographic Information System (GIS) A system designed to capture, store, manipulate, analyze, manage, and visually present all types of location-specific data.

Geographic Information System (GIS) A system designed to capture, store, manipulate, analyze, manage, and visually present all types of location-specific data.

Gigabits per second (Gbps) A common unit of measure for data transmission over high-speed networks.

Gigabyte 2^{30} or 1,073,741,824 bytes.

GIS See: Geographic Information System.

Global processing Computations on the entire image, without regard to specific regional content.

GO See: Gene Ontology.

Gold-standard test The test or procedure whose result is used to determine the true state of the subject—for example, a pathology test such as a biopsy used to determine a patient's true disease state.

Google A commercial search engine that provides free searching of documents on the World Wide Web.

GPS A system for calculating precise geographical location by triangulating information obtained from satellites and/or cell towers.

GPU See: Graphics processing unit.

Grammar A mathematical model of a potentially infinite set of strings.

Graph In computer science, a set of *nodes* or circles connected by a set of *edges* or lines.

Graphical user interface (GUI) A type of environment that represents programs, files, and options by means of icons, menus, and dialog boxes on the screen.

Graphics processing unit (GPU) A computer hardware component that performs graphic displays and other highly parallel computations.

Gray scale A scheme for representing intensity in a black-and-white image. Multiple bits per pixel are used to represent intermediate levels of gray.

Guardian Angel Proposal A proposed structure for a lifetime, patient-centered health information system.

GUI See: Graphical user interface.

Guidance In a computer-based education program, proactive feedback, help facilities, and other tools designed to assist a student in learning the covered material.

Guideline Element Model (GEM) An XML specification for marking up textual documents that describe clinical practice guidelines. The guideline-related XML tags make it possible for information systems to determine the nature of the text that has been marked up and its role in the guideline specification.

GWAS See: Genome-Wide Association Studies.

Haptic feedback A user interface feature in which physical sensations are transmitted to the user to provide a tactile sensation as part of a simulated activity.

Haptic sensation The sensation of touch or feel. It can be applied to a simulation of such sensation as presented within a virtual or augmented reality scenario.

Hard disk A magnetic disk used for data storage and typically fixed in the disk-drive unit.

Hardware The physical equipment of a computer system, including the central processing unit, memory, data-storage devices, workstations, terminals, and printers.

Harmonic mean An average of a set of weighted values in which the weights are determined by the relative importance of the contribution to the average.

HCI See: Human-computer interaction.

HCO See: Healthcare organization.

Head word The key word in a multi-word phrase that conveys the central meaning of the phrase. For example, a phrase containing adjectives and a noun, the noun is typically the head word.

Header (of email) The portion of a simple electronic mail message that contains information about the date and time of the message, the address of the sender, the addresses of the recipients, the subject, and other optional information.

Health Evaluation and Logical Processing [HELP] On of the first electronic health record systems, developed at LDS Hospital in Sal Lake City, Utah. Still in use today, it was innovative for its introduction of automated alerts.

Health informatics Used by some as a synonym for biomedical informatics, this term is increasingly used solely to refer to applied research and practice in clinical and public health informatics.

Health information and communication technology (HICT) The broad spectrum of hardware and software used to capture, store and transmit health information.

Health Information exchange (HIE) The process of moving health information electronically among disparate health care organizations for clinical care and other purposes; or an organization that is dedicated to providing health information exchange.

Health Information Infrastructure (HII) The set of public and private resources, including networks, databases, and policies, for collecting, storing, and transmitting health information.

Health Information Technology (HIT) These of computers and communications technology in healthcare and public health settings.

Health Information Technology for Economic and Clinical Health (HITECH) Also referred to as HITECH Act. Passed by the Congress as Title IV of the American Recovery and Reinvestment Act of 2009 (ARRA) in 2009, established four major goals that promote the use of health information technology: (1) Develop standards for the nationwide electronic exchange and use of health information; (2) Invest $20B in incentives to encourage doctors and hospitals to use HIT to electronically exchange patients' health information; (3) Generate $10B in savings through improvements in quality of care and care coordination, and reductions in medical errors and duplicative care and (4) Strengthen Federal privacy and security law to protect identifiable health information from misuse. Also codified the Office of the National Coordinator for Health Information Technology (ONC) within the Department of Health and Human Services.

Health Insurance Portability and Accountability Act (HIPAA) A law enacted in 1996 to protect health insurance coverage for workers and their families when they change or lose their jobs. An "administrative simplification" provision requires the Department of Health and Human Services to establish national standards for electronic healthcare transactions and national identifiers for providers, health plans, and employers. It also addresses the security and privacy of health data.

Health Level Seven (HL7) An ad hoc standards group formed to develop standards for exchange of health care data between independent computer applications; more specifically, the health care data messaging standard developed and adopted by the HL7 standards group.

Health literacy A constellation of skills, including the ability to perform basic reading, math, and everyday health tasks like comprehending prescription bottles and appointment slips, required to function in the health care environment.

Health Maintenance Organization (HMO) A group practice or affiliation of independent practitioners that contracts with patients to pro- vide comprehensive health care for a fixed periodic payment specified in advance.

Health on the Net[HON] A private organization establishing ethical standards for health information published on the World Wide Web.

Health Record Bank (HRB) An independent organization that provides a secure electronic repository for storing and maintaining an individual's lifetime health and medical records from multiple sources and assuring that the individual always has complete control over who accesses their information.

Healthcare Effectiveness Data and Information Set (HEDIS) Employers and individuals use HEDIS to measure the quality of health plans. HEDIS measures how well health plans give service to their members. HEDIS is one of health care's most widely used performance improvement tools. It is developed and maintained by the National Committee for Quality Assurance.

Healthcare organization (HCO) Any healthrelated organization that is involved in direct patient care.

Healthcare team A coordinated group of health professionals including physicians, nurses, case managers, dieticians, pharmacists, therapists, and other practitioners who collaborate in caring for a patient.

HEDIS See: Healthcare Effectiveness Data and Information Set.

HELP See Health Evaluation and Logical Processing.

HELP sector A decision rule encoded in the HELP system, a clinical information system that was developed by researchers at LDS Hospital in Salt Lake City.

Helper (plug- in) An application that are launched by a Web browser when the browser downloads a file that the browser is not able to process itself.

Heuristic A mental "trick" or rule of thumb; a cognitive process used in learning or problem solving.

Heuristic evaluation (HE) A usability inspection method, in which the system is evaluated on the basis of a small set of well-tested design principles such as visibility of system status, user control and freedom, consistency and standards, flexibility and efficiency of use.

HICT See: Health information and communication technology. From standard of care practices), so as to provide the basis for comparison to data sets derived from participants who have received an experimental intervention under study.

HIE See: Health Information Exchange.

HIE See: Health Information Exchange.

Hierarchical An arrangement between entities that conveys some superior-inferior relationship, such as parent–child, whole-part etc.

Hierarchical Task Analysis Task analytic approach that involves the breaking down of a task into sub-tasks and smaller constituted parts (e.g., sub-sub-tasks). The tasks are organized according to specific goals.

High-bandwidth An information channel that is capable of carrying delivering data at a relatively high rate.

Higher-level process A complex process comprising multiple lower-level processes.

HII See: Health Information Infrastructure.

HII See: Health Information Infrastructure.

Hindsight bias The tendency to over-estimate the prior predictability of an event, once the events has already taken place. For example, if event A occurs before event B, there may be an assumption that A predicted B.

HIPAA See: Health Insurance Portability and Accountability Act.

HIPAA See: Health Insurance Portability and Accountability Act.

HIS See: Hospital information system.

Historical control In the context of clinical research, historical controls are subjects who represent the targeted population of interest for a study. Typically, their data are derived from existing resources in a retrospective manner and that represent targeted outcomes in a non-interventional state (often resulting among humans and other elements of a system, and the profession that applies theory, principles, data, and other methods to design in order to optimize human well-being and overall system performance.

Historically controlled study See: before-after study.

HIT See: Health Information Technology.

HITECH See: Health Information Technology for Economic and Clinical Health.

HITECH regulations The components of the Health Information Technology for

Economic and Clinical Health Act, passed by the Congress in 2009, which authorized financial incentives to be paid to eligible physicians and hospitals for the adoption of "meaningful use" of EHRs in the United States. The law also called for the certification of EHR technology and for educational programs to enhance its dissemination and adoption.

HIV See: Human immunodeficiency virus.

HL7 See: Health Level 7.

HMO See: Health maintenance organization.

Home Telehealth The extension of telehealth services in to the home setting to support activities such as home nursing care and chronic disease management.

HON See: Health on the Net.

Hospital information system (HIS) Computer system designed to support the comprehensive information requirements of hospitals and medical centers, including patient, clinical, ancillary, and financial management.

Hot fail over A secondary computer system that is kept in constant synchronization with the primary system and that can take over as soon as the primary fails for any reason.

Hounsfield number The numeric information contained in each pixel of a CT image. It is related to the composition and nature of the tissue imaged and is used to represent the density of tissue.

HRB See: Health Record Bank.

HTML See HyperText.

HTTP See: HyperText Transfer Protocol.

Human factors The scientific discipline concerned with the understanding of interactions.

Human Genome Project An international undertaking, the goal of which is to determine the complete sequence of human deoxyribonucleic acid (DNA), as it is encoded in each of the 23 chromosomes.

Human immunodeficiency virus (HIV) A retrovirus that invades and inactivates helper T cells of the immune system and is a cause of AIDS and AIDS-related complex.

Human-computer interaction (HCI) Formal methods for addressing the ways in which human beings and computer programs exchange information.

Hyper Text markup language (HTML) The document specification language used for documents on the World Wide Web.

Hypertext Text linked together in a non sequential web of associations. Users can traverse highlighted portions of text to retrieve additional related information.

HyperText Transfer Protocol (HTTP) The client–server protocol used to access information on the World Wide Web.

Hypothesis generation The process of proposing a hypothesis, usually driven by some unexplained phenomenon and the derivation of a suspected underlying mechanism.

Hypothetico-deductive approach A method of reasoning made up of four stages (cue acquisition, hypothesis generation, cue interpretation, and hypothesis evaluation) which is used to generate and test hypotheses. In clinical medicine, an iterative approach to diagnosis in which physicians perform sequential, staged data collection, data interpretation, and hypothesis generation to determine and refine a differential diagnosis.

Hypothetico-deductive reasoning Reasoning by first generating and then testing a set of hypotheses to account for clinical data (i.e., reasoning from hypothesis to data).

ICANN See: Internet Corporation for Assigned Names and Numbers.

ICD-9-CM See: International Classification of Diseases, 9th Edition, Clinical Modifications.

ICMP See: Internet Control Message Protocol.

Icon In a graphical interface, a pictorial representation of an object or function.

ICT See: Information and communications technology.

IDF See: Inverse document frequency.

IDN See: Integrated delivery network.

Image acquisition The process of generating images from the modality and converting them to digital form if they are not intrinsically digital.

Image compression A mathematical process for removing redundant or relatively unimportant information from an electronic image such that the resulting file appears the same (lossless compression) or similar (lossy compression) when compared to the original.

Image content representation Makes the infor-mation in images accessible to machines for processing.

Image database An organized collection of clinical image files, such as x-rays, photographs, and microscopic images.

Image enhancement The use of global processing to improve the appearance of the image either for human use or for subsequent processing by computer.

Image interpretation/computer reasoning The process by which the individual viewing the image renders an impression of the medical significance of the results of imaging study, potentially aided by computer methods.

Image management/storage Methods for storing, transmitting, displaying, retrieving, and organizing images. The application of methods for storing, transmitting, displaying, retrieving, and organizing images.

Image metadata Data about images, such as the type of image (e.g., modality), patient that was imaged, date of imaging, image features (quantitative or qualitative), and other information pertaining to the image and its contents.

Image processing The transformation of one or more input images, either into one or more output images, or into an abstract representation of the contents of the input images.

Image quantitation The process of extracting useful numerical parameters or deriving calculations from the image or from ROIs in the image.

Image reasoning Computerized methods that use images to formulate conclusions or answer questions that require knowledge and logical inference.

Image rendering/visualization A variety of techniques for creating image displays, diagrams, or animations to display images more in a different perspective from the raw images.

Imaging informatics A subdiscipline of medical informatics concerned with the common issues that arise in all image modalities and applications once the images are converted to digital form.

IMIA See: International Medical Informatics Association.

Immersive and virtual environments A computer-based set of sensory inputs and outputs that can give the illusion of being in a different physical environment.

Immersive environment A computer-based set of sensory inputs and outputs that can give the illusion of being in a different physical environment; see; Virtual Reality.

Immersive simulated environment A computer-based set of sensory inputs and outputs that can give the illusion of being in a different physical environment.

Immersive simulated environment A teaching environment in which a student manipulates tools to control simulated instruments, producing visual, pressure, and other feedback to the tool controls and instruments.

Immunization Information System (IIS) Confidential, population based, computerized databases that record all immunization doses administered by participating providers to persons residing within a given geopolitical area. Also known as Immunization Registries.

Immunization Registry Confidential, population based, computerized databases that record all immunization doses administered by participating providers to persons residing within a given geopolitical area. Also known as Immunization Information Systems.

Implementation science Implementation science refers to the study of socio-cultural, operational, and behavioral norms and processes surrounding the dissemination and adoption of new systems, approaches and/or knowledge.

Inaccessibility A property of paper records that describes the inability to access the record by more than one person or in more than one place at a time.

Incrementalist An approach to evaluation that tolerates ambiguity and uncertainty and allows changes from day-to-day.

Independent Two events, A and B, are considered independent if the occurrence of one does not influence the probability of the occurrence of the other. Thus, $p[A \mid B] = p[A]$. The probability of two independent events A and B both occurring is given by the product of the individual probabilities: $p[A,B] = p[A] \times p[B]$. (See conditional independence.).

Independent variable In a correlational or experimental study, a variable thought to determine or be associated with the value of the dependent variable (q.v.).

Index In information retrieval, a shorthand guide to the content that allows users to find relevant content quickly.

Index Medicus The printed index used to catalog the medical literature. Journal articles are indexed by author name and subject heading, then aggregated in bound volumes. The Medline database was originally con- structed as an online version of the Index Medicus.

Index test The diagnostic test whose performance is being measured.

Indexing In information retrieval, the assignment to each document of specific terms that indicate the subject matter of the document and that are used in searching.

Indirect-care Activities of health professionals that are not directly related to patient care, such as teaching and supervising students, continuing education, and attending staff meetings.

Inductive reasoning Involves an inferential process from the observed data to account for the unobserved. It is a process of generating possible conclusions based on available data. For example, the fact that a patient who recently had major surgery has not had any fever for the last 3 days may lead us to conclude that he will not have fever tomorrow or in the immediate days that follow. The power of inductive reasoning lies in its ability to allow us to go beyond the limitations of our current evidence or knowledge to novel conclusions about the unknown.

Inference engine A computer program that reasons about a knowledge base. In the case of rule-based systems, the inference engine may perform forward chaining or backward chaining to enable the rules to infer new information about the current situation.

Inflectional morpheme A morpheme that creates a different form of a word without changing the meaning of the word or the part of speech (e.g., *–ed*, *–s*, *–ing* as in *activated*, *activates*, *activating*.).

Influence diagram A belief network in which explicit decision and utility nodes are also incorporated.

Infobutton A context-specific link from health care application to some information resource that anticipates users' needs and provides targeted information.

Infobutton manager Middleware that provides a standard software interface between infobuttons in an EHR and the documents and other information resources that the infobuttons may display for the user.

***info*RAD** The information technology and computing oriented component of the very large exhibition hall at the annual meeting of the Radiological Society of North America.

Information Organized data from which knowledge can be derived and that accordingly provide a basis for decision making.

Information and communications technology (ICT) The use of computers and communications devices to accept, store, transmit, and manipulate data; the term is roughly a synonym for information technology, but it is used more often outside the United States.

Information blocking A practice or position that interferes with exchange or accessibility of patient data or electronic health information. This was defined by the 21st Century Cures Act.

Information extraction Methods that process text to capture and organize specific information in the text and also to capture and organize specific relations between the pieces of information.

Information model A representation of concepts, relationships, constraints, rules, and operations to specify data semantics for a chosen domain of discourse. It can provide sharable, stable, and organized structure of information requirements for the domain context.

Information need In information retrieval, the searchers' expression, in their own language, of the information that they desire.

Information resource Generic term for a computer-based system that seeks to enhance health care by providing patient-specific information directly to care providers (often used equivalently with "system").

Information retrieval (IR) Methods that efficiently and effectively search and obtain data, particularly text, from very large collections or databases. It is also the science and practice of identification and efficient use of recorded media. See also Search.

Information science The field of study concerned with issues related to the management of both paper-based and electronically stored information.

Information theory The theory and mathematics underlying the processes of communication.

Information visualization The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.

Ink-jet printer Output device that uses a moveable head to spray liquid ink on paper; the head moves back and forth for each line of pixels.

Input and Output Devices, such as keyboards, pointing devices, video displays, and laser printers, that facilitate user interaction and storage or just

Input The data that represent state information, to be stored and processed to produce results (output).

Inspection method Class of usability evaluation methods in which experts appraising a system, playing the role of a user to identify potential usability and interaction issues with a system.

Institute of Medicine The health arm of the National Academy of Sciences, which provides unbiases, authoritative advice to decision makers and the public. Renamed the National Academy of Medicine in 2016.

Institutional Review Board (IRB) A committee responsible for reviewing an institution's research projects involving human subjects in order to protect their safety, rights, and welfare.

Integrated circuit A circuit of transistors, resistors, and capacitors constructed on a single chip and interconnected to perform a specific function.

Integrated delivery network (IDN) A large conglomerate health-care organization developed to provide and manage comprehensive health-care services.

Integrated Service Digital Network (ISDN) A digital telephone service that allows high-speed network communications using conventional (twisted pair) telephone wiring.

Integrative model Model for understanding a phenomenon that draws from multiple disciplines and is not necessarily based on first principles.

Intellectual property Software programs, knowledge bases, Internet pages, and other creative assets that require protection against copying and other unauthorized use.

Intelligent system See: knowledge-based system.

Intelligent Tutor A tutoring system that monitors the learning session and intervenes only when the student requests help or makes serious mistakes.

Interactome The set of all molecular interactions in a cell.

Interface engine Software that mediates the exchange of information among two or more systems. Typically, each system must know how to communicate with the interface engine, but not need to know the information format of the other systems.

Intermediate effect process of continually learning, re-learning, and exercising new knowledge, punctuated by periods of apparent decrease in mastery and declines in performance, which may be necessary for learning to take place. People at intermediate levels of expertise may perform more poorly than those at lower level of expertise on some tasks, due to the challenges of assimilating new knowledge or skills over the course of the learning process.

Internal validity In the context of clinical research, internal validity refers to the minimization of potential biases during the design and execution of the trial.

International Classification of Diseases, 9th Edition, Clinical Modifications A US extension of the World Health Organization's International Classification of Diseases, 9th Edition.

International Medical Informatics Association (IMIA) An international organization dedicated to advancing biomedical and health informatics; an "organization of organizations", it's members are national informatics societies and organizations, such as AMIA.

International Organization for Standards (ISO) The international body for information and other standards.

Internet A worldwide collection of gateways and networks that communicate with each other using the TCP/IP protocol, collectively providing a range of services including electronic mail and World Wide Web access.

Internet address See Internet Protocol Address.

Internet Control Message Protocol (ICMP) A network-level Internet protocol that provides error correction and other information relevant to processing data packets.

Internet Corporation for Assigned Names and Numbers (ICANN) The organization responsible for managing Internet domain name and IP address assignments.

Internet of Things (IoT) A system of interconnected computing devices that can transfer data and be controlled over a network. In the consumer space, IoT technologies are most commonly found in the built environment where devices and appliances (such as lighting fixtures, security systems or thermostats) can be controlled via smartphones or smart speakers, creating "smart" homes or offices.

Internet protocol The protocol within TCP/IP that governs the creation and routing of data packets and their reassembly into data messages.

Internet Protocol address A 32-bit number that uniquely identifies a computer connected to the Internet. Also called "Internet address" or "IP address".

Internet service provider (ISP) A commercial communications company that supplies fee-for-service Internet connectivity to individuals and organizations.

Internet standards The set of conventions and protocols all Internet participants use to enable effective data communications.

Internet Support Group (ISG) An on-line forum for people with similar problems, challenges or conditions to share supportive resources.

Interoperability The 21st Century Cures Act defines interoperability as health information technology that—(A) enables the secure exchange of electronic health information with, and use of electronic health information from, other health information technology without special effort on the part of the user; (B) allows for complete access, exchange, and use of all electronically accessible health information for authorized use under applicable State or Federal law; and (C) does not constitute information blocking.

Interpreter A program that converts each statement in a high-level program to a machine- language representation and then executes the binary instruction(s).

Interventional radiology A subspecialty of radiology that uses imaging to guide invasive diagnostic or therapeutic procedures.

Intrinsic evaluation An evaluation of a component of a system that focuses only on the performance of the component. See also Extrinsic Evaluation.

Intuitionist–pluralist or de-constructivist A philosophical position that holds that there is no truth and that there are as many legitimate interpretations of observed phenomena as there are observers.

Inverse document frequency (IDF) A measure of how infrequently a term occurs in a document collection.

$$\mathrm{IDFi} = \log\left(\frac{\text{number of documents}}{\text{number of documents with term}}\right)+1$$

IOM See: Institute of Medicine.

IP address See: Internet Protocol Address.

IR See: Information retrieval.

IRB See: Institutional Review Board.

ISDN See: Integrated Service Digital Network.

ISG See: Internet Support Group.

ISO See: InternationalOrganization for Standards.

Iso-semantic mapping A relationship between an entity in one dataset or model and an

entity in another dataset or model where the meaning of the two entities is identical, even if the syntax or lexical form is different.

ISP See: Internet service provider.

Job A set of tasks submitted by a user for processing by a computer system.

Joint Commission (JC) An independent, not-for-profit organization, The Joint Commission accredits and certifies more than 19,000 health care organizations and pro- grams in the United States. Joint Commission accreditation and certification is recognized nationwide as a symbol of quality that reflects an organization's commitment to meeting certain performance standards. The Joint Commission was formerly known as JCAHO (the Joint Commission for the Accreditation of Healthcare Organizations).

Just-in-time adaptive interventions (JITAIs) An intervention design that aims to provide the type of support that is most likely to be helpful in a particular context at times when users are most likely to be receptive to that support, by adapting intervention provision to an individual's changing internal and contextual state.

Just-in-time learning An approach to providing necessary information to a user at the moment it is needed, usually through anticipation of the need.

Kernel The core of the operating system that resides in memory and runs in the background to supervise and control the execution of all other programs and direct operation of the hardware.

Key field A field in the record of a file that uniquely identifies the record within the file.

Key Performance Indicator (KPI) A metric defined to be an important factor in the success of an organization. Typically, several Key Performance indicators are displayed on a Dashboard.

Keyboard A data-input device used to enter alphanumeric characters through typing.

Keyword A word or phrase that conveys special meaning or to refer to information that is relevant to such a meaning (as in an index).

Kilobyte 2^{10} or 1024 bytes.

Knowledge Relationships, facts, assumptions, heuristics, and models derived through the formal or informal analysis (or interpretation) of observations and resulting information.

Knowledge acquisition The information-elicitation and modeling process by which developers interact with subject-matter experts to create electronic knowledge bases.

Knowledge base A collection of stored facts, heuristics, and models that can be used for problem solving.

Knowledge graph A kind of knowledge representation in which entities are encoded as nodes in a graph and relationships among entities are encoded as links between the nodes.

Knowledge-based information Information derived and organized from observational or experimental research.

Knowledge-based system A program that symbolically encodes, in a knowledge base, facts, heuristics, and models derived from experts in a field and uses that knowledge to provide problem analysis or advice that the expert might have provided if asked the same question.

KPI See: Key Performance Indicator.

Laboratory function study Study that explores important properties of an information resource in isolation from the clinical setting.

Laboratory user effect study An evaluation technique in which a user is observed when given a simulated task to perform.

LAN See: Local-area network.

Laser printer Output device that uses an electromechanically controlled laser beam to generate an image on a xerographic surface, which then is used to produce paper copies.

Latency The time required for a signal to travel between two points in a network.

Latent failures Enduring systemic problems that make errors possible but are less visible or not evident for some time.

Law of proximity Principle from Gestalt psychology that states that visual entities that are close together are perceptually grouped.

Law of symmetry Principle from Gestalt psychology that states that symmetric objects are more readily perceived.

LCD See: Liquid crystal display.

Lean A management strategy that focuses only on those process that are able to contribute specific and measurable value for the end customer. The LEAN concept originated with Toyota's focus on efficient manufacturing processes.

Learning Content Management System A software platform that allows educational content creators to host, manage, and track changes in content.

Learning health system A proposed model for health care in which outcomes from past and current patient care provide are systematically collected, analyzed and then fed back into decision making about best practices for future patient care.

Learning healthcare system See: Learning health system.

Learning Management System An LMS is a repository of educational content, and interface for delivering courses and content to learners, and a vehicle for faculty to track learner usage and performance.

LED See: Light-emitting diode.

Lexemes A minimal lexical unit in a language that represents different forms of the same word.

Lexical-statistical retrieval Retrieval based on a combination of word matching and relevance ranking.

Lexicon A catalogue of the words in a language, usually containing syntactic information such as parts of speech, pluralization rules, etc.

Light-emitting diode (LED) A semiconductor device that emits a particular frequency of light when a current is passed through it; typically used for indicator lights and computer screens because low power requirement, minimal heat generated, and durability.

Likelihood ratio (LR) A measure of the discriminatory power of a test. The LR is the ratio of the probability of a result when the condition under consideration is true to the probability of a result when the condition under consideration is false (for example, the probability of a result in a diseased patient to the probability of a result in a non-diseased patient). The LR for a positive test is the ratio of true-positive rate (TPR) to false-positive rate (FPR).

Link-based An indexing approach that gives relevance weight to web pages based on how often they are cited by other pages.

Linux An open source operating system based on principles of Unix and first developed by Linus Torvalds in 1991.

Liquid crystal display (LCD) A display technology that uses rod-shaped molecules to bend

light and alter contrast and viewing angle to produce images.

Listserver A distribution list for electronic mail messages.

Literature reference database See: bibliographic database.

Local-area network (LAN) A network for data communication that connects multiple nodes, all typically owned by a single inst tution and located within a small geographic area.

Logical Observations, Identifiers, Names and Codes (LOINC) A controlled terminology created for providing coded terms for observational procedures. Originally focused on laboratory tests, it has expanded to include many other diagnostic procedures.

Logical positivist A philosophical position that holds that there is a single truth that can be inferred from the right combination of studies.

Logic-based A knowledge representation method based on the use of predicates.

LOINC See: Logical Observations, Identifiers, Names and Codes.

Longitudinal Care Plan A holistic, dynamic, and integrated plan that documents important disease prevention and treatment goals and plans. A longitudinal plan is patient-centered, reflecting a patient's values and preferences, and is dependent upon bidirectional communications.

Long-term memory The part of memory that acquires information from short-term memory and retains it for long periods of time.

Long-term storage A medium for storing information that can persist over long periods with- out the need for a power supply to maintain data integrity.

Lossless compression A mathematical technique for reducing the number of bits needed to store data while still allowing for the re- creation of the original data.

Lossy compression A mathematical technique for reducing the number of bits needed to store data but that results in loss of information.

Low-level processes An elementary process that has its basis in the physical world of chemistry or physics.

LR See: Likelihood ratio.

Machine code The set of primitive instructions to a computer represented in binary code (machine language).

Machine language The set of primitive instructions represented in binary code (machine code).

Machine learning A computing technique in which information learned from data is used to improve system performance.

Machine translation Automatic mapping of text written in one natural language into text of another language.

Macros A reusable set of computer instructions, generally for a repetitive task.

Magnetic disk A round, flat plate of material that can accept and store magnetic charge. Data are encoded on magnetic disk as sequences of charges on concentric tracks.

Magnetic resonance imaging (MRI) A modality that produces images by evaluating the differential response of atomic nucleli in the body when the patient is placed in an intense magnetic field.

Magnetic resonance spectroscopy A noninvasive technique that is similar to magnetic resonance imaging but uses a stronger field and is used to monitor body chemistry (as in metabolism or blood flow) rather than anatomical structures.

Magnetic tape A long ribbon of material that can accept and store magnetic charge. Data are encoded on magnetic tape as sequences of charges along longitudinal tracks.

Magnetoencephalography (MEG) A method for measuring the electromagnetic fields generated by the electrical activity of the neurons using a large arrays of scalp sensors, the output of which are processed in a similar way to CT in order to localize the source of the electromagnetic and metabolic shifts occurring in the brain during trauma.

Mailing list A set of mailing addresses used for bulk distribution of electronic or physical mail.

Mainframe computer system A large, expensive, multi-user computer, typically operated and maintained by professional computing personnel. Often referred to as a "mainframe" for short.

Malpractice Class of litigation in health care based on negligence theory; failure of a health professional to render proper services in kee ing with the standards of the community.

Malware Software that is specifically design to cause harm to computer systems by disrupting other programs, damaging the machine, or gaining unauthorized access to the system or the data that it contains.

Management The process of treating a patient (or allowing the condition to resolve on its own) once the medical diagnosis has been determined.

Mannequin A life size plastic human body with some or many human-like functions.

Manual indexing The process by which human indexers, usually using standardized terminology, assign indexing terms and attributes to documents, often following a specific protocol.

Markov cycle The period of time specified for a transition probability within a Markov model.

Markov model A mathematical model of a set of strings in which the probability of a given symbol occurring depends on the identity of the immediately preceding symbol or the two immediately preceding symbols. Processes modeled in this way are often called Markov processes.

Markov process A mathematical model of a set of strings in which the probability of a given symbol occurring depends on the identity of the immediately preceding symbol or the two immediately preceding symbols.

Markup language A document specification language that identifies and labels the components of the document's contents.

Massively Online Open Course (MOOC) In a traditional MOOC, the teacher's content is digitally recorded and made available online, freely, as a sequence of lectures with supporting learning material.

Master patient index (MPI) A database that is used across a healthcare organization to maintain consistent, accurate, and current demographic and essential clinical data on the patients seen and managed within its various departments.

Mean average precision (MAP) A method for measuring overall retrieval precision in which precision is measured at every point at which a relevant document is obtained, and the MAP measure is found by averaging these points for the whole query.

Mean time between failures (MTBF) The average predicted time interval between anti ipated operational malfunctions of a system, based on long-term observations.

Meaningful use The set of standards defined by the Centers for Medicare & Medicaid

Services (CMS) Incentive Programs that governs the use of electronic health records and allows eligible providers and hospitals to earn incentive payments by meeting specific criteria. The term refers to the belief that health care providers using electronic health records in a meaningful, or effective, way will be able to improve health care quality and efficiency.

Measurement study Study to determine the extent and nature of the errors with which a measurement is made using a specific instrument (cf. Demonstration study).

Measures of concordance Measures of agreement in test performance: the true-positive and true-negative rates.

MedBiquitous A healthcare-specific standards consortium led by Johns Hopkins Medicine.

Medical computer science The subdivision of computer science that applies the methods of computing to medical topics.

Medical computing The application of methods of computing to medical topics (see medical computer science).

Medical entities dictionary (MED) A compendium of terms found in electronic medical record systems. Among the best known MEDs is that developed and maintained by the Columbia University Irving Medical Center and Columbia University. Contains in excess of 100,000 terms.

Medical errors Errors or mistakes, committed by health professionals, that hold the potential to result in harm to the patient.

Medical home A primary care practice that will maintain a comprehensive problem list to make fully informed decisions in coordinating their care.

Medical informatics An earlier term for the biomedical informatics discipline, medical informatics is now viewed as the subfield of clinical informatics that deals with the management of disease and the role of physicians.

Medical Information Bus (MIB) A data-communication system that supports data acquisition from a variety of independent devices.

Medical information science The field of study concerned with issues related to the management and use of biomedical information (see also biomedical informatics).

Medical Literature Analysis and Retrieval System (MEDLARS) The initial electronic version of Index Medicus developed by the National Library of Medicine.

Medical Logic Module (MLM) A single chunk of medical reasoning or decision rule, typically encoded using the Arden Syntax.

Medical record committees An institutional panel charged with ensuring appropriate use of medical records within the organization.

Medical Subject Headings (MeSH) Some 18,000 terms used to identify the subject content of the biomedical literature. The National Library of Medicine's MeSH vocabulary has emerged as the de facto standard for biomedical indexing.

Medication A substance used for medical treatment, typically a medicine or drug.

MEDLARS Online (MEDLINE) The National Library of Medicine's electronic catalog of the biomedical literature, which includes information abstracted from journal articles, including author names, article title, journal source, publication date, abstract, and medical subject headings.

Medline Plus An online resource from the National Library of Medicine that contains health topics, drug information, medical dictionaries, directories, and other resources, organized for use by health care consumers.

Megabits per second (Mbps) A common unit of measure for specifying a rate of data transmission.

Megabyte 2^{20} or 1,048,576 bytes.

Member checking In subjectivist research, the process of reflecting preliminary findings back to individuals in the setting under study, one way of confirming that the findings are truthful.

Memorandum of understanding A document describing a bilateral or multilateral agreement between two or more parties. It expresses a convergence of will between the parties, indicating an intended common line of action.

Memory Areas that are used to store programs and data. The computer's working memory comprises read-only memory (ROM) and random access memory (RAM).

Memory sticks A portable device that typically plugs into a computer's USB port and is capable of storing data. Also called a "thumb drive" or a "USB drive".

Mendelian randomization (MR) A technique used to provide evidence for the causality of a biomarker on a disease state in conditions in which randomized controlled trials are difficult or too expensive to pursue. The technique uses genetic variants that are known to associate with the biomarker as instrument variables.

Mental images A form of internal representation that captures perceptual information recovered from the environment.

Mental models A construct for describing how individuals form internal models of systems. They are designed to answer questions such as "how does it work?" or "what will happen if I take the following action?".

Mental representations Internal cognitive states that have a certain correspondence with the external world.

Menu In a user interface, a displayed list of valid commands or options from which a user may choose.

Merck Medicus An aggregated set of resources, including Harrison's Online, MDConsult, and DXplain.

Meta-analysis A summary study that combines quantitatively the estimates from individual studies.

Metabolomic Pertaining to the study of small-molecule metabolites created as the end-products of specific cellular processes.

Metadata Literally, data about data, describing the format and meaning of a set of data.

Metagenomics Using DNA sequencing technology to characterize complex samples derived from an environmental sample, e.g., microbial populations. For example, the gut "microbiome" can be characterized by applying next generation sequencing of stool samples.

Metathesaurus One component of the Unified Medical Language System, the Metathesaurus contains linkages between terms in Medical Subject Headings (MeSH) and in dozens of controlled vocabularies.

MIB See Medical Information Bus.

Microarray chips A microchip that holds DNA probes that can recognize DNA from samples being tested.

Microbiome The microorganisms in a particular environment (including the body or a part of the body) or the combined genomes of those organisms.

Microprocessor An integrated circuit that contains all the functions of a central processing unit of a computer.

Microsimulation models Individual-level health state transition models that provide a means to model very complex events flexibly over time.

MIMIC II Database See Multiparameter Intelligent Monitoring in Intensive Care.

Minicomputers A class of computers that were introduced in the 1960s as a smaller alternative to mainframe computers. Minicomputers enabled smaller companies and departments within organizations (like HCOs) to implement software applications at significantly less cost than was required by mainframe computers.

Mistake Occurs when an inappropriate course of action reflects erroneous judgment or inference.

Mixed-initiative dialog A mode of interaction with a computer system in which the computer may pose questions for the user to answer, and vice versa.

Mixed-initiative systems An educational program in which user and program share control of the interaction. Usually, the program guides the interaction, but the student can assume control and digress when new questions arise during a study session.

Mobile health (mHealth) The practice of medicine and public health supported by mobile devices. Also referred to as mHealth or m-health.

Model organism database Organized reference databases that combine bibliographic databases, full text, and databases of sequences, structure, and function for organisms whose genomic data has been highly characterized, such as the mouse, fruit fly, and Sarcchomyces yeast.

Model organism databases Organized reference databases the combine bibliographic databases, full text, and databases of sequences, structure, and function for organ- isms whose genomic data has been highly characterized, such as the mouse, fruit fly, and Sarcchomyces yeast.

Modem A device used to modulate and demodulate digital signals for transmission to a remote computer over telephone lines; converts digital data to audible analog signals, and vice versa.

Modifiers of interest In natural language processing, a term that is used to describe or otherwise modify a named-entity that has been recognized.

Molecular imaging A technique for capturing images at the cellular and subcellular level by marking particular chemicals in ways that can be detected with image or radiodetection.

Monitoring tool The application of logical rules and conditions (e.g., range-checking, enforcement of data completion, etc.) to ensure the completeness and quality of research-related data.

Monotonic Describes a function that consistently increases or decreases, rather than oscillates.

Morpheme The smallest unit in the grammar of a language which has a meaning or a linguistic function; it can be a root of a word (e.g., *–arm*), a prefix (e.g., *re-*), or a suffix (e.g., *–it is*).

Morphology The study of meaningful units in language and how they combine to form words.

Morphometrics The quantitative study of growth and development, a research area that depends on the use of imaging methods.

Mosaic The first graphical web browser credited with popularizing the World Wide Web and developed at the National Center for Supercomputing Applications (NCSA) at the University of Illinois.

Motion artifact Visual interference caused by the difference between the frame rate of an imaging device and the motion of the object being imaged.

Mouse A small boxlike device that is moved on a flat surface to position a cursor on the screen of a display monitor. A user can select and mark data for entry by depressing buttons on the mouse.

Multi-axial A terminology system composed of several distinct, mutually exclusive term sub- sets that care combined to support post-coordination.

Multimodal interface A design concept which allows users to interact with computers using multiple modes of communication or tools, including speaking, clicking, or touchscreen input.

Multiparameter Intelligent Monitoring in Intensive Care (MIMIC-II) A publicly and freely available research database that encompasses a diverse and very large population of ICU patients. It contains high temporal resolution data including lab results, electronic documentation, and bedside monitor trends and waveforms.

Multiprocessing The use of multiple processors in a single computer system to increase the power of the system (see parallel processing).

Multiprogramming A scheme by which multiple programs simultaneously reside in the main memory of a single central processing unit.

Multiprotocol label switching (MPLS) A mechanism in high-performance telecommunications networks that directs data from one network node to the next based on short path labels rather than long network addresses, avoiding complex lookups in a routing table.

Multiuser system A computer system that shares its resources among multiple simultaneous users.

Mutually exclusive State in which one, and only one, of the possible conditions is true; for example, either A or not A is true, and one of the statements is false. When using Bayes' theorem to perform medical diagnosis, we generally assume that diseases are mutually exclusive, meaning that the patient has exactly one of the diseases under consideration and not more.

Myocardial ischemia Reversible damage to cardiac muscle caused by decreased blood flow and resulting poor oxygenation. Such ischemia may cause chest pain or other symptoms.

Naïve Bayesian model The use of Bayes Theorem in a way that assumes conditional independence of variables that may in fact be linked statistically.

NAM See: National Academy of Medicine.

Name Designation of an object by a linguistic expression.

Name authority An entity or mechanism for controlling the identification and formulation of unique identifiers for names. In the Internet, a name authority is required to associate common domain names with their IP addresses.

Named-entity normalization The natural language processing method, after finding a named entity in a document, for linking (normalizing) that mention to appropriate database identifiers.

Named-entity recognition In language processing, a subtask of information extraction that seeks to locate and classify atomic elements in text into predefined categories.

Name-server In networked environments such as the Internet, a computer that converts a host name into an IP address before the message is placed on the network.

National Academies The collective name for the National Academy of Engineering, National Academy of Medicine and National Academy of Sciences which are private, non-profit institutions that provide expert advice on some of the most pressing challenges facing the nation and the world. The work of the National Academies helps shape sound policies, inform public opinion, and advance the pursuit of science, engineering, and medicine.

National Academy of Medicine (NAM) An independent organization of eminent professionals from diverse fields including health and medicine; natural, social, and behavioral sciences and more. Established in 1970 as the Institute of Medicine (IOM), and in 2016 the name was changed to the National Academy of Medicine (NAM).

National Center for Biotechnology Information (NCBI) Established in 1988 as a national resource for molecular biology information, the NCBI is a component of the National Library of Medicine that creates public databases, con- ducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information.

National Committee on Quality Assurance (NCQA) An independent 501 nonprofit organization in the United States that works to improve health care quality through the administration of evidence-based standards, measures, programs, and accreditation.

National Guidelines Clearinghouse A public resource, coordinated by the Agency for Health Research and Quality, that collects and distributes evidence-based clinical practice guidelines (see ► www.guideline.gov).

National Health Information Infra–structure (NHII) A comprehensive knowledge-based network of interoperable systems of clinical, public health, and personal health information that is intended to improve decision-making by making health information available when and where it is needed.

National Health Information Network (NHIN) A set of standards, services, and policies that have been shepherded by the Office of the National Coordinator of Health Information Technology to enable secure health information exchange over the Internet.

National Information Standards Organization (NISO) A non-profit association accredited by the American National Standards Institute (ANSI), that identifies, develops, maintains, and publishes technical standards to manage information (see ► www.niso.org).

National Institute for Standards and Technology (NIST) A non-regulatory federal agency within the U.S. Commerce Department's Technology Administration; its mission is to develop and promote measurement, standards, and technology to enhance productivity, facilitate trade, and improve the quality of life (see ► www.nist.gov).

National Library of Medicine (NLM) The government-maintained library of biomedicine that is part of the US National Institutes of Health.

National Quality Forum A not-for-profit organization that develops and implements national strategies for health care quality measurement and reporting.

Nationwide Health Information Network (NwHIN) A set of standards, services, and policies that have been shepherded by the Office of the National Coordinator of Health Information Technology to enable secure health information exchange over the Internet.

Natural language Unfettered spoken or written language. Free text.

Natural language processing (NLP) Facilitates tasks by enabling use of automated methods that represent the relevant information in the text with high validity and reliability.

Natural language query A question expressed in unconstrained text, from which meaning must somehow be extracted or inferred so that a suitable response can be generated.

Naturalistic Describes a study in which little if anything is done by the evaluator to alter the setting in which the study is carried out.

NCBI Entrez global query A search interface that allows searching over all data and information resources maintained by NCBI.

NCI Thesaurus A large ontology developed by the National Cancer Institute that describes entities related to cancer biology, clinical oncology, and cancer epidemiology.

NCQA See National Committee on Quality Assurance.

Needs assessment A study carried out to help understand the users, their context and their needs and skills, to inform the design of the information resource.

Negative dictionary A list of stop words used in information retrieval.

Negative predictive value (PV–) The probability that the condition of interest is absent if the result is negative—for example, the probability that specific a disease is absent given a negative test result.

Negligence theory A concept from tort law that states that providers of goods and services are expected to uphold the standards of the community, thereby facing claims of negligence if individuals are harmed by substandard goods or services.

Nestedstructures In natural language processing, a phrase or phrases that are used in place of simple words within other phrases.

Net reclassification improvement (NRI) In classification methods, a measure of the net fraction of reclassifications made in the correct direction, using one method over another method without the designated improvement.

Network access provider A company that builds and maintains high speed networks to which customers can connect, generally to access the Internet (see also Internet service provider).

Network Operations Center (NOC) A centralized monitoring facility for physically distributed computer and/or telecommunications facilities that allows continuous real-time reporting of the status of the connected components.

Network protocol The set of rules or conventions that specifies how data are prepared and transmitted over a network and that governs data communication among the nodes of a network.

Network stack The method within a single machine by which the responsibilities for network communications are divided into different levels, with clear interfaces between the levels, thereby making network software more modular.

Neuroinformatics An emerging subarea of applied biomedical informatics in which the discipline's methods are applied to the management of neurological data sets and the modeling of neural structures and function.

Next Generation Internet Initiative A federally funded research program in the late 1990s and early in the current decade that sought to provide technical enhancements to the Internet to support future applications that currently are infeasible or are incapable of scaling for routine use.

Next generation sequencing methods Technologies for performing high throughput sequencing of large quantities of DNA or RNA. Typically, these technologies determine the sequences of many millions of short segments of DNA that need to be reassembled and interpreted using bioinformatics.

NHIN Connect A software solution that facilitates the exchange of healthcare information at both the local and national level. CONNECT leverages eHealth Exchange standards and governance and Direct Project specifications to help drive interoperability across health information exchanges throughout the country. Initially developed by federal agencies to support specific healthcare-related missions, CONNECT is now available to all organizations as downloadable open source software.

NHIN Direct A set of standards and services to enable the simple, direct, and secure transport of health information between pairs of health care providers; it is a component of the Nationwide Health Information Network and it complements the Network's more sophisticated components.

NHIN See: National Health Information Network.

Noise The component of acquired data that is attributable to factors other than the underlying phenomenon being measured (for example, electromagnetic interference, inaccuracy in sensors, or poor contact between sensor and source).

Nomenclature A system of terms used in a scientific discipline to denote classifications and relationships among objects and processes.

Nosocomial hospital-acquired infection An infection acquired by a patient after admission to a hospital for a different reason.

NQF See: National Quality Forum.

Nuclear magnetic resonance (NMR) spectroscopy A spectral technique used in chemistry to characterize chemical compounds by measuring magnetic characteristics of their atomic nuclei.

Nuclear medicine imaging A modality for producing images by measuring the radiation emitted by a radioactive isotope that has been attached to a biologically active compound and injected into the body.

Nursing informatics The application of biomedical informatics methods and techniques to problems derived from the field of nursing. Viewed as a subarea of clinical informatics.

NwHIN Direct A set of standards and services to enable the simple, direct, and secure transport of health information between pairs of health care providers; it is a component of the Nationwide Health Information Network and it complements the Network's more sophisticated components.

Nyquist frequency The minimum sampling rate necessary to achieve reasonable signal quality. In general, it is twice the frequency of the highest-frequency component of interest in a signal.

Object Any part of the perceivable or conceivable world.

Object Constraint Language (OCL) A textual language for describing rules that apply to the elements a model created in the Uniform Modeling Language. OLC specifies constraints on allowable values in the model. OCL also supports queries of UML models (and of models constructed in similar languages). OCL is a standard of the Object Modeling Group (OMG), and forms the basis of the GELLO query language that may be used in conjunction with the Arden Syntax.

Objectivist approaches Class of evaluation approaches that make use of experimental designs and statistical analyses of quantitative data.

Object-oriented database A database that is structured around individual objects (concepts) that generally include relationships among those objects and, in some cases, executable code that is relevant to the management and or understanding of that object.

Odds-ratio form An algebraic expression for calculating the posttest odds of a disease, or other condition of interest, if the pretest odds and likelihood ratio are known (an alternative formulation of Bayes' theorem, also called the odds-likelihood form).

Office of the National Coordinator for Health Information Technology (ONC) An agency within the US Department of Health and Human Services that is charged with supporting the adoption of health information technology and promoting nationwide health information exchange to improve health care.

Omics A set of areas of study in biology that use the suffix "-ome", used to connote breadth or completeness of the objects being studied, for example genomics or proteomics.

-omics technologies High throughput experimentation that exhaustively queries a certain biochemical aspect of the state of an organism. Such technologies include proteomics (protein), genomics (gene expression), metabolomics (metabolites), etc.

On line analytic processing (OLAP) A system that focuses on querying across multiple patients simultaneously, typically by few users for infrequent, but very complex queries, often research.

On line transaction processing (OLTP) A system designed for use by thousands of simultaneous users doing repetitive queries.

Ontology A description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. In biomedicine, such ontologies typically specify the meanings and hierarchical relationships among terms and concepts in a domain.

Open access publishing (OA) An approach to publishing where the author or research funder pays the cost of publication and the article is made freely available on the Internet.

Open consent model A legal mechanism by which an individual can disclose their own private health information or genetic information for research use. This mechanism is used by the Personal Genome Project to enable release of entire genomes of identified individuals.

Open source An approach to software development in which programmers can read, redistribute, and modify the source code for a piece of software, resulting in community development of a shared product.

Open standards development policy In standards group, a policy that allows anyone to become involved in discussing and defining the standard.

OpenNotes An international movement that urges doctors, nurses, therapists, and other clinicians to invite patients to read notes that clinicians write to describe a visit. OpenNotes provides free tools and resources to help clinicians and healthcare systems share notes with patients.

Operating system (OS) A program that allocates computer hardware resources to user programs and that supervises and controls the execution of all other programs.

Optical Character Recognition (OCR) The conversion of typed text within scanned documents to computer understandable text.

Optical coherence tomography (OCT) An optical signal acquisition and processing method. It captures micrometer-resolution, three-dimensional images from within optical scattering media (e.g., biological tissue).

Optical disk A round, flat plate of plastic or metal that is used to store information. Data are encoded through the use of a laser that marks the surface of the disc.

Order entry The use of a computer system for entering treatments, requests for lab tests or radiologic studies, or other interventions that the attending clinician wishes to have performed for the benefit of a patient.

Orienting issues/questions The initial questions or issues that evaluators seek to answer in a subjectivist study, the answers to which often in turn prompt further questions.

Outcome data Formal information regarding the results of interventions.

Outcome measurements Using metrics that assess the end result of an intervention rather than an intervening process. For example, remembering to check a patient's Hemoglobin A1C is a process measure, whereas reducing the complications of diabetes is an outcome measure.

Outcome variable Similar to "dependent variable," a variable that captures the end result of a health care or educational process; for example, long-term operative complication rate or mastery of a subject area.

Outpatient A patient seen in a clinic rather than in the hospital setting.

Output The results produced when a process is applied to input. Some forms of output are hardcopy documents, images displayed on video display terminals, and calculated values of variables.

P4 medicine P4 medicine: a term coined by Dr. Leroy Hood for healthcare that strives to be personalized, predictive, preventive and participatory.

Packets In networking, a variable-length message containing data plus the network addresses of the sending and receiving nodes, and other control information.

Page A partitioned component of a computer users' programs and data that can be kept in temporary storage and brought into main memory by the operating system as needed.

Pager One of the first mobile devices for electronic communication between a base station (typically a telephone, but later a computer) and an individual person. Initially restricted to receiving only numeric data (e.g., a telephone number), pagers later incorporated the ability to transmit a response (referred to as "two way pagers") as well as alpha characters so that a message of limited length could be transmitted from a small keyboard. Pagers have been gradually replaced by cellular phones because of their greater flexibility and broader geographical coverage.

PageRank (PR) algorithm In indexing for information retrieval on the Internet, an algorithmic scheme for giving more weight to a Web page when a large number of other pages link to it.

Parallel processing The use of multiple processing units running in parallel to solve a single problem (see multiprocessing).

Parse tree The representation of structural relationships that results when using a grammar (usually context free) to analyze a given sentence.

Partial parsing The analysis of structural relationships that results when using a grammar to analyze a segment of a given sentence.

Partial-match searching An approach to information retrieval that recognizes the inexact nature of both indexing and retrieval, and attempts to return the user content ranked by how close it comes to the user's query.

Participant calendaring Participant calendaring refers to the capability of a CRMS to support the tracking of participant compliance with a study schema, usually represented as a calendar of temporal events.

Participant screening and registration participant screening and registration refers to the capability of a CTMS to support the enrollment phase of a clinical study.

Participants The people or organizations who provide data for the study. According to the role of the information resource, these may include patients, friends and family, formal and informal carers, the general public, health professionals, system developers, guideline developers, students, health service managers, etc.

Part-of-speech tags Assignment of syntactic classes to a given sequence of words, e.g., determiner, adjective, noun and verb.

Parts of speech The categories to which words in a sentence are assigned in accordance with their syntactic function.

Patent A specific legal approach for protecting methods used in implementing or instantiating ideas (see intellectual property).

Pathognomonic Distinctively characteristic, and thus, uniquely identifying a condition or object (100% specific).

Patient centered care Clinical care that is based on personal characteristics of the patient in addition to his or her disease. Such characteristics include cultural traditions, preferences and values, family situations and lifestyles.

Patient centered medical home A team-based health care delivery model led by a physician, physician's assistant, or nurse practitioner that provides comprehensive, coordinated, and continuous medical care to patients with the goal of obtaining maximized health outcomes.

Patient engagement Participation of a patient as an active collaborator in his or her health care process.

Patient generated health data Health-related data that are recorded or collected directly by patients.

Patient portal An online application that allows individuals to view health information and otherwise interact with their physicians and hospitals.

Patient record The collection of information traditionally kept by a health care provider or organization about an individual's health status and health care; also referred to as the patient's chart, medical record, or health record, and originally called the "unit record".

Patient safety The reduction in the risk of unnecessary harm associated with health care to an acceptable minimum; also the name of a movement and specific research area.

Patient triage The process of allocating patients to different levels or urgency of care depending upon the complaints or symptoms displayed.

Patient-specific information Information derived and organized from a specific patient.

Patient-tracking applications Monitor patient movement in multistep processes.

Pattern check A procedure applied to entered data to verify that the entered data have a required pattern; e.g., the three digits, hyphen, and four digits of a local telephone number.

Pay for performance Payments to providers that are based on meeting pre-defined expectations for quality.

Per diem Payments to providers (typically hospitals) based on a single day of care.

Perimeter definition Specification of the boundaries of trusted access to an information system, both physically and logically.

Personal clinical electronic communication Web-based messaging solutions that avoid the limitations of email by keeping all interactions within a secure, online environment.

Personal computers A small, relatively inexpensive, single-user computer.

Personal Digital Assistants (PDA) A small, mobile, handheld device that provides computing and information storage and retrieval capabilities for personal or business use. PDAs can typically run third-party applications.

Personal grid architecture A security methodology that prevents large-scale data loss from a central repository by separately storing and encrypting each person's records. While searching across records must be sequential, reasonable response times can be achieved by massive parallelization of the search process in the cloud.

Personal health application Software for computers, tablet computers, or smart phones that are intended to allow individual patients to monitor their own health or to stimulate their own personal health activities.

Personal health informatics The area of biomedical informatics based on patient-centered care, in which people are able to access care that is coordinated and collaborative.

Personal health record (PHR) A collection of information about an individual's health status and health care that is maintained by the individual (rather than by a health care provider); the data may be entered directly by the patient, captured from a sensing device, or transferred from a laboratory or health care provider. It may include medical information from several independent provider organizations, and may also have health and well-being information.

Personal Internetworked Notary and Guardian (PING) An early personally controlled health record, later known as Indivo.

Personalized medicine Also often call individualized medicine, refers to a medical model in which decisions are custom-tailored to the patient based on that individual's genomic data, preferences, or other considerations. Such decisions may involve diagnosis, treatment, or assessments of prognosis. Also known as precision medicine.

Personally controlled health record (PCHR) Similar to a PHR, the PCHR differs in the nature of the control offered to the patient, with such features as semantic tags on data elements that can be used to determine the subsets of information that can be shared with specific providers.

Petabyte A unit of information equal to 1000 terabytes or 10^{15} bytes.

Pharmacodynamics program (PD) The study of how a drug works, it's mechanism of action and pathway of achieving its affect, or "what the drug does to the body".

Pharmacogenetics The study of drug-gene relationships that are dominated by a single gene.

Pharmacogenomic variant A particular genetic variant that affects a drug-genome interaction.

Pharmacokinetic program Pharmacokinetics or PK is the study of how a drug is absorbed, distributed, metabolized and excreted by the body, or "what the body does to the drug".

Pharmacovigilance The pharmacological science relating to the collection, detection, assessment, monitoring, and prevention of adverse effects with pharmaceutical products.

Phase In the context of clinical research, study phases are used to indicate the scientific aim of a given clinical trial. There are 4 phases (Phase I, Phase II, Phase III, and Phase IV).

Phase I (clinical trial) Investigators evaluate a novel therapy in a small group of participants in order to assess overall safety. This safety assessment includes dosing levels in the case of non-interventional therapeutic trials, and potential side effects or adverse effects of the therapy. Often, Phase I trials of non-interventional therapies involve the use of normal volunteers who do not have the disease state targeted by the novel therapy.

Phase II (clinical trial) Investigators evaluate a novel therapy in a larger group of participants in order to assess the efficacy of the treatment in the targeted disease state. During this phase, assessment of overall safety is continued.

Phase III (clinical trial) Investigators evaluate a novel therapy in an even larger group of participants and compare its performance to a reference standard which is usually the current standard of care for the targeted disease state. This phase typically employs a randomized controlled design, and often a multi-center RCT given the numbers of variation of subjects that must be recruited to adequately test the hypothesis. In general, this is the final study phase to be performed before seeking regulatory approval for the novel therapy and broader use in standard- of-care environments.

Phase IV (clinical trial) Investigators study the performance and safety of a novel therapy after it has been approved and marketed. This type of study is performed in order to detect long-term outcomes and effects of the therapy. It is often called "post-market surveillance" and is, in fact, not an RCT at all, but a less formal, observational study.

PheKB.org A web site that houses EHR-based algorithms for determining phenotypes.

Phenome characterization Identification of the individual traits of an organism that characterize its phenotype.

Phenome-wide association scan A study that derives case and controls populations using the EMR to define clinical phenotypes and then examines the association of those phenotypes with specific genotypes.

Phenome-wide association study (PheWAS) A study that tests for association between a particular genetic variant and a large number of phenotypic characteristics.

Phenotype The observable physical characteristics of an organism, produced by the interaction of genotype with environment.

Phenotype definition The process of determining the set of observable descriptors that characterize an organism's phenotype.

Phenotype risk score (PheRS) A calculation of the likelihood of a particular genetic variant being present based on a weighed score of one or more phenotypic characteristics.

Phenotypic Refers to the physical characteristics or appearance of an organism.

Picture Archive and Communication Systems (PACS) An integrated computer system that acquires, stores, retrieves, and displays digital images.

Pixel One of the small picture elements that makes up a digital image. The number of pixels per square inch determines the spatial resolution. Pixels can be associated with a single bit to indicate black and white or with multiple bits to indicate color or gray scale.

Placebo In the context of clinical research, a placebo is a false intervention (e.g., a mock intervention given to a participant that resembles the intervention experienced by individuals receiving the experimental intervention, except that it has no anticipated impact on the individual's health or other indicated status), usually used in the context of a control group or intervention.

Plain old telephone service (POTS) The standard low speed, analog telephone service that is still used by many homes and businesses.

Plastination A method of embalming a part of a human body using plastic to suffuse human tissue.

Plug-in A software component that is added to web browsers or other programs to allow them a special functionality, such as an ability to deal with certain kinds of media (e.g., video or audio).

Pointing device A manual device, such as a mouse, light pen, or joy stick, that can be used

to specify an area of interest on a computer screen.

Polygenic risk score (PRS) See Genetic risk score.

Population Health is not universally defined but is a commonly used term to organize activities performed by private or public entities for assessing, managing, and improving the well-being and health outcomes of a defined group of individuals. Population may be defined by a specific geographic community or region; enrollees of a health plan; persons residing in a health systems catchment area; or an aggregation of individuals with specific conditions. Population health is based on the underlying assumption that multiple common factors impact the health and well-being of specific populations, and that focused interventions early in the causal chain of disease may save resources, and prevent morbidity and mortality.

Population management Health care practices that assist with a large group of people, including preventive medicine and immunization, screening for disease, and prioritization of interventions based on community needs.

Positive predictive value (PV+) The probability that the condition of interest is true if the result is positive—for example, the probability that a disease is present given a positive test result.

Positron emission tomography A tomographic imaging method that measures the uptake of various metabolic products (generally a com- bination of a positron-emitting tracer with a chemical such as glucose), e.g., by the functioning brain, heart, or lung.

Postcoordination The combination of two or more terms from one or more terminologies to create a phrase used for coding data; for example, "Acute Inflammation" and "Appendix" combined to code a patient with appendicitis. See also, precoordination.

Posterior probability The updated probability that the condition of interest is present after additional information has been acquired.

Postgenomic database A database that com- bines molecular and genetic information with data of clinical importance or relevance. *Online Mendelian Inheritance in Man* (OMIM) is a frequently cited example of such a database.

Post-test probability The updated probability that the disease or other condition under consideration is present after the test result is known (more generally, the posterior probability).

Practice management system The software used by physicians for scheduling, registration, billing, and receivables management in their offices. May increasingly be linked to an EHR.

Pragmatics The study of how contextual information affects the interpretation of the underlying meaning of the language.

Precision The degree of accuracy with which the value of a sampled observation matches the value of the underlying condition, or the exactness with which an operation is performed. In information retrieval, a measure of a system's performance in retrieving relevant information (expressed as the fraction of relevant records among total records retrieved in a search).

Precision Medicine The application of specific diagnostic and therapeutic methods matched to an individual based on highly unique information about the individual, such as their genetic profile or properties of their tumor.

Precoordination A complex phrase in a terminology that can be constructed from multiple terms but is, itself, assigned a unique identifier within the terminology; for example, "Acute Inflammation of the Appendix." See also, postcoordination.

Predatory journal A name given to journals that publish under the OA model and have no to minimal peer review of submitted papers.

Predicate The part of a sentence or clause containing a verb and stating something about the subject.

Predicate logic In mathematical logic, the generic term for symbolic formal systems like first-order logic, second-order logic, etc.

Predictive value (PV) The posttest probability that a condition is present based on the results of a test (see positive predictive value and negative predictive value).

Preparatory phase In the preparatory phase of a clinical research study, investigators are involved in the initial design and documentation of a study (developing a protocol document), prior to the identification and enrollment of study participants.

President's Emergency Plan for AIDS Relief (PEPFAR) The United States government's response to the global HIV/AIDS epidemic, and represents the largest commitment by any nation to address a single disease in history. PEPFAR is intended to save and improved millions of lives, accelerating progress toward controlling and ultimately ending the AIDS epidemic as a public health threat. PEPFAR collects and uses data in the most granular manner (disaggregated by sex, age, and at the site level) to do the right things, in the right places, and right now within the highest HIV-burdened populations and geographic locations.

Pretest probability The probability that the dis- ease or other condition under consideration is present before the test result is known (more generally, the prior probability).

Prevalence The frequency of the condition under consideration in the population. For example, we calculate the prevalence of disease by dividing the number of diseased individuals by the number of individuals in the population. Prevalence is the prior probability of a specific condition (or diagnosis), before any other information is available.

Primary knowledge-based information The original source of knowledge, generally in a peer reviewed journal article that reports on a research project's results.

Prior probability The probability that the condition of interest is present before additional information has been acquired. In a population, the prior probability also is called the prevalence.

Privacy A concept that applies to people, rather than documents, in which there is a presumed right to protect that individual from unauthorized divulging of personal data of any kind.

Probabilistic context free grammar A context free grammar in which the possible ways to expand a given symbol have varying probabilities rather than equal weight.

Probabilistic relationship Exists when the occurrence of one chance event affects the probability of the occurrence of another chance event.

Probabilistic sensitivity analysis An approach for understanding how the uncertainty in all (or a large number of) model parameters affects the conclusion of a decision analysis.

Probability Informally, a means of expressing belief in the likelihood of an event. Probability is more precisely defined mathematically in terms of its essential properties.

Probalistic causal network Also known as a Bayesian network, a statistical model built of directed acyclic graph structures (nodes) that are connected through relationships (edges). The strength of each of the relationships is defined through conditional probabilities.

Probes Genetic markers used it genetic assays to determine the presences or absence of a particular variant.

Problem impact study A study carried out in the field with real users as participants and real tasks to assess the impact of the information resource on the original problem it was designed to resolve.

Problem space The range of possible solutions to a problem.

Problem-based learning Small groups of students, supported by a facilitator, learned through discussion of individual case scenarios.

Procedural knowledge Knowledge of how to perform a task (as opposed to factual knowledge about the world).

Procedure An action or intervention undertaken during the management of a patient (e.g., starting an IV line, performing surgery). Procedures may also be cognitive.

Procedure trainer (Also Part-task trainer). An on-screen simulation of a surgical or other procedure that is controlled using physical tools such as an endoscope. It allows repeated practice of a specific skill.

Process integration An organizational analysis methodology in which a series of tasks are reviewed in terms of their impact on each other rather than being reviewed separately. In a hospital setting, for example, a process integration view would look at patient registration and scheduling as an integrated workflow rather than as separate task areas. The goal is to achieve greater efficiency and effectiveness by focusing on how tasks can better work together rather than optimizing specific areas.

Prodrug A chemical that requires transformation *in vivo* (typically by enzymes) to produce its active drug.

Product An object that goes through the processes of design, manufacture, distribution, and sale.

Prognostic scoring system An approach to prediction of patient outcomes based on formal analysis of current variables, generally through methods that compare the patient in some way with large numbers of similar patients from the past.

Progressive caution The idea that reason, caution and attention to ethical issues must govern research and expanding applications in the field of biomedical informatics.

Propositions An expression, generally in language or other symbolic form, that can be believed, doubted, or denied or is either true or false.

Prospective study An experiment in which researchers, before collecting data for analysis, define study questions and hypotheses, the study population, and data to be collected.

Prosthesis A device that replaces a body part—e.g., artificial hip or heart.

Protected memory An segment of computer memory that cannot be over-written by the usual means.

Protein Data Bank (PDB) A centralized repository of experimentally determined three dimensional protein and nucleic acid structures.

Proteomics The study of the protein products produced by genes in the genome.

Protocol A standardized method or approach.

Protocol analysis In cognitive psychology, methods for gathering and interpreting data that are presumed to reveal the mental processes used during problem solving (e.g., analysis of "think-aloud" protocols).

Protocol authoring tools A software product used by researchers to construct a description of a study's rationale, guidelines, endpoints, and the like. Such descriptions may be structured formally so that they can be manipulated by trial management software.

Protocol management Protocol management refers to the capability of a CRMS to support the preparatory phase of a clinical study.

Provider-profiling system Software that utilizes available data sources to report on patterns of care by one or several providers.

Pseudo-identifier A unique identifier substituted for the real identifier to mask the identify but can under certain circumstances allow linking back to the original person identifier if needed.

Public health The field that deals with monitoring and influencing trends in habits and disease in an effort to protect or enhance the health of a population, from small communities to entire countries.

Public health informatics An application area of biomedical informatics in which the field's methods and techniques are applied to problems drawn from the domain of public health.

Public health informatics The systematic application of informatics methods and tools to support public health goals and outcomes, regardless of the setting.

Public Health Surveillance The ongoing systematic collection, analysis, and interpretation of data (e.g., regarding agent/hazard, risk factor, exposure, health event) essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those responsible for prevention and control. ► http://www.aphl.org/Pages/default.aspx. Also see Biosurveillance and Surveillance.

Public Library of Science (PLoS) A family of scientific journals that is published under the open-access model.

Publication type One of several classes of articles or books into which a new publication will fall (e.g., review articles, case reports, original research, textbook, etc.).

Public-key cryptography In data encryption, a method whereby two keys are used, one to encrypt the information and a second to decrypt it. Because two keys are involved, only one needs be kept secret.

Public-private keys A pair of sequences of characters or digits used in data encryption in which one is kept private and the other is made public. A message encrypted with the public key can only be opened by the holder of the private key, and a message signed with the private key can be verified as authentic by anyone with the public key.

PubMed A software environment for searching the Medline database, developed as part of the suite of search packages, known as Entrez, by the NLM's National Center for Biotechnology Information (NCBI).

PubMed Central (PMC) An effort by the National Library of Medicine to gather the full-text of scientific articles in a freely accessible database, enhancing the value of Medline by providing the full articles in addition to titles, authors, and abstracts.

QRS wave In an electrocardiogram (ECG), the portion of the wave form that represents the time it takes for depolarization of the ventricles.

Quality assurance A means for monitoring and maintaining the goodness of a service, product, or process.

Quality Data Model An information model that describes the relationships between patient data and clinical concepts in a standardized format. The model was originally proposed to enable electronic quality-performance measurement and it is now aligned with CDS standards.

Quality management A specific effort to let quality of care be the goal that determines changes in processes, staffing, or investments.

Quality measurements Numeric metrics that assess the quality of health care services. Examples of quality measures include the portion of a physician's patients who are screened for breast cancer and 30-day hospital readmission rates. These measurements have tradition- ally been derived from administrative claims data or paper charts but there is increasing interest in using clinical data form electronic sources.

Quality-adjusted life year (QALY) A measure of the value of a health outcome that reflects both longevity and morbidity; it is the expected length of life in years, adjusted to account for diminished quality of life due to physical or mental disability, pain, and so on.

Quasi-experiments A quasiexperiment is a non-randomized, observational study design in which conclusions are drawn from the evaluation of naturally occurring and non-controlled events or cases.

Query The ability to extract information from an EHR based on a set of criteria; e.g., one could query for all patients with diabetes who have missed their follow-up appointments.

Query and Reporting Tool Software that supports both the planned and ad-hoc extraction and aggregation of data sets from multiple data forms or equivalent data capture instruments used within a clinical trials management system.

Query-response cycle For a database system, the process of submitting a single request for information and receiving the results.

Question answering (QA) A computer-based process whereby a user submits a natural language question that is then automatically answered by returning a specific response (as opposed to returning documents).

Question understanding A form of natural language understanding that supports computer-based question answering.

Radiology The medical field that deals with the definition of health conditions through the use of visual images that reflect information from within the human body.

Radiology Information System (RIS) Computer-based information system that supports radiology department operations; includes management of the film library, scheduling of patient examinations, reporting of results, and billing.

Random-access memory (RAM) The portion of a computer's working memory that can be both read and written into. It is used to store the results of intermediate computation, and the programs and data that are currently in use (also called variable memory or core memory).

Randomized clinical trial (RCT) A prospective experiment in which subjects are randomly assigned to study subgroups to compare the effects of alternate treatments.

Randomly Without bias.

Range check A procedure applied to entered data that detects or prevents entry of values that are out of range; e.g., a serum potassium level of 50.0 mmol/L—the normal range for healthy individuals is 3.5–5.0 mol/L.

Ransomware Malicious software that blocks access to a computer system or its data until a sum of many is paid to the perpetrators.

Read-only memory (ROM) The portion of a computer's working memory that can be read, but not written into.

Really simple syndication (RSS) A form of XML that publishes a list of headlines, article titles or events encoded in a way that can be easily read by another program called a news aggregator or news reader.

Real-time acquisition The continuous measurement and recording of electronic signals through a direct connection with the signal source.

Real-time feedback This is feedback to the learner in response to each action taken by the learner. Real time feedback is particularly useful in the initial steps of learning a topic. As the learner becomes more experienced with a topic, real time feedback is often withdrawn and summative feedback is provided at the end of a session.

Recall In information retrieval, the ability of a system to retrieve relevant information (expressed as the ratio of relevant records retrieved to all relevant records in the database).

Receiver In data interchange, the program or system that receives a transmitted message.

Receiver operating characteristic (ROC) A graphical plot that depicts the performance of a binary classifier system as its discrimination threshold is varied.

Records In a data file, a group of data fields that collectively represent information about a single entity.

Reductionist approaches An attempt to explain phenomena by reducing them to common, and often simple, first principles.

Reductionist biomedical model A model of medical care that emphasizes pathophysiology and biological principles. The model assumes that diseases can be understood purely in terms of the component biological processes that are altered as a consequence of illness.

Reference Information Model (RIM) The data model for HL7 Version 3.0. The RIM describes the kinds of information that may be transmitted within health-care organizations, and includes *acts* that may take place (procedures, observations, interventions, and so on), relationships among acts, the manner in which health-care personnel, patients, and other entities may participate in such acts, and the roles that can be assumed by the participants (patient, provider, specimen, and so on).

Reference resolution In NLP, recognizing that two mentions in two different textual locations refer to the same entity.

Reference standard See gold standard test.

Referential expression A sequence of one or more words that refers to a particular person, object or event, e.g., "she," "Dr. Jones," or "that procedure".

Referral bias In evaluation studies, a bias that is introduced when the patients entering a study are in some way atypical of the total population, generally because they have been referred to the study based on criteria that reflect some kind of bias by the referring physicians.

Region of interest (ROI) A selected subset of pixels within an image identified for a particular purpose.

Regional Extension Centers (RECs) In the con- text of health information technology, the 60+ state and local organizations (initially funded by ONC) to help primary care providers in their designated area adopt and use EHRs through out-reach, education, and technical assistance.

Regional Health Information Organization (RHIO) A community-wide, multi-stakeholder organization that utilizes information technology to make more complete patient information and decision support available to authorized users when and where needed.

Regional network A network that provides regional access from local organizations and individuals to the major backbone networks that interconnect regions.

Registers In a computer, a group of electronic switches used to store and manipulate numbers or text.

Registry A data system designed to record and store information about the health status of patients, often including the care that they

receive. Such collections are typically organized to include patients with a specific disease or class of diseases.

Regular expression A mathematical model of a set of strings, defined using characters of an alphabet and the operators concatenation, union and closure (zero or more occurrences of an expression).

Regulated Clinical Research Information Management (RCRIM) An HL7 workgroup that is developing standards to improve information management for preclinical and clinical research.

Relations among named entities The characterization of two entities in NLP with respect to the semantic nature of the relationship between them.

Relative recall An approach to measuring recall when it is unrealistic to enumerate all the relevant documents in a database. Thus the denominator in the calculation of recall is redefined to represent the number of relevant documents identified by multiple searches on the query topic.

Relevance judgment In the context of information retrieval, a judgment of which documents should be retrieved by which topics in a test collection.

Relevance ranking The degree to which the results are relevant to the information need specified in a query.

Reminder message A computer-generated warning that is generated when a record meets prespecified criteria, often referring to an action that is expected but is frequently forgotten; e.g., a message that a patient is due for an immunization.

Remote access Access to a system or to information therein, typically by telephone or communications network, by a user who is physically removed from the system.

Remote Intensive Care Use of networked communications methods to monitor patients in an intensive care unit from a distance far removed from the patients themselves. See remote monitoring.

Remote interpretation Evaluating tests (especially imaging studies) by having them delivered digitally to a location that may be far removed from the patient.

Remote monitoring The use of electronic devices to monitor the condition of a patient from a distant location. Typically used to refer to the ability to record and review patient data (such as vital signs) by a physician located in his/her office or a hospital while the patient remains at home. See also remote intensive care.

Remote-presence health care The use of video teleconferencing, image transmission, and other technologies that allow clinicians to evaluate and treat patients in other than face-to-face situations.

Report generation A mechanism by which users specify their data requests on the input screen of a program that then produces the actual query, using information stored in a database schema, often at predetermined intervals.

Representation A level of medical data encoding, the process by which as much detail as possible is coded.

Representational effect The phenomenon by which different representations of a common abstract structure can have a significant effect on reasoning and decision making.

Representational state A particular configuration of an information-bearing structure, such as a monitor display, a verbal utterance, or a printed label, that plays some functional role in a process within the system.

Representativeness A heuristic by which a person judges the chance that a condition is true based on the degree of similarity between the current situation and the ste-

reotypical situation in which the condition is true. For example, a physician might estimate the probability that a patient has a particular disease based on the degree to which the patient's symptoms matches the classic disease profile.

Request for Proposals A formal notification of a funding opportunity, requiring application through submission of a grant proposal.

Research protocol In clinical research, a prescribed plan for managing subjects that describes what actions to take under specific conditions.

Resource Description Framework (RDF) An emerging standard for cataloging metadata about information resources (such as Web pages) using the Extensible Markup Language (XML).

RESTful API A "lightweight" application programming interface that enables the transfer of data between two Web-based software systems.

Results reporting A software system or subsystem used to allow clinicians to access the results of laboratory, radiology, and other tests for a patient.

Retrieval A process by which queries are com- pared against an index to create results for the user who specified the query.

Retrospective chart review The use of past data from clinical charts (classically paper records) of selected patients in order to perform research regarding a clinical question. See also retrospective study.

Retrospective study A research study performed by analyzing data that were previously gathered for another purpose, such as patient care. See also retrospective chart review.

Return on investment A metric for the benefits of an investment, equal to the net benefits of an investment divided by its cost.

Review of systems The component of a typical history and physical examination in which the physician asks general questions about each of the body's major organ systems to discover problems that may not have been suggested by the patient's chief complaint.

RFP See: Request for Proposals.

Ribonucleic acid (RNA) Ribonucleic acid, a nucleic acid present in all living cells. Its principal role is to act as a messenger carrying instructions from DNA in the production of proteins.

Rich text format (RTF) A format developed to allow the transfer of graphics and formatted text between different applications and operating systems.

RIM See Reference Information Model.

Risk attitude A person's willingness to take risks.

Risk-neutral Having the characteristic of being indifferent between the expected value of a gamble and the gamble itself.

Role-limited access The mechanism by which an individual's access to information in a database, such as a medical record, is limited depending upon that user's job characteristics and their need to have access to the information.

Router/switch In networking, a device that sits on the network, receives messages, and for- wards them accordingly to their intended destination.

RS-232 A commonly used standard for serial data communication that defines the number and type of the wire connections, the voltage, and the characteristics of the signal, and thus allows data communication among electronic devices produced by different manufacturers.

RSS feed A bliographic message stream that provides content from Internet sources.

Rule engine A software component that implements an inference engine that operates on production rules.

Rule-based system A kind of knowledge-based system that performs inference using production rules.

Sampling rate The rate at which the continuously varying values of an analog signal are measured and recorded.

Scenario A method of teaching that presents a clinical problem in a story format.

Schema In a database-management system, a machine-readable definition of the contents and organization of a database.

Schemata Higher-level kinds of knowledge structures.

SCORM Shareable Content Object Reference Model, a standard for interoperability between learning content objects.

Script In software systems, a keystroke-by-keystroke record of the actions performed for later reuse.

SDO See: Standards development organizations.

Search A synonym for information retrieval.

Search See Information retrieval.

Search engine A computer system that returns content from a search statement entered by a user.

Secondary knowledge–based information Writing that reviews, condenses, and/or synthesizes the primary literature (see primary knowledge-based information).

Secret-key cryptography In data encryption, a method whereby the same key is used to encrypt and to decrypt information. Thus, the key must be kept secret, known to only the sender and intended receiver of information.

Secure Sockets Layer (SSL) A protocol for transmitting private documents via the Internet. It has been replaced by Transport Layer Security. By convention, URLs that require an SSL connection start with https: instead of http:

Security The process of protecting information from destruction or misuse, including both physical and computer-based mechanisms.

Segmentation In image processing, the extraction of selected regions of interest from an image using automated or manual techniques.

Selectivity In data collection and recording, the process that accounts for individual styles, reflecting an ongoing decision-making process, and often reflecting marked distinctions among clinicians.

Self-experimentation Experiments in which experimenters themselves are subjects of their research.

Semantic analysis The study of how symbols or signs are used to designate the meaning of words and the study of how words combine to form or fail to form meaning.

Semantic class In NLP, a broad class that is associated with a specific domain and includes many instances.

Semantic grammar A mathematical model of a set of sentences based on patterns of semantic categories, e.g., patient, doctor, medication, treatment, and diagnosis.

Semantic network A knowledge source in the UMLS that provides a consistent categorization of all concepts represented in the Metathesaurus in which each concept is assigned at least one semantic type.

Semantic patterns The study of the patterns formed by the co-occurrence of individual words in a phrase of the co-occurrence of the associated semantic types of the words.

Semantic relations A classification of the meaning of a linguistic relationship, e.g., "treated in 1995" signifies time while "treated in ER" signifies location.

Semantic sense In NLP, the distinction between individual word meaning of terms that may be in the same semantic class.

Semantic types The categorization of words into semantic classes according to meaning. Usually, the classes that are formed are relevant to specific domains.

Semantic Web A future view which envisions the Internet not only as a source of content but also as a source of intelligently linked, agent-driven, structured collections of machine-readable information.

Semantics The meaning of individual words and the meaning of phrases or sentences consisting of combinations of words.

Semi structured interview Where the investigator specifies in advance a set of topics that he would like to address but is flexible about the order in which these topics are addressed, and is open to discussion of topics not on the pre-specified list.

Sender In data interchange, the program or system that sends a transmitted message.

Sensitivity (of a test) The probability of a positive result, given that the condition under consideration is present—for example, the probability of a positive test result in a person who has the disease under consideration (also called the true-positive rate).

Sentence boundary In NLP, distinguishing the end of one sentence and the beginning of the next.

Sentiment analysis The study of how symbols or signs are used to designate the meaning of words and the study of how words combine to form or fail to form meaning.

Sequence alignment An arrangement of two or more sequences (usually of DNA or RNA), highlighting their similarity. The sequences are padded with gaps (usually denoted by dashes) so that wherever possible, columns contain identical or similar characters from the sequences involved.

Sequence database A database that stores the nucleotide or amino acid sequences of genes (or genetic markers) and proteins respectively.

Sequence information Information from a database that captures the sequence of component elements in a biological structure (e.g., the sequence of amino acids in a protein or of nucleotides in a DNA segment).

Sequential Bayes A reasoning method based on a naïve Bayesian model, where Bayes' rule is applied sequentially for each new piece of evidence that is provided to the system. With each application of Bayes' rule, the posterior probability of each diagnostic possibility is used as the new prior probability for that diagnosis the next time Bayes' rule is invoked.

Server A computer that shares its resources with other computers and supports the activities of many users simultaneously within an enterprise.

Service An intangible activity provided to consumers, generally at a price, by a (presumably) qualified individual or system.

Service oriented architectures (SOA) A software design framework that allows specific processing or information functions (services) to run on an independent computing platform that can be called by simple messages from another computer application. Often considered to be more flexible and efficient than more traditional data base architectures. Best known example is the Internet which is based largely on SOA design principles.

Set-based searching Constraining a search to include only documents in a given class or set (e.g., from a given institution or journal).

Set-top box A device, such as a cable box, that converts video content to analog or digital television signals.

Shallow parsing See partial parsing.

Shielding In cabling, refers to an outer layer of insulation covering an inner layer of conducting material. Shielded cable is used to reduce electronic noise and voltage spikes.

Short-term/working memory An emergent property of interaction with the environment; refers to the resources needed to maintain information active during cognitive activity.

Signal processing An area of systems engineering, electrical engineering and applied mathematics that deals with operations on or analysis of signals, or measurements of time-varying or spatially-varying physical quantities.

Simple Mail Transport Protocol (SMTP) The standard protocol used by networked systems, including the Internet, for packaging and distributing email so that it can be processed by a wide variety of software systems.

Simple Object Access Protocol (SOAP) A protocol for information exchange through the HTTP/HTTPS or SMTP transport protocol using web services and utilizing Extensible Markup Language (XML) as the format for messages.

Simulation A system that behaves according to a model of a process or another system; for example, simulation of a patient's response to therapeutic interventions allows a student to learn which techniques are effective without risking human life.

Simulation center Specialized type of learning center, though its governance may reside in an academic department such as anesthesiology or surgery depending on the center's origin and history.

Simultaneous access Access to shared, computer-stored information by multiple concurrent users.

Simultaneous controls Use of participants in a comparative study who are not exposed to the information resource. They can be randomly allocated to access to the information resource or in some other way.

Single nucleotide polymorphism (SNP) A DNA sequence variation, occurring when a single nucleotide in the genome is altered. For example, a SNP might change the nucleotide sequence AAGC**C**TA to AAGC**T**TA. A variation must occur in at least 1% of the population to be considered a SNP.

Single-photon emission computed tomography A nuclear medicine tomographic imaging technique using gamma rays. It is very similar to conventional nuclear medicine planar imaging using a gamma camera. However, it is able to provide true 3D information. This information is typically presented as cross-sectional slices through the patient, but can be freely reformat- ted or manipulated as required.

Single-user systems Computers designed for use by single individuals, such as personal computers, as opposed to servers or other resources that are designed to be shared by multiple people at the same time.

Six sigma A management strategy that seeks to improve the quality of work processes by identifying and removing the causes of defects and minimizing the variability of those processes. Statistically, a six sigma process is one that is free of defects or errors 99.99966%, which equates to operating a process that fits six standard deviations between the mean value of the process and the specification limit of that process.

Slip A type of medical error that occurs when the actor selects the appropriate course of action, but it was executed inappropriately.

Slots In a frame-based representation, the elements that are used to define the semantic characteristics of the frame.

SMART See: Substitutable Medical Applications and Reusable Technologies.

SMART on FHIR An open, standards-based platform for medical apps to access patients' data from electronic medical records. SMART on FHIR builds on two technology efforts: the Substitutable Medical Applications, Reusable Technologies (SMART) Platforms Project and Fast Health Information Resources (FHIR).

Smart phones A mobile telephone that typically integrates voice calls with access to the Internet to enable both access to web sites and the ability to download email and applications that then reside on the device.

Smartwatch A type of wearable computer in the form of a wristwatch. Typically provides health monitoring features, ability to run simple third-party apps, and WiFi or Bluetooth connectivity, in addition to telling time.

SMS messaging The sending of messages using the text communication service component of phone, web or mobile communication system-Short Message Service.

SNOMED Systematized Nomenclature of Medicine—A set of standardized medical terms that can be processed electronically; useful for enhancing the standardized use of medical terms in clinical systems.

SNOMED-CT The result of the merger of an earlier version of SNOMED with the Read Clinical Terms.

SNP See Single nucleotide polymorphism.

Social determinants of health Conditions in which people live, learn, work, and play. Negative examples include: poverty, poor access to healthy foods, substandard education, unsafe neighborhoods.

Social networking The use of a dedicated Web site to communicate informally (on the site, by email, or via SMS messages) with other members of the site, typically by posting messages, photographs, etc.

Sociotechnical systems An approach to the study of work in complex settings that emphasizes the interaction between people and technology in workplaces.

Software Computer programs that direct the hardware how to carry out specific automated processes.

Software development life cycle (SDLC) or software development process A framework imposed over software development in order to better ensure a repeatable, predictable process that controls cost and improves quality of a software product.

Software oversight committee A groups within organizations that is constituted to oversee computer programs and to assess their safety and efficacy in the local setting.

Software psychology A behavioral approach to understanding and furthering software design, specifically studying human beings' interactions with systems and software. It is the intellectual predecessor to the discipline of Human-Computer interaction.

Solid state drive (SSD) A data storage device using integrated circuit assemblies as memory to store data persistently. SSDs have no moving mechanical components, which distinguish them from traditional electromechanical magnetic disks such as hard disk drives (HDDs) or floppy disks, which contain spinning disks and movable read/write heads.

Spamming The process of sending unsolicited email to large numbers of unwilling recipients, typically to sell a product or make a political statement.

Spatial resolution A measure of the ability to distinguish among points that are close to each other (indicated in a digital image by the number of pixels per square inch).

Specialist Lexicon One of three UMLS Knowledge Sources, this lexicon is intended to be a general English lexicon that includes many biomedical terms and supports natural language processing.

Specificity (of a test) The probability of a negative result, given that the condition under consideration is absent—for example, the probability of a negative test result in a person who does not have a disease under consideration (also called the true-negative rate).

Spectrum bias Systematic error in the estimate of a study parameter that results when the study population includes only selected subgroups of the clinically relevant population—for example, the systematic error in the estimates of sensitivity and specificity that results when test performance is measured in a study population consisting of only healthy volunteers and patients with advanced disease.

Speech recognition Translation by computer of voice input, spoken using a natural vocabulary and cadence, into appropriate natural language text, codes, and commands.

Spelling check A procedure that checks the spelling of individual words in entered data.

Spirometer An instrument for measuring the air capacity of the lungs.

Standard of care The community-accepted norm for management of a specified clinical problem.

Standard order sets Predefined lists of steps that should be taken to deal with certain recurring situations in the care of patients, typically in hospitals; e.g., orders to be followed routinely when a patient is in the post-surgical recovery room.

Standard-gamble A technique for utility assessment that enables an analyst to determine the utility of an outcome by comparing an individual's preference for a chance event when compared with a situation of certain outcome.

Standards development organizations An organization charged with developing a standard that is accepted by the community of affected individuals.

Static In patient simulations, a program that presents a predefined case in detail but which does not vary in its response depending on the actions taken by the learner.

Stemming The process of converting a word to its root form by removing common suffixes from the end.

Stop words In full-text indexing, a list of words that are low in semantic content (e.g., "the", "a", "an") and are generally not useful as mechanisms for retrieving documents.

Storage devices A piece of computer equipment on which information can be stored.

Store-and-forward A telecommunications technique in which information is sent to an intermediate station where it is kept and sent at a later time to the final destination or to another intermediate station.

Strict product liability The principle that states that a product must not be harmful.

Structural alignment The study of methods for organizing and managing diverse sources of information about the physical organization of the body and other physical structures.

Structural informatics The study of methods for organizing and managing diverse sources of information about the physical organization of the body and other physical structures. Often used synonymously with "imaging informatics".

Structure validation A study carried out to help understand the needs for an information resource, and demonstrate that its proposed structure makes sense to key stakeholders.

Structured data entry A method of human-computer interaction in which users fill in missing values by making selections from predefined menus. The approach discretizes user input and makes it possible for a computer system to reason directly with the data that are provided.

Structured encounter form A form for collecting and recording specific information during a patient visit.

Structured interview An interview with a schedule of questions that are always presented in the same words and in the same order.

Structured Query Language (SQL) A commonly used syntax for retrieving information from relational databases.

Structured reports A report where the content of the report has coded values for the key information in each pre-specified part of the report, enabling efficient and reliable computation on the report.

Study arm in the context of clinical research, a study arm represents a specific modality of an experimental intervention to which a participant is assigned, usually through a process of randomization (e.g., random assigned in a balanced manner to such an arm). Arms are used in clinical study designs where multiple variants of a given experimental intervention are under study, for example, varying the timing or dose of a given medication between arms to determine an optimal therapeutic strategy.

Study population The population of subjects—usually a subset of the clinically relevant population—in whom experimental outcomes (for example, the performance of a diagnostic test) are measured.

Subheadings In MeSH, qualifiers of subject headings that narrow the focus of a term.

Subjectivist approaches Class of approaches to evaluation that rely primarily on qualitative data derived from observation, interview, and analysis of documents and other artifacts. Studies under this rubric focus on description and explanation; they tend to evolve rather than be prescribed in advance.

Sublanguage Language of a specialized domain, such as medicine, biology, or law.

Substitutable Medical Applications and Reusable Technologies (SMART) A technical platform enables EHR systems to behave as "iPhone-like platforms" through an application programming interface (API) and a set of core services that support easy addition and deletion of third party apps, such that the core system is stable and the apps are substitutable.

Summarization A computer system that attempts to automatically summarize a larger body of content.

Summary ROC curve A composite ROC curve developed by using estimates from many studies.

Summative evaluation after the product is in use, is valuable both to justify the completed project and to learn from one's mistakes.

Supervised learning An approach to machine learning in which an algorithm uses a set of inputs and corresponding outputs to try to learn a model that will enable prediction of an output when faced with a previously unseen input.

Supervised learning technique A method for determining how data values may sug-

gest classifications, where the possible classifications are enumerated in advance, and the performance of a system is enhanced by evaluating how well the system classifies a training set of data. Statistical regression, neural networks, and support vector machines are forms supervised learning.

Supervised machine learning A machine learning approach that uses a gold standard set as input to learn classifiers.

Surveillance The ongoing collection, analysis, interpretation, and dissemination of data on health conditions (e.g., breast cancer) and threats to health (e.g., smoking prevalence). In a computer-based medical record system, systematic review of patients' clinical data to detect and flag conditions that merit attention. Also see public health surveillance and biosurveillance.

Symbolic-programming language A programming language in which the program can treat itself, or material like itself, as data. Such programs can write programs (not just as character strings or texts, but as the actual data structures that the program is made of). The best known and most influential of these languages is LISP.

Syndromic surveillance A particular type of public health surveillance. It is an ongoing process of monitoring clinical data, generally from public health, hospital, or outpatient resources, or surrogate data indicating early illness (e.g., school or work absenteeism) with a goal of early identification of outbreaks, new conditions, health threats, or bioterrorist events.

Synonyms Multiple ways of expressing the same concept.

Syntax The grammatical structure of language describing the relations among words in a sentence.

System programs The operating system, compilers, and other software that are included with a computer system and that allow users to operate the hardware.

Systematic review A type of journal article that reviews the literature related to a specific clinical question, analyzing the data in accordance with formal methods to assure that data are suitably compared and pooled.

Systems biology Research on biological networks or biochemical pathways. Often, systems biology analyses take a comprehensive approach to model biological function by taking the interactions (physical, regulatory, similarity, etc.) of a set of genes as a whole.

Tablet Generally refers to a personal computing device that resembles a paper tablet in size and incorporates features such as a touch screen to facilitate data entry.

Tactile feedback In virtual or telepresence environments, the process of providing (through technology) a sensation of touching an object that is imaginary or otherwise beyond the user's reach (see also haptic feedback).

TCP/IP Transmission Control Protocol/Internet Protocol—A set of standard communications protocols used for the Internet and for net- works within organizations as well.

Teleconsultation The use of telemedicine techniques to support the interaction between two (or more) clinicians where one is providing advice to the other, typically about a specific patient's care.

Telegraphic In NLP, describes language that does not follow the usual rules of grammar but is compact and efficient. Clinical notes written by hand often demonstrate a "telegraphic style".

Telehealth The use of electronic information and telecommunications technologies to support long-distance clinical health care, patient and professional health-related education,

public health and health administration. See telemedicine.

Telehome care The use of communications and information technology to deliver health services and to exchange health information to and from the home (or community) when distance separates the participants.

Tele-ICU See remote intensive care.

Telemedicine A broad term used to describe the delivery of health care at a distance, increasingly but not exclusively by means of the Internet.

Teleophthalmology The use of telemedicine methods to deliver ophthalmology services.

Telepresence A technique of telemedicine in which a viewer can be physically removed from an actual surgery, viewing the abnormality through a video monitor that displays the operative field and allows the observer to participate in the procedure.

Telepsychiatry The use of telemedicine methods to deliver psychiatric services.

Teleradiology The provision of remote interpretations, increasing as a mode of delivery of radiology services.

Telesurgery The use of advanced telemedicine methods to allow a doctor to perform surgery on a patient even though he or she is not physically in the operating room.

Temporal resolution A metric for how well an imaging modality can distinguish points in time that are very close together.

Terabyte A unit of information equal to one million million (10^{12}) or strictly, 2^{40} bytes.

Term A word or phrase.

Term Designation of a defined concept by a linguistic expression in a special language. In information retrieval, a word or phrase which forms part of the basis for a search request.

Term frequency (TF) In information retrieval, a measurement of how frequently a term occurs in a document.

Term weighting The assignment of metrics to terms so as to help specify their utility in retrieving documents well matched to a query.

Terminal A simple device that has no processing capability of its own but allows a user to access a server.

Terminology A set of terms representing the system of concepts of a particular subject field.

Terminology authority An entity or mechanism that determines the acceptable term to use for a specific entity, descriptor, or other concept.

Terminology services Software methods, typically based on computer-based dictionaries or language systems, that allow other systems to determine the locally acceptable term to use for a given purpose.

Test collection In the context of information retrieval, a collection of real-world content, a sampling of user queries, and relevance judgments that allow system-based evaluation of search systems.

Test-interpretation bias Systematic error in the estimates of sensitivity and specificity that results when the index and gold-standard test are not interpreted independently.

Test-referral bias Systematic error in the estimates of sensitivity and specificity that results when subjects with a positive index test are more likely to receive the gold-standard test.

Tethered personal health record An EHR portal that is provided to patients by an institution and can typically be used to manage

information only from that provider organization.

Text generation Methods that create coherent natural language text from structured data or from textual documents in order to satisfy a communication goal.

Text mining The use of large text collections (e.g., medical histories, consultation reports, articles from the literature, web-based resources) and natural language processing to allow inferences to be drawn, often in the form of associations or knowledge that were not previously apparent. See also data mining.

Text processing The analysis of text by computer.

Text readability assessment and simplification An application of NLP in which computational methods are used to assess the clarity of writing for a certain audience or to revise the exposition using simpler terminology and sentence construction.

Text REtrieval Conference (TREC) Organized by NIST, an annual conference on text retrieval that has provided a testbed for evaluation and a forum for presentation of results. (see ► trec.nist.gov).

Text summarization Takes one or several documents as input and produces a single, coherent text that synthesizes the main points of the input documents.

Text-comprehension A process in which text can be described at multiple levels of realization from surface codes (e.g., words and syntax) to deeper level of semantics.

TF*IDF weighting A specific approach to term weighting which combines the inverse document frequency (IDF) and term frequency (TF).

Thesaurus A set of subject headings or descriptors, usually with a cross-reference system for use in the organization of a collection of documents for reference and retrieval.

Thick-client A computer node in a network or client–server architecture that provides rich functionality independent of the central server. See also thin client.

Thin client A program on a local computer system that mostly provides connectivity to a larger resource over a computer network, thereby providing access to computational power that is not provided by the machine, which is local to the user.

Think-aloud protocol In cognitive science, the generation of a description of what a person is thinking or considering as they solve a problem.

Thread The smallest sequence of programmed instructions that can be managed independently by an operating system scheduler.

Three-dimensional printing Construction of a physical model of anatomy or other object by laying down plastic versions of a stack of cross-sectional slices through the object.

Three-dimensional structure information In a biological database, information regarding the three-dimensional relationships among elements in a molecular structure.

Time-sharing networks An historical term describing some of the earliest computer networks allowing remote access to systems.

Time-trade-off A common approach to utility assessment, comparing a better state of health lasting a shorter time, with a lesser state of health lasting a longer time. The time-trade-off technique provides a convenient method for valuing outcomes that accounts for gains (or losses) in both length and quality of life.

Tokenization The process of breaking an unstructured sequence of characters into larger units called "token," e.g., words, numbers, dates and punctuation.

Tokens In language processing, the composite entities constructed from individual char-

acters, typically words, numbers, dates, or punctuation.

Top-down In search or analysis, the breaking down of a system to gain insight into its compositional subsystems.

Topology In networking, the overall connectivity of the nodes in a network.

Touch screen A display screen that allows users to select items by touching them on the screen.

Track pad A computer input device for controlling the pointer on a display screen by sliding the finger along a touch-sensitive surface: used chiefly in laptop computers. Also called a touchpad.

Transaction set In data transfer, the full set of information exchanged between a sender and a receiver.

Transcription The conversion of a recording of dictated notes into electronic text by a typist.

Transcriptomics The study of the set of RNA transcripts that are produced by the genome and the context (specific cells or circumstances) in which transcription occurs.

Transition matrix A table of numbers giving the probability of moving from one state in a Markov model into another state or the state that is reached in a finite-state machine depending on the current character of the alphabet.

Transition probability The probability that a person will transit from one health state to another during a specified time period.

Translational Bioinformatics (TBI) According to the AMIA: the development of storage, analytic, and interpretive methods to optimize the transformation of increasingly voluminous biomedical data, and genomic data, into proactive, predictive, preventive, and participatory health.

Translational medicine Translational medicine: the process of transferring scientific discoveries into preventive practice and clinical care.

Transmission control protocol/internet protocol (TCP/IP) The standard protocols used for data transmission on the Internet and other common local and wide-area networks.

Transport Layer Security (TLS) A protocol that ensures the privacy of data transmitted over the Internet. It grew out of Secure Sockets Layer.

Treatment threshold probability The probability of disease at which the expected values of withholding or giving treatment are equal. Above the threshold treatment is recommended; below the threshold, treatment is not recommended and further testing may be warranted.

Trigger event In monitoring, events that cause a set of transactions to be generated.

True negative In assessing a situation, an instance that is classified negatively and is subsequently shown to have been correctly classified.

True positive In assessing a situation, an instances that is classified positively and is subsequently shown to have been correctly classified.

True-negative rate (TNR) The probability of a negative result, given that the condition under consideration is false—for example, the probability of a negative test result in a patientwho does not have the disease under consideration (also called specificity).

True-negative result (TN) A negative result when the condition under consideration is false—for example, a negative test result in a

patient who does not have the disease under consideration.

True-positive rate (TPR) The probability of a positive result, given that the condition under consideration is true—for example, the probability of a positive test result in a patient who has the disease under consideration (also called sensitivity).

True-positive result (TP) A positive result when the condition under consideration is true—for example, a positive test result in a patient who has the disease under consideration.

Turn-around-time The period for completing a process cycle, commonly expressed as an average of previous such periods.

Tutoring A computer program designed to provide self-directed education to a student or trainee.

Tutoring system A computer program designed to provide self-directed education to a student or trainee. (Also Intelligent Tutoring System).

Twisted-pair wires The typical copper wiring used for routine telephone service but adaptable for newer communication technologies.

Type-checking In computer programming, the act of checking that the types of values, such as integers, decimal numbers, and strings of characters, match throughout their use.

Typology A way of classifying things to make sense of them, for a certain purpose.

Ubiquitous computing A form of computing and human-computer interaction that seeks to embed computing power invisibly in all facets of life.

Ultrasound A common energy source derived from high-frequency sound waves.

UMLS See: Unified Medical Language System.

UMLS Knowledge Sources Components of the Unified Medical Language System that support its use and semantic breadth.

UMLS Semantic Network A knowledge source in the UMLS that provides a consistent categorization of all concepts represented in the Metathesaurus. Each Metathesaurus concept is assigned at least one semantic type from the Semantic Network.

Unicode Represents characters needed for foreign languages using up to 16 bits.

Unified Medical Language System (UMLS) Project A terminology system, developed under the direction of the National Library of Medicine, to produce a common structure that ties together the various vocabularies that have been created for biomedical domains.

Unified Modeling Language (UML) A standardized general-purpose modeling language developed for object-oriented software engineering that provides a set of graphic notation techniques to create visual models that depict the relationships between actors and activities in the program or process being modeled.

Uniform resource identifier (URI) The combination of a URN and URL, intended to provide persistent access to digital objects.

Uniform resource locator (URL) The address of an information resource on the World Wide Web.

Uniform resource name (URN) A name for a Web page, intended to be more persistent than a URL, which often changes over time as domains evolve or Web sites are reorganized.

Unique health identifier (UHI) A government-provided number that is assigned to an individual for purposes of keeping track of their health information.

Universal Serial Bus(USB) A connection technology for attaching peripheral devices to a computer, providing fast data exchange.

Unobtrusive measures Measures made using the records accrued as part of the routine use of the information resource, including, for example, user log files.

Unstructured interview An interview where there are no predetermined questions.

Unsupervised machine learning A machine learning approach that learns patterns from the data without labeled training sets.

URAC An organization that accredits the quality of information from various sources, including health-related Web sites.

Usability The quality of being able to provide good service to one who wishes to use a product.

Usability testing A class of methods for collecting empirical data of representative users performing representative tasks; considered the gold standard in usability evaluation methods.

User authentication The process of identifying a user of an information resource, and verifying that the user is allowed to access the services of that resource. A standard user authentication method is to collect and verify a username and password.

User-centered design An iterative process in which designers focus on the users and their needs in each phase of the design process. UCD calls for involving users throughout the design process via a variety of research and design techniques to increase the likelihood that the product will be highly usable by its intended users.

User-interface layer The architectural layer of a software environment that handles the interface with users.

Utility In decision making, a number that represents the value of a specific outcome to a decision maker (see, for example, quality-adjusted life-years).

Validity check A set of procedures applied to data entered into an EHR intended to detect or prevent the entry of erroneous data; e.g., range checks and pattern checks.

Value-based reimbursement In health care, an alternative to traditional fee-for-service reimbursement, aimed at rewarding quality rather than quantity of services.

Variable Quantity measured in a study. Variables can be measured at the nominal, ordinal, interval, or ratio levels.

Vector mathematics In the context of information retrieval, mathematical systems for measuring and comparing vector representations of documents and their contents.

Vector-space model A method of full-text indexing in which documents can be conceptualized as vectors of terms, with retrieval based on the cosine similarity of the angle between the query and document vectors.

Vendor-neutral archives (VNA) A technology in which images (and potentially any file of clinical relevance) is stored (archived) in a standard format with a standard interface (e.g., DICOM), such that they can be accessed in a vendor-neutral manner by other systems.

Vertically integrated Refers to an organizational structure in which a variety of products or services are offered within a single chain of command; contrasted with horizontal integration in which a single type of product is offered in different geographical markets. A hospital that offers a variety of services from obstetrics to geriatrics would be "vertically integrated." A diagnostic imaging organization with multiple sites would be "horizontally integrated".

Veterinary informatics The application of biomedical informatics methods and techniques to problems derived from the field of veterinary medicine. Viewed as a subarea of clinical informatics.

Video-display terminal (VDT) A device for displaying input signals as characters on a screen, typically a computer monitor.

View In a database-management system, a logical submodel of the contents and structure of a database used to support one or a subset of applications.

View schemas An application-specific description of a view that supports that program's activities with respect to some general database for which there are multiple views.

Virtual address A technique in memory management such that each address referenced by the CPU goes through an address mapping from the virtual address of the program to a physical address in main memory.

Virtual medical record A standard model of the data elements found in EHR systems. The virtual medical record approach assumes that, even if particular EHR implementations adopt nonstandard data dictionaries and disparate ways for storing clinical data, mapping the contents of each EHR to a canonical model greatly simplifies interoperability with CDS Systems and other applications that may need to access the data.

Virtual memory A scheme by which users can access information stored in auxiliary memory as though it were in main memory. Virtual memory addresses are automatically translated into actual addresses by the hardware.

Virtual patient A digital representation of a patient encounter that can range from a simple review of clinical findings to a realistic graphical view of a person who can converse and can be examined for various clinical symptoms and laboratory tests.

Virtual Private Network (VPN) A private communications network, usually used within a company or organization, or by several different companies or organizations, communicating over a public network. VPN message traffic is carried on public networking infrastructure (e.g., the Internet) using standard (often insecure) protocols.

Virtual reality A collection of interface methods that simulate reality more closely than does the standard display monitor, generally with a response to user maneuvers that heighten the sense of being connected to the simulation.

Virtual world A three-dimensional representation of an environment such as a hospital, a clinic or a home-care location. The represented space usually includes a virtual patient, and interactive equipment and supplies that can be used to examine and care for the patient. Some virtual worlds are multi-user and allow multiple learners to manifest themselves as characters in the virtual world for interaction with each other and the patient.

Virus/worm A software program that is written for malicious purposes to spread from one machine to another and to do some kind of damage. Such programs are generally self-replicating, which has led to the comparison with biological viruses.

Visual-analog scale A method for valuing health outcomes, wherein a person simply rates the quality of life with a health outcome on a scale from 0 to 100.

Vocabulary A dictionary containing the terminology of a subject field.

Volatile A characteristic of a computer's memory, in that contents are changed when the next program runs and are not retained when power is turned off.

Volume rendering A method whereby a computer program projects a two-dimensional

image directly from a three-dimensional voxel array by casting rays from the eye of the observer through the volume array to the image plane.

vonNeuman machine A computer architecture that comprises a single processing unit, computer memory, and a memory bus.

Voxel A volume element, or small cubic area of a three-dimensional digital image (see pixel).

Washington DC Principles for Free Access to Science An organization of non-profit publishers that aims to balance wide access with the need to maintain sustainable revenue models.

Wearables In the context of mobile health, wearables refer to a range of electronic devices that can be incorporated into clothing or worn on the body, such as smartwatches, activity trackers, and physiological sensors, that are used to collect health-related data and provide health interventions. Also referred to as wearable devices or wearable technologies.

Web browser A computer program used to access and display information resources on the World Wide Web.

Web catalog Web pages containing mainly links to other Web pages and sites.

Web Services Discovery Language (WSDL) An XML-based language used to describe the attributes of a web service, such as a SOAP service.

Web-based technologies Computer capabilities that rely on the architecture principles of the Internet for accessing data from remote servers.

Weblogs/blogs A type of Web site that provides discussion or information on various topics.

WebMD An American company that provides web-based health information services.

Whole Slide Digitization The process of capturing an entire specimen on a slide into a digital image. Compared with capturing images of a single field of view from a microscope, this captures the entire specimen, and can be millions of pixels on a side. This allows subsequent or remote review of the specimen without requiring capture of individual fields.

Wide-area networks (WANs) A network that connects computers owned by independent institutions and distributed over long distances.

Wi-Fi A common wireless networking technology (IEEE 802.11x.) that uses radio waves to provide high-speed connections to the Internet and local networks.

Word In computer memory, a sequence of bits that can be accessed as a unit.

Word sense disambiguation (WSD) The process of determining the correct sense of a word in a given context.

Word senses The possible meanings of a term.

Word size The number of bits that define a word in a given computer.

Workstation A powerful desktop computer system designed to support a single user. Workstations provide specialized hardware and software to facilitate the problem-solving and information-processing tasks of professionals in their domains of expertise.

World Intellectual Property Organization (WIPO) An international organization, headquartered in Geneva and dedicated to promoting the use and protection of intellectual property.

World Wide Web (WWW or Web) An application implemented on the Internet in which multimedia information resources are made accessible by any of a number of protocols, the most common of which is the HyperText Transfer Protocol (HTTP).

Worm A self-replicating computer program, similar to a computer virus; a worm is self-contained and does not need to be part of another program to propagate itself.

xAPI Experience Application Programming Interface goes beyond interoperability standards, such as SCORM, and supports collection of data about the learner's experience while using the learning object.

XML A metalanguage that allows users to define their own customized markup languages. See Extensible Markup Language.

X-ray crystallography A technique in crystallography in which the pattern produced by the diffraction of x-rays through the closely spaced lattice of atoms in a crystal is recorded and then analyzed to reveal the nature of that lattice, generally leading to an understanding of the material and molecular structure of a substance.

Name Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Subject Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z