ASSOCIATION OF AMERICAN MEDICAL COLLEGES

Information Technology Enabling Clinical Research

Findings and Recommendations from a Conference Sponsored by the Association of American Medical Colleges with Funding from the National Science Foundation

ASSOCIATION OF AMERICAN MEDICAL COLLEGES

Information Technology Enabling Clinical Research

Findings and Recommendations from a Conference Sponsored by the Association of American Medical Colleges with Funding from the National Science Foundation

October 30-31, 2002

To request additional copies of this publication, please contact:

Morgan Passiment
Association of American Medical Colleges
2450 N Street, NW
Washington, DC 20037-1134
Phone: 202-828-0476 • Fax: 202-828-1125
E-mail: mpassiment@aamc.org

www.aamc.org

Copyright 2003 by the Association of American Medical Colleges. All rights reserved.

Table of Contents

Executive Summary 1

Summary Recommendations 4

Information Requirements for Clinical Research 8

Advance Clinical Information Systems 20

Clinical Data Standards 36

The Roles of Industry and Government 48

Glossary of Acronyms 64

With funding from the National Science Foundation, the Association of American Medical Colleges (AAMC) sponsored a conference on October 30-31, 2002, in Leesburg, Virginia, to discuss information technology (IT) and the future of clinical research.

In the eight months prior to the conference, a distinguished steering committee outlined issues to be addressed across five panels, each representing an important constituency:

- Clinical research, including patient oriented research, genetics, clinical trials, and population-based research (including public health and health services research)
- IT leaders from major healthcare organizations
- Leaders in the area of standards development for healthcare IT
- The principal major commercial healthcare software vendor companies
- Representatives of major government agencies with a commitment to healthcare IT

Early in their deliberations, the steering committee was approached by staff of the Clinical Research Roundtable (CRR), a program sponsored by the Institute of Medicine (IOM), to provide guidance on surveying the state of IT applications for clinical research at academic medical centers (medical schools and teaching hospitals). The staff of the CRR agreed to present their findings at the conference. Their findings, as well as an overview paper by John Glaser of Partners Health Care, have been incorporated into Chapter 2 of this document.

The steering committee identified the panelists in each area. A steering committee member chaired each of the five panels; having previously identified the panel members and helped them to focus in on issues and develop their presentations before the conference. Following the conference, panelists reviewed the summaries of their presentations. This report represents a summary and synthesis of the work of the conference. It has been reviewed and endorsed by each of the members of the steering committee.

Steering Committee Membership
Roger E. Meyer, MD, AAMC, Washington, DC, chair; Marion J. Ball, EdD, Healthlink, Baltimore, MD; Christopher G. Chute, MD, DrPH, Mayo Foundation, Rochester, MN; James J. Cimino, MD, Columbia University, New York, NY; Milton Corn, MD, National Library of Medicine, Bethesda, MD; William E. Hammond, PhD, Duke University, Durham, NC; Stephanie L. Reel, Johns Hopkins Medicine, Baltimore, MD

Panel 1: The Needs of Different Clinical Research Areas
Stephanie L. Reel, John Hopkins, chair; Robert M. Califf, MD, Duke Clinical Research Institute, Durham, NC (Clinical Trials); Frank Hartel, PhD, National Cancer Institute, Bethesda, MD (Cancer); Bruce R. Korf, MD, PhD, University of Alabama Birmingham Center for Genetics and Genomics, Birmingham, AL (Human Genetics and Genomics); Steven L. Solomon, MD, Centers for Disease Control, Atlanta, GA (Public Health and Health Services Research); Gordon H. Williams, MD, Brigham and Women's Hospital, Boston, MA (Translational Research)

Panel 2: Clinical Information Systems: State of the Art
Milton Corn, MD, National Library of Medicine, chair; Homer L. Chin, MD, Kaiser Permanente Northwest, Portland, OR; Carol Hanges, Merck & Co., Inc., Rahway, NJ; Robert M. Kolodner, MD, Veterans Health Administration, Washington, DC; Daniel R. Masys, MD, University of San Diego Medical School, San Diego, CA; Randolph A. Miller, MD, Vanderbilt University Medical Center, Nashville, TN

Panel 3: Standards Development
Christopher G. Chute, MD, DrPH, Mayo Foundation, Rochester, MN, and James J. Cimino, MD, Columbia University, New York, NY, co-chairs; David W. Bates, MD, MSc, Brigham and Women's Hospital, Boston, MA; Stanley Huff, MD, Intermountain Health Care, Salt Lake City, UT; Clement J. McDonald, MD, Regenstrief Institute, Indianapolis, IN

Panel 4: The Views of the Vendor Community
William E. Hammond, PhD, Duke, Durham, NC, chair; Nick Beard, MD, MS, IDX Systems, Seattle, WA; Barry P. Chaiken, MD, MPH, McKesson Corporation; Floyd Eisenberg, MD, MPH, Siemens Medial Solutions, Malvern, PA; Paul Gertman, MD, Eclipsys Corporation, Boston, MA; David P. McCallie, Jr.,MD, Cerner Corporation, Kansas City, MO

Panel 5: The Views of Government Agencies
Roger E. Meyer,MD, AAMC, Washington, DC, chair; Claire V. Broome, MD, Centers for Disease Control, Atlanta, GA; Gene A. Carruth, Blue Cross and Blue Shield, Phoenix, AR; Donald A.B. Lindberg, MD, National Library of Medicine, Bethesda, MD; Eduardo Ortiz, MD, MPH, Agency for Healthcare Research and Quality, Rockville, MD; William Rollow MD, MPH, Centers for Medicare and Medicaid Services, Baltimore, MD; Judith L. Vaitukaitis, MD, National Center for Research Resources, Bethesda, MD

Additional Speakers
John P. Glaser, PhD, Partners Health Care System, Boston, MA; Alexander Ommaya, Sc.D., Institute of Medicine, Washington, DC

Other Participants
William F. Crowley, MD, Massachusetts General Hospital, Boston, MA; Jorge Ferrer, MD, MBA, Centers for Medicare and Medicaid Services, Baltimore, MD; Jay A. Freedman, PhD, Department of Veterans Affairs, Washington, DC; Janet Marchibroda, eHealth Initiative, Washington, DC; Robert Nussenblatt, National Eye Institute, Bethesda, MD; Stephen J. Rosenfeld, MD, National Institutes of Health Clinical Center, Bethesda, MD; Dean F. Sittig, MD, Center for Health Research, Portland, OR

Staff Support
Morgan L. Passiment and Stacia Gueniguian, AAMC; Judith V. Douglas, project writer

Executive Summary

"Clinical research is a component of medical and health research intended to produce knowledge valuable for understanding human disease, preventing and treating illness, and promoting health. Clinical research embraces a continuum of studies involving interactions with patients, diagnostic clinical materials or data, or populations in a number of categories including: disease mechanisms; translational research; clinical knowledge, detection, diagnosis and natural history of disease; therapeutic interventions including clinical trials; prevention and health promotion; behavioral research; epidemiology; and community-based and managed care-based research."[1]

While clinical research is a defining characteristic of medical schools and teaching hospitals, for the most part these institutions have not focused their information technology (IT) development efforts to meet clinical research needs. Rather, like other healthcare institutions, they have applied IT to financial reporting and payer requirements. More recently they have targeted clinical care for IT enhancements that enable them to comply with new federal regulations, specifically the Health Insurance Portability and Accountability Act (HIPAA), and to respond to national initiatives in the areas of patient safety, healthcare quality, and biosurveilance. Importantly, clinical research is a significant element in improving health care quality, strengthening patient safety, and dealing with bioterrorism. If clinical research IT needs are addressed within these critical national and local IT initiatives, the health of the American people will benefit.

Although all clinical research disciplines share the need for complex, consistently defined data, individual research fields and projects have widely varying IT needs. For example, translational research requires highly specialized data collection in a small number of subjects in environmentally controlled clinical research facilities using real time acquisition and analyses. The data may vary from semiquantitative self-reports in rating scales, to standard laboratory test or X ray results, to cutting edge biological markers of disease states and/or organ function. Genetics research has emerged in the last decade with great promise because of the completion of the Human Genome Project. Genetics research is heavily dependent upon IT in relation to genotyping capabilities and repositories; and the ability to link genetic information to clinically derived phenotypes, environmental exposures, and differential outcomes in clinical trials.

[1] This definition of clinical research was developed in the course of a Consensus Development Conference at Graylyn, North Carolina, November 20-22, 1998, as part of the Clinical Research Summit Process sponsored by the AAMC, AMA and the Wake Forest University School of Medicine, and supported by the Pew Charitable Trusts, the Commonwealth Fund, the Robert Wood John Foundation, the John D. and Catherine T. MacArthur Foundation, the Burroughs Wellcome Fund, the Merck Foundation, and the Ethics and Leadership Fund at Wake Forest University.

Population-based investigators in public health and health services research share a need for high quality clinical information that can be accessed by researchers at distal locations. This research would be greatly facilitated by the creation of federated databases. Clinical trials researchers would benefit enormously from clinical IT systems that might help them to identify potential subjects for clinical trials, and to track long-term outcomes through IT systems capable of post-clinical trial and post-marketing surveillance, consistent with federal regulations and institutional review board requirements. If clinical research is not one of the drivers for IT enhancements, and/or if the worlds of clinical systems IT and of clinical research continue to operate in separate domains within the same institutions, then the promise of IT and of clinical research to the future of clinical care will be significantly compromised.

It is technologically possible to incorporate IT enhancements for clinical research into advanced clinical information systems (CIS) and other high-priority applications developed to support clinical care. Clinicians and clinical researchers all depend upon access to complete, consistent, and comprehensive patient data, but they require different functionalities. Clinicians require access to data on an individual patient. Clinical researchers often need to group and compare information from many records that may involve multiple sites. Optimal advanced CIS can enable both researchers and clinicians to do their jobs easier, faster, and better. Clinical research IT requires the creation of institutional data warehouses and repositories to hold patient-focused data, in which data derived from clinical care are stored for research. The data warehouse remains distinct from the clinical information system that meets the operational needs of clinicians. Planning and operation of data warehouses are essential elements in IT support for clinical research. They need to be adequately funded and staffed, including clinical informaticians who can serve as “data mediators” or “translators” to “modify” clinical data in the warehouse/repository for use by different clinical research end-users for varied purposes.

An optimal clinical information system for patient care and clinical research is dependent upon the adoption of common standards and terminologies across health care and upon the development of off-the-shelf solutions that can be readily adapted at the local level. Existing standards such as Health Level Seven (HL7) for message data interchange and Logical Observation Identifier Names and Codes (LOINC) for laboratory results have the potential to make data more consistent and comparable. Many health care institutions have yet to adopt these standards, due in part to costs and in part to institutional inertia. While most federal agencies have been participating in standards development, there is a need to bridge efforts at the Food and Drug Administration with those of other federal agencies to assure compatibility of vocabulary between the data gathered in sponsored clinical trials and in other healthcare applications. Clinical research should not create new standards; rather, it should build its requirements into the standards now being developed, using the capabilities they provide.

A critical linchpin in the development of healthcare IT applications is a small group of commercial vendors that have developed software already in operation at many health care institutions, for healthcare operations and financial management. That the vendor community serves the needs of multiple health systems brings potential advantages of interoperability and

connectivity across health systems. However, these vendors have insufficent resources to undertake general IT research for a community that has not yet enforced the power of standards. Because the sector is high-risk and attracts little venture capital, new product development tends to be totally market driven. In the absence of clear evidence of a market for software to support both patient care and clinical research, the vendor sector lacks the incentive to develop such large-scale, complex software.

Teaching hospitals and faculty practice plans are dependent upon vendors to help them to address their clinical IT needs and to comply with government and other regulatory reporting requirements. In these times, most institutions lack the capital to replace legacy IT systems because of cost. They do not value the need for common standards, and do not see a mandate for specific standards (particularly terminology) that would justify the expediture for change. They have insufficent reserves to support general research in IT and they do not offer modules for clinical research. The existing grant mechanisms at the National Institutes of Health, the Agency for Healthcare Research and Quality, and the Centers for Disease Control, exclude or limit funding for vendors. Although governmental agencies like the Centers for Medicare and Medicaid Services (CMS) might be able to leverage their influence on clinical IT development as major purchasers of health care, and by encouraging standards development, a true leadership role requires that federal efforts be brought together into a coherent program of support.

At this juncture, there are some promising developments and opportunities that highlight possible IT enhancements for clinical research as well as government/industry/academic collaborations, but there is no overarching vision or funding that would lead to the type of system change necessary to bring the full benefits of IT to clinical research and to our national health care agenda. The conference identified a number of the promising initiatives such as the planning for clinical care, operations and clinical research in the new IT systems for the Department of Veterans Affairs, within a part of the Kaiser Permanente health system, at the Centers for Disease Control and Prevention, the Centers for Medicare and Medicaid Services, the Agency for Healthcare Research and Quality, and the National Institutes of Health.

There is a real opportunity for government agencies to work together more closely to advance an IT agenda and to strengthen clinical research and clinical IT resources at teaching hospitals and medical schools. In the end, given the magnitude of the tasks and the very uneven distribution of high quality healthcare-related IT resources across the country, the Congress and the Executive Branch need to consider whether the large national agenda around health care IT initiatives calls for a program with a grand vision and funding to match the vision. When faced with the need for a broad hospital modernization program in an earlier era, Congress passed and funded the Hill Burton program. The challenges, in relation to clinical and clinical research IT enhancements, call for a similar vision and commitment to bring the healthcare system and clinical research into the 21st century.

Summary Recommendations

1. **The Association of American Medical Colleges, the Clinical Research Roundtable of the Institute of Medicine, and clinical research advocacy groups should lead an advocacy effort to advance the development of information systems for clinical research purposes, and to implement the recommendations of this Report.**

2. **Incorporate the IT requirements of clinical research investigators into the evolving health information technology infrastructure.**

   Integrate features that support clinical research into clinical information systems used for patient care.

   Improve the quality of clinical data gathered during routine care through the further development of the electronic medical record and the adoption of computerized physician order entry.

   Develop a common set of specifications for clinical information systems which academic health centers would agree to accept without customization, allowing vendors to focus on systems development. Exploit capabilities now in place by funding health care institutions to take those capabilities to the next level.

   Engage the clinical research community in the development and evolution of clinical information systems under the direction of the vendor community and the IT leadership at medical schools and teaching hospitals. Bring them into the process early to address both human factors and technical issues.

   Align planning and advocacy for clinical research IT enhancements to other parts of the national healthcare IT agenda, such as patient privacy, patient safety, biosurveilance and quality improvement.

3. **Clinical research IT should use and build upon existing health care IT standards. Standards adoption must be coordinated across all healthcare communities.**

   Support the use of existing standards, including Health Level Seven (HL7) for messaging, Digital Imaging and Communications in Medicine (DICOM) for images, National Council for Prescription Drugs (NCPDP), Systematized Nomenclature of Medicine (SNOMED) for clinical vocabulary, Logical Observation Identifier Names and Codes (LOINC) for laboratory test names, and Accredited Standards Committee (ASC) X12 for electronic exchange of billing information.

   The federal government should be playing a major leadership role in endorsing the implementation of standards, and assuring that funding is available to enable institutions to bring their CIS into compliance with the standards.

Actively involve data users, including researchers, healthcare providers, and public health officials, in efforts to promote the widespread adoption and implementation of these standards.

Bridge the apparent terminology disconnect between the Food and Drug Administration and the pharmaceutical industry and their use of the Medical Dictionary for Regulatory Activities (MedDRA), with the standards such as SNOMED and HL7 accepted by other health agencies and healthcare systems[2].

Ensure data validity by coordinating efforts such as those by the Patient Safety Data Standards Committee to maximize the utility of data in electronic databases by standardizing data collection methods and by developing/implementing criteria and methods to evaluate and monitor data quality.

4. **Design systems for effective storage and retrieval of clinical data to support clinical research.**

Create institutional data warehouses and clinical-research focused warehouses that contain aggregated data from more than one institution as repositories to hold patient-focused data.

Store data derived from clinical care in a data warehouse distinct from the clinical information system meeting operational needs of clinicians.

Determine under whose auspices and what funding sources data warehouses will be created, and maintained.

Empower clinical informaticians to act as "data mediators" or "translators" to "modify" clinical data in the warehouse/repository for use by different end-users for varied purposes.

5. **The IT Needs of the Clinical Research Communities:**

General
Identify features that can ease the regulatory burdens associated with clinical care and clinical research.

Determine methodologies, consistent with federal regulations and IRB requirements, to identify potential subjects for specific intensive investigations.

Translational Research
Support development of highly specialized data collection in small numbers of individual subjects in environmentally controlled clinical research facilities, using real time acquisition and analyses to allow appropriate protocol adjustments.

[2] Since the conference there has been global endorsement of RX/NORM. RX/NORM is a joint project of the VA (NDF/RT), the FDA, the NLM/UMLS (dailymed), and HL7.

Genetic Research
Address the increasing use of genomic information in clinical medicine through the provision of clinical information systems capable of managing such information efficiently and confidentially.

- Incorporate controlled vocabulary into clinical database development to describe phenotypes of large groups or individuals, derived as much as possible from medical data collected in the course of routine care.
- Make available tools for physicians to record and analyze family history.
- Provide test results with decision support tools enabling physicians to make informed treatment decisions for their patients based on individual genetic profiles.
- Provide secure access to DNA and, ideally, fresh tissue samples from members of large patient cohorts that can be used in genetic linkage and association studies.
- Develop mechanisms to support large-scale clinical outcomes studies integrating clinical and genetic/genomic data and bioinformatic tools to sort the volumes of genetic data to help identify genetic markers for health and disease.

Public Health and Health Services Research
Create federated databases for outcomes research and public health functions by building upon the existence of such repositories and something resembling an aggregated, patient-centric electronic health record with an infrastructure in place to create a national database.

When planning information systems for clinical research, design for the interdependency of local and federated databases.

Clinical Trials
Determine methodologies, consistent with federal regulations and IRB requirements, to identify potential subjects for clinical trials.

Clinical trials need a common terminology that results from the integration of multiple terminologies including SNOMED, LOINC, RX/NORM and others[3].

6. **Improve mechanisms and encourage coordination to optimize available funding and address future funding.**

The vision, requirements and functionalities of the proposed patient care and clinical research IT products need to be clearly established and the role of, and impact on, each stakeholder needs to be clearly stated. Motivation and value must be understood and accepted by each stakeholder before this approach will be fully embraced by the key players.

[3] Since the conference there has been global endorsement of RX/NORM. RX/NORM is a joint project of the VA (NDF/RT), the FDA, the NLM/UMLS (dailymed), and HL7.

Leverage the influence of large federal purchasers of health care (Centers for Medicare and Medicaid Services, major corporations, and health plans and HMOs) to provide incentives to health care providers that implement major IT enhancements for clinical care and clinical research.

Facilitate collaboration and sharing of research enhancements, developed separately by the NIH Clinical Center and by the Department of Veterans Affairs, with the extramural research community through the GCRC program. Use GCRC funds available for IT enhancements as seed funding to enable sites to adapt these clinical research enhancements to make better use of limited funds.

Stimulate vendors to participate in the development of clinical research systems by identifying governmental funding mechanisms as incentives for innovative collaborative models involving vendors, government and the research community.

Define the "big ticket" projects that are critical leverage points and are too expensive to be funded by any one source or existing mechanisms.

Initiate new federal investments in IT infrastructure to support health care systems (including clinical research) to prepare for 21st century medicine just as the Hill-Burton Act enabled communities across the country to modernize their non-profit hospitals to prepare for developments in the decades following the discovery of antibiotics.

Information Requirements for Clinical Research

Clinical research, regardless of discipline or area, requires the analysis of vast volumes of complex data. Obtaining data is an arduous and time-consuming task, as researchers struggle to extract information and integrate data from multiple sources. Once the data are collected, researchers rely upon information technology (IT) tools to store and retrieve the data, and to correlate, manipulate, and assess the relevance and relationship among individual data elements. Disciplines and areas within clinical research have made tremendous strides in developing and using these tools, yet they often struggle to identify potential research subjects rapidly and to accurately use the information available in clinical systems. Data used routinely in the clinical care of patients are not easily available for research purposes

The requirements for clinical research are rigorous. Information must be explicitly and consistently defined, consistently used, consistently available over time, and consistently available across various settings. As the criteria for quality in health care are being defined, it is evident that clinical care data and clinical research data will increasingly share these criteria, which cannot be met without standards for terminology and data definitions. Moreover, if they are not met, integration across our systems cannot occur and (sharing and aggregation) reusability of data cannot be achieved. Both are key to aggregating data gathered by different parties for different purposes and creating databases that support clinical research functions.

Technical complexity, the new Health Information Privacy Rule, and cost militate against clinical systems generating research repositories. Yet clinical systems hold patient care data essential to all areas of medicine, including clinical research. The information they hold can populate research repositories and identify research subjects for patient oriented research (POR). Clinical systems focus on real time treatment and patient safety, while research repositories stress utility for retrospective analysis. Despite these differences, subsets of clinical data can be relevant to research.

Discipline-Specific Requirements

The rapid rate of scientific discovery has given rise to new knowledge within the different clinical research disciplines. With this new knowledge have come new forms of data, including soft data (for example, some rating scales), quantitative data (laboratory tests), and specialized data (comparative functional imaging). These new forms of data demand new storage and retrieval capabilities as well as new linkages to other types of data. In and of themselves, they constitute new challenges for clinical research and for integration strategies within and across research disciplines. In addition to these capacities, IT offers potentially cost effective ways to work with the regulatory requirements governing clinical research and the protection of human participants involved in these studies. (See Chapter on Advanced Clinical Information Systems.) Beyond these generic needs shared to varying degrees across all branches of clinical research, information requirements vary by discipline. In the following sections, the IT needs of population-based researchers (public health and health services research, translational researchers, genetics researchers and clinical trialists) are described in greater detail.

Population-Based Research

Public Health Research: Public health research encompasses a number of activities that address the collection, aggregation, analysis, interpretation, and dissemination of data used to improve the health of populations. These activities include the following:

- describing the clinical and epidemiologic characteristics of new syndromes and diseases;
- defining risk factors for disease—both extrinsic risk factors, such as infectious agents and toxins, and intrinsic risk factors, such as genetic diseases and behaviors which lead to or increase the risk of disease; and
- evaluating the effectiveness of treatments, interventions, and prevention measures, such as vaccines, treatments that modify risk factors (e.g., cholesterol-lowering drugs), behavioral interventions (e.g., changes in diet) or the use of screening tests, in the primary, secondary, or tertiary prevention of disease, disability, and death in defined populations.

Health Services Research: Outcomes research studies the relationship between healthcare outcomes and healthcare processes over various time frames—from short term to long term—and measures the results of healthcare in terms of an individual patient's overall state of health or the health status of a demographically defined population. Measures may include morbidity or disability, and mortality. The expenditure of resources involved in healthcare is often explicitly measured as an outcome. Healthcare processes may be defined as the organization, structure, practice and implementation of healthcare services as well as the effect of specific treatments or procedures, including drugs, various types of therapies and the use of diagnostic tests. Health services research often overlaps with both public health research and other research on health outcomes. Virtually all the data for studies in these areas are gathered from clinical sources.

Both public health research and outcomes research are most productive when studies include large numbers of subjects. Population-based data are important when effects are relatively small or are observed over long time periods or when a large number of confounding or modifying factors are present. However, for newly identified or poorly understood diseases, both clinical and epidemiologic descriptions as well as hypothesis generation regarding risk factors and causation may require collecting quite detailed information longitudinally for a relatively small number of patients.

The need for large numbers of observations over time on a small or large sample, or on an entire population of subjects can be greatly facilitated by the use of electronic data in computerized systems. The advent of the electronic medical record (EMR) holds great promise for simplifying and speeding access to critical information in clinical settings. Access to clinical information in paper records is limiting, very time consuming, and plagued by lost or missing charts. By permitting large amounts of data to be collected, stored, and managed, computerization has allowed the use of more sophisticated analytical techniques into the complex rela-

tionships between healthcare practices and healthcare outcomes. Thus, the EMR holds the promise of allowing studies with greater epidemiologic validity and statistical power to be conducted at significantly lower cost. In addition, the EMR, especially in conjunction with computerized clinical decision support systems and the facilitated exchange of clinical data, has been shown to have a demonstrable effect on improving the quality of care. However, access to EMR data is limited by legal and regulatory concerns as well as technical problems. Thus, both researchers and clinicians often find that obtaining and using electronic data can be as frustrating and time-consuming as working from paper records.

From a research (and clinical) perspective, there are four principal types of technical limitations or barriers to the use of electronic data:

- Lack of standardization in coding, vocabulary and nomenclature (including a common reference information model, common data types and common units)
- Lack of standardized messaging for transmitting data
- Lack of infrastructure capacity for inputting, storing, and retrieving data
- Lack of a "gold standard" for the validity of data contained in electronic systems
- Lack of use of algorithrims to enhance the validity of data

Clinicians in different settings use a variety of vocabularies to identify clinical events and findings, which may be imprecisely described. Even objective and reproducible findings, such as laboratory results, need to be coded for inputting into electronic databases. Numerous coding schemes exist (e.g., SNOMED, LOINC, ICD, CPT, MedDRA, etc.). These coding systems are highly variable and individualized and some, which are proprietary, require the user to pay significant fees for their use. Considerable effort is being expended to standardize these vocabularies but success to date has been limited. A more detailed discussion of these issues appears in the chapter on Clinical Data Standards.

Sending messages from one computer or one system to another is a second area of difficulty. This lack of interoperability is a major hindrance to collecting and aggregating data across healthcare sites and venues, even when facilities have electronic databases. Coding and messaging problems can limit the utility and accessibility of data even within a single institution. In some hospitals the pharmacy computer system can not communicate with the laboratory information system nor can it routinely exchange information with other hospital systems (except on occasion for billing purposes). Progress is being made in this area, such as through the development and use of messaging standards such as Health Level Seven (HL7). However, barriers to data sharing remain a significant weakness in information technology applications for public health and outcomes research.

Many facilities that have installed or purchased computerized systems lack the infrastructure to optimally manage these systems. In some hospitals in-house staff may have limited familiarity with the purchased system and the system may lack flexibility. Some systems serve a single intended purpose but cannot be easily modified should the needs of the facility change. Such systems may provide pre-programmed reports, but obtaining data that is aggregated or cross-tabulated in a different format may be difficult or impossible. In some facilities even finding

the appropriate documentation to fully understand the system may be more difficult than would be expected. The need for maintenance and the continued cost for maintenance is not appreciated or planned for by the leadership of many healthcare facilities.

Data may be entered into electronic databases through a variety of means. For example, input may be automatic from laboratory equipment, manually entered by the provider (e.g., encounter data, computerized physician order entry of medications), downloaded from another database, or it may find its way into the database from some other source. Human entry may be direct in real-time, or indirect by the source writing on paper for later entry into the system by a third party. In the first instance, real-time feedback can considerably influence the accuracy, consistency and completeness of the data. In the second case, interpretation of what is meant, including errors in reading handwriting, can lead to significant errors. Once entered into the system, the data may be manipulated in ways that confound its value for research purposes, e.g., the regrouping of ICD codes to maximize prospective payment. It must be noted that coding of data for various proposes leads to loss of informational content. ICD coding is grossly inadequate for clinical research. Clinical research data requires a finely granular coding that provides the greatest level of detail. What is entered must be retained in the coding. No real "gold standard" or criteria exist for assessing how data has been collected and entered or downloaded into a computerized database. Thus, the validity of data from some sources may be suspect if the provenance of that data is insufficiently known and described.

The information technology needs of public health and outcomes research are as follows:

- Standardization of coding and vocabularies. The National Library of Medicine has begun to address this need by establishing the Unified Medical Language System, and the Institute of Medicine has convened a Patient Safety Data Standards Committee. Such efforts should be supported, expanded, and coordinated. The creation of a single, integrated terminology set using the large set of existing terminologies such as SNOMED, LOINC, MEDCIN, RX/NORM and others must be accomplished immediately before any significant sharing and aggregation of data can occur. This action must be accompanied by plan for long term maintenance and distribution with adequate funding.
- Development and widespread implementation of messaging standards. These efforts are underway, but considerable effort from data users, including researchers, healthcare providers, and public health officials are needed to promote the widespread adoption and implementation of these standards.
- Improved information technology infrastructure. Considerable capital investment may be required to put in place sufficient capacity to handle the data needs of research, clinical care and public health. However, from a societal perspective, the anticipated improvement in healthcare quality and patient outcomes would, over time, more than recoup the cost of that initial investment.
- Ensuring data validity. The necessary requirements to maximize the utility of data obtained from electronic databases are standardized data collection methods and the development and implementation of criteria and methods for evaluating and monitoring data quality.

Translational Research

Perhaps the oldest form of clinical research, translational research was known as "integrative physiology" in the middle to latter third of the 20th century, and encompassed studies in intact humans and animals. With the substantial increase in information coming from cellular and molecular biology laboratories, there is a critical need to assess how much of these data are applicable to humans and human disease. This has led to the re-creation of the human physiologist as a "translational" or "patient oriented researcher." As the bridge between bench research and new treatments, diagnostic techniques, and preventive measures, *translational research* focuses on small numbers of subjects, yet produces huge volumes of data. Translational investigators usually study these individuals intensely to identify pathways and mechanisms involved in human physiology and pathophysiology. Most translational research relies heavily on human genetics and is the branch of research that identifies human phenotypes of genetic processes. Because of the substantial influence of environmental factors on physiologic and pathophysiologic processes, the translational researcher relies heavily on a facility that can identify and control as many of those factors as possible so that the underlying physiologic and disease mechanisms can be accurately identified. Translational investigations are usually quite expensive on a per subject basis. Often they are conducted in a facility that is equivalent to a hospital's intensive care unit in terms of personnel and specialized equipment. At the same time they require a level of bioinformatic support that rivals the most complicated outcomes study.

While the approaches used by the translational investigator vary from one discipline to another, in general they fall into three broad areas. The most common approach is to identify an issue raised because of conclusions generated by animal model, molecular or cellular research. For example, in the mid 20th century, various medical treatments and governmental regulations were created based on three assumptions: humans, in contrast to animals, have a limited capacity to reset their circadian clock, that light does not influence the resetting process in humans as it does in animals; and that the human "clock" is a 24-hour one. Research in a highly controlled environment where tens of thousands of data points were captured on less than 50 individuals, documented that all three assumptions were incorrect. These studies have lead to substantial changes in the regulation of shift workers, health care professionals and in the treatment of psychiatric conditions and sleep disorders.

The second approach has as its starting point questions raised, but unanswered, by the results of epidemiologic or outcomes research or clinical trials. These patient-oriented-research studies primarily focus on why a specific outcome was observed. Just as mechanistic studies cannot inform clinical therapeutics with confidence, population based studies can only generate mechanistic hypotheses, because of the substantial number of uncontrolled variables present in these types of studies. It is in the intensive investigation of a few subjects under highly controlled conditions, as performed by the translational investigator, that such hypotheses can be proven or discarded.

The third approach, that can have its beginnings from data derived either in the basic research laboratory or from a population or clinical trial study, has resulted from the advances in genetics

and genomics. While the study of simple mendelian diseases often have been done in the absence of a specifically designed approach, the study of complex diseases has and will require a more structured design. In most cases these common complex diseases with genetic components, e.g., hypertension, asthma, depression, osteoporosis, diabetes, are actually syndromes with a variety of mechanisms leading to the final entity that is defined as the "disease." Thus, in identifying the underlying mechanisms and their individual genetic components, the syndromes need to be divided into more homogeneous subgroups—intermediate phenotypes. Since many of these syndromes are produced by a gene/environment interaction, to identify the gene requires a tight control of environmental factors. Thus, to accomplish the goals of this approach, the design the translational investigator uses requires obtaining a substantial volume of data from a limited number of subjects in a highly controlled environment using highly specific and precise perturbations.

The IT requirements for the translational investigator include the following:

- Highly specialized data collection and environmentally controlled facilities in medical schools and teaching hospitals to intensively study individual subjects. Ideally these should be structured so that real time acquisition and analyses are performed to allow appropriate protocol adjustments.
- Common nomenclature for clinical parameters.
- Specific bioinformatic support to allow genotype/phenotype correlations independent of any clinical care support.
- Databases and programs to identify potential subjects for specific intensive investigations and IT approaches to deal with the increasingly numerous regulatory requirements to conduct clinical research.

Genetics and Genomics Research

Though a relative newcomer to the clinical research enterprise, genetics and genomics research is one of the major engines for progress in understanding the basis for both health and disease. Genetic and genomic research requires access to refined information systems; indeed, genomics can be viewed as *the information science* of the organism. There is a long history of genetic research in medicine; but until recently, most of this focused on management of rare disorders due to mutations in individual genes. The sequencing of the human genome and characterization of human genetic variation has expanded the scope of genetics and genomics research to more common disorders such as hypertension and cancer. In the short term, new diagnostic tools are emerging; in the long run, insights into pathogenesis are likely to suggest new avenues for treatment.

The relationship of genetics and genomics to clinical research can be conceptualized in three phases, each of which places major demands on IT support. First is the identification of genes that contribute to both health and disease. This work is critically dependent on access to clinical information about large populations, identifying cohorts who can participate in studies that relate specific patterns of genetic variation to clinical status. Such studies require both accurate

clinical data ("phenotype") and samples of DNA ("genotype"). In principle, the medical data being collected as a component of routine care should provide a rich source of phenotypic information. If this could be coupled with access to DNA from blood or tissue samples, while maintaining the privacy of information, a resource of incalculable value would be made available. Common disorders – or traits that predispose to good health – are likely to be the outcome of diverse factors, both genetic and non-genetic. These factors will not only differ in different disorders, but even in different populations with the same disorder. Unraveling the genetics of health and of complex disorders will require study of multiple, clinically well characterized populations.

The second phase occurs mostly in the laboratory, and focuses on understanding the cellular and developmental pathways that underlie health and disease. Although pathways of intermediary metabolism have been known for a long time, developmental and physiological pathways are just coming to light. Although genetic studies do not in themselves reveal these pathways, they can indicate the major components of the pathway, opening the way to further study of mechanisms. The 30,000 or so genes encode several times that number of proteins, which in turn have complex interactions in space and time. Revealing these interactions for all of development will take a generation or more, but modest success in specific disorders, for example cancer biology, have already had a great impact on diagnosis and treatment of disease. An example is the development of Gleevec used in treatment of chronic myelogenous leukemia (CML). The drug was developed on the basis of knowledge of the genetics of CML and represents the first effective therapy for this disease.

The third phase brings advances back into the clinic with studies of the outcomes of new testing and treatment approaches. Studies of clinical efficacy will require following large cohorts and documenting clinical outcomes. Physicians, moreover, will require assistance with diagnostic and therapeutic guidelines, taking into account results of genetic and other tests before implementation of treatment. The volume of information about genetics and genomics is rapidly outpacing human ability to remain current. Physicians will need computer-based tools for decision analysis to integrate the growing body of individual and family-based medical and genetic data. They will also need support in the ethical use of this information to improve the quality of care rather than subject their patients to deleterious outcomes such as loss of employment or insurance.

IT needs to support the integration of genetics and genomics into clinical research can be summarized as follows:

- Clinical databases using controlled vocabulary to describe the phenotype of large groups on individuals, derived as much as possible from medical data collected in the course of routine care.
- Access to DNA, and, ideally, fresh tissue samples from members of large patient cohorts that can be used in genetic linkage and association studies.
- Bioinformatic tools to sort the immense volume of genetic data and help identify genetic markers for health and disease.

- Mechanisms to support large scale clinical outcomes studies integrating clinical and genetic/genomic data.
- Tools for physicians to use to record and analyze family history and the results of genetic tests for their patients and to make treatment decisions based on individual patient genetic profiles.

Clinical Trials

The effort required to mount a clinical trial, whether for epidemic disease or low-prevalence disease, is enormous today because we lack point-of-care data collection mechanisms and a common terminology. Except in the special circumstances of the NIH Cooperative Groups, each time a new trial is to be started, a case report form is developed, a data collection system is organized, and everyone must be educated about the choice of definitions by the study leaders. These activities can consume up to 30 percent of the budget of a clinical trial.

With an ongoing data repository that uses a common terminology, the study organization could simply harvest the baseline characteristics and outcomes data at a markedly lower cost, with additional funds going to collect data not routinely collected in clinical practice. Not only would this enhance trial start-up but it also would markedly reduce the need to "clean" the data as the trial progresses. Such an approach could reduce the cost of clinical trials by 50 percent.

For large-scale clinical trials, the needs include reusability of data; this is critical to creating high quality metadata. Point-of-care data collection and standard terminology could resolve these problem areas, thereby expediting the trials and cutting their time significantly. Already on the horizon, new technologies such as the Web and XML will be used to facilitate data transfers.

From the above, it seems clear that clinical trials research would be greatly advanced by CIS incorporating terminology standards that are common to clinical care and to clinical trials. Moreover, a critical piece that is not being systematically addressed in the drug development process—post-marketing surveillance—could be substantially advanced by compatible data standards. At the conference, as described in the chapter on Advanced Clinical Information Systems, there appeared to be a serious disconnect between IT planning by the pharmaceutical industry (which is influenced by the preferences of FDA) and the rest of the health care system. Apparently with the encouragement of the FDA, the pharmaceutical industry had accepted MeDRA as the standard for terminology for drug-related advise events and industry forums like the Clinical Data Interchange Standards Commission (CDISC) had been working to establish standard formats and coding structures for data exchanges apart from the federal effort at setting standards among NIH, CDC, DVA, DoD, CMS and AHRQ. Until recently, the FDA has stood outside of this federal agency effort[1]. In order to assure an optimal environment for future IT applications related to clinical trials, the problem of apparent disconnects in nomenclature standards development need to be addressed.

[1] Since the conference, there has been global endorsement of RX/NORM. RX/NORM is a joint project of the VA (NDF/RT), the FDA, the NLM/UMLS (dailymed), and HL7.

Clinical Systems and Research Repositories

The ability to leverage investments in existing systems is limited by several factors. Many of the IT systems were launched in the absence of comprehensive standards and mutually compatible IT architectures. In addition, critical investments in point-of-care data collection, physician order entry, and a common terminology have yet to be made in many institutions. Clinical systems are still far from being comprehensive, and systems lack standardized interfaces that would allow them to represent the full range of a patient's care and history. Very little aggregation of data for a single patient, even in a single institution, takes place. Finally, some of the commercial system vendors who have developed the legacy systems and/or subsequent enhancements have relied on business models that reduce the flexibility and accessibility of data for re-use. Customization of systems for each institution limits the ability to economically upgrade a system. A related problems is that institutions run different versions of systems from a mixture of vendors.

For the most part, clinical systems currently available do not provide the clinical data researchers require. Clinical systems do not, for example, hold the detailed histories of familial factors and diseases that translational and genetics researchers need to identify appropriate subjects. Data for population-based studies suffer from inconsistent definitions, non-standard terminology, and the inability to link databases from multiple sources within and across institutions.

The creation of research repositories faces significant barriers. On a technical level, data are not yet all electronic, and data formats are variable and difficult to catalog. Social and cultural barriers include issues regarding the ownership and management of data and concerns about the accuracy and reliability of data. These non-technical issues can impede efforts to use existing technology to support cross-institutional patient registries and to pool resources among multiple research groups within and among institutions.

Benefits from Existing Clinical Systems

Academic Medical Centers are beginning to appreciate and address the insatiable demand for information technologies to support patient safety, compliance and financial performance. Investments in these areas have grown and results are being realized. These investments are notable, and may provide much needed infrastructure for growth, collaboration and integration; but, in their present state, they were not designed to address complex issues such as clinical research. The needs of clinical investigators need to be addressed through additional investments, and new designs. Four questions come to mind:

1. Can we successfully leverage what we have without sacrificing what we need?
2. Is collaboration among end-users of clincial information an appropriate solution in a culture that values individual achievement?
3. Is collaborative design even a possibility?
4. Can we improve the efficiency and effectiveness of our clinical research processes through the introduction of information technologies?

The answers to each of these questions must be "yes"; but, we will in fact require new methodology for capturing data, aggregating data, segregating information, empowering patients, protecting privacy and federating databases. As described in the chapter on Advanced Clinical Information Systems, for most organizations, the needs of clinical research are best served by storing data derived from clinical care in a data warehouse distinct from the operating system serving the clinicians. The functions of the data warehouses within the CIS in the Department of Veterans Affairs and at Kaiser Permanente are described in that chapter. It is implied, but not stated that the creation of a data warehouse will vary by institution. Each institution has to determine under whose auspices and cost center a data warehouse will be created, and maintained.

In circumstances where the purpose for data collected in a clinical setting varies substantially from the need for the same data in clinical research, there would seem to be a need for a "data mediator" or a "translator" in the middle of the repository that can "modify" the data usefulness to different end-users. In this regard, institutions should consider adding clinical informaticians who could bridge the worlds of IT personnel and the clinical investigators.

Federated vs. Local Databases

Federated databases allow data to continue to reside in widely distributed networks and locations, while information about where to find the data is easily accessible. Such databases are prerequisites for outcomes research and public health functions. An optimal federated database depends upon the existence of patient-focused repositories and something resembling a "national" patient record. It is clear that for data to be aggregated as in a federated database, it needs to be identified at least when the data is aggregated. Such a requirement demands a unique person identifier, if the data is to be aggregated without identificaton errors.

While federated databases might help to aid recruitment for studies involving genetic or translational research, they are not detailed enough to meet the data requirements for these research studies, which continue to be dependent upon local, researcher controlled, databases. Efforts to bring data from local, researcher-controlled databases into federated databases would likely get some resistance from the researchers at this stage. The reality is that this area is ripe for a convergence of interests. Local databases can be derived from a federated one. Researchers can identify subjects on a federated database and move the appropriate data to a local, more sophisticated database. In like manner, researchers can keep detailed local databases as long as required nomenclature is used and a common dataset is collected in a manner that can be mapped to the national standard. In planning for IT for clinical research, it is critical to plan for interdependency of local and federated databases. This need exists across the full span of clinical research.

The CDC hopes to optimize the interdependency of local databases to feed its federated surveillance networks. The success of this model will depend on the reliability of the network and the level of customer service given to those who contribute to the joint database. Data that have been locally collected must be accessible to that site, and the site's ability to access, manage and analyze the data should be time and resource efficient, at little or no additional cost. The network of General Clinical Research Centers (GCRCs) in academic health centers across the

country could serve as a pilot to test the feasibility and problems associated with the implementation of interdependent local and federated databases in relation to translational, genetic and NIH-funded clinical trials research. The NIH Clinical Center is currently investing in a CIS that can meet the needs of translational, genetic and clinical trials researchers across disease categories and organ systems. In an optimal world, the efforts of the NIH Clinical Center, the GCRC program, and the IT investments by medical schools and teaching hospitals could be brought together for joint planning efforts that would bring an advanced clinically focused CIS to the Clinical Center and modular enhancements for clinical research to the academic sites.

Issues in the Creation of a Comprehensive Clinical Repository

Building a comprehensive clinical repository will require a focused effort, with significant government involvement. Providers, delivery systems, and the medical products industry all have major roles to play, but no single private constituency is powerful enough to make it happen. If federal agencies agreed on common standards and a system of nomenclature, they would make a comprehensive clinical research repository possible. The actual repository could be maintained in a "virtual" manner according to the needs of the organization using it, and most of the work in the repositories could be organized by public-private partnerships. The federal government needs to take the leadership role in those matters that require a national degree of coordination, leadership, and funding. The creation of a National Healthcare Informatics Infrastructure requires government leadership. Identificaton and acceptance of clnical data standards, terminologies, data models, etc. requires a parntership of public and private sectors. Funding of the development of data standards should also be a partnership. Creating the nuts and bolts of the system will be the responsibility of the healthcare delivery system and the medical products and information systems industries.

Databases of aggregated, summary, patient data should occur at a regional level to ensure scalability. Analysis could be accomplished at the institutional, regional, and national levels. Barriers to creation of a comprehensive research repository are social and cultural. No new technologic advances or developments are required to create combined or "federated" databases. However, creation of combined databases, even in a single institution, requires a significant investment of resources from IT professionals, clinicians, epidemiologists, laboratory scientists, administrators, and other managers and content experts. The principal technologic barriers involve software incompatibility, communications problems between medical and IT staffs, difficulties in linking discrete systems, and, not infrequently, fundamental disparities in the system design goals or "systems thinking" inherent to the clinical/research/public health function, the administrative function, and the IT function. The goal must be to reduce the power attributable to "possessing the data" and increase the premium on *using* the data to advance knowledge and to improve health care. Incentives must reward providers and health systems in the initial phases for collecting the data in the specified manner, and reward medical-products companies for creating compatible systems.

The need to foster mutual comprehension between diverse research disciplines and clinical disciplines is immediate and continual. At the present time, healthcare professionals must adhere

to standards of CMS and Medicaid when billing for services. Academic researchers must adhere to standards of NIH, AHRQ, and CDC. Industry sponsored research is tightly regulated by the FDA. The work of the Health Level Seven (HL7) clinical trials technical committee and the Clinical Data Interchange Standards Consortium (CDISC) collaboration with HL7 are both exemplary efforts of bringing different constituencies together. (Chapter 3)

The quality care imperative, emphasized in recent reports from the Institute of Medicine, will occur only when computerized, point-of-care, clinical systems are in place. The reality is that the effort to make these systems work has barely started. Given that there is no tangible return on energy spent on paperwork, however, incentives should succeed in encouraging systems and providers to work toward the goal of computerized systems.

The Potential Impact of the Health Privacy Rule Under HIPAA: A Post Script

The implementation date for the Privacy Rule under Health Insurance Portability and Accountability Act of 1996 (HIPAA) was April 14, 2003. The HIPAA Privacy Rule was originally described as a measure intended to protect the privacy of personal information in the health record. However, it was not supposed to discourage the use of IT applications designed to improve quality and reduce medical errors in health care systems. It may be that these goals (absolute privacy protection and IT quality enhancements) are incompatible or financially non-feasible for many health care systems.

According to HIPAA, clinical research was not deemed a specific covered function, like treatment, billing, or health care operations. Under the rule, protected health information (PHI) may be used *within* an organization that is a "covered entity" (this designation does not apply to pharmaceutical firms which will require Business Associate agreements). Disclosure is required when that information *moves outside* the walls of the organization. Clinical trials create PHI; other research projects use PHI created by others but do not create it de novo. The Privacy Rule requires patient authorization for use of information that is outside of the scope of treatment, payment, and healthcare operations (TPO). Waivers may be granted by Institutional Review Boards (IRBs) or, in some instances, by Privacy Boards whose makeup is specified by the Privacy Rule, under specifically identified criteria. In a change from past practice, protections under PHI continue after an individual is deceased.

There is great concern that the HIPAA Privacy Rule may result in reluctance on the part of institutions and researchers to share information across different networks. Clearly, this will have a chilling effect on all types of clinical research that depend on the flow of relevant patient information. It will be important for the clinical research communities to remain vigilant and attentive to emerging problems stemming from HIPAA, and to bring these to the attention of responsible federal officials. Countervailing forces calling for public health, regulatory and fiscal oversight, and quality improvement may offset HIPAA's potentially negative impact. If the Privacy Rule, as presently written, proves to be problematic in terms of costs, critical clinical operations, or clinical research activities, pressure may build to secure federal funding for enhancements in CIS that can enable these systems to meet their goals and federal requirements.

Advanced Clinical Information Systems

Context: Status of IT at Six Leading Institutions

The Survey

Early in their planning, the steering committee was approached by staff of the Clinical Research Roundtable (CRR), a program sponsored by the Institute of Medicine (IOM), to provide guidance on surveying the state of IT applications for clinical research at academic medical centers (medical schools and teaching hospitals). Steering committee members advised IOM staff to narrow the survey focus, and helped to develop the format and methodology. Committee members also identified 14 institutions considered to be leaders in clinical research. From this list, the CRR selected six: Brigham and Women's Hospital, Mayo Rochester, Regenstrief/Indiana University, University of Pennsylvania, Sloan-Kettering, and Vanderbilt. The survey involved 15 to 30 minute semi-structured interviews with 27 individuals, including eight IT support staff, seven informatics researchers, seven health services/outcomes researchers, four clinical trialists, and one administrator. It was conducted and reported by David Dowdy, ScM, now a MD/PhD candidate at Johns Hopkins. The staff of the CRR agreed to present their findings at the conference. These findings, as well as an overview paper by John Glaser of Partners Health Care, have been incorporated into the following sections of this chapter under Impressions and Top Ten Needs. Some of the identified impressions and needs echo the needs identified by the researchers in Panel 1 (Chapter 1).The remaining sections of this chapter reflect the input by the members of Panel 2 (Clinical Information Systems: State of the Art).

Impressions

Governance. At the institutions surveyed, the IT systems for clinicians and the IT systems for researchers are often separate, as are the infrastructures that support each system. Communications between the research support systems and the clinical operation systems tend to be limited. The General Clinical Research Centers (GCRC) focus on IT that support their activities only, without much communication beyond their borders. Clinical informatics departments try to act as a glue to bring information resources together, but these departments are not usually large enough to provide all the support that is needed.

Data model and management. Each of these flagship institutions reported that it should not be difficult to secure the technology needed to meet the needs of a substantial majority of clinical researchers, but that putting together a suitable data model and management structure remains a major challenge. The key to successful implementation of a clinical research database is institutional buy-in from the highest administrative levels and from the clinical research communities.

IT impact on research. Information technology (IT) is already having an impact on the design of clinical research studies at these six institutions. The electronic medical record (EMR) is proving to be a useful tool for cohort retrieval and sample size. Guidance for regulatory compliance is

now becoming available online. IT is also affecting data collection and protocol management, through Web-based databases and patient recruitment, clinical trial management (scheduling, etc.), and hardware/software support of data collection.

Top Ten Needs

1. Clinical data. Current CIS, particularly the EMR, serve clinicians better than researchers. Of those surveyed, 53% believed electronic data collected at their institutions were highly useful to clinicians, while only 21% considered them of high utility for researchers. Researchers required data collected and recorded according to rigorous methodologies, but most CIS were designed for clinicians and included unstructured free-text data. Improved data quality would facilitate research and support operations (coding, compliance, quality, and safety programs) within teaching hospitals and medical schools. Consulting with researchers during the design of CIS could dramatically improve system utility for clinicians and managers, as well as for clinical investigators.

2. Patient privacy. No issue was mentioned more than privacy, due in part to apprehension regarding HIPAA regulations. Of the respondents, 57% considered HIPAA privacy requirements a very serious problem, and 43% rated it "something of a problem." In their view, patient information can never be fully de-identified without compromising the quality or utility of data. Because regulations and patient concerns are certain to evolve over time, systems development must stress flexibility and the capability of systems to change. "Role-based" security could provide different degrees of de-identification, depending on role, such as patient's provider or researcher on IRB-approved study.

3. Standardization. Clinicians find viewing scanned information useful, but clinical researchers require formatted data. Standards used should be broad enough to capture the full range of medical terms and flexible enough to account for different terms denoting the same concept. Existing standards include HL7 for the communications of clinical information, SNOMED for pathological naming, DICOM for transmission of digital images, and LOINC for laboratory test names. These models can be improved upon, IT specialists agree, but the most pressing need is to put existing standards into widespread use and to develop standardized interfaces.

4. Organizational support. The need for investigators with multidisciplinary backgrounds and intimate knowledge of an institution's informatics infrastructure is growing. Increasingly, clinical researchers interested in using an institution's EMR or other information sources need to collaborate with investigators who know how to obtain the appropriate information from a vast collection of data. Institutions should initiate or enhance departments of informatics and inter-departmental research centers. Fostering collaboration among individuals with expertise in various fields will contribute to the development of IT solutions that serve both medical practice and clinical research.

5. Genomic revolution. As population-level studies are undertaken utilizing genomic information on each individual participant, an unprecedented amount of information will be generated. IT support staff need guidance from clinical researchers as to what their data-handling needs are likely to be as genetic information is incorporated into their research. Inter-institutional, and perhaps national, repositories of genetic information are likely to be extremely useful to clinical researchers seeking to carry out studies with sufficient power to detect gene-environment or gene-gene interactions.

6. Web-based applications. The Web enables investigators to carry out studies without being tied to a single physical machine or data-entry platform. Web-centered studies are often easier to manage because only a single server needs to be maintained. The majority of institutions surveyed are in the process of integrating the Web into their IT support structures, both by supporting investigators in design of Web-based applications and by maintaining Web-based interfaces to the EMR and other sources of data.

7. Communications. Despite feeling their relationships were generally positive, IT support staff and clinical researchers express some frustration with each other. Clinical researchers are disappointed they can not make greater use of the electronic data available, while IT specialists consider researchers' expectations unrealistic at times. Researchers are not always aware of the IT support available to them.

8. Integration of IT solutions. When looking at IT solutions for health care, uniformity should never be assumed, anywhere. Intra-institutional variability and the lack of collaborative research tools loom large as barriers. Without support by high-level administrators, individual departments and investigators are reluctant to abandon traditional data systems. The lack of systems integration inhibits the ability of researchers to perform studies on broader patient populations. It also undermines the value of the EMR to clinicians.

9. Broader implementation of EMRs. In the view of those surveyed, institutions that have been developing their EMR systems for the longest periods of time are leaders in the field. Nonetheless, obstacles to broader implementation remain. As long as administrators remain unconvinced that EMRs provide a return on investment, they will hesitate to implement such systems. Despite an abundance of EMR vendors, no single EMR system is accepted as an industry standard. Many are not sufficiently flexible and robust to meet organizational needs. Moreover, when the lack of EMR is acceptable to clinicians, related clinical research efforts also stall.

10. Funding. Although only 10% of the interviewees consider funding a "very serious problem" and 38% indicate it is "something of a problem," the remaining 52% classify it as "not a problem" at their institutions. Nonetheless, there is agreement that funding mechanisms need greater flexibility for IT support. Grant funding is inadequate, with insufficient support for informatics staff under the GCRC program, or for enhancement of existing IT support under investigator initiated NIH grants. Furthermore, investigators often under-budget their IT support needs.

While funding was not a problem for a slight majority of respondents to the survey, the huge costs of these systems remain a very real barrier at most teaching hospitals and medical schools. It will be very difficult for institutions that do not have successful EMR systems in place to duplicate the level of commitment, both in terms of time and resources, that institutions with successful EMR systems have made over time. This will be a major barrier to their involvement in the next generation of clinical research, as well as for the practice of 21st century health care.

Functionalities and Characteristics of state of the art clinical information systems

To support clinicians as they give care, CIS must be intuitive and easy to use, with tools that support medical decision making, including alerts, feedback, and views of information across patients. CIS must also facilitate billing and insurance transactions. In the 24/7 world of clinical care, CIS must make patient data quickly, and reliably accessible in health care venues, while assuring privacy, confidentiality, and security of patient information.

To support researchers, CIS should permit computer-based comparisons among patients. The system, necessarily, must represent patient data in the form of coded information that is descriptive, non-ambiguous, clear, and reproducible across users and systems. CIS tools can help meet these data criteria for re-usability by assisting users in finding appropriate codes, embedding or providing easy access to definitions, and using standard vocabularies wherever possible. These functionalities can improve accuracy, quality, and inter-rater reliability.

To provide the functionalities required by clinicians and clinical researchers, optimal advanced CIS should be the following:

- ***Complete.*** The entire patient record must be electronic, including multimedia information such as pathology slides, cardiology examination results, laboratory, X-Ray, and endoscopies.

- ***Comprehensive.*** The patient record must contain the full range of data, both inpatient and outpatient.

- ***Standard-based.*** All data must meet commonly accepted standards. Consistently defined data elements and data sets are essential to producing clinical data for research analysis and for optimal care.

- ***Automatically transferred.*** Clinical data must move from the CIS into the research record easily and quickly. The data must be accessible to researchers, without disruption to the CIS.

- ***Secure and confidential.*** Patient data must be protected in the clinical setting. When used for research, data, including blood and tissue samples, must be de-identified and/or anonymized to comply with HIPAA. If identifiable data are required, the investigator must justify this to the IRB, and secure patient consent.

At present, CIS fulfill these criteria to varying degrees. Relatively advanced CISs are rare, however, and usually available only in organizations that can command the skills and funds needed to develop a cutting-edge system. By virtue of their governance, mission, and funding they want more from their CIS and they can afford to make the critical investments. There is no one CIS solution for all health care organizations. The 'right' system for an enterprise should be linked to its mission. Described below are some sophisticated CISs as developed to serve the needs of a federal department, a health maintenance organization, a private university, a public university, and a pharmaceutical corporation.

A Federal System

The Veterans Health Administration

The Veterans Health Administration (VHA) delivers care to more than 4 million veterans a year at over 1,300 points of care, employing about 15,000 physicians, 50,000 nurses, and 33,000 allied health professionals. It is the nation's largest laboratory for multi-site studies, with a research budget of $1.301 billion from multiple sources, and the world's largest provider of training for healthcare professionals.

Clinical Information System. Today the VHA's clinical information system supports day-to-day operations at all local VA healthcare facilities, including medical centers, outpatient clinics, nursing homes, and others. Known as VistA (Veterans Information Systems and Technology Architecture), it brings together over 100 applications, such as pharmacy, nursing, and laboratory, and provides clinicians with a single interface for reviewing and updating a patient's medical record.

At most VA medical centers, clinicians have access to a multimedia patient record that includes medical images of all kinds, from pathology slides and scanned documents, to cardiology exam results and wound photos. Clinicians can access clinical information on patients treated elsewhere in the VHA or by the Department of Defense (DoD) prior to separation from active duty.

Separate Research Database. Clinical data from local systems are systematically transferred into national aggregate databases and registries. The subset of clinical data made available for research purposes is stored in a national data warehouse separate from the VHA's clinical information systems. Research makes extensive use of both local systems and the national databases as in, for example, the Quality Enhancement Research Initiative. Maintaining a separate database for research purposes optimizes the data for aggregate analysis, without putting clinical response time at risk.

The original system architecture, now over 20 years old, makes adding new data elements slow and arduous. The lack of a data model and data standardization results in variation across sites. Complex storage structures tend to impede the extraction and transfer of data, with adverse impacts on the performance of production systems used for patient care. Ease of access varies from site to site.

Planning the System. Traditionally, the VHA has involved clinical researchers, many of whom are also clinicians, in managing and planning for clinical information systems. Over the past few years, the VHA extended its focus to include the needs of health services researchers, providing them with new capabilities when possible, and involving them early in the course of planning for the next generation health system, Health*e*Vet-VistA.

New Health Data Systems. Clinical data will reside in a new health data repository (HDR), fed by multiple provider systems. To create this repository, the VHA is adopting data standards and standardizing information at all its VHA's local sites and what are now 128 local VistA systems. The system configuration includes one primary national HDR, with a minimum of one offsite "hot backup" and one backup for analytic use. As the core of the new health data systems, the HDR will thus support both clinical and research functions, by easing access and shared use and reducing the burden on local sites. The current timetable calls for population of the data repository in 2004 and full operation by October 2005.

Federal Effort to Link Data Standards. In their effort to converge their next generation systems, the VA and Department of Defense (DoD) are collaborating with all health-related federal agencies to identify, select, adopt, and implement common health data and communications standards. Led by the Center for Medicaid and Medicare Services, this effort establishing Federal health information interoperability standards could potentially lead to their voluntary adoption by the private sector.

Additional Capabilities. VA's My Health*e*Vet project provides patients with an internet-based, secure personal health record; plans call for developing this record further, to include images and even genomic information. The ability to share genomic information with one's relatives and to link the electronic records of family members will make possible new methods and strategies for genetic research. In the future, other capabilities may be added, for example, voluntarily revealing characteristics of the individual in a de-identified manner so that requested services can be provided to the individual. One such application could be a service that matches the individual with clinical trials.

Health Maintenance Organization

Kaiser Permanente Northwest

A nonprofit, group-practice health maintenance organization, Kaiser Permanente Northwest (KPNW) serves over 450,000 members in northwest Oregon and southwest Washington. It includes 18 medical offices, one general hospital, and formal affiliations with six community and academic hospitals and medical centers. Members have unique permanent health record numbers and receive most of their care within the system.

KPNW has a Center for Health Research (CHR). Its agenda is shaped by being part of an integrated comprehensive capitated system with a single system of ancillary services, single ownership of most systems, and aligned incentives between the medical group and health plan

to improve care. The Center maintains its own research data systems and draws data from KPNW's online clinical, ancillary, and administrative systems.

Clinical Information System. Today KPNW has a complete and comprehensive electronic medical record (EMR). All elements are electronic, with any residual paper documents scanned in. The record includes complete inpatient and outpatient data, transcribed reports, registry information, cost information, telephone advice, and radiology, pharmacy, and laboratory data, etc. All clinicians providing care must be able to access a minimum dataset from the record with all relevant data.

Research Support. The clinical system requires speed and support for day-to-day transactional activities, clinical workflow processes, and medical decision -making. Because these capabilities cannot be in any way compromised, a separate data warehouse was created to support research functions. The data warehouse is optimized for query capability across patients, with multidimensional online analytical processing (OLAP) databases. Designed for easy end-user access and use, it supports data manipulation and data marts. Researchers can aggregate information from multiple sources, reconcile clinical data with financial data, de-identify and/or encrypt patient identity, and check for data consistency and scrub if necessary. In addition, they can extract and store subsets of data, or have outreach features conduct telephone surveys.

The data warehouse supports researchers in multiple research functions, such as early hypothesis testing, prevalence and incidence determination, study planning, cohort identification, building complex models, and data mining. At KPNW, it enables researchers to study ways to provide and improve patient care, using a population-based approach. It provides the ability to track quality and utilization and identify variations in costs, quality, and outcomes. It is even used to help identify situations in need of early intervention or more intense monitoring.

A data warehouse, which is geared to meet the needs of clinical researchers, depends upon the use of the EMR to capture coded information, i.e., more detailed and more consistent data. At KPNW, researchers are studying how the EMR and features such as alerts, etc., affect clinician behavior and medical decision-making. In addition, researchers are exploring "frontier issues," such as patient access to their EMR, direct patient entry of information into the EMR, and patient-physician electronic messaging.

Input from Researchers in Planning for CIS. Involving clinical researchers in discussions when designing, managing, and improving CIS helps set expectations. For example, it helps identify standard datasets that are realistically capable of being reliably captured at the point of care. It facilitates the use of standard terminology when possible and built-in data capture that meets research needs. In addition, consultation with researchers is critical to structuring capabilities needed to export data into research databases, de-identify the data, and randomize interventions.

Automatic Transfer of CIS Data. As care delivery advances, there are often multiple locations where care can be delivered, such as inpatient, primary care, specialty care, home health, etc. The capability to aggregate information from different sources is key to getting a "complete picture" of the patient. Kaiser Permanente Northwest, by virtue of its integrated comprehensive healthcare system has significant advantages over most other healthcare organizations in data aggregation. Most other healthcare organizations have issues around unique patient identifiers and common vocabulary that need to be resolved. They also have greater issues with the differing quality of the information from different systems . Finally, there are political and privacy issues that need to be overcome when data are located in different organizations.

Challenges. In the implementation of their EMR, data warehouse and other system changes, the IT leadership at KPNW has identified two main types of challenges. The first involves organizational issues and human factors, including communication and coordination, resources for implementation, agility, decision making processes, and the need to create and track the effect of interventions. The second category involves data standards, quality, and definitions, and the tasks of aggregating data, agreeing on what to collect (extent of collection as well as which data to collect), establishing priorities for data collection, and providing incentives to collect research data. The lack of clear "line of sight" benefits for the work involved intensifies the difficulties.

Planning the System. Researchers at KPNW have been involved in quality improvement committees. They have helped to set the direction for information technology development in the support of quality health care initiatives. Because the needs of the researchers are somewhat different from those in clinical operations; there is occasional tension between the groups. Whereas research has the goal of advancing knowledge through fundable research projects, clinical operations focus on improving quality and service in the day-to-day care of patients. Clinical operations strive for results that are "actionable" and can be implemented quickly. It is closer to the core mission of KPNW.

Despite the occasional tensions, the two areas are aligned in building systems to support epidemiological population-based research and in applied research areas like medical informatics. Many of these shared interests have resulted in significant improvements in clinical practice.

A Private Research University System

Vanderbilt University Medical Center

Governance, infrastructure, funding and intensive communication are critical to achieving a strategic informatics vision. The presentation on CIS developments at Vanderbilt focused on these organizational issues. Without top-level organizational support and vision for IT investments for clinical research, and a high degree of interaction between the informatics team and the end users, the needs of clinical researchers (or of clinicians) cannot be met.

Leadership and Governance. A large academic medical center, Vanderbilt University Medical Center has integrated all institutional informatics units administratively, into the Vanderbilt

Informatics Center. All are led by a chief information officer (CIO) who is one of five associate vice chancellors (AVCs) reporting to the health center vice chancellor who, in turn, reports to the chancellor of the university. The AVC/CIO also reports to the medical school, thereby functioning within the inner circle of medical school governance.

Faculty Funding Model. The Vanderbilt model for informatics faculty is analogous to the traditional model for clinical faculty, who are expected to "earn their keep" through revenue from patient care activities, sponsored research, or a combination of both. At Vanderbilt, informatics faculty use institutionally budgeted monies for salary support (equivalent to "clinical service") to develop tools that help the institution meet its strategic informatics missions. The result is a hospital-and-clinic based laboratory in which real-world innovations are developed, implemented, and evaluated, facilitating acquisition of grant support for additional work.

Organizational Structure. The Vanderbilt Informatics Center currently has more than 300 full-time employees and an annual budget between $25 and $30 million. It consists of the biomedical informatics department in the medical school, the medical library, and the hospital-and-clinic-based applied information technology group. Literally dozens of programmers within the Informatics Center support and develop the clinical software environment at Vanderbilt University Hospital and outpatient clinics as a part of their ongoing daily work, often in conjunction with biomedical informatics faculty.

The core offices of the Informatics Center and biomedical informatics department are located on the fourth floor of the library, and members of all of the units of the Informatics Center have been linked into a cohesive team with diverse talents. The applied information technology group is responsible for software and hardware across Vanderbilt's clinical facilities, and includes network and management groups. It has significantly evolved from what once was a strictly hospital-based information systems group.

All of the units of the Informatics Center collaborate closely, under the direction of the AVC/CIO, so that academic software projects, once mature, can be delivered directly to hospital wards or clinics in conjunction with programmers and managers from the applied technology group—the individuals who will be responsible for the long-term maintenance of such projects. Conversely, members of the hospital-and-clinic based group regularly involve members of the academic unit in their planning, development, and troubleshooting activities.

Technological Infrastructure. This coordinated and centralized approach has led to a number of resources useful for clinical research, including the following:

- An overall systems architecture coordinates system interactions rather than creating a
- tower of Babel" with disparate, incompatible systems.
- A generic interface engine, developed at Vanderbilt, coordinates all messages among systems (HL-7, proprietary vendor-related, and home grown varieties) and has data definitions for all fields of all messages. This allows "repackaging" of message contents between systems that require different fields and formats.

- Data standards are in place in all systems.
- A clinical data repository "siphons" information from primary systems (laboratory, radiology, medical transcription, etc.) and stores copies of data from these systems in a standard format with consistent header information. All of a patient's electronic medical record (EMR) can be logically arranged and presented, despite having multiple points of origin.
- Institution-specific business logic is kept outside of proprietary systems, so that upgrades and switching vendors does not cause loss, or the need for manually intensive translation efforts.

Clinical Research Needs. Data in the EMR and the care provider order entry (CPOE) can be directly used for research as they are, or combined to address new information needs. The CPOE system sends a copy of each order for each patient to the EMR system, where a series of configurable Perl programs parse the orders into their component fields. With approval from the Institutional Review Board (IRB), researchers who want to identify patients eligible for clinical studies give criteria to Informatics Center programmers, who generate a secure Web-page, available only to specified researchers, listing patients meeting those criteria.

For pharmacogenomics and clinical trials research, the system produces a real-time display, updated hourly, of all patients who have had their first dose of a researcher-specified medication ordered, making study enrollment far more efficient than by manual means. Other applications of the same system can locate all patients with any kind of isolation order, for example, or identify subjects with specified combinations of admission diagnoses, reasons for performing x-rays and electrocardiograms, and medications. Information-rich environments facilitate clinical research.

Lessons Learned. Today, clinical informaticians appreciate the utility of basing systems development on clinical information needs. In like manner, systems developed to support clinical researchers must be predicated upon a clear understanding of their information needs. Software development must involve collaboration from the start.

There is almost always a *cognitive dissonance* between developers and specialized end-users. Developers ask, "tell me how you do your work and how we can help you," and end-users say, "tell me what the system will do for me after it is installed." End-users cannot fully describe their workflows, and developers cannot anticipate all system interactions with those workflows. Developers must be convinced to "live with" and support the new system directly after it goes live. By studying end-users as they use the new system and seeing it through their eyes, developers can determine how to fix and improve the system, and then use rapid programming methods to do so.

The collaboration needs to be long-term, with ongoing feedback. The Vanderbilt Informatics Center hosts weekly free pizza lunches to foster informal dialogue with clinical users. End-users identify problems they have recently encountered, including ones not reported to the help desk or ones the help desk had too little information about to address. The Informatics

Center discusses planned changes to the system and offers "live mock-ups" to learn firsthand from the users where problems are likely to occur and how to address them. Once end-users realize that the system developers actually respond to their complaints and suggestions, they become avid system developers in their own right. At Vanderbilt, end-users have provided some of the best ideas for new system features and functionality implemented to date.

Public Research University

University of California, San Diego

The presentation on IT developments at UC San Diego focused on the ways in which the informatics leadership at that university has been able to develop successful approaches to regulatory compliance issues in clinical research. Clinical research in academic health centers is conducted in a thicket of federal, state, and institutional policies and procedures. These include requirements for review of proposed projects with respect to ethics and funding mechanisms, as well as specialized reviews for environmental health and safety, biosafety, radiation safety, conflict of interest, utilization of medical center resources, and risk management. As seen by regulatory offices within academic health centers, the mandated review processes generate large volumes of paper that are a logistical nightmare.

At the University of California, San Diego each year approximately 1400 new projects are submitted for Institutional Review Board (IRB) review, comprising over 125,000 pages of information that must be distributed to 45 committee members on four different committees. The IRB office receives 12,000 documents per year and generates another 12,000 in response. In support of this blizzard of paper, the IRB office required two full time positions for handling the input processing, copying, collation, and distribution of paper documents, and IRB members used luggage with wheels to transport materials to and from committee meetings.

A Systems Approach. In 2001 a planning effort was undertaken to devise a systems approach to IRB reviews that would improve the efficiency of the process and reduce the burdens placed on IRB members and faculty investigators. The approach focused on improving the infrastructure of the IRB office for creating, archiving and providing access to large numbers of research-related documents. Other goals included enhancing the responsiveness of the IRB staff to inquiries, facilitating the coordination of related administrative processes throughout the organization and strengthening the systemic capacity for oversight of on-going clinical research.

The approach to achieving the functional goals listed above has relied upon several widely used and proven information technologies. As seen by a researcher submitting a new research application, a protocol amendment, or an adverse event notification, the process begins by going to the IRB Web site (http://irb.ucsd.edu) using any standard web browser, and authenticating oneself. Authenticated users select the type of transaction they wish to initiate, and are prompted through a series of data input screens that include data previously entered on paper forms such as the new project face page. Must-complete items and items requiring valid ranges are automatically checked at the time of online submission, and incomplete submissions are not accepted.

At the completion of the online data entry pages, if a document is associated with the requested service, such as a new protocol submission, users are prompted to select the word processor file that contained the project plan, amendment or other document at their own local workstation. Clicking on the submit button on the Web browser uploads this word processor document to the IRB server, where it is automatically converted to an Acrobat PDF file, assigned the relevant document tracking keywords and codes, and stored for inclusion on the agenda of the relevant committee meeting. Since there is no widely implemented standard for digital signatures, documents requiring a signature are printed at the local workstation of the person submitting the document, and this single page summary of the submitted documents is mailed or faxed to the IRB office for archival purposes, where it also is scanned into PDF format. New and revised consent documents are maintained in an editable word processor format, so that minor consent changes can be made by the IRB committee without requiring project staff to submit revised consents.

Support for IRB Meetings. The process of assembling the agenda of an IRB meeting involves the creation of an electronic hypertext index (i.e., an agenda whose items are links to the actual documents for review). This electronic agenda is available to IRB committee members at a secure private committee Web site.

The review of documents by IRB members requires that they have state-of-the-art, lightweight laptop computers, Internet connections, and high-speed printers with associated print supplies. The IRB office purchases the computer equipment and provides it to committee members. The standard package for an IRB member includes a high speed, large screen laptop with wired and wireless network capability, an annual telecommunications subsidy for broadband Internet service, a laser printer and print supplies such as paper and toner cartridges. Committee members who serve 24 months may keep their existing equipment and be eligible for an additional laptop and printer on a bi-annual basis. Regardless of length of service, committee members are free to use their computer and telecommunications equipment in any way they see fit, as long as they use it also for the IRB committee duties.

IRB committee members bring their laptop computers to the IRB conference room for meetings, where there is a high-speed wireless network. The meeting is conducted with all committee members stepping through the agenda, calling up documents and their own submitted review comments via the private Web site. Rather than printing and mailing letters, the IRB office creates digitally watermarked PDF files that are the image of a signed letter on office letterhead. The e-document system that produces these files is compliant with FDA provisions for electronic signatures that are required by 21 CFR 11(c). These PDF files are sent as e-mail attachments to investigators. The original file remains in the electronic archive, hyperlinked to the office project tracking system so that at the display of the "life history" of a project is the complete set of all documents sent and received relevant to the project. Each of the documents descriptions in the tracking system is a hyperlink, which when clicked retrieves the PDF image of the actual full text correspondence.

The "e-IRB" prototype system was deployed in May of 2002 with the creation of a new IRB consisting of volunteers from existing IRB committees and newly recruited committee members. As of this writing, four monthly review cycles have been completed, and the full text document management system contains just over 9,000 documents comprising about 1.5 Gigabytes of information.

Lessons Learned. Not unlike electronic medical records, where the difficulties are 15 percent technology and 85 percent sociology, the rate of implementation has been limited by the willingness of investigators and their staffs to use electronic submission methods, and the capacity of the IRB to adapt to reviews done and submitted online rather than on paper.

The greatest challenges, not anticipated at the time of deployment, have been the interfaces to other regulatory offices of the University, most of which are still firmly entrenched in manual paper-based methods. The IRB office has responded with a set of secure private Web interfaces for these offices that enable them to view and print for their own use copies of documents submitted to or produced by the IRB office. Lists of projects newly submitted are available to them via dynamic database queries from the Web, and each project number in a list is a hyperlink to retrieve the project summary and full text documents associated with that project.

Next Step. To accommodate the heterogeneous technical sophistication of users and other offices, the next step in the evolution of the IRB document management policy will be to supplement the current infrastructure with a high speed scanning workstation capable of accepting very large documents such as sponsor-supplied master protocols, and automatically converting them into PDF page images. This technology will enable the IRB office to work in electronic document mode regardless of the format of submissions received, and will provide a rich, web-accessible archive of all documents associated with all research projects.

Inter-process Coordination. For such technical initiatives to work well, an institution that engages in clinical research needs an institution-wide governance structure and a common information infrastructure that enables each specific regulatory function such as human subjects, animal subjects, radiation safety, biosafety, conflict of interest, and health center resource management offices to share information that investigators supply once rather than submitting in duplicate on a profusion of domain-specific forms. Inter-process coordination is the major challenge faced by the organization, and it requires both a broad view of the overlapping responsibilities of regulatory offices, and skills in knowledge engineering to coordinate their activities via shared information systems.

Pharmaceutical Industry

Merck & Co., Inc., Research Information Services

A research-driven pharmaceutical products and services company, Merck has introduced 17 new medicine and vaccines since 1995, and had $47 billion in sales in 2001.

Today, a significant portion of all clinical research—clinical trials on hundreds of thousands of subjects—is sponsored by pharmaceutical companies, which are governed by standards set by the Food and Drug Administration (FDA). More than half of the data involved in clinical research is out-sourced. The research done by these entities must satisfy requirements detailed by the FDA in the document referred to as 21 CFR Part 11.

Pharmaceutical firms rely on CIS functionalities for the collection, transfer, review, and reporting of data. Their CIS products must meet regulatory requirements, including worldwide safety reporting, data and document consolidation, clinical trial administration, regulatory tracking and correspondence, and New Drug Application (NDA) submissions. Unlike other sectors of the healthcare system, and with the apparent encouragement of the FDA, the pharmaceutical industry has embraced the Medical Dictionary for Regulatory Activities (MeDRA) as its standard for nomenclature.

External Data Management

In addition to traditional data collected in the investigators' offices, the patient's research record must integrate data from multiple external sources, including laboratory test results, genomics data, digital instrument data, patient diary and questionnaire information, data collected by clinical research organizations (CROs), and data collected by interactive voice response systems.

Current Process. Information systems that receive and load these data are able to process a variety of formats, technologies, and transfer mechanisms. Once a clinical trial establishes a need for particular types of data, potential suppliers are identified, and formats and transfer mechanisms are negotiated. When a supplier is selected, detailed specifications are developed, and external data are mapped to the metadata defined for the receiving database. The next step is to develop transfer systems and validate them. Tests are conducted with sample data, and the supplier's systems are reviewed for validating evidence. To insure data accuracy, computer quality assurance staff visit new suppliers and conduct an audit to confirm that the internal systems used to store and analyze clinical research results conform to the requirements detailed by the FDA in 21 CFR Part 11. This is a time-consuming process: 14-65 working days (WD) to establish contract, 20-40 WD to establish load and mapping mechanisms, and 15 WD to test and validate those mechanisms. Some central laboratories maintain as many as 1,200 different formats and systems to supply data to clinical research firms; often up to 50 of these formats are maintained for a single research firm.

Once the systems are in place and the study begins, external data transfers can begin. The automated process picks up the file, formats it, maps its contents, and loads the data. An automated review process identifies content or format errors and generates error information for communication back to the supplier. When revision records are received, they are identified, processed, and added to the overall audit trail of the clinical records.

Experience to Date. Clinical research systems have followed the workflow described above and handled it quite well. Traditionally the main sources of external data were professional labora-

tories that received all of the samples from patients in the studies, performed the prescribed tests, and delivered the results back to the study sponsors through automated transfer mechanisms. These processes were established in the beginning of a clinical trial through negotiations with each supplier. Data types tended to be similar, and records consisted of patient and protocol identifiers, date, test descriptors, and test results. Suppliers were often reused, and the process became fairly routine. In the past, data were transferred only once or twice during a trial, and the data management team handled the review process efficiently.
Today, there are many more kinds of suppliers and types of data. Loading and review are performed closer to real time with transfers occurring weekly or even daily. External transfer data are essential to the conduct of the trial and reviewed with patient data regularly by clinical research staff as well as data management.

These changes and the challenges they present to data management and IT staff involved in transfer activities can be accommodated by an optimal advanced CIS. Identifying external sources early in the protocol development process will expedite contract negotiations and system development work. General-purpose receiving tools able to accept a variety of formats and mechanisms facilitate frequent transfers.

Industry forums like the Clinical Data Interchange Standards Commission (CDISC) are working to establish standard formats and coding structures for these exchanges. New technologies such as the Web and XML will be used to facilitate transfers; and, metrics will be maintained to monitor frequent transfers, to quantify and track discrepancies, and to alert contract administrators to error-prone suppliers. Two-way communication mechanisms will allow faster error resolution, and enhanced authentication services will insure security of the transfers. Transfers and loading of data will be automated to eliminate delays caused by manual processes.

Genomics Data and Clinical Research

The potential use of genomics data in clinical trials research falls into three general areas: pharmacogenomics (the study of how variations in human genetics affect individual response to drugs), target validations (the identification of target populations for clinical trials), and sample repositories. A number of pharmaceutical firms have begun to collect blood and samples routinely from subjects in clinical trials. Collected this way, patient samples cost far less and are accompanied by much richer demographic and clinical data than samples purchased in the open market. As the data accumulate over time, their statistical power will grow.

The “Statement on DNA Sampling: Control and Access,” published by the Ethics Committee of the International Human Genome Organisation (HUGO), defines industry standards for collecting and banking genetic samples. The identifiers, or codes, used for samples must be completely different from the identifiers used in clinical trials to protect the anonymity of patients. The code attached to the sample accompanies the sample when it is sent to a sample bank. After the sample is processed and genetic assays are conducted, the genotyping result is transmitted back to the sponsor with that code only. To comply with HUGO, the sponsor’s IT system must be able to generate a unique genetic sample code for each clinical identifier.

Information systems must support users at the sponsor's organization in being able to link clinical trial data and genotype data in association studies. Security mechanisms must be in place to ensure that patient identifiable information has been stripped from the data. Thus, in addition to being coded, the samples must be anonymized. Any necessary demographic and clinical data that may accompany an anonymized sample must be encoded in such a way as to preclude users from tracing back such information to an individual patient.

Because patients must be offered the right to withdraw samples after collection, the information system must keep track of the samples collected and the type of consent patients have signed. If the sample is coded, the system retrieves it from storage and destroys it. If the sample is anonymized, the system informs the user that the sample cannot be destroyed since the sample identifier has been stripped.

Standardization of Genomics Information. As a result of high throughput technologies, genomics data are generated at numerous biotech and genomics companies, government labs, and non-profit research institutions around the world. The growth of genomics data has been exponential. Each organization in the genomics community has its own proprietary data format that is difficult to exchange with other organizations. It sometimes appears that scientists around the world *spend more time on reformatting, mapping, and parsing data from various sources than analyzing the data for research purposes*. In this post-genome era, the industry urgently needs standards to govern the biological entity naming convention, genomics data inter-exchange, and transmission protocols. Reducing the hurdle of genomics data exchange will enable researchers in all sectors to spend resources on discovery and development rather than on controlling or preventing the sharing of scientific information.

Research scientists studying a given drug target use many types of data. Today, they must retrieve data for analysis manually, including the target gene's sequence data from GenBank, common polymorphism data from dbSNP, sequence variation data from a vendor, gene expression data from another vendor, and signal transduction pathway data from pathway databases. One standards setting organization for the life sciences is the Interoperable Informatics Infrastructure Consortium, or I3C. This group is developing a common open platform that uses XML and Java technology to exchange genomics and other life sciences data across many organizations. Standardization efforts will facilitate the sharing of various types of information, greatly increasing the power and speed of discovery. Researchers will be able to retrieve all the data from many sources effortlessly, by specifying the target name and setting a few parameters.

Data Analysis. Analyzing genomics data demands computing power. For example, a CPU with 1 GB memory can only process a haplotype that contains up to 20 single nucleotide polymorphisms (SNP). Today, the industry is looking for more scalable and low cost alternatives, such as running genomics applications on a Linux cluster. Advanced CIS will have to provide analysis tools to enable development scientists to review these data; these are yet to be defined.

Long Term Retention of Electronic Clinical Data
Given the inevitable obsolescence of computer technology, today's electronic records are in danger of being rendered less usable, potentially unintelligible, or potentially misleading in more modern computing environments. No solution now exists. A proposal being prepared by industry for the regulatory agencies suggests that it be permissible to (1) retain the enormous volumes of raw data for shorter times than the key parameters and final results, (2) reduce the levels of detail in retained data (e.g., lower resolution for images, lower sampling rates) when the meaning and value of the data can be retained, and (3) off-load some electronic records to paper or microfiche for long-term retention (>10 years).

Overall Impression of the State of Advanced Clinical IT Systems:

The characteristics of a well-developed information system depend on what the users want it to do; current technology permits a wide variety of reasonable solutions. Some useful points can be gleaned from a review of some of the better systems now in operation.

Today's model systems incorporate off-the-shelf components to a greater or lesser extent, but in almost all instances, significant additional time and expense are necessary to adapt the system to enterprise needs; "turn-key" systems, particularly if clinical research is also to be supported, are not yet available. Successful models must take into consideration the human factors involved in development and evolution of a CIS, It is essential that system users be involved early in the process.

For most organizations, the needs of clinical research are best served by storing data derived from clinical care in a data warehouse distinct from the operating system serving the clinicians. The rapidly increasing use of genomic information in clinical medicine greatly increases the demands on an information system, but concurrently makes more imperative the need for an efficient CIS if such data is to be managed efficiently and confidentially. Long-term storage of health data in information systems is a problem for which no good solution yet exists.

Well-designed CIS features can ease considerably the increasing regulatory burden associated with clinical care and clinical research and improve quality of care. Although the literature provides evidence for considerable cost-savings when decision support from CIS is used to reduce clinical errors, those involved with good working systems stress the advantages of increased efficiency, better decision support, and improved support of research and administration, rather than dollar return on investment.

Clinical Data Standards

As a structured source of patient presentations and recorded outcomes, electronic medical records (EMRs) can help meet the needs of clinical research, and improve quality and safety of patient care. EMRs cannot fulfill these needs without standards, conventions, and consistency. Absent information standards such as vocabularies and structural rules, the information that is put up by an EMR cannot be shared among the end users of the data within a health system, or between health systems. Such an EMR is *not* interoperable and lacks the capacity to consistently exchange coded data. At this time, while most EMRs are intended to inform a clinician in real-time about the status and data pertaining to a particular patient, they are not standards-based. As such, the information in these systems cannot provide automated decision support, because computer-based decision support requires a shared semantic basis with the record that can only be achieved with standards. This chapter provides a review of the state of the art in standards development, the barriers to implementation, the consequences of our failure to implement them in the US (in contrast with a number of other developed countries), and the success of one health system at implementing a standards-based EMR.

Representing Medical Concepts

Events, observations, and interventions, in the context of episodic or ongoing medical care or research, require language that is capable of nuance, detail, and unrestricted by expression semantics or contents of a given vocabulary. Ideally, the language used in medical care or research should be current, natural, friendly, and established, requiring codes that are concise, structured, and consistent, so that they can be analyzed and manipulated. However, as spoken by humans, natural language is often ambiguous and unpredictable, most notably from the perspective of a machine trying to interpret it.

The International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) restricts the combinations available, and does not permit code modifications. For example, it does not permit modifying the ICD term "pneumonia" with the term "severe." In ICD-9, combinations must exist as pre-coordinations or be listed as two co-occurring codes. In reality, codes are rigid, tedious, and high maintenance. The limited semantic model around ICD and most codes is a problem inherited by developers of EMR. Until medical vocabularies are able to capture a robust terminology model, this problem will persist. The Health Level Seven (HL7) Clinical Templates Level 3 developed for the communication of clinical information can, when fully specified, provides such a high terminology model.

Terminology. From the perspective of nomenclature, EMR terminology needs to retain detail and specificity, such as "Stage IIIb adenocarcinoma of the ascending colon with c-myc oncogene expression," while capturing symptoms, findings, events, and interventions. The terminology also needs to support high-level classifications and summary codes, such as ICD-9 codes. In particular, it needs to support a network model in which findings, events, and interventions link to purpose-specific aggregations, for example, in the areas of decision support and error detection, public health and surveillance, reimbursement and management, and

outcome research and epidemiology. One single set of data cannot serve all purposes, giving rise to the need for a Web-based model of aggregations.

Classification. From the perspective of aggregated classification, EMR terminology needs to enable accurate, reproducible groupings used in fiscal, reimbursement, and management activities, as well as in public health and statistical efforts. Doing so requires a Web-like continuum of aggregation, including small clusters of like conditions or events, broad groups of related concerns, and groupings for other purposes. In public health, for example, all bacteriologic sub-types of beta strep throat could be lumped into one category, "Beta-Strep Throat Infections." Alternatively, all stages and histologies of new breast cancers could be grouped into the category of "New Dx of Breast Cancer."

National initiatives in error reduction and evidence-based medicine require access to detailed patient data, consistent and comparable recordings, facile and familiar data entry, and real-time provider interaction. These requirements define the agenda of health data standards developers. Research protocols assembled from EMR data standards are more likely to harvest useful information from EMRs. Nonetheless, most medical records, in the natural course of being completed, do not capture much information needed for research. For example, drug sensitivity studies by Beta-strep serotype may not be readily available in the some records.

Terminology plays a strategic role as an interface to knowledge resources, including guidelines, critical paths, and reminders, as well as decision support and references. Terminology is key to the support of practice analysis, recognizing like "risk factors" in patient records even though they may be expressed by different words or phrases. In the area of infection control quality improvement, for example, it can help by assigning AIDS patients, cancer patients who are experiencing the height of chemotherapeutic immunosuppression as a side effect, and transplant patients on immunosuppressive therapy to a high-level category of "Immunosuppressed patients" for infection control quality improvement.

Terminology Systems. The relationships among terminology systems are defined by their functional roles. Entry terminology involves humans; reference terminology is specific to machines; and aggregate and administrative terminology drives applications. However, there are multiple abstraction layers, with an inverse relationship between the extent of abstraction and the richness of detail. The lowest level of abstraction is molecular, genomic, and cellular information. Next, in terms of abstraction, come clinical observations and patient records. Machine abstracted and algorithmically derived classifications include surveys, scales, and databases, on to clinical classifications, and, from there, to higher level groups, such as diagnostic related groupings (DRGs). Aggregations for specific purposes such as outcomes research, public health and surveillance, and decision support all require linkages to findings, events, and interventions.

Structure and Context. Terminology standards must also accommodate structure and context. For example, the existence of a term such as *heart disease* in a patient record does not necessarily mean the patient has those conditions; it may appear within a family history, fundamentally changing the meaning of the term to *family history of heart disease*. These are equivalent ways of

representing the same concepts. In other words, the term "Stroke" popped into a family history "box" in an EMR is equivalent to "family hx of stroke" in a patient note. These expressions relate the notion of family history by very different mechanisms, the first by structure, the second by modifying terminology. The semantics, however, are identical. That inter-convertibility needs to be recognized as equivalent content.

Dual Use. The challenge is to create information that can serve both clinical care and research protocols, with a common underlying infrastructure. Research abstractions can be based on clinical standard components, including structure, terminology, messaging, and laboratory data. Initiatives like the National Cancer Institute's Common Data Elements (CDE) show that this is possible. However, what is relevant to a clinical encounter may not overlap with the requirements of a research protocol. The level of detail and specificity for a given research protocol is likely to exceed the data requirements for most clinical episodes. Focusing on and addressing the intersection of these needs will be key to creating data usable for research and clinical care.

The prerequisite to harvesting data from the record is negotiating a standard set of patient descriptions useful for patient care and research. Without standards, separate data collection forms and abstractions will continue to be necessary. The goal is to minimize the need for these, lessening the burden and cost for clinical research. It is likely, however, that researchers will need to collect some specialized data separately.

Barriers to Realizing the Potential of Standards

Health information standards have the *potential* to make data more consistent and comparable in some areas, especially format (HL7 for messaging) and some areas of content (Logical Observation Identifier Names and Codes, or LOINC, for laboratory test names). Yet these standards are still not in widespread use, and many provider organizations have yet to make the transition to existing standards, due in part to inertia and the costs involved. These organizations will not appreciate any significant value from adopting standards until they have invested up to and beyond some critical threshold, after which they derive efficiencies. Getting past that threshold is the inertia. The expenses are manifest in the acquisition of new EMR systems and modules, and in the costs related to the need for new hires, for personnel re-training, and for orienting the end users to make full use of the system. These costs are high in money, in time, and in human frustration.

Lack of Consistency and Comparability. In some terminology areas, standards are simply not available to satisfy the requirements for consistent and comparable data. In the area of diagnosis, for example, ICD-9-CM lacks the level of detail that would make it reusable for patient care or research. Its categories are too big and "lumpy." For example, researchers can make little from the ICD-9 classification of "Other bowel polyposis." In the case of diabetes mellitus, it offers ten different diagnoses, one without complications and nine with complications, but it does not define severity or allow the use of more than one code. Moreover, the tendency of ICD-9 to change the meanings of its codes and to offer "Not Elsewhere Classified" as a catchall term results in data that are not comparable between sites and patients.

The individual data in an EMR can meet the needs of clinical care. Certain research arenas can make use of aggregations of data across patients at the same site and across sites. However, some areas of research, notably genomics and translational research, require highly detailed and specific data not present in aggregated data.

Proprietary Standards add to Costs. Although diagnostic codes in the Systematized Nomenclature of Medicine (SNOMED) are well developed, the system is proprietary. The proprietary nature of these standards presents several problems. In the past, the College of American Pathologists (CAP), which developed SNOMED, was not open or consistent about its pricing policy; it could demand what the traffic would bear. According to one anecdote related in the standards community, the usage fees that CAP would have charged on a per-code basis for a community pathology coding demonstration project under National Cancer Institute (NCI) funding would have exceeded the total amount of the grant, leaving no money to do the research. In the future, with a site license for use in the United States, this problem should become moot. Proprietary issues have also affected the standards for procedures coding. Current licensing restrictions make Current Procedural Terminology 4 (CPT4) too costly to use except for functions that generate bills.

There are some who hold that standards should not be proprietary; and, organizations that have developed them should either sell them for a fair price to the government, or the government should put development of alternatives out for competitive bid. Free and open access to the periodic table of elements enabled science to advance and develop practical applications in chemistry. No proprietary organization can charge every time the symbol C is used for carbon, or Na for sodium. Yet this is precisely what proprietary organizations are trying to do for health care: charge for the functional equivalent of health care's periodic table.

Lack of Incentives. Where standards do exist, there is little incentive to make the investment needed to incorporate them into the clinical information infrastructure. Because solutions (such as terminologies, interfaces and decision rules) are often "hard wired," the adoption of external standards is often difficult or, in the case of some turnkey systems, impossible. Maintenance of local solutions is often a simple matter, while coping with changes in external standards may require resources not typically available to health systems. If CPT were to redefine its codes to require sophisticated new capabilities (description logic inferencing engines, for example), some health systems would find it difficult to accommodate such changes.

Inflexibility. Even when implemented, existing standards may fall short of user needs because they may be unresponsive to changes, such as the recognition of a new disease like AIDS. Sometimes users have local requirements that would not, in any case, be met by standards, but their systems lack a way to account for local extensions. For example, at Mayo, it is an ongoing struggle to keep track of ICD codes versus codes used locally in the ICD slots.

Existing standards can also impede the adoption of higher-quality standards. Requirements for producing data encoded in CPT4 and ICD-9-CM usually results in efforts to capture data

in those terminologies and divert resources away from capturing clinical data. Clinical research would be better served if requirements focused on capturing detailed patient data and then mapping to the "lumpier" ICD-9 and CPT4 classifications as two separate steps.

Actions Needed

Evaluation. To date, studies have only shown that standards do *not* work well for multiple purposes, or sometimes even for a single purpose. Other standards, for example for medications, are not yet sufficiently well developed for widespread use. According to one classic study, "No classification captured all concepts, although SNOMED did notably the most complete job. The systems in major use in the United States, ICD-9-CM and CPT, failed to capture substantial clinical content…. The major clinical classifications in use today incompletely cover the clinical content of patient records; thus analytic conclusions that depend on these systems may be suspect."[2]

There must be demonstration projects that show how standards fare when attempts are made to use them to record clinical data for patient care and research simultaneously, or for any intention with later re-use for other purposes. Evaluation must be an iterative process, with feedback to standards developers who must be responsive to the needs of users. Standards can be iteratively improved, and it would be of major benefit to make selections and then improve those that are chosen. In early 2003, a consortium proposal for Open Health Terminology (OHT) was still pending before an unnamed funding organization. Essentially, it was a proposal for an open source like Linux dedicated to EMR-related applications.

Endorsement. Detailed patient data standards, namely SNOMED, if endorsed, would make an enormous difference from both the clinical and research perspectives. If the Centers for Medicare and Medicaid Services (CMS) mandated that all data exchanged with it were in standard formats, other insurers would follow suit. Common use drives standards to be more comprehensive and inclusive. For example, if CMS mandated the use of RxNorm, now under development by the National Library of Medicine, this standard would be better positioned to receive the funding needed for completion. For standards, lack of full readiness does not represent as big a problem as lack of endorsement.

Controlled Terminologies[3]

Adequacy and Usability. Often regarded as a special case of information standards, controlled terminologies are important in both clinical and research contexts. However, issues always arise when data are used for varying purposes. Research, for example, often requires a much higher degree of granularity than clinical medicine. Controlled terminologies are sometimes adequate for specific tasks, but none has been shown to work well for re-use of data for disparate purposes. Terminologies that are acceptable for coders often do not work for clinicians.

[2] Chute CG, Chon SP, Campbell KE, Oliver DE, Campbell JR. "The content coverage of clinical classifications. For the Computer-Based Patient Record Institute's Work Group on Codes and Structures." JAMIA 1996 May-Jun;3(3):224-33.

[3]Terminology in this context is a finite, enumerated set of terms intended to convey information unambiguously.

Controlled terminologies are usually licensed, and are available subject to licensing restrictions. As of this report, terminology developers do not publish anything approaching an "aggregation logic" with their terminologies. Doing so could hasten progress toward a network model of purpose-specific aggregations. It is likely that, subject to iterative formative evaluation and in concert with the terminology developers, controlled terminologies could be made broadly functional.

Barriers. Barriers to the development and validation of controlled terminologies include resources to carry out the evaluations and to secure the cooperation of developers. While the LOINC model supports open use of terminologies, other licenses are based on number of computers, applications, and/or users.

Implementation of standard terminologies in commercial systems is often a barrier to acceptance. Such systems may make use of flat sets of codes and terms that can at best be replaced by new versions of the terminology, with no way to use classifications and no way to account for changed or retired terms. Local extensions are either disallowed (limiting the expressiveness of the system) or allowed without limitation (reducing the usefulness and sharability of the data).

Reducing barriers will require improving terminology content, improving licensing such that it is nominal and enterprise-wide, and adapting commercial systems to utilize terminology servers rather than hard-coded local term lists. With an open source as suggested in the Open Health Terminology (OHT) proposal, there would be no barriers to anyone using standard terminology to describe patients and diseases.

Research vs. Patient Care

Differences in data content and use requirements. There is no simple relationship between the granularity of research and patient care data. The formats are not necessarily different, although the data collected for research are typically much more specific than those routinely captured by a clinician. For example, a clinician may note that a patient's appetite is decreased, but not record whether the decrease is mild, moderate, or severe, while a researcher may need to know level of severity. In other cases, researchers may be interested in general categories of conditions (such as "does the patient have lung disease?"), while a clinician would never record that a patient has simply "lung disease."

Reconciling Differences. These differences are not irreconcilable. In some cases, it is simply a matter of making sure that the terminology is capable of representing concepts at all desired levels of granularity and that there is sufficient mapping between the levels, such as "is-a" hierarchies. (e.g.: "asthma" "is-a" "lung disease.") Data collection programs will need to become more sophisticated to assure that data that are captured for patient care have the level of detail required for clinical research. For example, a clinician may not record the patient's appetite, let alone the "degree of decrease," during a typical office visit for, say, hypertension. Medical record systems will need to know when such specific data are required as part of a

patient's research protocol (for example, to record side effects of an investigational antihypertensive agent). Once the terminologies are reconciled, it will be up to software applications to match the concepts of research with those of patient care.

Even if there is no convergence over time between the requirements of clinicians and clinical researchers, more standardization of clinical data and in particular EMRs could still facilitate clinical research. The United States is rapidly falling behind Europe, Australia and Canada in its capability to apply CIS to clinical research. The United Kingdom has adopted SNOMED and placed it into wide-scale use, and Canada has huge public health databases of consistently coded information for all patients, not just a subset, as is the case in the United States with Medicare.

Interoperability

Sharing information within large databases is predicated upon interoperability, or the ability of multiple organizations and multiple systems to "talk" to one another, that is, to exchange consistently coded data. The key step in achieving interoperability is the endorsement of standards for the most important types of clinical data, including messaging, and around content such as conditions, procedures, laboratory results, and radiological data. For most of these types of data, the standards are reasonably well understood. There needs to be a process for selecting standards, supporting further development, and determining when development is sufficiently advanced that standards are ready for use.

For more complex data, such as guideline and rule representation, additional work needs to be done to select what appears to be the best of the available standards, and then these will need to be built out. For dual use by clinicians and clinical investigators, clinical applications should specify for clinicians when additional detail is required to allow data to be reused in research.

The major barrier to the full development of interoperable EMR systems has been the lack of standards. Healthcare systems (hospitals and group practices) have been reluctant to invest with vendors because of a concern that they may select the wrong one. Adoption of standards for key types of clinical data—which has occurred in countries such as the United Kingdom and Australia—would substantially mitigate this risk. Access to even limited clinical data in standard format—for example demographics, diagnoses, medication and laboratory data—would be of enormous value to researchers.

A Roadmap to Adoption of Standards for EMR Systems that will Support Clinical Research. The first step is to inventory available terminologies, establish relationships with their developers, and devise mechanisms for evaluation and improvement. This includes collection of terminologies used in clinical research, such as drug trial protocols, case report forms, and the like. The next step is to initiate an ongoing cycle for evaluation and improvement for each terminology, and to coordinate those efforts. For example, the terminology for laboratory studies needs to be coordinated with the terminology for laboratory findings so procedures can be ordered, results reported, and research conducted on resource utilization and epidemiology. An additional requirement is to establish terminology services to support the common needs for

information systems. These include term look-up (e.g., for user data entry), class-based queries ("is X a lung disease" and "get me all the lung diseases"), and those based on other attributes ("what are the lab tests that measure Y?"). It is likely that vendors of information systems will need tools and resources to help them adapt their systems to make use of such terminology services.

The above steps all refer to terminology aspects. But even as the "terminology problem" is "solved," there will still be a need to develop clinical systems that can serve the needs of patient care and research. For example, mechanisms will be needed to specify data collection protocols that help clinicians know when additional level of detail is required in clinical data capture to support re-use in research. It will not be sufficient to "solve the terminology problem," though it is necessary. We also need to standardize how research data relates to patient records.

Patient Identifier

It is not possible to avoid fragmentation of care and fragmentation of data unless the issue of a unique patient identifier is addressed. Because of concerns among some patients and policy makers, this may need to be a voluntary program, with incentives, such as quicker hospital check-ins for patients who opt for a unique identifier.

Denmark has a national identifier that ties together encounter records, diagnoses, procedures, medication usage (in two provinces), and a number of registries. Danish researchers are able to conduct critical public health and health outcomes research—without any technical magic. The Canadian provinces of Saskatchewan, Manitoba, Ontario, and British Columbia now have something very close to a provincial medical record, with all drugs and all labs, fully coded and available, using HL7 and LOINC. Great Britain is using its national database to create a 500,000-patient cohort that will collect clinical data and history.

In comparison to the population-based research capacities of these other countries, the approach in the United States is at best haphazard. Matching algorithms on a national scale to match a patient against a death noted in the Social Security system are an unsatisfactory solution. They suffer, first, from data-entry errors on all sides and, second, from sheer numbers. Using birth date and name may render, for example, 15,000 John Does, with five to ten of them having the same birth date. Pharmaceutical research databases do not do link to health-care data, limiting our capacity to conduct effective post-marketing surveillance of drugs, or to link serious adverse events to previously unknown biological markers of vulnerability. Even the National Cancer Institute's data model has no intentional connections between patients across studies. It identifies a subject within a given study only; it does not recognize a patient involved in more than one study.

Research Databases

Data Set Formats. Clinical databases are more standardized than research databases, with the possible exception of some cardiology research databases. Most researchers tend to define data sets as hard-copy documents. The data entry fields are unconnected and limited in space.

Thus, research databases are flat data sets, in which specific questions are viewed as units, rather than hierarchical data sets, governed by field definitions. This works for the first study, but the addition of one more variable requires a whole new database. Combining the data becomes difficult in the absence of a conceptual framework that allows a given question to stand independently, outside of this data set, for reuse as is or as adapted. Over time, biostatisticians try to merge, integrate, and manipulate these data.

In the clinical care setting, systems—whether for laboratory, pharmacy, inventory, or billing—do not have a field for each identifiable variable. Rather, they have a master file code that points to different master files in different cases. Because the information is structured, it can be linked to other information, such as who recorded it, where it was recorded, what time it was recorded—whatever kind of detail is needed, making it easier to expand and to enrich by including what is very often more raw data. Using hierarchical, structured data sets will enable clinical researchers to assemble and manipulate data, without having to locate and match up data in multiple flat data sets.

Free Text Data. Unlike data that are already structured in the system, narrative data are not readily usable for clinical research. Physicians, like other professionals, think and work in narrative text. Whatever codes are developed, physicians will not be inclined to code in their entries. That said, coding some of the information, especially what is really important, would be a reasonable investment and could yield a reasonable return. But the code should not be just a code; it should be a question, with a specified set of answers. A reference catalog of questions would result in more common data in various systems. First, there must be research into what constitutes a good question. For example, developers of the Ottawa guidelines for ankle x-rays identified around 500 questions (e.g., a bruise in a certain place that is more than 5 centimeters), reduced them to 100, tested them on 1000 patients, and analyzed the outcomes to determine which questions yielded predictive information. There needs to be more experimental and empirical work of this nature, testing and proving the value of questions.

The terminology problem is not a problem of simply atomic codes or phrases. For example, the Bethesda standard for describing a cervical Pap smear is like a small sentence: "not this, plus that." Survey-instrument questions are also sentences: "Have you, in the last three weeks, been bothered by chest pain enough that you stopped doing what you were doing?" The challenge is to decide what dimension to measure. In some instances, there are alternatives that are not allowed under current systems. For example, the ICD-9 code requires that physicians categorize diabetes as under control or out of control. Entering the hemoglobin A1c would give an objective measure and better data, but that option is not available.

Linking to Registries. Linking clinical trial data with data in tumor registries, for instance, can complement the strengths and weaknesses of the two, for epidemiologic work, and for doing clinical trials. Follow-up in the registry can help get additional information or ask extra questions.

Standardizing Data at Intermountain Health Care

A 22-hospital, 1.4-million member system, Intermountain Health Care (IHC) has a component-based information architecture. Departmental systems come from a variety of vendors and communicate with each other via interfaces, most of which are either HL7 for the clinical data or X12 for the billing/patient administration data. All data are entered into the electronic medical record system, or clinical information system (CIS). From the clinical database they are moved into a data warehouse. A data dictionary is used for coding to ensure that data are consistent and comparable across all information systems. As of October 2002, IHC was entering close to 2 million new observations a day into the CIS. Trends to data suggest that growth will be exponential.

As in translational research, the number of data points in some parts of the clinical care system can be high. For example, a patient in the intensive care unit (ICU) who is on a ventilator, cardiovascular hemodynamic monitoring equipment, and IVs (all of which are automated) may have up to 2000 data points a day for blood gases, chemistry data, etc., including 750 for q15-minute vital signs. With a standard stay of 14 days in the ICU, this amounts to over 100,000 data points per patient. This volume must be accommodated in CIS. At times, the responses to requests to "see vital signs report" were extremely slow. This resulted in the recognition of the volumes of data involved and subsequent improvements in the ability of the system to handle them.

The lack of consistent modeling and terminology related to the data constitute an obstacle that has limited the amount of data available for research. For instance, an arbitrary code may combine diastolic and systolic blood pressure, or a description of heart rhythm with heart rate, as if they were text expressions, rather than standard terminology. The process of understanding these arbitrary codes, defining them, and putting them in the data repository takes from 12 minutes to an hour. An estimated 30,000 such concepts in the data repository amounts to many hours of work. Today, clinical research at IHC involves extensive manual work and overhead, with about 20 people mapping data and configuring interfaces.

Vertical interfaces present a problem. IHC communicates with multiple entities, from the Food and Drug Administration (FDA) to the Society of Thoracic Surgeons. However, it is clearly unscalable for IHC to support technology and individual terminologies for everyone—each government agency, each professional society, each with their own systems. In fact, IHC was driven to HL7 and standards because it was unsupportable and unscalable to do otherwise.

Today, IHC is putting automated processes in place that can filter and process data that end up in both research databases and the electronic medical record. The intent is to avoid having to support another standalone application with a different operating system, a different desktop, and its own environment. Rather, the goal is to have one single user interface support both clinical research and patient care. Although the data might be entirely different, the vision is to have the interfaces be the same, for order entry, laboratory, etc. This is scalable and can be done with the right staff and training.

The goal at IHC is to use common codes within two kinds of coding schemes, namely, observation identifiers and the actual values of those observations. This requires coordination between clinical terminology and the structure in which it is used. Terminology development cannot be done in a vacuum, but must be based on an understanding of how the terminology fits into the data slots, columns and rows within a given database or message.

Clinical Research Support

To support clinical trials, IHC uses these systems to identify eligible candidates, to enroll subjects, and to use order entry and other kinds of systems to capture information about them. The infrastructure and terminology, data-exchange standards, and knowledge standards can all be re-used. Clinical research can also be done on existing data using automated data extraction and submitting data to national databases in which IHC participates. For a number of years, IHC has studied adverse drug events. Data from the EMR support genetic epidemiological studies, with appropriate security and IRB approval. For example, researchers can find records of patients with asthma, link those records to ones about drug usage, and link patients together in family trees to discover the profile of asthma within families.

Comparing clinical data across all 22 hospitals allows IHC to assess outcomes. For example, when researchers looked at transurethral prostatectomy (TURP) and length of stay, they found some hospitals had a length of stay that was 1.5 to 2 days longer than others. Analysis of the data showed that pulling the catheter a day earlier allowed the patient to go home a day earlier, while leaving the catheter in another day meant a longer stay, without much difference in terms of clinical outcomes. When these findings were announced, clinicians modified their practice so patients could go home earlier. As other studies take advantage of data-mining, more improvements to current practices will be discovered.

Forms of User Resistance at IHC

Although there are very real technical barriers, the primary barriers are not technology but complacency and divergent goals. Making consistent standards and terminology is immensely important and practical. Still, people give a host of reasons for not using standards.

"My code isn't in there." HL7 accommodates requests from people who are trying to exchange data, doing what is necessary to make the system usable.

"It's hard to implement." The first one is the hardest. Each additional one gets easier, because reusing technology and terminologies gives economies of scale.

"We're unique." In a variant of the "not invented here" syndrome, users believe their needs are so special that standards are not applicable. However, they can specialize the standards to acknowledge any differences and allow for reuse of the terminologies and infrastructure involved by adding an implementation guide.

"I've got this great new XML thing." Syntax is not the same as the specification of content. Both Extensible Markup Language (XML) and HL7 use vertical bars, but XML is content-free and

HL7's work is 10 percent about syntax and 90 percent about what must be communicated to meet clinical-care needs and business needs in an environment. Because XML has no open process, no consistent organization to support it, and no consistent terminology underneath it, using it does not lead to standard data exchange.

The Benefit of Standards

Over time, the ability to make detailed observations is the basis of many scientific advances. Theories about heat and entropy evolved after the development of thermometers and accurate readings. The ability to measure the speed of light and the search for ether led to theories of relativity. In medicine today, everyone is making measurements that are not comparable across institutions and thus are not generally usable. Standardized terminologies and data models will make it possible to see and hear what cannot be observed today. This can be done for a fraction of the amount that we are spending on other kinds of research—essentially a rounding error in the budget for the National Aeronautics and Space Administration.

General CIS standards in health care will simplify the process of training people about the technology and how to use the terminology. General standards will also provide the opportunity to use commercial tools, to find consultants who are familiar with the model and with the terminologies. In short, general standards promise tremendous economies of scale. The federal government should be playing a major leadership role in endorsing the implementation of standards, and assuring that funding is available to enable institutions to bring their CIS into compliance with the standards.

The Roles of Industry and Government

The final two panels presented the perspectives of the principal clinical IT software vendors and of government agencies. The gestalt of these two panels conveyed the impression that solutions to the IT needs of the clinical research community operating within medical schools and teaching hospitals will not be easily achieved because of the limited financial resources of each academic center, the nature of the business models and the marginal financial strength of the modest-sized vendor companies, and the limited types of granting models and limited funds available from the principal federal agencies.

To date, whatever progress has been made has involved collaborative projects involving one or more vendors, government agencies, and academic medical centers or clinical academic societies. These projects have occurred at the margin, and produced marginal successes related to a limited set of goals. While the separate projects certainly affirm the possibility of collaboration, they have not begun to approach the scope or scale necessary to bring about incremental change. If we are to achieve real systemic progress, the government will need to re-think its traditional approaches to funding clinical IT and related clinical research enhancements through existing grant mechanisms at health-related agencies. An alternative model of government/industry/academic collaboration should be considered, if the promise of clinical IT to improved patient safety, public health and bioterrorism surveillance, cost-effective patient care, and clinical research is to be achieved. As noted in the previous chapter, the federal government should be playing a major leadership role in endorsing the implementation of standards, and assuring that funding is available to enable institutions to bring their CIS into compliance with the standards.

Healthcare institutions are dependent upon vendors to help them address their clinical IT needs and to comply with government and other regulatory reporting requirements. In these times, most institutions would not contemplate replacing legacy IT systems because of cost. They rely on a relatively small number of undercapitalized vendors—vendors that are generally discouraged from applying for federal support for system development. Vendors have insufficient reserves to support general research in IT, and they do not offer clinical research modules.

In general healthcare institutions that perform research rely on a relatively small number of vendors. That these vendors operate across a large number of health care institutions offers a greater promise of common standards and connectivity than can be achieved by separate, independent IT development at each health care organization. The challenge to the clinical research community, and to government agencies that fund clinical research, will be to put the IT needs of clinical researchers into the planning efforts of vendors and of academic medical centers.

Within the federal government, the Centers for Medicare and Medicaid Services, the Veterans Administration, and the Department of Defense can exert substantial influence on IT development as large purchasers and/or providers of health care. Other governmental agencies can foster IT development through targeted initiatives such as healthcare quality (Agency for Health Research and Quality), bioterrorism surveillance (Centers for Disease Control), and high-end

instrumentation (National Institutes of Health). At present, given limited funding, these agencies are attempting to align their critical healthcare missions with information technology (IT) enhancements. However, there is no real precedent or mechanism for governmental involvement with vendors of the magnitude necessary to change healthcare information systems enterprise-wide.

The sections that follow highlight the issues and opportunities facing vendors and government agencies. At the end of this chapter, the focus is on a recent model of collaboration, the SAGE (Sharable Active Guideline Environment) Project. This project is unique in its new federal funding source, the Department of Commerce NIST/ATP, and in its ability to engage multiple vendors and academic sites in development and feasibility testing. The emergence of this model gives hope that the federal government, when faced with significant opportunities, challenges and/or crises, is able to craft solutions that support innovation, infrastructure and program development in unique partnerships of government, academia and free-enterprise.

The Vendor Sector

Undercapitalized and customer-driven

Software companies that develop clinical information systems (CIS) for healthcare institutions are a small subset of the large and volatile information technology (IT) industry. Representing this sector at the workshop were five vendors whose CIS have a major presence in health care. They include Cerner, with Millenium™; Eclipsys, with Sunrise Clinical Manager™; IDX, with CareCast™; McKesson, with Horizon™; and Siemens, with Soarian™. Each of the five has a particular marketing focus—for example, Cerner stresses clinical documentation, IDX emphasizes patient safety, and Siemens highlights workflow—but all offer their own version of a fully functional CIS. In supporting their products, each vendor works with its own clients, sometimes developing additional tools to take advantage of the system's capabilities.

In the absence of clear evidence of a market for software to support both patient care and clinical research, the vendor sector lacks the incentive to develop such large-scale, complex software. Because the sector is high-risk and attracts little venture capital, new product development tends to be totally market driven. Although vendors sometimes follow their own ideas in deciding what to build and when to build it, more often they are driven by either state or federal mandates or by requests for proposal from prospective clients. When development fails, there is no return on investment; when development succeeds, return can be slow, as it has been for computerized physician order entry (CPOE) systems. To date, the vendor community has not seen a market of sufficient size or financial reward to justify the expenditures needed to develop systems that support both clinical care and clinical research. Estimates as to the development costs vary, from $10 to $20 million up to $50 million. The vendor with the more costly estimate explained that once a system is demonstrable, the amount it will cost to put in production and the charge to customers is much easier to estimate; moreover, it is possible to obtain customer contractual commitments to cover the additional costs.

Clients are driven by the need to maintain and to improve their daily operations, including direct patient care, and to respond to outside requirements. The latter include utilization and quality measures set by government agencies such as the Centers for Medicare and Medicaid Services, by insurance company and/or managed care organizations, and by the Joint Commission on Accreditation of Healthcare Organizations. Most clients want to improve operations without replacing their legacy systems. Today, however, they must do so in the absence of an accepted model for determining costs. Estimates range, for every $1 spent to purchase a system (hardware and software), a healthcare organization will spend between $2 to $10 in implementation. One chief information officer (CIO) at a major academic medical center calculates that software costs can account for 15% of overall implementation costs, making the figure $7 for implementation and hardware for every $1 spent on software.

An alternate scenario for computing costs, according to the same CIO, is to envision that an organization with a $100 million clinical research budget might spend $3 million to acquire and implement a system to support that research. Amortized over five years, this capital investment would be $600,000 per year, or 0.6% of the clinical research budget. He postulates that this is not an unreasonable portion of the clinical research budget to spend on IT, although it does not include operational costs, such as IT staff. Given limited capital and the absence of strong products, $600,000 is difficult to get because there are no offsetting costs or revenue gains to help the return on investment. Moreover, the clinical research community at most academic medical centers is fragmented and politically impotent, meaning that there is no one at the capital budget table to argue forcefully for these investments.

The fact remains that most software development projects entail a business case where a positive return on investment is gained—and such returns are rarely guaranteed. Both electronic medical records (EMRs) and CPOE are risky initiatives because their success depends on widespread adoption and because they impact physician workflow. If physicians and other health care workers are resistant to the new system, it will fail, no matter how "perfect" the design of the software.

There is some evidence that low risk, high benefit projects can foster adoption. For example, Web-based "portals" can save physicians precious time by giving them access to clinical information from outside the hospital. When such portals are incorporated into practice, they may foster a higher level of enthusiasm for subsequent, more complex, initiatives. External pressure can also bring about change, as the Leapfrog Group has done in the area of patient safety by insisting on CPOE. In like manner, HIPAA is forcing movement on standards for privacy and security in regard to electronic data interchange (EDI). These initiatives are aimed at institutional improvements through broad systemic change. Regrettably, there is no current source of focused and sustained pressure on behalf of the IT needs of the clinical research community and its many diverse components.

Progress to Date

As described at the outset of this chapter, progress to date has been "at the margin." In working to develop systems that meet their clients' needs, vendors have had to address problems outside their expertise. For example, one vendor (Cerner), like their competitors, has created a fully structured clinical note on a single screen at the bedside, eliminating the cost of dictation and increasing the richness of data collection. Development was laborious and entailed building a nomenclature model based on existing standards, creating a "clinical grammar," and developing structured templates that map back to a common data dictionary. The end result is an effective clinical documentation tool, a descriptive record with the data that the clinician chose to capture. The tool helps the physician render a diagnosis, place orders, and set care goals for downstream processing. However, the data are not necessarily what a researcher may need, and the system does not have internal algorithms to detect inconsistencies before extracting the data into a research database.

Most of the major vendors have developed similar clinical documentation capabilities. One critical problem is how to codify entry of clinical findings. In developing their systems, Cerner and several other vendors worked toward standardization by using the Unified Medical Language System (UMLS). They did this with input from the academic community, and at great expense and effort. Subsequently, the National Library of Medicine discouraged the use of UMLS for such purposes. Today, the issues around licensing and openness seem to be prompting a move back to the UMLS. The pace of product development by software vendors is ultimately hostage to the adoption of standards that carries the imprimatur of the federal government, as proposed by Panel 3. Software companies understand clinical information systems, control terminologies, and data warehouses. While they do understand controlled medical vocabularies or systems for patient care, they do not have experience in the vocabularies, terminology standards, and similar issues for supporting clinical research. There is minimal consistency in how information management is handled in clinical practice and clinical research. The problem here is terminology, and terminology is not a software vendor's problem. As an end user of clinical data, the research community should have some input to the deliberations on standards development, as to vendor product development.

Collaborations

Improving Medical Outcomes: Professional Associations and Vendors

The American Academy of Orthopaedic Surgeons and Eclipsys' SunClinical Data Institute are developing a national joint replacement registry that is similar to an operational model in Sweden. By tracking almost all joint replacements, the goal is to lower the rate of complications and the need for post-hoc revisions by identifying devices and operative procedures with lower "survival" rates. Orthopaedic replacement procedures are very complex, involving multiple device components, different operative procedures, and varying adjunct treatments. Very few institutions have a clinical information system that collects the necessary procedure details. The plan is to collect data on key items (e.g., device and lot numbers, operative approach, biologic adjuvants, etc.) using a wireless Web tool in the OR, thereby avoiding the need for hospitals to upgrade their general OR clinical systems to support the needs of the registry.

These vital details, perhaps a dozen pieces of data linked to general patient information, could make a substantial difference towards improving the health outcomes of patients receiving joint replacements.

An Academic Medical Center, a Vendor, and a Federal Agency
The University of Illinois at Chicago and Cerner, with funding from the AHRQ, are creating a repository of HIV genotype data. This early effort successfully integrates genotype information into the EMR and clinical data repository. When a physician in the HIV clinic orders an anti-retroviral drug, the system will check the order against all previous genotypic analyses of that particular patient's virus to determine whether or not the drug is appropriate. A sequencer from Visible Genetics produces the HIV viral gene sequence and sends that genotype, using an HL7 message, to the patient's EMR in the clinical system where it is stored. An estimated 60 or 70 rules will represent the knowledge captured by the sequencer. Decision support alerts will indicate that the patient is or may be resistant and give the clinician the opportunity to select an alternative drug.

Not yet in clinical production, the product will define what type of genetic information to store—the full sequence, the amino acid substitutions, or the interpretation of resistance generated by the sequencer. In this instance, HL7 is used to take the genetic information and place it in existing records. These messages pass through another vendor's system before arriving at the clinical data repository. Additional refinements in the HL7 process will create more specific ways to hold genetic information.

The Centers for Disease Control (CDC), Denver Health and Siemens Medical Solutions are collaborating to study the effectiveness of electronic patient records (EPR) and clinical decision support systems on tuberculosis (TB) screening and prevention in the ambulatory setting. Beginning with the patient registration process, "trigger" factors enable physicians to initiate the CDC latent TB screening guideline. Through public-private partnerships, Siemens is exploring the development of standard alogorithms to capture these data elements in their systems. These will assist researchers with surveillance for certain data elements that indicate a patient might meet the inclusion criteria for a particular clinical trial.

Vendor Initiated Federated Databases as a Model of Collaboration

As noted earlier in this report, the creation of federated databases is of particular concern to clinical researchers. Vendors have been creating federated databases using clinical data from clients at multiple institutions. Most vendor-initiated warehouse work involves anonymized data, with special exceptions for public health requirements and for the patient-consented systems and sample tracking used in pharmaceutical research. With regard to the latter, vendors are focusing on their ability to support multi-institutional trials that involve academic and non-academic sites. Such efforts allow the vendors to use data from existing clients to populate databases and create a new revenue source as well.

Health Facts. Health Facts, a federated database created by Cerner, pools data from about 30 of that vendor's clients in a comparative health outcomes research database. It now has data on about 1.5 million hospital inpatients, 3.5 million admissions, 7 million medication orders, and 3,200 brand name drugs with NDC coded information about the dose, route and frequency and so forth. Compiled primarily from billing data, from clients that may not have comprehensive electronic medical records, Health Facts extracts the limited data set demographics allowed under HIPAA. To the extent possible, data are aggregated within the individual institutions, where the use of identifiers is allowed, before the data are uploaded over a secure link to the database where they are de-identified or encrypted. This allows individual clients to drill in and decrypt the medical record numbers of patients from their institution while it also protects data from other institutions.

As part of this effort, Cerner is working with a subset of clients who have agreed to share their data. These "smart clients" are basically those who are implementing clinical systems most aggressively; they benefit through closed loop analysis. Health Facts and the institute are both funded by the sale of de-identified data to pharmaceutical companies, which use the databases to investigate prescribing patterns and side effects of medications, including their duration and frequency of administration. Useful information can be extracted from what might be considered fairly shallow data. In this instance, it is possible to answer certain questions without complete data.

SunClinical Data Institute. A new division at Eclipsys, the SunClinical Data Institute is pooling financial, transcription, and clinical data from hospitals that agree to participate in the Institute as members. It organizes these data into a standardized format and creates three databases: an encrypted analytic file for use by the member that has not had data previously linked from its different systems; a national benchmarking file; and a large anonymous file for use by academic scientists and commercial organizations. It is scheduled to deliver benchmarking capabilities to its customers in 2003, and will sell analytic products and services to pharmaceutical companies and device manufacturers as well. SunClinical is making a considerable investment in how to handle free text data, using natural language processing tools to extract and codify information, because in the short run most of the clinically interesting information is not codified. With daily data from most sources (and appropriate approvals from institutional review boards), cases can be flagged in near real time, allowing researchers to request supplemental data on a patient who may be a candidate for specialized data collection.

Public Health and Bioterrorism Surveillance

Creating longitudinal databases for public health surveillance without creating extra workflow for clinicians presents special challenges to vendors. Early detection and rapid response require the detection of patterns. From the standpoint of public health practice and research, data collection involves hospitals, clinics, nursing homes and other facilities that are often geographically disparate and unrelated in terms of data systems and vendor software. Doing this without redundant data entry and with many different reporting methodologies is not an easy task. Many institutions still lack an even rudimentary EMR.

Surveillance is based on known patterns and algorithms. Algorithms need to be standardized. For each vendor to develop algorithms independently is neither cost effective, nor functional for a surveillance system. Moreover, the method of computing the measurement should not be proprietary. Vendors should not be judged on their algorithms, but on how they handle the standards.

HealthSentry
A public health and bioterrorism surveillance system, HealthSentry was created after the events of 9/11/2001. The system uploads data from 22 microbiology laboratories in the Kansas City, Missouri area. Since March 2002, it has recorded one million encounters. The fact that the system went from nothing to deployment in 90 days illustrates that vendors can move rapidly with clearly defined goals. The system has dramatically improved the accuracy and timeliness of reporting, and the vendor Cerner is working with CDC to make the system compliant with the National Electronic Disease Surveillance System (NEDSS) developed by the Centers for Disease Control.

eHealth Initiative
To determine whether their clients could do public health surveillance without creating extra workflow for clinicians, Siemens collaborated with the private sector group, e*Health Initiative, to identify the data elements available from emergency department visits. They assessed geographic coverage, ease and speed of data collection, and current standards, all with the intent of identifying what answers the data could provide. The vendor looked at facilities known to have local repositories with interface engines that pull data from various systems, including their own and some from other vendors. This required taking multiple clinical information systems,and capturing, filtering, deidentifying, and encrypting the data.

Without a mandate from the Centers for Disease Control or other public health agencies, separate memoranda of understanding had to be negotiated with every facility, addressing legal and operational issues. Siemens established the feasibility of the model using bioterrorism surveillance at the Olympic Games in Salt Lake City by enabling the transmission of real-time data from existing systems. Within six weeks of the signing of the memoranda, connections were established among Utah hospitals, including those with other vendors' systems, to collect syndromic information using HL7. The data pulled from these systems were electronically transmitted to an analysis system, which had algorithms to detect patterns. The vendor did not create these algorithms, but sent the data for analysis to the University of Pittsburgh. Since that time, Siemens has entered into an agreement with the Pennsylvania Department of Health which will link up to 225 hospital emergency departments across the Commonwealth of Pennsylvania for the purpose of bioterrorism surveillance. The project demonstrated the ability to aggregate and analyze data, taken from emergency department records, namely the chief complaints and triage information compiled for billing purposes, from multiple vendor systems. Concerns about bioterrorism, and the efficacy of our public health surveillance systems could serve as a stimulus to strengthen the information system capacity to serve public health practice and research, and drawing in the interests of vendors in developing the necessary products.

Government

Governmental agencies are responsible for a wide range of activities: acting as purchaser of health care, supporting infrastructure development, and funding targeted programs in areas such as bioterrorism, patient safety, and quality improvement. At this juncture, government efforts are largely disconnected and unfocused. Grant mechanisms within funding agencies are not suitable for the required effort, and funding levels are too low. For example, NLM and AHRQ are able to fund private sector companies through the Small Business Innovative Research Program (SBIR). Because funding levels for the initial period cannot exceed $100,000, they are not really suitable to leverage the software vendors to develop products relevant to clinical research.

As in the description above of the efforts by vendors to address institutional IT needs, in the absence of an overarching federal program initiative, the successes that have been achieved with federal funding have had a relatively marginal effect. At the same time, the successes that have been achieved may suggest successful government approaches to systemic change.

Research in Clinical Informatics, and Training of Informaticians (NLM)

For several decades, the National Library of Medicine at NIH has been the principal source of funding in the U.S. for research into the application of information systems and telecommunications to health care delivery and clinical research. NLM has also pioneered in initiating and supporting training programs for the development of informaticians capable of carrying out research in informatics and in applying informatics to biomedicine. Eighteen such training programs are currently supported across the country. These programs have produced the majority of the informaticians in American academic informatics departments.

Leveraging Its Position as a Large Purchaser of Health Care: The Centers for Medicare and Medicaid Services (CMS)

As a major purchaser of health care, CMS has influence over healthcare providers. As part of their Quality Improvement Organization (QIO) Program, CMS is fostering electronic connectivity to eliminate the need for costly chart abstraction. CMS is working to create a broader, electronically-based data set that can give better (and more) information for quality improvement and research. This is very good news for health services researchers who have been highly reliant on the Medicare database.

The QIO Program is being funded at approximately $1 billion over a three-year contracting cycle, with QIOs in each state. Currently, the QIOs are focusing on quality measures in four clinical areas: hospital, physician office, home health, and nursing homes. For the first two, data remain confidential; for the second two, data are publicly reported. In late 2002, CMS published “Nursing Home Compare” on www.medicare.gov. This site provides information on 10 different quality measures.

The data collection that formed the basis for the quality measures was a requirement set by CMS as a condition for payment for services. In the six states where this public reporting

activity was piloted, nursing home requests for technical assistance from CMS-funded QIOs increased tenfold. Approximately half of the facilities in each of the pilot states requested assistance in such areas as transmissions standards and clinical vocabularies, affirming their commitment to the program.

For physician offices, the QIOs have collected state-level data on mammography screening, required immunizations, and selected measures on diabetes screening and treatment for which claims data were available. The goal now is to develop a richer data set, using data electronically available in ambulatory practice. CMS and other stakeholder organizations, including the AMA, are collaborating to develop a set of physician office measures that go beyond what is available from claims data—or constrained by costly medical record abstraction.

Through such activities, the QIOs have stimulated the adoption of a functional IT infrastructure and the adoption of data standards and clinical vocabularies. Providers and practitioners in the pilot program have grown to expect the functionalities offered by computerized physician order entry (CPOE), electronic medical records, electronic prescribing, and case registries. The latter two services require IT connections to the outside world.

Practitioners and providers base purchase decisions on the return on investment. This may involve financial incentives and differential reimbursement under the statutory mandate for CMS. Differential reimbursement is attractive because it gets the resources to the providers, including health insurance and managed care. By providing this type of financial incentive, the government can foster investment in and use of IT systems to provide clinical care. In addition, the government can expedite the processing and payment of claims for organizations providing industry-standard messages. Quality incentives can include making the case through research, demonstration, and dissemination that IT can improve safety, outcomes, and quality of care. The QIO program has influenced investment decisions in the adoption of standards and in IT infrastructure through these direct and indirect positive and negative incentives.

CMS characterizes these activities as "pull strategies" for coordinating practitioner and provider purchases around functional expectations. For clinical researchers, the CMS initiative promises access to richer datasets, as well as technical assistance on the types of data that are available and the best ways to access it. It would be optimal, at this stage, if CMS could engage the input of health services researchers as it upgrades its IT systems and requirements.

Targeting Patient Safety and Outcomes: The Agency for Healthcare Research and Quality (AHRQ)

Medical informatics is one of many important components necessary to improve the quality of health care in the United States. Research funded by the Agency for Healthcare Research and Quality (AHRQ) over the past 30 years has given clinicians solutions to many problems they face in patient care, and provided the basis for several clinical research data repositories. AHRQ remains committed to supporting new and innovative technologies, information/knowledge management, improved communication between patients and providers, shared decision

making, identifying and overcoming barriers to the use of computers in health care, and new challenges posed by the threat of bioterrorism.

National Patient Safety Database. The Agency for Healthcare Research and Quality (AHRQ) is the "Lead Agency" for the federal government on national patient safety. The effort involves developing the infrastructure and incentives for collecting data on medical errors, "near misses," and adverse events, and building a data repository of patient safety incidents for analysis. This requires a standards-based taxonomy for medical errors, as well as clinical vocabulary and reference terminology that relate to existing codes and common medical terms.

Outcomes Research. AHRQ is also the lead agency for outcomes research. The goal of AHRQ-funded studies is to identify interventions that improve outcomes, interventions that do not improve outcomes, and to encourage wider application of effective interventions while discouraging the use of ineffective treatments. The focus of the research is on measurable and sustainable impacts: overall mortality, disease-specific mortality, hospitalizations, and disability. It is not on biochemical markers, physiologic parameters, and other surrogate markers that may or may not correlate well with clinical or quality of life end points. Outcomes research should lead to new knowledge and interventions that improve patients' lives in meaningful ways. This requires access to a wider range of clinical data than is generally available today. AHRQ has long supported public-private initiatives to improve quality of care and address the infrastructure needs of the healthcare community as demonstrated by AHRQ-sponsored projects with Siemens and Cerner.

Targeting Bioterrorism Surveillance and Public Health Practice: The Centers for Disease Control (CDC)

The Centers for Disease Control (CDC) is using electronic data in its surveillance, tracking, and planning activities. The centerpiece of CDC's current programs is the Public Health Information Network (PHIN), a real-time interoperable network for data and communication exchange based on national and industry standards for interoperability. The network will function to detect what is happening and communicate it, as appropriate, to CDC's partners in health care, emergency management, law enforcement, and the public. Thus, these systems are designed to be highly secure while permitting bi-directional communication. To this end, CDC has developed a messaging tool for secure transport using HL7 standards. It will be independently certified and available for use by CDC's partners for point-to-point messaging. Technical assistance will be provided to those partners as well. CDC is working with the FDA to develop an HL7 message that can be used for reporting notifiable conditions to CDC or adverse event information from clinical providers to the FDA. When fully operational, the system will offer providers a highly secure, authenticated, appropriate tool for the exchange of sensitive information.

The National Electronic Disease Surveillance System (NEDSS). CDC is using NEDSS with state and local health partners to get the surveillance data needed for health tracking. NEDSS is replacing the cards traditionally filled out by physicians and sent to the health department with a Web

interface for entering this kind of information. This also eliminates the hundreds of different forms that were previously required. NEDSS also accepts automatic standard HL7 messages from clinical laboratories, most of which are required by law to report conditions of public health importance. In pilot projects, NEDSS has had a dramatic impact on the efficacy and efficiency of surveillance: the number of cases reported more than doubled, and the reports were made much more rapidly and with more complete information. Follow-up investigations were expedited by the quality and speed of the reporting system, transforming the practice of public health. Building on its ability to diminish the burden on its clinical partners, CDC is currently working with multi-jurisdictional national laboratories to implement NEDSS. At this time, CDC plans to work with a broader range of clinical data and to use the upgraded message format when it becomes available.

In addition to providing high-level security compatible with what has been proposed for HIPAA, NEDSS is built on an open systems platform concept that can readily accommodate commercial products; vendor-developed applications are now functioning. CDC has developed a NEDSS compatible solution for use at state health departments; 22 states have been funded to deploy this system, and deployment of version 1.0 started in October 2002. All 50 states have received federal funding for NEDSS as part of the public health preparedness effort.

Ultimately, NEDSS can link clinical sites—including hospitals, providers, laboratories, pharmacies, etc.—with each other and with public health authorities for early detection of outbreaks and for monitoring the effectiveness of interventions. A specific goal has been to reduce the amount of time spent in data collection and reporting so that more time will be available for direct intervention to prevent and control disease. As a mechanism for sharing population-based data, NEDSS can greatly augment the ability to monitor and respond to emerging health problems on a local, regional, or national scale. Recent experience with West Nile Virus infection, the development of bioterrorism detection and response capacity, contaminated tissue allografts, and the spread of antimicrobial resistant microorganisms all demonstrate the extraordinary promise of the full deployment of NEDSS.

Public Health Preparedness. The ability to detect an unannounced bioterrorist event is critically dependent upon access to data available in the clinical care sector, as in emergency departments. Optimally, such data must be available in real time. Electronic data reporting is critical for preparedness. There is no alternative to a national electronic database linking clinical care and public health systems. It complements alert clinicians, backstops the system, and eases the burden of public health reporting requirements. Over the past year, CDC has been working to accelerate the use of data already in electronic form by implementing standard messages using HL7. With the use of standards-based HL7 messages, health system providers can provide reports of public health threats to their state health department or CDC (when appropriate), reports of serious adverse events to FDA, required documentation to the Joint Commission on Accreditation of Health Care Organizations, as well as data to research repositories. CDC hopes that, by participating in open standards consortia, they can help to improve existing standards—not to create standards de novo.

Initiating Research Infrastructure Development

With a budget of just under $1.2 billion, the National Center for Research Resources (NCRR) is responsible for developing and sustaining the infrastructure for health-related research. That includes a network of GCRCs which host nearly 11,000 investigators who conduct about 7,500 projects annually. The General Clinical Research Centers (GCRCs) comprise a national network of 80 centers that provide optimal settings for medical investigators to conduct safe, controlled, state-of-the-art, in-patient and out-patient studies of both children and adults. A GCRC grant supports the components essential to clinical research: operating expenditures, hospitalization and ancillary laboratory costs, and salaries of key personnel including nurses, research bionutritionists, administrators, core laboratory staff, biostatisticians, and computer personnel. Funds for renovation and equipment also may be provided.

The GCRCs provide locally competitive support for pilot projects to be conducted with GCRC resources. Separately, NCRR provides support for outpatient-based clinical research centers (CRCs), located at six minority-serving medical schools through the Research Centers in Minority Institutions program. All GCRC sites are provided funding for biostatisticians and all but 1-2 sites have an NCRR-funded systems manager to maintain software systems and to assist clinical investigators in the use of IT applications. In that regard, the Panel learned of the major effort underway at the NIH Clinical Center to develop a CIS that would serve the needs of clinical investigators across Institutes. If this technology development could be leveraged by NCRR and shared with institutions housing a GCRC, the NCRR could substantially strengthen its funding for IT support in the GCRC program. Alternatively, both the NIH intramural program and academic centers might gain from the substantial investment that the Department of Veterans Affairs has made in its CIS system which was designed to serve the requirements of quality health care, and of VA-funded clinical research.

The NCRR Division of Biomedical Technology supports several national resources that are part of the spectrum of research technologies that complement the infrastructure provided through GCRCs and RCMI-supported CRCs. Those includes resources for bioinformatics, scalable computing up to the teraflop level, imaging technologies and many more. To address the rapidly evolving need of investigators for bioinformatics tools and access to a wide array of technologies, NCRR has undertaken the building of the Biomedical Informatics Research Network (BIRN). The BIRN involves several partners, including the National Science Foundation (NSF), the San Diego Supercomputer Center, several universities, and Internet2. It began by creating a test bed for neuroscientists, who are facile in using large databases and manipulating data.

To date NCRR has invested $20 to $25 million in designing and building BIRN in close collaboration with end users. By the end of the first year, BIRN had over 100 investigators involved, collaborating and sharing data. The goal is to develop an intertwined, interactive national network that links networked high performance computing resources and bioinformatics tools with NCRR's network of approximately 200 research resources and repositories. Additional IT tools are to be developed along with support networks to facilitate multisite collaborations,

clinical trials and remote access to technologies and resources. To this end, NCRR is interacting with institutes within NIH to provide access to virtual laboratories and resources and databases, such as those offered by the National Cancer Institute. Needs identified through interactions with investigators drive the resources placed into these dual networks.

NCRR is especially proud of its collaboration with the Cystic Fibrosis Foundation on a clinical trials network to help get novel therapies for rare diseases being developed by biotechnology companies more quickly to market. The Web-based data entry system they developed allows for rapid accrual of data at ten sites across the country and increasing access to larger numbers of patients with cystic fibrosis. Within ten days to two weeks after completing a phase 1 or 2 trial, data are ready to go to the FDA, a remarkably speedy turnaround. Upon learning about this network, pharmaceutical and biotech companies asked the Foundation to start putting their drugs and devices through the system. For small companies especially, time is money. Thus, the network provides an incentive for them to become involved in the development of innovative drugs or devices. This effort has been extended to many more sites and now includes phase 3 trials.

A major initiative at NCRR and NIH will be to educate investigators on this new approach to multi-site collaboration through high-end computers and high-speed networks. These tools allow investigators to maintain their own data, while data mining techniques make it possible to treat all the data as one database.

Model for Collaboration: The SAGE Project

The SAGE (Sharable Active Guideline Environment) Project is multi-site, collaborative, academic-industrial research and development (R&D) partnership that offers a model demonstrating how multiple entities can work together to define information standards for health care. The goal of this three-year, $18M project is to develop a comprehensive technology infrastructure that will enable encoding and dissemination of interoperable, computable medical knowledge in the form of clinical practice guidelines.

Project Origins and Funding

The SAGE research agenda was conceived at the November 2000 Symposium of the American Medical Informatics Association and enabled nine months later by a $9.5 million competitive award from the National Institute of Standards (NIST) Advanced Technology Program (ATP). The NIST/ATP program, under the auspices of the Department of Commerce, has a unique mission of encouraging high-risk technology R&D with significant potential for benefits on a national scale. The industry partners in the project are IDX, a leading vendor of advanced healthcare information systems, and Apelon, a software company that deals with controlled medical vocabularies. The academic healthcare partners include Intermountain Health Care, the Mayo Clinic, the University of Nebraska, and Stanford University Medical Center.

SAGE Objectives and Approach

The output of SAGE will be threefold. First, a guideline model developed in close collaboration

with HL7 will provide the capability to take a guideline, encode it, and place it on a file that can be distributed via a Website or on a CD. Second, a guideline authoring/encoding workbench produced by Stanford's Medical Informatics Department will be available as open source software at minimal cost. (This workbench was developed using Protégé, an open source Java tool used to create customized knowledge base tools.) Third, a guideline deployment system now under development will be put into the public domain, allowing the guidelines to be readily distributed across organizations. Currently, the project team is focusing most of their energy on the deployment system; according to project leaders, this is "the really hard part."

The SAGE Project is an R&D consortium aiming at building an infrastructure for the distribution of clinical guidelines. It is not in the business of creating clinical guidelines content. By analogy, the SAGE Project is more akin to the developers of the CD player and not to the music companies that record music on the CDs.

To develop interoperable guidelines, SAGE is using standard information models and standard medical terminology that allow the guidelines to be encoded. The idea here is to enable guidelines that are not simply static text or PDF files, but executable objects to be distributed among heterogeneous clinical systems. The guideline logic will be integrated into the clinician's workflow via functions of the local clinical information system.

It explicitly recognizes that clinical guidelines, however they are represented, are distinct from clinical trials protocols. The organizers of the SAGE Project also recognized that there might be meaningful overlap between clinical guidelines and clinical trials protocols, and they plan to address the opportunities in the latter area over time.

Implications for Health Care

The SAGE Project is solving technological and organizational problems in order to create infrastructure to make a distribution infrastructure and market for computable guidelines. The notion is that someone somehow somewhere—perhaps a private organization—will use this technology, and create a national guideline clearinghouse. With the clearinghouse in place, a healthcare delivery organization will be able to use its own local copy of the workbench, download the guideline file, import the guideline, and localize it by making any required change.

SAGE Project: Lessons Learned

Initiating the SAGE Project required:

- Identifying the necessary blend of expertise in guideline modeling science, applied medical informatics, and clinical quality improvement.
- Building new and complex business relations on which to found the R&D consortium.
- Obtaining a major commitment from the partners involved.
- Seeking out and obtaining significant funding from federal sources.

Collaboration in the project is focused by aligning the interests and expertise of each partner with an R&D methodology that combines requirements-driven development with iterative cycles of prototyping. Expertise from each partner is channeled via coordinated workgroups, staffed from members across the partnership, and integrated via a central leadership team. SAGE may represent a template for optimal collaboration among government, industry and academia. If it proves successful, and the developers of clinical guidelines and healthcare systems adopt the new technology, then we may have a clear idea of the elements necessary for *the* major collaborative effort that will be necessary to tackle the formidable challenges and barriers to systemic change that would bring EMR and the clinical research applications of IT into the 21st century.

Conclusion

It is technologically possible to incorporate IT enhancements for clinical research into advanced clinical information systems (CIS) and other high-priority applications developed to support clinical care. Optimal advanced CIS can enable both researchers and clinicians to do their jobs easier, faster, and better.

Although the characteristics of a well-developed information system depend on what the users want it to do, current technology permits a wide variety of reasonable solutions. Clinical research IT requires the creation of institutional data warehouses and repositories to hold patient-focused data, in which data derived from clinical care is stored for research. The rapidly increasing use of genomic information in clinical medicine greatly increases the demands on an information system, but concurrently makes more imperative the need for an efficient CIS if such data is to be managed efficiently and confidentially. To support clinical researchers, CIS must permit computer-based comparisons among patients.

The CIS must represent patient data in the form of coded information that is descriptive, non-ambiguous, clear, and reproducible across users and systems. Observational data can serve as the evidence base for new medical knowledge, helping in the construction of clinical guidelines and expert systems. As a structured source of patient presentations and recorded outcomes, electronic medical records (EMRs) can help meet the needs of clinical research, and improved quality and safety of patient care. An optimal clinical information system for patient care and clinical research is dependent upon the adoption of common standards and terminologies across health care and upon the development of off-the-shelf solutions that can be readily adapted at the local level. The federal government should be playing a major leadership role in endorsing the implementation of standards, and assuring that funding is available to enable institutions to bring their CIS into compliance with the standards.

To date, whatever progress has been made has involved collaborative projects involving one or more vendors, government agencies, and academic medical centers or clinical academic societies. These projects have occurred at the margin, and produced marginal successes related

to a limited set of goals. While the separate projects certainly affirm the possibility of collaboration, they have not begun to approach the scope or scale necessary to bring about incremental change. If we are to achieve real systemic progress, the government will need to re-think its traditional approaches to funding clinical IT and related clinical research enhancements through existing grant mechanisms at health-related agencies. "Big ticket" projects that are critical leverage points, and are too expensive to be funded by any one source or existing grant mechanisms, need to be identified, with the recognition that in the absence of a major new federal investment in IT infrastructure to support healthcare systems (including clinical research) the country is not prepared for 21st century medicine.

At a similar transition point near the middle of the last century, the Hill-Burton Act enabled communities across the country to modernize their non-profit hospitals to prepare for developments in the decades following the discovery of antibiotics. Because of the multiplicity of players involved in CIS development and implementation, new models of government/industry/academic collaboration should be considered, to reach the promise to improve patient safety, public health and bioterrorism surveillance, have cost-effective patient care, and clinical research that can only come from the application of advanced information technology to the healthcare system.

Glossary of Acronyms

21 CRF 11	Title 21 Code of Federal Regulations Part 11
AAMC	Association of American Medical Colleges
AHRQ	Agency for Healthcare Research and Quality
AMA	American Medical Association
AMIA	American Medical Informatics Association
ASC	Accredited Standards Committee
AVC	Associate vice chancellor
BIRN	Bioinformatics Research Network
CAP	College of American Pathologist
CD	Compact disk
CDC	Centers for Disease Control and Prevention
CDISC	Clinical Data Interchange Standards Consortium
CHR	Center for Health Research
CIO	Chief information officer
CIS	Clinical information systems
CMS	Centers for Medicare and Medicaid Services
CML	Chronic myelogenous leukemia
CPOE	Computerized physician order entry (alternately, care provider order entry)
CPT	Current Procedural Terminology
CRO	Contract research organization
CRR	Clinical Research Roundtable
DARPA	Defense Advanced Research Projects Agency
DbSNP	Database for single nucleotide polymorphisms
DICOM	Digital Imaging and Communications in Medicine
DoD	Department of Defense
DRG	Diagnostic related groupings
Dx	Diagnosis
EDI	Electronic data interchange
EMR	Electronic medical record
FDA	Food and Drug Administration
GCRC	General Clinical Research Center
HDR	Health data repository
HIPAA	Health Insurance Portability and Accountability Act of 1996
HL7	Health Level Seven
HUGO	International Human Genome Organisation
I3C	Interoperable Informatics Infrastructure Consortium
ICD	International Classification of Diseases
ICD-9-CM	ICD, 9th Revision, Clinical Modifications
ICU	Intensive care unit
IHC	Intermountain Health Care
IOM	Institute of Medicine

IRB	Institutional Review Board
IT	Information technology
IV	Intravenous
JCAHO	Joint Commission on Accreditation of Healthcare Organizations
KPNW	Kaiser Permanente Northwest
LOINC	Logical Observation Identifier Names and Codes
MeDRA	Medical Dictionary for Regulatory Activities
NCI	National Cancer Institute
NCPDP	National Council for Prescription Drug Program
NCRR	National Center for Research Resources
NDA	New Drug Application
NDC	National Drug Code
NEDSS	National Electronic Disease Surveillance System
NHII	National Health Infrastructure Initiative
NIH	National Institutes of Health
NIST/ATP	National Institute for Standards and Technology/Advanced Technology Program
NLM	National Library of Medicine
NSF	National Science Foundation
OHT	Open Health Technology
OLAP	Online analytic processing
OR	Operating room
PHI	Protected Health Information
PHIN	Public Health Information Network
POR	Patient oriented research
QIO	Quality Improvement Organization
R & D	Research and development
ROI	Return on investment
SAGE	Sharable Active Guideline Environment
SBIR	Small Business Innovative Research
SEER	Surveillance, Epidemiology, and End Results
SNOMED	Systematized Nomenclature of Medicine
SNP	Single nucleotide polymorphisms
TPO	Treatment, payment, and healthcare operations
TURP	Transurethral prostatectomy
UCUM	Unified Code for Units of Management
UMLS	Unified Medical Language System
VA	Veterans Administration
VHA	Veterans Health Administration
VistA	Veterans Information Systems and Technology Architecture
WD	Working days
XML	Extensible Markup Language