

NCBI Bookshelf. A service of the National Library of Medicine, National Institutes of Health.

Institute of Medicine (US) Committee on Data Standards for Patient Safety; Aspden P, Corrigan JM, Wolcott J, et al., editors. Patient Safety: Achieving a New Standard for Care. Washington (DC): National Academies Press (US); 2004.

4 Health Care Data Standards

CHAPTER SUMMARY

Data standards are the principal informatics component necessary for information flow through the national health information infrastructure. With common standards, clinical and patient safety systems can share an integrated information infrastructure whereby data are collected and reused for multiple purposes to meet more efficiently the broad scope of data collection and reporting requirements. Common data standards also support effective assimilation of new knowledge into decision support tools, such as an alert of a new drug contraindication, and refinements to the care process. This chapter provides both a short overview introducing data standards to the lay reader and a more technical review of the specific data standards required for the informatics-oriented professional. Please note that in the technical portion of the paper, once a standard is introduced it will be referred to in its acronym form due to the number of data standards involved. Readers may refer to the list of acronyms in Appendix B for assistance as needed.

OVERVIEW OF HEALTH CARE DATA STANDARDS

Although much of the data needed for clinical care, patient safety, and quality improvement resides on computers, there is as yet no means to transfer these data easily and economically from one computer to another, despite the availability of the communications technologies to support such data exchange. The chief obstacle to achieving this capability has been the haphazard adoption of data standards for organizing, representing, and encoding clinical information so that the data can be understood and accepted by the receiving systems (Hammond, 2002). At the level of the health care organization, the lack of common data standards has prevented information sharing between commercial clinical laboratories and health care facilities, between pharmacies and health care providers regarding prescriptions, and between health care organizations and payers for reimbursement (Hammond, 2002). The lack of standards has also prevented the reuse of clinical data to meet the broad range of patient safety and quality reporting requirements, shown in Table 4-1. The first column of this table lists the data sources often associated with an electronic health record (EHR); the second, those associated with clinical information systems, decision support tools, and external data sources; the third, state, regulatory, and private-sector patient safety reporting systems; and the fourth, federal reporting systems. The fact that there is no standard means of representing the data for any of these datasets or requirements is astonishing and highlights the amount of unnecessary work performed by health care and regulatory organizations to prepare, transmit, and use what amount to custom reports. The federal government has recognized this problem and is moving forward with the integration of its safety-related systems. This study goes further by recommending common standards for the clinical and patient safety data that span the full range of data sources listed in Table 4-1. Many of the data standards required are already available; others need further development.

TABLE 4-1

Comprehensive List of Health Care Data Sources and Reporting

What Are Data Standards?

In the context of health care, the term *data standards* encompasses methods, protocols, terminologies, and specifications for the collection, exchange, storage, and retrieval of information associated with health care applications, including medical records, medications, radiological images, payment and reimbursement, medical devices and monitoring systems, and administrative processes (Washington Publishing Company, 1998). Standardizing health care data involves the following:

- *Definition of data elements*—determination of the data content to be collected and exchanged.
- *Data interchange formats*—standard formats for electronically encoding the data elements (including sequencing and error handling) (Hammond, 2002). Interchange standards can also include document architectures for structuring data elements as they are exchanged and information models that define the relationships among data elements in a message.
- *Terminologies*—the medical terms and concepts used to describe, classify, and code the data elements and data expression languages and syntax that describe the relationships among the terms/concepts.
- *Knowledge Representation*—standard methods for electronically representing medical literature, clinical guidelines, and the like for decision support.

At the most basic level, data standards are about the standardization of data elements: (1) defining what to collect, (2) deciding how to represent what is collected (by designating data types or terminologies), and (3) determining how to encode the data for transmission. The first two points apply to both paper-based and computer-based systems; for example, a laboratory test report will have the same data elements whether paper or electronic. A data element is considered the basic unit of information, having a unique meaning and subcategories of distinct units or values (van Bommel and Musen, 1997). In computer terms, data elements are objects that can be collected, used, and/or stored in clinical information systems and application programs, such as patient name, gender, and ethnicity; diagnosis; primary care provider; laboratory results; date of each encounter; and each medication. Data elements of specific clinical information, such as blood glucose level or cholesterol level, can be grouped together to form datasets for measuring outcomes, evaluating quality of care, and reporting on patient safety events.

Associated with data elements are data types that define their form. Simple data types include date, time, numeric, currency, or coded elements that rely on terminologies (Hammond, 2002). Examples of complex data types are names (a structure for names) and addresses. For comparability and interchange, data types must be universal and must be carried through all uses of the data. The designation of common scientific units is also necessary. Units (e.g., kilograms, pounds) must be specified as another measure to prevent adverse events such as those related to dosing errors. Until recently, each institution or organization defined independently the data it wished to collect and the units employed, did not use data types, and created local vocabularies, resulting in fragmentation that prevented reuse.

For data elements that rely on terminologies and their codes for definition, merely referencing a terminology alone does not provide enough specificity. To ensure data comparability, specific codes must be identified within each terminology set to represent the data elements. This becomes a major issue for some of the larger clinical terminologies, which may have hundreds or thousands of terms. It is also a major issue given the amount of data that must be collected for the data sources and requirements listed in Table 4-1 and that will be encompassed by the national health information infrastructure (NHII). Common data standards are essential to simplify and streamline data requirements and allow the information systems that carry the data to function as an integrated enterprise.

TECHNICAL REVIEW OF HEALTH CARE DATA STANDARDS¹

This section provides a detailed technical review of the three primary areas in which standards for health care data need to be developed: data interchange, terminologies, and knowledge representation. The final subsection addresses the implementation of these data standards.

Data Interchange Standards

In the area of data interchange, standards are needed for message format, document architecture, clinical templates, user interface, and patient data linkage.

Message Format Standards

Message format standards facilitate interoperability through the use of common encoding specifications, information models for defining relationships between data elements, document architectures, and clinical templates for structuring data as they are exchanged. In March 2003, the Consolidated Health Informatics (CHI) initiative announced its requirement that all federal health care services agencies adopt the primary clinical messaging format standards (i.e., the Health Level Seven [HL7] Version 2.x [V2.x] series for clinical data messaging, Digital Imaging and Communications in Medicine [DICOM] for medical images, National Council for Prescription Drug Programs [NCPDP] Script for retail pharmacy messaging, Institute of Electrical and Electronics Engineers [IEEE] standards for medical devices, and Logical Observation Identifiers, Names and Codes [LOINC] for reporting of laboratory results) (Office of Management and Budget, 2003). It is worth noting that HL7, through its Laboratory Automated Point-of-Care Special Interest Group, has also developed messaging standards for the devices used in laboratory automation (e.g., robots and laboratory instruments) and point-of-care test devices (e.g., blood glucose monitors). These standards are in the process of being incorporated into IEEE standards and eventually will become standards at the International Organization for Standardization (ISO).

The HL7 V2.x series is the primary data interchange standard for clinical messaging and is presently adopted in 90 percent of large hospitals (American National Standards Institute, 2002). However, there have been a number of technical problems with the standard that have been difficult to resolve. For one, “conditional optionality” was built into the framework such that it permits a number of terminologies to represent a data element (e.g., Systemized Nomenclature of Human and Veterinary Medicine [SNOMED], LOINC) without being precise about the specific codes (i.e., allowable values) within the terminology (Hammond, 2002). The “openness” of the optionality has led to discrepancies in the application of the standard and misunderstanding of the specifications due to different vendor information models. Also, although V2.x does not support Web-based protocols (e.g., Simple Object Access Protocol [SOAP]), it can be sent over

the Internet and expressed as an extensible markup language (XML) syntax standard (Hammond, 2002). However, V2.x does not incorporate an information model that is needed for more advanced messaging of clinical information.

Resolving these issues would be time consuming and labor intensive and could easily be accomplished by completion and implementation of HL7 Version 3.0 (V3), in which few of the data fields are open to interpretation. Currently, the scope of the V3 standard remains the same as that of V2.x; however, the initial release of V3 did not include the domains for patient referrals, patient care, or laboratory automation, all of which are important to patient safety (Health Level Seven, 2001). To move forward, the first step is to accelerate the completion of V3 and develop implementation guides for the use of both V2.x and V3 and interoperability between the two standards, with clear definition of the standard specifications. Both V2.x and V3 require a controlled terminology specified at the data element level to support interoperability.

Additionally, since V2.x will probably continue to be widely used for some time, it is important that any difficulties with interoperability between this standard and others be resolved. For example, V2.x is used for medication order messages in the inpatient and outpatient settings, while the NCPDP Script standard is used by retail pharmacies. Ensuring interoperability between these two standards is necessary for high-functioning systems.

HL7 V3 differs from V2.x in that it incorporates a Reference Information Model (RIM) for setting up the messaging format. The RIM is based on object-oriented modeling, expressing the classes of information required and the properties of those classes, including attributes, relationships, and states (Health Level Seven, 2001). The structured specifications of the information requirements minimize optionality by clearly defining each aspect of the RIM (Van Hentenryck, 2001). Data fields are populated with explicit controlled vocabulary, increasing semantic interoperability among various code sets (Van Hentenryck, 2001). Additionally, HL7 V3 messages are encoded using XML (as are Versions 2.4 and 2.5), which is easy to use, extensible, capable of representing complex data, and Internet compatible (Van Hentenryck, 2001). XML is used to exchange the data quickly and simply, but the RIM is needed to provide the necessary semantic interoperability.

At the core of the RIM are four high-level classes from which all other classes are derived—entity, role, participation, and act. Figure 4-1 is a simplified depiction of the structural relationships encompassed by the RIM that should aid in understanding the basis of the model.



FIGURE 4-1

HL7 reference information model. SOURCE: Hammond, 2002.

Information modeling facilitates recognition of high-risk procedures having a direct impact on patient safety (Russler, 2002). Both safe active patient care and retrospective analysis for a patient safety event depend on proper information relationships (Russler, 2002). To this end, the information model must facilitate the process of care such that the link from entities to their intentional actions can support the information relationships used in analyzing patient safety issues, as well as larger issues of cost and quality improvement (Russler, 2002). Using an analogy from aviation, examination of the link between a precipitating event and an adverse event is as important as comparing the data from a flight data recorder with the data from the voice recording in the cockpit in the case of an airline accident (Russler, 2002).

HL7 V3 and the RIM are particularly important to the advancement of integrated clinical systems because they provide the backbone for the next set of standards needed for the EHR including those required for the use of concept-oriented terminologies, document architectures, clinical templates, alerts and reminders, and automated clinical guidelines, all of which would result in improved interoperability and structuring of clinical and patient data.

Document Architecture

A method for representing electronically clinical data such as discharge summaries or progress notes and patient safety reports requires a standardized document architecture. This need stems from the desire to access the considerable content currently stored in free-text clinical notes and to enable comparison of content from documents created on information systems of widely varying characteristics (Dolin et al., 2001). The architecture should be designed as a markup standard (Dolin et al., 2001) so that clinical documents can be revised as needed or appended to existing documents. It should also be able to accommodate the desire for rich narrative text that makes up a significant portion of patient safety information from voluntary and mandatory reports.

One example is the HL7 Clinical Document Architecture (CDA), a defined, complete information object that can include text, images, sounds, and other multimedia content (Dolin et al., 2001). The CDA provides a hierarchical set of specifications for the structure of clinical documents and derives its semantic content from the RIM (Dolin et al., 2001). Initial specifications define the document header in detail (i.e., identifying document name, type, source, author, date–time, and the like, including an area for narrative text), while the document body is structured to represent narrative clinical notes. This structure minimizes technical barriers to the adoption of the standard in that it intentionally lacks some of the complex semantics used in HL7 V3 messages. The initial specifications lay the foundation for future specifications that will incorporate clinical templates and additional RIM-derived markup, enabling the clinical content to be expressed more formally to the extent that it can be encompassed fully in the RIM or V3 message (Dolin et al., 2001). Again, because both HL7 V2.x and V3 will be in use for the short term and midterm, implementation protocols should include the ability of systems to translate CDA documents to and from V2.x and V3.

Clinical Templates

HL7 V3 provides the mechanism to specify further constraints on the optionality of the data elements through the use of templates that can be applied against a V3 message or document. The HL7 V3 messages maintain moderate optionality, although the RIM provides some constraints. For greater precision in standardization of clinical data, more targeted specifications of the allowable values for the data elements must be applied. A template in the broadest sense is simply a constraint on a more generic model that permits, among other things, the definition of a complex object, such as a blood chemistry measurement or a heart murmur (Hammond, 2002). For example, an HL7 message format for laboratory observations may specify that the data elements for a complete blood count test must include measurements for hemoglobin, hematocrit, and platelets. The design of constraints will be left to the discretion of health care organizations and providers, as HL7 provides the mechanisms and technical specifications for their use. Clinical templates will be important in the development of electronic structure for the collection and analysis of clinical and patient safety data, particularly those related to the 20 priority areas identified by the Institute of Medicine (2003).

User Interface

The medical device industry is well versed in developing user interfaces that make devices safer, more effective, and easier to use by employing a voluntary standard for human factors design established by the Association for the Advancement of Medical Instrumentation (AAMI) and approved by the American National Standards Institute (ANSI) (Association for the Advancement of Medical Instrumentation, 2001). This standard—the ANSI/ AAMI HE74 Human Factors Design Process for Medical Devices—establishes tools and techniques to support the analysis, design, testing, and evaluation of both simple and complex systems; these tools and techniques have been applied for many years in the engineering of consumer products, military applications, aviation equipment, and nuclear power systems. Consideration of the HE74 standard may provide insight into the processes employed for designing and developing user-friendly clinical information systems, including electronic patient safety reporting systems.

An overview of the human factors engineering process that governs HE74 is provided in Figure 4-2. The specific activities at each step in the cycle vary with the particular development effort (Association for the Advancement of Medical Instrumentation, 2001). The cycle in Figure 4-2 emphasizes the iterative nature of the development process, whereby the outcomes (i.e., outputs) of one step provide input to the next step, but also, as needed, the output of some steps feeds back to previous steps. Although entry into the cycle can begin at any step, involving users at the early stages of development is critical. Once user needs and the consequent concept for the device (system) have been well defined, it becomes possible to address the design criteria/requirements that define the operating conditions, user characteristics, functions, and potential hazards of the device/system. The hardware and software designers can then craft the necessary technical requirements and specifications. Next, structured evaluation of the resulting device/system can ensure that the design is technically sound (i.e., verification) and that it also meets the user's needs and intended uses (i.e., validation). The last step—implementation and deployment—is related to the manufacturing, marketing, sales, and regulatory aspects of the device/system, including postmarket surveillance and vigilance reporting that provide critical data on design strengths and shortcomings (Association for the Advancement of Medical Instrumentation, 2001). Whether this model or another is developed, the committee urges further research on the development of user interfaces for integrated systems.



FIGURE 4-2

Human factors engineering process governing the HE74 standard. SOURCE: Association for the Advancement of Medical Instrumentation, 2001.

HL7 has taken an approach to data integration at the visual level by way of the user interface. These applications and standards facilitate the integration of multiple independent applications from many different systems through interface standards (Van Hentenryck, 2001). An example is the ANSI-certified HL7 Context Manager standard, illustrated in Figure 4-3. The context manager standard establishes the primary architecture, a core set of data definitions, rules for application user interfaces, security specifications, and translation of the architecture for interoperability with applications in a way that is technology neutral (Van Hentenryck, 2001). With the Context Manager, an organization can give providers a single sign-on capability so they do not have to log on for each separate clinical application program they need (Seliger, 2003). Likewise, the Context Manager provides a single patient selection that, similar to a single

clinician sign-on, allows all patient data in multiple applications to be readily available for use as needed by the clinician (Seliger, 2003). This context management gives users the experience of using a single system, when in fact they are accessing multiple applications simultaneously (Seliger and Royer, 2001).



FIGURE 4-3

Overview of the HL7 Context Manager. SOURCE: Seliger and Royer, 2001.

For example, to review patients of immediate importance, a physician might inspect a patient list in a scheduling application. To further the understanding of each patient, the physician might also wish to view laboratory test results via a laboratory application, view computed axial tomography (CAT) scans via a picture archiving and communications system (PACS) application, and order new tests or medications via an order entry application (Seliger and Royer, 2001). The physician's selection of patient Jane Smith via the scheduling application would cause the other applications to tune into the same. In this way, the laboratory application, the PACS application, and the order entry application would all be synchronized with the physician's clinical context, in this case, Jane Smith as the patient currently of interest (Seliger and Royer, 2001).

Along with identifying users and patients, the Context Manager can identify concepts for unifying the availability of clinical data across applications. Two concepts for which the Context Manager specifications have been developed are the clinical encounter and clinical observation (Seliger, 2003). Thus a clinician can use an encounter or observation identifier to access multiple applications for information related to that encounter or observation.

User interface standards such as the Context Manager provide a mechanism to begin the process of achieving interoperability at the level of the user interface. Data linkage allows the users to create tiles of the different applications and compose them into a single visual window utilizing the data currently held by the organization. Data are viewed in an integrated manner while the organization progressively builds a truly integrated, comprehensive clinical information system at the back end.

Patient Data Linkage

While not a data standard in the traditional sense, being able to link a patient's health care data from one departmental location or site to another unambiguously is important for maintaining the integrity of patient data and delivering safe care. The administrative simplification provisions of the Health Insurance Portability and Accountability Act (HIPAA) originally mandated the implementation of a unique health identifier for individuals. However, Congress withheld funding of the implementation pending adequate federal privacy protection. Now that the HIPAA privacy rules have been implemented nationwide, means to link patient data across organizations should be revisited.

In the meantime, pragmatic approaches to linking patient data have been emerging within the provider community. One approach used by many health care systems is the enterprise master patient index, which essentially creates a local unique patient identifier for persons cared for within a single health care system. Since most health care is local, and relationships among patients, physicians, and specific hospitals are ongoing, this approach has served as a viable

interim solution; however, it is costly to maintain, does not address the issue of data coming from other systems of care, and requires the development of matching algorithms to solve such problems as patients with similar names. Because no algorithm is perfect, a small percentage of attempted matches will result in errors that can be recognized and reconciled only through human intervention.

Another approach under study, developed by the Patient Safety Institute (PSI) for its project to link health care providers statewide, is based on the Visa credit card network system which allows for connections among doctors' offices and hospitals (Carper, 2003). In PSI's network, PSI manages the automated system as a master patient index of only patient names and their identification numbers. Each hospital/clinic's method for identifying the patient data is mapped to this index, which is maintained by PSI. All medical data are retained by the health care organizations behind secure firewalls. Initially, authentication was accomplished through established hospital procedures. A three-key approach to authentication is in the process of being implemented: card and password for patient, digital certificate and password for physician, and permission key for the hospital or clinic. As of this writing, control features have not been implemented; access to medical information is "all or nothing." PSI is now in the process of implementing controls for two levels: general medical information and sensitive information (e.g., mental health, drug rehabilitation, HIV). Patients voluntarily opt into the program to allow their physicians access to past diagnoses, laboratory results, medications, allergies, and immunizations (Carper, 2003). Sensitive medical data (e.g., HIV status) are excluded from level-one access. PSI has also developed a set of guiding principles to safeguard access and limit exposure of patients' electronic medical data. These principles include avoiding use of a patient's social security number as their unique identifier, never releasing a patient's medical records to anyone without the patient's express authorization, physically separating a patient's clinical and demographic data, and using cutting-edge encryption technology and secure private networks (Carper, 2003).

The committee believes that the careful examination and development of innovative methods for patient data linkage should be undertaken against a background of changing technology, illness patterns, and consumer attitudes. In particular, changing demographics have resulted in the growth of chronic care conditions that involve multiple providers and data sources, making it more difficult to maintain and integrate relevant patient information. Consumers will be more involved in their self-care and disease management and will require the capability to utilize a personal health record and engage in electronic communication with their provider(s). Likewise, as they continue to become more savvy in accessing and understanding health information on the Internet, the demand for tools to incorporate this information into their care protocols and personal health records will likely increase. With HIPAA security rules in place, it is also possible to create patient data linkages in a manner that empowers patients to permit access to some of their data while restricting access to other, more sensitive data (e.g., mental health).

Terminologies

Standardized terminologies facilitate electronic data collection at the point of care; retrieval of relevant data, information, and knowledge (i.e., evidence); and data reuse for multiple purposes, such as automated surveillance, clinical decision support, and quality and cost monitoring. To promote patient safety and enable quality management, standardized terminologies that represent the focus (e.g., medical diagnosis, nursing diagnosis, patient problem) and interventions of the variety of clinicians involved in health care as well as data about the patient (e.g., age, gender,

ethnicity, severity of illness, preferences, functional status) are necessary. Significant efforts during the last quarter-century have resulted in the development of standardized terminologies for the core phenomena of clinical practice: (1) diagnoses, symptoms, and observations (e.g., medical diagnoses, nursing diagnoses, problem list); (2) interventions, procedures, and treatments, including those focused on prevention and health promotion; and (3) health outcomes (e.g., disability, functional status, symptom status, quality of life) (Wang et al., 2002a). Although standardized measures for health outcomes have been developed, the incorporation of such measures into standardized terminologies has lagged behind that of measures for problems and interventions. Additionally, standardized terms for patient goals (i.e., expected outcomes) have been addressed only minimally and almost exclusively by the nursing community (Johnson et al., 2000). While no single current terminology has the breadth and depth needed for health care data, the National Library of Medicine (NLM) houses the world's largest database of standardized terminologies from a broad array of digital knowledge sources—the Unified Medical Language System (UMLS) (see Chapter 3). The terminology resources available through the UMLS are critical to initiatives to establish the NHII and to corresponding use of the EHR and patient safety systems.

Technical Criteria and Representation of Clinical Domains

Standardized terminologies vary along many dimensions; most important is the primary purpose of the terminology, as well as the extent to which it is concept oriented and possesses the semantic structures that enable computer (algorithmic) processing (Ingenerf, 1995; Rossi et al., 1998). To achieve the integrated approach to patient safety envisioned by the committee, the terminology must serve the purposes of decision support tools, the EHR, and knowledge resources (Chute et al., 1998). Terminology efforts for the EHR have focused on how to represent the history, findings, diagnoses, management, and outcomes of patients in a way that can preserve clinical detail and identify characteristics that enable improved risk adjustment, the development of common guidelines, aggregate outcome analyses, and shared decision support rules.

While a number of diverse terminologies are required for clinical care, patient safety, and other aspects of biomedicine, a central group of terminologies can serve as the backbone of clinical information systems. A number of technical criteria must be met for terminologies to function in a way that serves these purposes. The most basic criteria for a controlled medical vocabulary are identified by Cimino (1998); they include domain completeness, nonredundancy, synonymy, nonambiguity, multiple classification, consistency of views, and explicit relationships. In 1998, the ANSI Health Informatics Standards Board went a step further and created a detailed framework of informatics criteria for the development and evolution of terminologies with high functionality (Chute et al., 1998). The National Committee on Vital and Health Statistics (NCVHS) used these informatics criteria to evaluate and select a core set of well-integrated, nonredundant clinical terminologies that will serve as the national standard for medical terminology for the EHR (Sujansky, 2003) (Table 4-2).

Required Technical Criteria	Description
Domain completeness	Domains of the vocabulary must include concepts, possibly with multiple representations, and be defined or defined relationships to other domains, categories, or types of concepts as needed (Ingenerf, 1995).
Nonredundancy	The meaning of each coded concept in a terminology must be based on a unique code that is assigned to a unique concept or code as defined in source (Ingenerf, 1995).
Consistency of views	Concepts must have explicit relationships. When a concept has multiple meanings, it must be represented into distinct concepts (Ingenerf, 1995).
Explicit relationships	Each instance of the relationship is represented by a unique identifier, and this can readily determine whether they are using the same set of data.

TABLE 4-2

Technical Criteria Used by the National Committee on Vital and Health Statistics for Evaluating and Selecting Terminologies.

Minimization of overlap in domain representation was another important criterion for selection of the NCVHS core terminology group. The CHI initiative is also evaluating the terminologies in this regard, as well as assessing their ability to meet the extensive data representation requirements for the common clinical domains that cut across the three dimensions of the NHII (i.e., provider health, personal health, and public health), points of overlap, and gaps in coverage. Issues related to data collection, sharing, and reuse are being addressed during the evaluations, as well as identification of the overlap and gaps in clinical representation. Table 4-3 provides an overview of the cross-cutting domains identified by CHI to date. The terminologies determined by CHI to best represent requirements of the clinical domain areas, after consultation with NCVHS, will be accepted for federal government-wide implementation. Additional areas within the clinical domains, including those relevant to patient safety, will be added as the process proceeds.

TABLE 4-3

Clinical Domain Areas of the Consolidated Health Informatics Initiative.

CHI is working rapidly and expects to make recommendations on terminologies to represent many, if not all, of the domain areas identified in Table 4-3 by late 2003. The first round of terminology evaluations includes laboratory results content, medications, demographics, immunizations, and interventions and procedures. Initially, CHI identified many of the domain areas that support the corresponding domains needed for patient safety reporting systems; however, the list is not comprehensive, and there will likely be a need to expand or extend the domains. For example, in the domain area for medications, CHI identifies clinical drugs, warnings, allergic reactions, and adverse drug events (ADEs) as primary areas for clinical representation. For patient safety, representation is also needed for subcategories, such as nutritional supplements and alternative medicines.

Expansion of the domain areas for comprehensive clinical and patient safety data is a subject for additional work. Appendix F provides a comprehensive listing of the additional clinical domain areas needed to represent patient safety.

Evolution and Development of New Terms

In cases where domain coverage of a terminology is inadequate, the best sources of data for the development of new terms to represent clinical and patient safety information are the clinical measures within standardized datasets for a health condition derived from evidence-based guidelines, documentation of physical findings, and narrative text of patient safety reports. Once comprehensive datasets have been identified, it may be possible to develop extensions of existing terminologies for those areas that are insufficient in representing clinical or safety data. In other cases, it may be necessary to develop new terminologies. Each data element (e.g., measure, finding) should include definitions of patient safety terms for near misses, basis for expanded clinical documentation in integrated systems. This approach takes into consideration the conceptual model for data integration discussed in Chapter 2.

Priority 3 Forthcoming Task Groups	Terminologies Used by Other Processes
---	--

Priority 3 Forthcoming Task Groups	Terminologies Used by Other Processes
Genes and proteins Multimedia, including but not limited to: Image Audio Waveforms Nursing, including: Diagnoses Interventions Goals and outcomes Physiology Supplies, including Ontology for the ordering physician Medical devices	Billing (HIPAA) Chemicals (UMLS) Disability (International Classification of Functioning, Disability and Health [ICF]) Scientific/fundamental (UMLS) Units (UMLS)

For the 20 priority areas identified by the IOM in its 2003 report *Priority Areas for National Action*, efforts to extend and create complete terminologies for clinical and safety data should follow a process that:

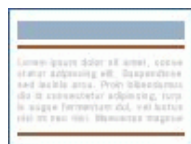
- Supports evidence-based clinical guidelines.
- Clearly defines the condition in terms of clinical measurements and actions, and define safety in terms of what could go wrong with those measures or actions.
- Evaluates datasets to determine whether they truly represent data elements necessary to measure outcomes, including safety.
- Continues refining the taxonomy for safety by asking about the reasons for errors of omission and commission (e.g., why a foot exam was not given), in addition to using information extracted through natural language processing from the rich narrative in textual reports and data from surveillance and decision support systems.
- Determines precursors to potential adverse events (e.g., a fall) where possible.

The need for multiple levels of granularity and cross-organizational terminologies means that a dataset will need to be either initially represented by or mapped to a concept-oriented clinical reference terminology and further mapped to high-level taxonomies for comparative analysis and research. The committee encourages further work on developing standardized datasets with the capability to represent patient safety information in all clinical areas.

Selection of the Core Terminology Group

The NCVHS core terminology group comprises a core set of medical terminologies that together are sufficiently comprehensive, technically sound, mutually consistent, and readily available to deliver most of the envisioned functionality of a national standard medical terminology for the EHR (Sujansky, 2003). Having a common clinical reference terminology is expected to reduce the cost, increase the efficiency, and improve the quality of data exchange, clinical research,

patient safety, sharing of computer guidelines, and public health monitoring. Terminologies to be included in the core group must have sufficient clinical granularity and serve multiple functions, including decision support, interoperability, aggregation and reporting, EHR data entry, order entry, indexing for data retrieval, and domain ontology. Supplemental terminologies should be mapped to the core terminologies to provide the functionalities associated with the use of data standards and information systems. [Box 4-1](#) provides a brief overview of these terminologies.



BOX 4-1

Overview of Core and Supplemental Terminologies. *CORE TERMINOLOGIES Systemized Nomenclature of Human and Veterinary Medicine, Clinical Terms (SNOMED CT)*—developed by the College of American Pathologists, SNOMED CT is an inventory of medical terms ([more...](#))

On November 13, 2003, NCVHS officially recommended that the Department of Health and Human Services (DHHS) adopt five medical terminologies for use by federal health care services programs: SNOMED, Clinical Terms CT; Laboratory LOINC; RxNORM; the National Drug File Clinical Drug Reference Terminology (NDF RT); and the Food and Drug Administration's terminology sets for drug ingredient name, dosage form, and package form for drugs (National Committee on Vital and Health Statistics, 2003b). NCVHS continues to study additional terminologies that it may recommend for adoption at a later date. Also of note, NCVHS recently voted to recommend that HHS adopt the International Classification of Disease, 10th revision, or ICD-10, as the new coding system under the HIPAA rule, replacing the current ICD-9 system (National Committee on Vital and Health Statistics, 2003b).

SNOMED CT SNOMED CT is the most well-developed concept-oriented terminology to date. A concept-oriented reference terminology can be defined as one that has such characteristics as a grammar that defines the rules for automated generation and classification of new concepts, as well as the combining of atomic concepts to form molecular expressions (Spackman et al., 1997). SNOMED CT is based on a formal terminology model that provides nonambiguous definitions of health care concepts and contains the most granular concepts for representing clinical and patient safety information. For example, the atomic concepts of “diabetes mellitus,” “self-management,” and “education” could be combined to form “diabetes self-management and education,” one of the priority areas identified by the IOM, as a precoordinated concept within a terminology or postcoordinated for a particular quality indicator report addressing errors of omission. SNOMED CT is designed to be the primary support for knowledge-based systems, the expression of clinical guidelines and datasets for the IOM priority conditions, and a key source for the development of new concepts for clinical and patient safety data. SNOMED CT's model was recently submitted to ANSI for approval as a standard. As part of the UMLS, it will serve as the core clinical reference terminology for the NHII (Department of Health and Human Services, 2003).

Laboratory LOINC Even with its comprehensiveness, SNOMED CT requires the support of additional terminologies to capture certain clinical data not currently available in the terminology with sufficient granularity or scope, namely laboratory, medication, and medical device data. LOINC has already been designated by CHI (in May 2003) as the terminology for representing laboratory test results and is a part of the NCVHS core terminology group (Consolidated Health Informatics Initiative, 2003). LOINC is the available terminology that most fully represents

laboratory data in terms of naming for tests (e.g., chemistry, hematology) and clinical observations (e.g., blood pressure, respiratory rate). The LOINC terms are composed of up to eight dimensions derived from component (e.g., analyte), type of property (e.g., mass concentration), timing (e.g., 24-hour specimen), specimen (e.g., urine), and method (American National Standards Institute, 1997). LOINC also contains information for clinical observations that is not included in the core terminology group at this time, since it may be possible to represent many of the observations with SNOMED CT, and one of the criteria for selection is to minimize overlap in terminology representation. However, Clinical LOINC currently is and will continue to be used by a number of organizations.

For LOINC clinical measures, the code usually includes identification of the organ system. In addition, with Clinical LOINC, many measurements are distinguished for estimated, reported, and measured values (e.g., patient's report of his/her body weight versus a measured result or a physician's estimate) (American National Standards Institute, 1997). Varying degrees of precoordination for an observation are also provided for (e.g., cardiac output based on the Fick method versus based on the 2D method) (American National Standards Institute, 1997). Both Laboratory LOINC and those portions of Clinical LOINC that do not overlap with SNOMED CT are important terminologies for patient safety, as well as for the EHR.

Drug terminologies Drug terminologies are an important part of the core group. NCVHS has been evaluating which drug (and device) terminologies best represent these areas. The process for determining drug terminologies is more complex than that for identifying a comprehensive reference terminology and laboratory terminology. Representation of drug information involves both definitional and knowledge-based information (National Committee on Vital and Health Statistics, 2003a). Definitional information serves the purpose of interoperability by providing standardized terms to represent clinical drugs in clinical information systems. Knowledge-based information provides terminology for such phenomena as drug interactions, allergies, and contraindications, thus supporting greater functionality of clinical systems (National Committee on Vital and Health Statistics, 2003a). For purposes of standardizing data elements for patient medical records information, the core terminology group will be focused on definitional terms.

NLM has developed a normalized (i.e., standard) form for clinical drugs and their components—the RxNORM terminology. RxNORM assigns a standardized name for the active ingredient (i.e., generic), strength and physical form as given to the patient (e.g., 120 milligrams), and standard dosage form (e.g., tablet) (Brown et al., 2003). The semantic form provides the ability to link drug concepts from disparate vocabularies with naming variations developed by different pharmacy knowledge base vendors and drug manufacturers to match more closely the actual form a physician would order for a patient (Nelson et al., 2002). RxNORM was developed to be fully compatible with the FDA's system that provides identifiers/codes for active and inactive ingredients—the Unique Ingredient Identifier (UNII) project (Brown et al., 2003). Preliminary research on incorporating RxNORM into actual systems indicates that some refinements are needed (e.g., a few drugs need to be added) for greater precision and comprehensiveness; however, it will be possible to begin implementing it for use with clinical systems in the near term (National Committee on Vital and Health Statistics, 2003b).

In addition to RxNORM, other drug-related terminologies under consideration for inclusion in the core terminology group are the UNII and National Drug Codes (NDCs), both managed by the FDA, and the NDF RT being developed by the Veterans Health Administration. Following CHI's evaluation of the terminologies for representing the medication domain and presentation of

findings in October 2003, NCVHS included these drug terminologies in its recommendation to DHHS.

Medical device terminologies A medical device terminology is also a must for the core terminology group. The two medical device terminologies being considered by NCVHS are the Global Medical Device Nomenclature (GMDN), developed by an international consortium to harmonize terms for regulatory purposes, and the Universal Medical Device Nomenclature System (UMDNS), developed and maintained by the Emergency Care Research Institute (ECRI). The terminology selected should be comprehensive in scope to cover the range of devices and their functions; capable of representing adverse events and malfunctions related to the devices; inclusive of emerging technologies used in investigative settings; sufficiently granular to capture essential data without losing critical information; and capable of being continuously maintained at a high level of technical quality, being mapped to other terms in use, and supporting high-quality translation to other languages for international use (Coates, 2003a).

Although the GMDN consortium initiated its activities using international standards and collaborated with six primary device terminology developers, the FDA found that the final resulting terminology did not meet the above criteria. The terminology that most closely meets these criteria is UMDNS. UMDNS provides a formal hierarchical system for representing complex medical device concepts, with content expressed in preferred terms and codes, entry terms, parent–child–sibling relationships, attributes, definitions, mappings, and linkages (Coates, 2003b). The process for maintaining the terminology is well developed at ECRI. In addition, ECRI functions as a collaborating center for the World Health Organization (WHO), an Agency for Healthcare Research and Quality (AHRQ) Evidence-Based Practice Center, a National Guideline Clearinghouse, and a National Quality Measures Clearinghouse, and it maintains an extensive patient safety reporting system (Coates, 2003b). These functions are important to the development of an integrated information infrastructure and the NHII, and the committee supports the inclusion of UMDNS in the NCVHS core terminology group. For international regulatory purposes, subsequent modifications and enhancements of the GMDN by the FDA may render it mappable to the terminologies in the core terminology group.

Mapping terminologies Mapping terminologies is a challenging task. The detailed terminologies of the core group and less granular classifications can be thought of as existing along a continuum of detail; for example, patient information can be expressed in a detailed nomenclature, such as SNOMED CT, funneling into a classification rubric, such as an ICD-9, Clinical Modification (CM) code (Chute, 2003). This is a limited one-way process in that once patient data have been expressed solely in the form of classifications, the original detail is lost and generally cannot be recovered. In many cases, this funneling process can be accomplished satisfactorily through a simple mapping or table that indicates which classification code subsumes a detailed description. However, such code-to-code mappings often fail since some terminologies incorporate complex criteria that can be reliably achieved only with rules for aggregating several patient details (Chute, 2003). Thus, such “aggregation logics” afford the automated and accurate mapping of detailed patient data into broader classifications, even for complex cases.

To satisfy the needs for the NHII and the EHR, computer-executable aggregation logic would stem from SNOMED CT and the other terminologies in the core group to the supplemental terminologies. It is also critical that the integration and mapping of the terminologies be based on the same information model as that of the data interchange standards—the HL7 RIM—to ensure

optimum system functionality and interoperability (National Committee on Vital and Health Statistics, 2003a).

The committee believes that several supplemental terminologies are necessary to support the requirements for an integrated information infrastructure that supports multiple methods of collecting, analyzing, disseminating, and incorporating patient safety data with consideration for the differences among health care settings. As noted earlier, the terminologies must support system functionality and knowledge-based activities such as automated chart reviews and surveillance, voluntary reporting, natural language processing of narrative text, decision support tools (e.g., alerts and reminders), and the use of computer-readable evidence-based clinical guidelines. The supplemental terminologies outlined in [Box 4-1](#) would be mapped through aggregation logic to the NCVHS core terminology group. These terminologies include HIPAA-designated code sets (i.e., ICD-9 CM, Current Procedural Terminology [CPT]-4, the Health Care Financing Administration Common Procedure Coding System [HCPCS], NDCs, Current Dental Terminology [CDT]), primary pharmacy knowledge bases (i.e., FirstDatabank National Drug Data File [NDDF]; plus MediSpan, Multum Lexicon), the Medical Dictionary for Drug Regulatory Affairs (MedDRA), UNII, International Society for Blood Transfusion [ISBT] 128, the Diagnostic and Statistical Manual for Mental Disorders [DSM-IV], and those nursing terminologies not already incorporated into SNOMED CT.

Terminologies for further investigation and research The NCVHS core terminology group and the supplemental terminologies support the basic functionalities of the conceptual model for integrated systems presented in [Chapter 2](#). However, the committee has determined that two additional terminologies are also needed and warrant further investigation and research—the International Classification of Functioning, Disability and Health (ICF) to represent outcomes data, and the International Classification of Primary Care (ICPC) to represent the data needs of the office practice clinician.

Promising sources for standardized representation of functional status and outcome reporting include the WHO International Classification of Functioning and Disability (WHO ICF) and nursing terminologies such as the Nursing Outcomes Classification (Johnson et al., 2000). Functional status can be regarded as the demonstrated or anticipated capacity of an individual to perform or undertake actions or activities deemed essential for independent living and physiological sustenance (Ruggieri et al., forthcoming). Computer formats for clinical data describing the functional status of patients will be in increasing demand for measuring the impact of health care interventions and gauging quality of life (Ruggieri et al., forthcoming). These outcome measures can be used not only to capture the effect of an intervention on health status but also to control symptoms of a chronic condition, supplement specific clinical findings, or understand the patient's perception of care (Nerenz and Neil, 2001). Information on functioning as a supplement to diagnosis provides a broader, more meaningful picture of individual or population health over time that can be used for clinical decision making (World Health Organization, 2001), reporting and surveillance, and research and analysis.

The Mayo Clinic is undertaking a study to determine how well ICF can represent functional status data as they emerge traditionally within the health care setting (Ruggieri et al., forthcoming). Preliminary findings suggest that in their current state, ICF terms lack unambiguous clarity, fidelity, and hence usability across the ranges of clinical data and granularity required for the varied and extensive use cases that rely on the representation of functional status data (Ruggieri et al., forthcoming). However, ICF provides an important foundation from which clinical modifications and extensions can be developed to support robust

functional status descriptions and representations in a broad spectrum of clinical domains and use cases (Ruggieri et al., forthcoming). Further study and development of outcome terminologies for patient safety applications, including nursing terminologies, are recommended.

ICPC was developed by the World Organization of National Colleges and Academic Associations of General Practice/Family Doctors (WONCA) to provide a system for classifying the broad range of symptoms, unease and difficulties, and conditions that make up those problems related to primary care that cannot be documented with the ICD codes (WONCA, 1998). More specifically, ICPC provides for simultaneous classification of the three elements of an encounter: the process of care, the reason for the encounter, and the health problem diagnosed (WONCA, 1998). Although ICPC is not widely used in the United States, it is the primary classification system used by much of the international community for electronic documentation of clinical practice in primary care or for reporting to national governments (Marshall, 2003). ICPC is used (in conjunction with ICD-9) extensively in the European Union and former U.K. countries, which have the most robust EHRs in the world. A study in Finland found repeatedly that ICPC permits coding of 95 percent or more of primary care visits (episodes of care), compared with 50 percent for ICD-9 (diagnosis) (Jamouille, 2001). The ability to monitor episodes of care would support concurrent surveillance efforts by permitting a longitudinal look at patient symptoms, encounters (including diagnoses and treatments), and outcomes.

Because ICPC captures episodes of care, it has also been used to produce probability tables for presenting symptoms and diagnoses. This function could support the development of triggers in data monitoring or data mining systems and could be the basis for a much more robust decision support function. The ability of U.S. primary care practitioners to evaluate their practice and compare it with those of other physicians around the world relates directly to their ability to use the ICPC terminology in association with the NCVHS core terminology group.

With regard to patient safety, the University of Colorado Department of Family Medicine and numerous other organizations are involved in a collaborative project entitled Applied Strategies for Improving Patient Safety. This project, sponsored by AHRQ to analyze the causes and effects of adverse events in primary care and reduce the incidence of errors, is using ICPC as its classification system (Pace, 2003). Preliminary results are not available at this time.

A conceptual diagram of the core terminology group and associated mappings to supplemental terminologies is presented in [Figure 4-4](#), which shows the possible relationships among the terminologies and the use of aggregation logic for mapping through various levels of granularity. This figure was developed as a modification of a presentation in August 2002 to NCVHS on clinical semantic interoperability by Dr. James R. Campbell of the University of Nebraska Medical Center (Campbell, 2002).



FIGURE 4-4

Conceptual diagram of the core terminology group and mappings to supplemental terminologies. SOURCE: Adapted from Campbell, 2002.

Knowledge Representation

Biomedical literature knowledge bases are powerful tools for clinical reference. These knowledge bases hold the vast body of medical research findings from both a historical perspective and the perspective of current best evidence-based practice. At present, most digital

sources of evidence are operating as stand alone systems without the ability to link to clinical information systems. With the development and use of common data standards, this linkage for enhanced access to medical knowledge bases can occur.

Clinical Guideline Representation Model

An earlier IOM report defines practice guidelines as systematically developed statements to assist practitioners and patients in making decisions about health care for specific circumstances (Institute of Medicine, 1992). The National Guideline Clearinghouse alone contains nearly 1,000 publicly accessible guidelines (Maviglia et al., 2003). There are gaps and inconsistencies in the medical literature supporting one practice versus another, as well as biases based on the perspective of the authors, who may be specialists, general practitioners, payers, marketers, or public health officials (Maviglia et al., 2003). Few national guidelines can be implemented in clinical information systems because of the lack of a way to represent the knowledge in machine-executable formats.

Automating guidelines requires a computer-readable format that is unambiguous and makes use of stored patient data. A number of computational models and tools for extracting, organizing, presenting, and sharing clinical guidelines are currently in developmental use. [Box 4-2](#) lists the most common of these.



BOX 4-2

Guideline Representation Models and Tools. Arden Syntax, Columbia University DILEMMA/PRESTIGE model, Europe

Few guidelines have been successfully translated and incorporated into real clinical settings (Advandi et al., 1999) because the language of which most text-based guidelines are composed is ambiguous. Eligibility criteria and severity of disease or symptoms often are not explicitly defined, and when they are defined, they may not map to computable data within an EHR (Maviglia et al., 2003) or other decision support systems. Simpler decision support that has worked successfully has been in the form of “if-then” rules triggered by EHR data that result in alert/reminder messages (Maviglia et al., 2003). The multitude of guideline models are dissimilar—they capture different features of a guideline and were created for different purposes. For example, guidelines can be used to support clinical work flow, to foster background utilization review and monitoring, to drive consultations, or to capture the process flow in a clinical protocol. As a consequence, no single model enables all of the features of the various models to be fully encoded.

One potential approach to data sharing is a model known as Guideline Interchange Format (GLIF), developed by the InterMed Collaboratory (comprising Harvard, Stanford, and Columbia universities), that encodes the essential features of guidelines common to all models (i.e., a maximal subset of features, not a superset) (Greenes et al., 2001). The goal of GLIF is to be able to (1) encode different requirements of clinicians during decision making, (2) support automatic verification and validation of guidelines, (3) facilitate standard approaches to guideline dissemination and local adaptation, and (4) be used as a template for the integration of automated clinical guidelines with clinical information systems (Wang et al., 2002b).

Following a workshop hosted by InterMed in March 2000, the collaboratory decided to develop a standard model for representing guidelines with HL7. Rather than pursue agreement on one

model, the group decided to focus on the building-block components that all guideline models must accommodate, such as a way to formulate queries and to express decision logic, a way to express the models' logical rules, a way to reference data (the data model), an approach to resolving terminology issues, and a way to represent the process flow/work flow of a guideline (Greenes, 2003). The common language for representing the components is GELLO (Guideline Expression Language, Object-Oriented), which was developed for GLIF. GELLO includes an object-oriented query language—that is, syntax for querying the EHR—thus specifying how one retrieves elements to be used in the logic expressions (Sordo et al., 2003). GELLO depends on the existence of an object-oriented data model (i.e., HL7 RIM). Additional work on GELLO is being undertaken by HL7 with the intent of making it an official standard. Sufficient resources should be made available for revisions to resolve specifications for GLIF and to complete GELLO.

Another aspect to consider with GLIF is recognition that guidelines most often are not executed in their entirety (Greenes, 2003). Instead, certain steps of a guideline may be implemented within different parts of a clinical information system (Greenes, 2003). For example, some steps may bear upon evaluation of clinical findings and may offer suggestions for diagnostic assessment or workup strategy; some may bear on the choice of particular medications or other procedures and may be implemented as order entry suggestions or templates; some may relate to the interpretation to be made and the action to be taken when an abnormal laboratory result is obtained and might be implemented as alerts; and some may trigger reminders or scheduling events (Sordo et al., 2003). Thus a guideline should be considered in terms of the application services or functions required by its various steps to be most effective (Sordo et al., 2003). These requirements will differ from one clinical information system to another based on functionality supported by the system (e.g., whether computerized physician order entry is present, or whether automatic alerts or time-driven scheduling reminders are supported), as well as institutional preferences about how to interface recommendations with actions (e.g., whether to offer them as suggestions or to trigger them as default actions that need to be overridden to be ignored, and what user interactions with the clinical information system will be affected by them) (Greenes, 2003). Thus work is proceeding on developing a taxonomy for application services that might be invoked by guidelines, as well as ways of marking up particular steps with the details of how the action is to be carried out in a specific clinical information system implementation (Greenes, 2003).

Representation of Medical Literature

The volume of medical information continues to grow exponentially, leaving some clinicians feeling that it has become unmanageable (Jerome et al., 2001). Yet the value of having the most recent medical literature for reference at the point of care is clear. Leveraging the efficiencies of information technology and expanding the services of medical libraries can facilitate this objective. Advances in communication technologies, application interface tools, and standardization in the representation of medical literature should allow such requests and retrievals to be completed through a fully automated system so that reliance on a librarian is not necessary, and information access is available to clinicians around the clock (Humphreys, 2003a). Since NLM holds the largest and most comprehensive database of medical literature, the development of application interface tools should initially be targeted to accessing the NLM databases. Automated data retrieval would require a direct connection to the various medical literature topics, rather than linkage through the NLM Web site as is now the case (Humphreys, 2003a). Such application interface tools would greatly enhance the usability of medical

knowledge bases and capabilities for information seeking at the point of care (Humphreys, 2003a). The committee recommends further study into what characteristics of information and what design of the interface tools would be most useful to clinicians in this regard.

In addition, resources should continue to be provided to NLM to maintain its services in making medical literature available to consumers through its MEDLINEplus program. MEDLINEplus identifies information that is easy to read for the consumer and makes more than 150 interactive tutorials available in English, which include voice corresponding to the information printed on the screen (Humphreys, 2003b). The interactive tutorials are a popular feature in part because they are also suitable for those with low literacy (Humphreys, 2003b). In fact, those who select material for inclusion in MEDLINEplus actively seek low-literacy materials. NLM also encourages the institutes of the National Institutes of Health and producers of patient and consumer health materials to both convert their existing materials to electronic form and produce more of these materials.

The Cochrane Collaboration—an international effort for preparing, maintaining, and promoting the accessibility of systematic reviews of the effects of health care interventions—is another important source of medical knowledge. Its database was designed to produce up-to-date summaries of the results of reliable research and is now considered one of the world's best sources of medical evidence on treatments, diagnostic techniques, and preventive interventions (Cochrane Collaboration, 2003).

Other types of knowledge bases are highly important to patient safety and decision support tools. Disease registries are special databases that contain information about people diagnosed with a specific disease. Most registries are either hospital or population based and are used for a number of purposes, such as patient outcome tracking, support for self-care, epidemiological research, and public health surveillance (New York State Department of Health, 1999). They are also used for direct patient care, such as providing reminders for follow-up visits.

Knowledge bases, such as those for pharmacology and pharmacokinetics, hold a vast amount of medical knowledge critical to the accurate prescription drugs and surveillance of drug reactions. These knowledge bases provide information on drug–body interactions to support decisions about what drugs to prescribe; drug–drug comparisons; advice on administration (Duclos-Cartolano and Venot, 2003); information on contraindications, interactions, or therapeutic strategies related to the physiological conditions of a particular disease; and listings of drugs according to some of their properties (Duclos-Cartolano and Venot, 2003). Common data standards can facilitate interconnections with bar-code medication administration systems, computerized physician order entry systems, and other decision support tools for the clinician and can support the self-care of patients by providing access to drug interaction checking programs.

Clinician and patient access to vital information about medications contained in labeling (i.e., package inserts) is also important to patient safety. NLM is playing a key role in the standardization of the information on medication package inserts so the information can be made available in electronic format over the Internet. The DailyMed database, as its name suggests, is intended to provide updates of medication information to the public on a daily basis. Labels are also being restructured so they will be easier to understand and useful to both nonprofessionals and information systems (Brown et al., 2003). A major innovation will be the inclusion of labeling highlights that include recent label changes, indications, usage, dosage, administration, how supplied, contraindications, warnings/precautions, drug interactions, and use in special

populations (Brown et al., 2003). Finally, NLM is working on the UNII project intended to code the molecular structure and other features of each new medication. With the pending market entry of about 1,500 new drugs in the next few years, the NLM Molfile database of molecular and manufacturing information on new drugs will augment pharmacy and pharmacogenetic databases by supplying more detailed information about medication functions and the prevention of adverse reactions.

Implementation of Data Standards

Implementing data standards is just as important as developing and selecting standards. In preparing to implement standards, several issues tend to arise that should be considered when establishing a mechanism for compliance; these issues include vendor readiness, organizational readiness, cost of compliance tools, unresolved issues related to terminologies and coding, identifiers for providers and patients, and interpretation of the implementation guides and standard specifications. Help in dealing with these issues is critical.

Most recently, in the implementation of HIPAA, the Workgroup on Electronic Data Interchange (WEDI) stepped forward to lead the Strategic National Implementation Process and provided guidance, assistance, and advice on the implementation of and compliance with HIPAA standards. The workgroup has developed a number of white papers that provide specific guidance on the technical aspects of implementing the associated code sets, messaging formats, security features, and privacy policies. It has also provided guidance on the testing and certification of clinical systems for compliance (i.e., conformity assessment) with the standards—testing organizational systems internally as well as testing systems externally with trading partners. The committee recommends that a similar entity be identified to assist with the implementation of clinical and patient safety data standards for the NHII. Such an entity might best be established with AHRQ as coordinator. The entity might assist organizations in increasing staff awareness and education; undertaking a gap analysis of current and desired standards; formulating a strategic plan, budget, and timeline to meet the CHI requirements; implementing the plan and certifying conformance; and providing an audit process for ongoing monitoring and enforcement. In contrast with HIPAA, however, self-certification should not be an option for compliance with clinical data standards.

In addition to the establishment of an oversight organization and a national implementation plan, a mechanism for assessing conformance with the data standards is needed. Conformity assessment, an integral part of the utilization of standards, is the comprehensive term for measures taken by manufacturers, their customers, regulatory authorities, and independent third parties to evaluate and determine whether products and processes conform to particular standards (National Research Council, 1995). The National Institute of Standards and Technology (NIST) could perhaps serve as the body supporting the implementation process as the developer of protocols for conformance tests, information assurance, and certification procedures to verify vendors' compliance with the standards.

Because the core terminology group for the EHR and other health-related applications will be housed for public availability within the UMLS, NLM will play a vital role in the coordination, mapping, and dissemination of the terminologies for national adoption. NLM will share responsibility for the maintenance and regular updating of the terminologies with the terminology developers. As the chief standards development organization for the EHR, HL7, in collaboration with government organizations (e.g., Centers for Medicare and Medicaid Services), will develop the specifications for the actual implementation of the terminologies.

As stated in Chapter 3, AHRQ can facilitate the standards adoption process by functioning as a coordinating body and provider of technical assistance for the efforts of CHI, NCVHS, NLM, FDA, and HL7 in the area of data standards and for the Quality Interagency Coordination Task Force, evidence-based practice centers, specialty societies, academic institutions, and professional organizations involved in the determination of best practices that become translated into electronic data systems. AHRQ should be fully funded to function in this capacity.

Assessing the costs related to the development, implementation, and dissemination of data standards will involve a coordinated set of evaluations by AHRQ and NLM. AHRQ would most likely have the responsibility for estimating the costs related to the establishment and operation of a WEDI-like entity for standards implementation and conformity assessment. NLM would have responsibility for estimating the costs related to the development and maintenance of the core terminology group and mappings to supplemental terminologies. Together, these organizations should engage in a comprehensive evaluation of the costs to provide the data standards needed for the NHII and patient safety systems.

REFERENCES

Advandi, A., S. Tu, M. O'Connor, R. Coleman, M. K. Goldstein, and M. Musen. 1999. *Integrating a Modern Knowledge-Based System Architecture with a Legacy VA Database: The ATHENA and EON Projects at Stanford*. Proceedings of the Annual American Medical Informatics Association Symposium. Philadelphia, PA: Hanley and Belfus. [PMC free article: PMC2232584] [PubMed: 10566440]

American National Standards Institute. 2002. *ANSI Procedures for the Development and Coordination of American National Standards*. Online. Available: <http://public.ansi.org/ansionline/Documents/Standards%20Activities/American%20National%20Standards/Procedures,%20Guides,%20and%20Forms/ansipro2002r.doc> [accessed June 1, 2003].

American National Standards Institute, Health Informatics Standards Board. 1997. *Inventory of Health Care Information Standards*. New York, NY: American National Standards Institute.

Association for the Advancement of Medical Instrumentation. 2001. *Human Factors Design Process for Medical Devices HE74*. Arlington, VA: Association for the Advancement of Medical Instrumentation.

Brown, S. H., R. Levin, M. J. Lincoln, R. M. Kolodner, and S. J. Nelson. 2003. *United States Government Progress Towards a Common Information Infrastructure for Medications*. Bethesda, MD: National Library of Medicine.

Campbell, J. R. 2002. *Presentation and Testimony on Converging the Clinical Care Model*. National Committee on Vital and Health Statistics, Subcommittee on Standards and Security.

Carper, T. 2003. *Health Information on Demand*. Online. Available: <http://www.ndol.org/print.cfm?contentid=251495> [accessed April 15, 2003].

Chute, C. 2003. *Terminology Mappings and Aggregation Logic*. Personal communication to Institute of Medicine's Committee on Data Standards for Patient Safety.

Chute, C. G., S. P. Cohn, and J. R. Campbell. 1998. A framework for comprehensive health terminology systems in the United States. *J Am Med Inform Assoc* 5 (6):503–510.

[PMC free article: PMC61331] [PubMed: 9824798]

Cimino, J. J. 1998. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med* 37 (4–5):394–403. [PMC free article: PMC3415631] [PubMed: 9865037]

Coates, V. H. 2003. a. *Medical Device Terminologies* . Personal communication to Institute of Medicine's Committee on Data Standards for Patient Safety.

———. 2003. b. *ECRI's Universal Medical Device Nomenclature System* . Presentation to the NCVHS Subcommittee on Standards and Security.

Cochrane Collaboration. 2003. *Cochrane Collaboration Brochure* . Online. Available: <http://www.cochrane.org/software/docs/newbroch.pdf> [accessed July 14, 2003].

Consolidated Health Informatics Initiative. 2003. *Standards Adoption Strategy and Portfolio* . Presentation to the NCVHS Subcommittee on Standards and Security.

Department of Health and Human Services. 2003. *HHS Launches New Efforts to Promote Paperless Health Care System* . Online. Available: <http://www.hhs.gov/news/press/2003pres/20030701.html> [accessed July 7, 2003].

Dolin, R. H., L. Alschuler, C. Beebe, P. V. Biron, S. L. Boyer, D. Essin, E. Kimber, T. Lincoln, and J. E. Mattison. 2001. The HL7 Clinical Document Architecture. *J Am Med Inform Assoc* 8 (6):552–569. [PMC free article: PMC130066] [PubMed: 11687563]

Duclos-Cartolano, C., and A. Venot. 2003. Building and evaluation of a structured representation of pharmacokinetics information presented in SPCs: From existing conceptual views of pharmacokinetics associated with natural language processing to object-oriented design. *J Am Med Inform Assoc* 10(3):271–280. [PMC free article: PMC342050] [PubMed: 12626375]

Greenes, R. A., M. Peleg, A. T. S. Boxwala, V. Patel, and E. H. Shortliffe. 2001. Sharable computer-based clinical practice guidelines: Rationale, obstacles, approaches, and prospects. *Medinfo* 10 (Pt 1):201–205. [PubMed: 11604733]

Greenes, R. A. 2003. *Electronic Representation of Clinical Guidelines (Attachments: 1)* . Personal communication to Institute of Medicine's Committee on Data Standards for Patient Safety.

Hammond, W. E. 2002. *Overview of Health Care Data Standards* . Commissioned paper for IOM Committee on Data Standards for Patient Safety.

Health Level Seven. 2001 (March). *HL7 V3 Standard: Backbone, Draft 1.0, Version 3.0*, K. Blyler, editor. , ed.

Humphreys, B. June 2, 2003a. *Medical Knowledgebases* . Personal communication to Institute of Medicine's Committee on Data Standards for Patient Safety.

———. June 20, 2003b. *Medical Knowledgebases* . Personal communication to Institute of Medicine's Committee on Data Standards for Patient Safety.

Ingenerf, J. 1995. Taxonomic vocabularies in medicine: The intention of usage determines different established structures. *MedInfo* 95 136–139. R. A. Greenes, editor; , H. E.

Peterson, editor; , and D. J. Protti, editor. , eds. Vancouver, British Columbia: Health Care Computing and Communications, Canada, Inc. [PubMed: 8591138]

Institute of Medicine. 1992. *Guidelines for Clinical Practice: From Development to Use* . Washington, DC: National Academy Press. [PubMed: 25121254]

———. 2003. *Priority Areas for National Action: Transforming Health Care Quality* . Washington, DC: The National Academies Press. [PubMed: 25057643]

Jamouille, M. 2001. *ICPC Use in the European Community* . Durban, South Africa: WONCA International Classification Committee at the 16th WONCA World Congress of Family Doctors.

Jerome, R. N., N. B. Giuse, K. W. Gish, N. A. Sathe, and M. S. Dietrich. 2001. Information needs of clinical teams: Analysis of questions received by the clinical information consult service. *Bulletin of the Medical Library Association* Apr 2001 177–185. [PMC free article: PMC31725] [PubMed: 11337949]

Johnson, M., editor; , M. Maas, editor; , and S. Moorehead, editor. , eds. 2000. *Nursing Outcomes Classification* . St. Louis, MO: Mosby.

Marshall, I. 2003. ICPC around the world. *WONCA News: An International Forum for Family Doctors* 29 (4):14–18.

Maviglia, S. M., R. D. Zielstorff, M. Paterno, J. M. Teich, D. W. Bates, and G. J. Kuperman. 2003. Automating complex guidelines for chronic disease: Lessons learned. *J Am Med Inform Assoc* 10:154–165. [PMC free article: PMC150369] [PubMed: 12595405]

National Committee on Vital and Health Statistics. August 19-20, 2003a. *Proceedings: Subcommittee on Standards and Security* . Personal notes of Institute of Medicine Staff to Committee on Data Standards for Patient Safety taken during proceedings.

———. 2003. b. *Letter to Tommy G. Thompson, Secretary of the Department of Health and Human Services, Recommendations on Uniform Data Standards for Patient Medical Record Information* . Online. Available: <http://www.ncvhs.hhs.gov/031105lt3.pdf> [accessed December 5, 2003].

National Research Council. 1995. *Standards, Conformity Assessment, and Trade: Into the 21st Century* . Washington, DC: National Academy Press.

Nelson, S. J., S. H. Brown, M. S. Erlbaum, N. Olson, T. Powell, B. Carlsen, J. Carter, M. S. Tuttle, and W. T. Hole. 2002. *A Semantic Normal Form for Clinical Drugs in the UMLS: Early Experiences with the VANDF* . Proceedings of the Annual American Medical Informatics Association Symposium. Philadelphia, PA: Hanley and Belfus, 557–561. [PMC free article: PMC2244318] [PubMed: 12463886]

Nerenz, D. R., and N. Neil. 2001. *Performance Measures for Health Care Systems* . Commissioned paper for the Center for Health Management Research.

New York State Department of Health. 1999. *Chronic Disease Teaching Tools—Disease Registries* . Online. Available: <http://www.health.state.ny.us/nysdoh/chronic/diseaser.htm> [accessed August 2003].

Office of Management and Budget. 2003. *Consolidated Health Informatics* . Online. Available: http://www.whitehouse.gov/omb/egov/gtob/health_informatics.htm [accessed

April 21, 2003].

Pace, W. 2003. *Applied Strategies for Improving Patient Safety: Making the Health Care System Safer*. Denver, CO: Agency for Healthcare Research and Quality.

Rossi, M., F. Consorti, and E. Galeazzi. 1998. Standards to support development of terminological systems for healthcare telematics. *Methods Inf Med* 37 (4–5):551–563. [PubMed: 9865053]

Ruggieri, A. P., S. V. Pakhomov, and C. G. Chute. Forthcoming. *A Corpus Driven Approach Applying the Frame Semantic Method for Modeling Functional Status Terminology*. Proceedings of the Annual American Medical Informatics Association Symposium. Philadelphia, PA: Hanley and Belfus. [PubMed: 15360850]

Russler, D. C. 2002. Chapter 5: Patient safety and the HL7 reference information model. In: Russell Lewis, editor. , ed. *The Impact of Information Technology on Patient Safety*. Chicago, IL: Healthcare Information and Management Systems Society. Pp.67–79.

Seliger, R. 2003. *HL7 Context Manager*. Personal communication to Institute of Medicine's Committee on Data Standards for Patient Safety.

Seliger, R., and B. Royer. 2001. *HL7: Standards for E-Health. CCOW Context Management Standard*. Online. Available: http://www.hl7.org/special/Committees/ccow_sigvi.htm [accessed September 5, 2003].

Sordo, M., O. Ogunyemi, A. A. Boxwala, and R. A. Greenes. 2003. *Software Specifications for GELLO: An Object-Oriented Query and Expression Language for Clinical Decision Support*. Boston, MA: Brigham and Women's Hospital and Harvard Medical School. [PMC free article: PMC1480304] [PubMed: 14728515]

Spackman, K. A., K. E. Campbell, and R. A. Cote. 1997. SNOMED RT: A reference terminology for health care. *Proc AMIA Annu Fall Symp* 640–644. [PMC free article: PMC2233423] [PubMed: 9357704]

Sujansky, W. 2003. *Summary and Analysis of Terminology Questionnaires Submitted by Developers of Candidate Terminologies for PMRI Standards: A Draft Report to the National Committee on Vital and Health Statistics Subcommittee on Standards and Security*. National Committee on Vital and Health Statistics Meeting 5.

van Bommel, J. H., and M. A. Musen. 1997. *Handbook of Medical Informatics*. Heidelberg, Germany: Springer-Verlag.

Van Hentenryck, K. 2001. *HL7: The Art of Playing Together*. Online. Available: www.medicalcomputingtoday.com [accessed September 25, 2001].

Wang, A. Y., J. H. Sable, and K. A. Spackman. 2002. a. The SNOMED clinical terms development process: Refinement and analysis of content. *Proc AMIA Symp* 845–849. [PMC free article: PMC2244575] [PubMed: 12463944]

Wang, D., M. Peleg, S. W. Tu, A. A. Boxwala, R. A. Greenes, V. L. Patel, and E. H. Shortliffe. 2002. b. Representation of primitives, process models and patient data in computer-interpretable clinical practice guidelines: A literature review of guideline representation models. *Int J Med Inf* 68:59–70. [PubMed: 12467791]

Washington Publishing Company. 1998. *Overview of Healthcare EDI Transactions: A Business Primer*. Online. Available: http://www.wpc-edi.com/Default_40.asp [accessed February 2002].

World Health Organization. 2001. ICF Introduction: International Classification of Functioning, Disability and Health. Geneva: World Health Organization.

WONCA—World Organization of National Colleges, Academies and Academic Associations of General Practitioners and Family Physicians. 1998. International Classification of Primary Care-2. Oxford, UK: Oxford University Press.

Footnotes

- 1 Numerous acronyms appear in the discussion in this section. We follow the convention of defining each upon its first appearance and using the acronym thereafter. All acronyms used here are defined in [Appendix B](#).

Copyright 2004 by the National Academy of Sciences. All rights reserved.

Bookshelf ID: NBK216088