Vocabulary and Health Care Information Technology: State of the Art

James J. Cimino
Department of Medical Informatics, Columbia University; Atchley Pavilion, Room 1310, Columbia-Presbyterian Medical Center, New York, NY 10032. Phone: 212-305-8127; Fax: 212-305-3302; E-mail: cimino@columbia.edu

Controlled medical vocabularies are at the heart of almost all health care computing applications. This article reviews the vocabularies available today and some of the reasons why they have failed to meet the needs of application developers. Also described is some of the current work being done to address shortcomings in previous vocabulary content and structure.

Medical information is complex: it often contains implicit attributes, internal intricacies, intentional ambiguities, and inaccuracies. One of the greatest challenges to medical computing is the representation of such information in a way that permits symbolic manipulation by computer programs. Clinical systems (defined here as those systems which collect data about and for direct patient care) represent information such as diagnoses using controlled vocabularies to support functions as mundane as billing and statistical reporting and as exotic as automated decision support. General medical information systems—which provide textual, bibliographic, or other reference and educational material—rely on controlled vocabularies for indexing their content. Medical expert systems—which apply medical knowledge to patient-specific diagnostic and therapeutic decisions—require controlled vocabularies which can match patient information with appropriate knowledge. In each case, system developers may choose to use an existing vocabulary, which may fit their needs imperfectly, or they may choose to create their own vocabulary, which requires additional effort and results in systems that tend to be isolated and difficult to integrate.

It may seem surprising that such a fundamental requirement for medical computing remains unresolved. But, in the brief history of the field of medical informatics, vocabulary has been a minor topic. As a result, the solutions available today for representing medical data are far less sophisticated than the programs which manipulate them. Fortunately, medical vocabularies are now being recognized as an important area for basic research. This article reviews the state of medical vocabularies, describes the inadequacies of currently available solutions, and identifies some of the key research being conducted.

© 1995 John Wiley & Sons, Inc.

Standard Vocabularies

Medical vocabularies have been around for at least a century, primarily for the purpose of classifying causes of death. The predominant coding system in this regard has been the *International Classification of Diseases* (ICD), published by the World Health Organization. The *Ninth Revision* (ICD-9) (World Health Organization, 1977) is the version currently in widespread use (see Fig. 1), but the *Tenth Revision* (ICD-10) (World Health Organization, 1992) has been published and is gradually replacing its predecessor as the official means for reporting mortality and morbidity statistics (see Fig. 2). The ICD uses a hierarchical coding system with groups of two- and three-digit integer codes for general categories of diseases, procedures, health status, disablements, and reasons for contact with health care providers. A decimal digit is used for many codes to specify more precise detail. This arrangement results in a strict, shallow hierarchy of terms.

ICD has been perceived as inadequate for the level of detail desired for statistical reporting in the United States (Kurtzke, 1979). In response, the United States National Center for Health Statistics published a set of "clinical modifications" (Commission on Professional and Hospital Activities, 1978). ICD-9-CM, as it is known, is compatible with ICD-9 but provides extra levels of detail in many places by adding fourth- and fifth-digit codes. Figure 3 shows a sample of the additional detail added to ICD-9 by ICD-9-CM. Most of the diagnoses assigned in the U.S. are coded in ICD-9-CM, allowing compliance with international treaty (by conversion to ICD-9) and supporting billing requirements (by

427.	Cardiac Dysrhythmias
427.0	Paroxysmal Supraventricular Tachycardia
427.1	Paroxysmal Ventricular Tachycardia
427.2	Paroxysmal Tachycardia, Unspecified
427.3	Atrial Fibrillation and Flutter
427.4	Ventricular Fibrillation and Flutter
427.5	Cardiac Arrest
427.6	Premature Beats
427.8	Other Specified Cardiac Dysrhythmias
427.9	Cardiac Dysrhythmia, Unspecified
785.0	Tachycardia, Unspecified

FIG. 1. Example of "Arrhythmia" terms from the World Health Organization's International Classification of Diseases, Ninth Edition (ICD-9). Note that "Tachycardia, Unspecified" appears as a cardiovascular symptom elsewhere in the hierarchy. The common medical term "Bradycardia" is not included in ICD-9, but maps to the code 427.8.

conversion to *Diagnosis-Related Groups*, or DRGs [3M Health Information Systems, annual]). No "clinical modifications" have been created for ICD-10, nor are any planned.

The National Library of Medicine (NLM) developed and maintains the *Medical Subject Headings* (MeSH) (National Library of Medicine, annual) for use in indexing its databases of citations to the medical literature (MEDLINE and others). MeSH terms cover many domains, including anatomy, chemicals, diseases, organisms, and procedures. The terms are arranged in a hierarchy and may have multiple contexts, or locations, in the hierarchy. Figure 4 shows a sample of the MeSH hierarchy.

The *Systematized Nomenclature of Human and Veterinary Medicine—SNOMED International* (Côté,

I46	Cardiac Arrest
I46.0	Cardiac Arrest with Successful Resuscitation
I46.1	Sudden Cardiac Death, So Described
I46.9	Cardiac Arrest, Unspecified
I47	Paroxysmal Tachycardia
I47.0	Re-entry Ventricular Arrhythmia
I47.1	Supraventricular Tachycardia
I47.2	Ventricular Tachycardia
I47.9	Paroxysmal Tachycardia, Unspecified
I48	Atrial Fibrillation and Flutter
I49	Other Cardiac Arrhythmias
I49.0	Ventricular Fibrillation and Flutter
I49.1	Atrial Premature Depolarization
I49.2	Junctional Premature Depolarization
I49.3	Ventricular Premature Depolarization
I49.4	Other and Unspecified Premature Depolarization
I49.5	Sick Sinus Syndrome
I49.8	Other Specified Cardiac Arrhythmia
I49.9	Cardiac Arrhythmia, Unspecified
R00	Abnormalities of Heart Beat
R00.0	Tachycardia, Unspecified
R00.1	Bradycardia, Unspecified
R00.2	Palpitations
R00.8	Other and Unspecified Abnormalities of Heart Beat

FIG. 2. Example of "Arrhythmia" terms from the World Health Organization's International Classification of Diseases, Tenth Edition (ICD-10). Notice that most of the "Arrhythmia" terms have "disease" codes, while some have "symptom and physical finding" (R) codes.

427.	Cardiac Dysrhythmias
427.0	Paroxysmal Supraventricular Tachycardia
427.1	Paroxysmal Ventricular Tachycardia
427.2	Paroxysmal Tachycardia, Unspecified
427.3	Atrial Fibrillation and Flutter
427.31	Atrial Fibrillation
427.32	Atrial Flutter
427.4	Ventricular Fibrillation and Flutter
427.41	Ventricular Fibrillation
427.42	Ventricular Flutter
427.5	Cardiac Arrest
427.6	Premature Beats
427.60	Premature Beats, Unspecified
427.61	Supraventricular Premature Beats
427.69	Other Premature Beats
427.8	Other Specified Cardiac Dysrhythmias
427.81	Sinoatrial Node Dysfunction
427.89	Other Specified Cardiac Dysrhythmias
427.9	Cardiac Dysrhythmia, Unspecified
785.0	Tachycardia, Unspecified

FIG. 3. Example of "Arrhythmia" terms from the United States National Center for Health Statistics's International Classification of Diseases, Ninth Edition with Clinical Modifications (ICD-9-CM). Compare this to the terms in Figure 1; the five-digit codes correspond to terms added in ICD-9-CM. As in ICD-9, "Tachycardia, Unspecified" appears as a cardiovascular symptom elsewhere in the hierarchy. The common medical term "Bradycardia" is not included in ICD-9-CM, but maps to the code 427.89.

Rothwell, Palotay, Beckett, & Brochu, 1993) is a recently-released vocabulary which has descended from a (relatively) long line of nomenclatures, including the *Standard Nomenclature of Diseases and Operations* (SNDO) (the New York Academy of Medicine, 1961), the *Standard Nomenclature of Pathology* (SNOP) (College of American Pathologists, 1971), and the *Systematized Nomenclature of Medicine* (SNOMED) (Côté & Rothwell, 1975). SNOMED International consists of a set of eleven axes, or taxonomies, for a specific set of concepts (organisms, diseases, procedures, etc.) and is intended for use in coding all contents of electronic medical records. The terms are coded with a semi-hierarchical code, as shown in Figure 5. Coding with SNOMED entails a "mix and match" process called postcoordination, in which terms from different axes are combined to represent the "utterances" that appear in the medical record.

The World Organization of National Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians (WONCA) provides the *International Classification of Primary Care* (ICPC) (Lambert & Wood, 1987) for use in coding clinical encounters (outpatient visits, hospital stays, operations, etc.). ICPC makes use of a combination of seven axes of terms and a structural framework for combining these terms. Like SNOMED, ICPC uses a postcoordination process. However, ICPC legislates how the terms can be combined within the framework.

The British National Health Service maintains a vocabulary called the *Read Clinical Codes* (NHS Centre for Coding and Classification, 1994a) which were de-

C14.280.67	Arrhythmia
C14.280.67.93	Arrhythmia, Sinus
C14.280.67.198	Atrial Fibrillation
C14.280.67.248	Atrial Flutter
C14.280.67.319	Bradycardia
C14.280.67.470	Extrasystole
C14.280.67.558	Heart Block
C14.280.67.558.137	Adams-Stokes Syndrome
C14.280.67.558.323	Bundle-Branch Block
C14.280.67.558.750	Sinoatrial Block
C14.280.67.565	Long QT Syndrome
C14.280.67.672	Parasystole
C14.280.67.780	Pre-Excitation Syndromes
C14.280.67.780.560	Lown-Ganong-Levine Syndrome
C14.280.67.780.770	Pre-Excitation, Mahaim-Type
C14.280.67.780.977	Wolff-Parkinson-White Syndrome
C14.280.67.829	Sick Sinus Syndrome
C14.280.67.845	Tachycardia
C14.280.67.845.695	Tachycardia, Paroxysmal
C14.280.67.845.880	Tachycardia, Supraventricular
C14.280.67.845.880.25	Accelerated Idioventricular Rhythm
C14.280.67.845.880.95	Tachycardia, Atrioventricular Nodal Reentry
C14.280.67.845.880.315	Tachycardia, Ectopic Atrial
C14.280.67.845.880.320	Tachycardia, Ectopic Junctional
C14.280.67.845.880.840	Tachycardia, Sinoatrial Nodal Reentry
C14.280.67.845.880.845	Tachycardia, Sinus
C14.280.67.845.880.900	Torsades de Pointes
C14.280.67.845.940	Tachycardia, Ventricular
C14.280.67.932	Ventricular Fibrillation

FIG. 4. Example of "Arrhythmia" terms from the National Library of Medicine's Medical Subject Headings (MeSH). Note that MeSH classifies terms such as "Heart Block," "Pre-Excitation Syndromes," "Parasystole," and their descendants as Arrhythmias; this differs from other controlled vocabularies which group these separately as "Conduction Abnormalities."

signed for use in coding electronic medical records. The terms cover broad domains, similar to SNOMED and ICPC, with terms arranged in a hierarchy, as shown in Figure 6. Like MeSH, the Read Codes can appear in multiple places in the hierarchy.

Uses of Vocabulary in Medical Computing

Developers of medical computing applications have been wary about adopting standard vocabularies, despite their availability. Some of this reticence is no doubt attributable to the "not invented here" effect. However, this does not account completely for the tendency of application developers to take on the additional work of vocabulary development. Additional reasons include problems with content in, conversion to, and organization of existing vocabularies. These effects are examined as they relate to the different application areas in medical computing (clinical, general information, and expert systems).

The predominant applications in clinical computing are those which support the billing systems in hospitals and office practices. As a result, the clinical information in such systems is often represented using ICD-9-CM and, for hospital systems, DRGs. However, these systems are now beginning to evolve into comprehensive electronic medical records (EMR) systems to support patient care functions. ICD-9-CM is simply inadequate for such purposes. Imagine, for example, being told by your physician that your diagnosis is "Other Specified Cardiac Dysrhythmias."

Adoption of standard terminologies for clinical systems appears to be working in some European countries. In the United Kingdom, the Read Codes have been mandated for use in electronic medical record keeping and anecdotal reports indicate good acceptance by system users. In the Netherlands, where doctors' practice systems are being standardized throughout the country, the ICPC is used to record diagnoses and reasons for office visits (van der Lei et al., 1993). However, in order to achieve user acceptance, extensive enhancements to ICPC have been made by the system developers.

In the U.S., most clinical information systems eschew standard vocabularies and make use of "home grown" terminologies. One reason for the failure to adopt standards has been that the available standards lack the domain coverage and adequate detail to capture clinical information in ways that support patient care functions. SNOMED International has attempted to address this issue with expanded content and is currently under consideration by many system developers. However, another

D3-30000	03-30-31 CARDIAC DYSRHYTHMIAS
D3-30000	Conduction disorder of the heart, NOS
D3-30010	Cardiac dysrhythmia, NOS
D3-30220	Paroxysmal tachycardia, NOS
D3-30800	Cardiac arrest
D3-30900	Premature beats, NOS
D3-30A00	Ectopic rhythm disorder
D3-31000	Sinoatrial node dysfunction
D3-31100	Sinus bradycardia
D3-31101	Severe sinus bradycardia
D3-31110	Persistent sinus bradycardia
D3-31120	Sick sinus syndrome
D3-31121	Tachycardia-bradycardia
D3-31124	Sinus tachycardia
D3-31130	Nodal rhythm disorder
D3-31140	Anomalous atrioventricular excitation
D3-31150	Pre-excitation atrioventricular conduction
D3-31300	Nonparoxysmal AV nodal tachycardia
D3-31340	Atrioventricular dissociation
D3-31350	Accelerated atrioventricular conduction
D3-31510	Atrial paroxysmal tachycardia
D3-31520	Atrial fibrillation
D3-31530	Atrial flutter
D3-31540	Atrial premature beats
D3-31560	Wandering atrial pacemaker
D3-31700	Ventricular tachycardia, NOS
D3-31710	Paroxysmal ventricular tachycardia
D3-31720	Ventricular fibrillation
D3-31730	Ventricular flutter
D3-31740	Ventricular premature beats
D3-31750	Paroxysmal supraventricular tachycardia
D3-31760	Supraventricular premature beats
D3-31800	Accessory atrioventricular conduction
D3-31810	Wolff-Parkinson-White syndrome
D3-31820	Lown-Ganong-Levine syndrome

FIG. 5. Example of "Arrhythmia" terms from the College of American Pathologists' Systematized Nomenclature of Human and Veterinary Medicine (SNOMED International). Note that terms such as "Accessory atrioventricular conduction," "Wolff-Parkinson-White syndrome," and "Lown-Ganong-Levine syndrome" are classified in SNOMED as Cardiac Dysrhythmias rather than as Heart Blocks (which have codes ranging from D3-32000 to D3-33470). This is similar to the classification in MeSH (see Fig. 4).

- G57 Disorder of heart rhythm
  - G576 Ectopic beats
    - G5761 Supraventricular ectopic beats
    - G5762 Ventricular ectopic beats
  - X2025 Paroxysmal tachycardia
    - G57y7 Sinus tachycardia
      - X2026 Sinoatrial node tachycardia
      - XaOD3 Atrial tachycardia
        - G5700 Paroxysmal atrial tachycardia
        - X2027 Ectopic atria tachycardia
        - X2028 Re-entrant atrial tachycardia
        - X2029 Incessant atrial tachycardia
        - X77BO Multifocal atrial tachycardia
      - XaOkY Atrioventricular tachycardia
        - X202A Ectopic atrioventricular node tachycardia
        - X202B Re-entrant atrioventricular node tachycardia
        - X202C Re-entrant atrioventricular tachycardia
        - G5701 Paroxysmal atrioventricular tachycardia
      - X202F SVT with functional bundle branch block
      - X202G His bundle tachycardia
        - X202H Congenital His bundle tachycardia
        - X202I Post-operative His bundle tachycardia
      - X202J Permanent junctional reciprocating tachycardia
      - G570 Paroxysmal supraventricular tachycardia
    - X202K Ventricular tachycardia
      - X202L Ventricular tachycardia with normal heart
      - X202M Familial ventricular tachycardia
      - X202N Induced ventricular tachycardia
      - X202O Incessant infant ventricular tachycardia
      - X202P Right ventricular outflow tract ventricular tachycardia
      - X77BS Wide QRS ventricular tachycardia
      - X77BT Narrow QRS ventricular tachycardia
      - X77BU Ventricular tachycardia, monomorphic
      - X77BV Ventricular tachycardia, polymorphic
        - X77BW Ventricular tachycardia, polymorphic, without QT prolongation
        - X77BX Ventricular tachycardia, polymorphic, with QT prolongation
  - X202Q Atrial dysrhythmia
    - G5731 Atrial flutter
    - G5730 Atrial fibrillation
      - X202R Lone atrial fibrillation
      - X202S Non-rheumatic atrial fibrillation
    - X202T Bradycardia-tachycardia syndrome
    - G57y3 Sick sinus syndrome
      - X202U Familial sick sinus syndrome
    - X202V Post-operative sino-atrial disease
    - X77BF Left atrial rhythm
  - X202W Ventricular dysrhythmia
    - G5741 Ventricular flutter
    - G5740 Ventricular fibrillation
    - X202X Paroxysmal familial ventricular fibrillation

FIG. 6. Example of "Disorder of heart rhythm" terms from the British National Health Service's Read Clinical Codes. Note that these terms are classified in Read Clinical Codes as "disorders." Many of these terms, along with several additional terms, can also be found classified under "observations."

problem faced by developers is the issue of providing continuity between legacy and new systems. The current trend in clinical systems is toward central repositories of patient data, collected from ancillary systems (e.g., laboratory, pharmacy, and radiology). These ancillary systems generally provide their information with codes in their own peculiar controlled vocabularies. While a central system developer might prefer to adopt a standard such as SNOMED, there are no adequate methods for converting the ancillary data into a standard form. Thus, the developers must either undertake the conversion process themselves or adopt the ancillary vocabulary as their method for representing information in their central system. Even at sites with significant system development resources, the controlled vocabularies remain home-grown conglomerates of ancillary vocabularies.

General medical information systems are less concerned with representing their content in coded form and more concerned with indexing the information so that it can be retrieved as needed. Indexing is often accomplished through simple keyword indexing, using the actual free-text words appearing in the text. This approach is straightforward and computationally simple, but it fails to account for the multiple meanings of words, topics discussed in the text but not specifically mentioned, and topics mentioned in the text but not specifically addressed. Indexing the material by manual or automated use of a coded vocabulary attempts to address these problems. When a controlled vocabulary is required in the medical domain, MeSH is the popular choice. However, for some purposes, such as indexing electronic textbooks, MeSH provides inadequate specificity, and developers must resort to developing their own, unique vocabularies for coding.

The vocabularies used in expert systems must be capable of bridging the difference between the relatively high-level concepts used for characterizing medical problems and the fine-grained terms used for representing the actual patient data which these systems process. Not surprisingly, no standard vocabulary has proved adequate for such systems. Most such systems are "stand alone," requiring the user to translate the specific patient data (used as input) into the higher level concepts used in the medical logic. For those systems in which the expert system is integrated directly into the clinical system, the vocabulary is invariably one of the home-grown varieties built from the terminologies of local ancillary systems (Pryor, Gardner, Clayton, & Warner, 1983; McDonald, Tierney, Overhage, Martin, & Wilson, 1992; Cimino, Clayton, Hripcsak, & Johnson, 1994).

Current Research

In the past, controlled medical vocabulary development has been viewed more often as a necessary task and less as a research topic. As a result, very little has appeared in the published literature about the issues and solutions addressed by vocabulary developers. Today, the situation is quite different, and vocabulary research is an active topic in the medical informatics literature. Whereas, in the past, vocabulary work consisted mainly of adding more terms until it was "complete," current research involves the techniques for representing vocabularies and the methods for developing vocabulary content in a coherent, consistent way. Certainly, "completeness" is an important characteristic for a good vocabulary, but it is by no means sufficient. Additional characteristics needed in a vocabulary which will support health care applications have been described (Cimino et al., 1994).

One of the most important is the eradication of redundant terms. For example, if a vocabulary contains two terms for the same diagnosis, then records about patients with such a diagnosis might be coded either way. But during retrieval, if only one code is searched for, then recall will be incomplete. Similarly, terms should not be ambiguous. For example, if a diagnosis term may have two meanings, then patients fitting either meaning might be coded together with the same code. When retrieval of one or the other meaning is attempted, patients with both meanings will be retrieved, impairing precision.

Another issue that arises is classification of terms. Most vocabularies organize their terms into some kind of hierarchical arrangement. However, *strict* hierarchies, the form most often used, are poor matches for medical terms, which often fall into more than one class. For example, most terms like pneumonia and lung cancer might be considered together in the same class of "Lung Disease," but they might also be grouped in the separate classes of "Infectious Disease" and "Cancer," respectively. Traditional controlled vocabularies usually choose a single organizational scheme, but different medical applications may have different needs. For example, a program which helps diagnose diseases discovered on chest x-ray might need to know about any lung diseases the patient might have, while a program that suggests antibiotic therapy might need to know about all infectious diseases.

Developers of existing vocabularies are now beginning to address these issues explicitly through a variety of techniques. The SNOMED developers are looking at ways to represent their vocabulary in a standardized data structure which better represents the meaning of each term (Rothwell, Côté, Cordeau, & Boisvert, 1993). The maintainers of the Read Codes are exploring similar methods as part of a concerted effort to expand their vocabulary's content (NHS Centre for Coding and Classification, 1994b).

Other researchers are looking at ways in which the definitions of the terms can be encoded such that the terms can be maintained with more sophisticated approaches. For example, by making term definitions explicit and well-structured, terms can be placed automatically in the classes to which they belong, while newly added terms can be compared to existing ones to determine if redundancy is being created. Work in this area is being conducted by researchers at the University of Utah (Rocha, Huff, Haug, & Warner, 1994), a group of independent researchers exploring collaborative vocabulary development (Friedman, Huff, Hersh, Pattison-Gordon, & Cimino, 1995), by researchers in the European Advanced Informatics in Medicine initiative (Rector, Glowinski, Nowlan, & Rossi-Mori, 1995), and by my own group at Columbia-Presbyterian Medical Center (Cimino et al., 1994).

Finally, the National Library of Medicine is trying to make some sense out of all these efforts by bringing the various terminologies together into a single Unified Medical Language System (UMLS) (Lindberg, Humphreys, & McCray, 1993). The UMLS was originally developed to support access to online information services by providing access to the controlled vocabularies used by these services. Enterprising researchers have used it for such purposes but also as a resource for automatically translating terms from one terminology to another. The NLM is now working with outside researchers to expand the content of the UMLS to cover clinical data coding.

Remaining Challenges

Despite the recent attention to vocabulary development, a number of significant issues remain to be addressed. The problem most often addressed in the past, comprehensive content, is becoming less of an issue as cumulative experience in compiling terms is gained. However, there are still some significant domains for which no generally accepted vocabulary exists, e.g., physical examination.

Another important challenge is the issue of precoordination versus postcoordination. Attempting to develop a coded term for every possible blend of term attributes rapidly leads to combinatorial explosion. Consider, for example, that bone fractures can be simple, compound, spiral, greenstick, comminuted, noncomminuted, and so on. When combined with the fact that any bone can be fractured and each bone may have several locations where fractures can occur, it makes sense to attempt to code fractures through the postcoordination of their attributes rather than create tens of thousands of fracture terms. On the other hand, identifying all information by its "atomic" attributes is tedious and overlooks many of the convenient syndrome terms used in medicine. For example, describing a fracture as "distal" + "right radius" + "distal" + "right ulna" ignores the existence of the term "Colle's Fracture" with its attendant implications for therapy. Postcoordination approaches can also lead to disputes over what constitutes an "atom" in the vocabulary. For example, is the term for the bone "right ulna" an atom, or is it more properly reduced by compounding to the term "ulna" with the attribute "right"?

A number of research issues are concerned with the maintenance of controlled vocabularies. Problems not easily addressed include deciding when new terms are needed, replacing existing terms with new terms (perhaps invalidating previously coded data), and disseminating new terms to users of the vocabulary. Research efforts by the National Library of Medicine (related to the UMLS), as well as other groups maintaining large controlled vocabularies have projects under way to address these problems.

Conclusions

Controlled vocabularies are crucial to almost all health care computing applications. Despite the availability of large standardized terminologies, most systems today use their own vocabularies, requiring extra effort on the part of the system developers and impeding the development of systems which speak a common language. Vocabulary research has been a small aspect of medical informatics research until just the past few years. Progress in this short time has been encouraging, but the continued application of disciplined approaches to vocabulary design, construction, and maintenance will be needed to provide system designers with vocabulary resources.

Acknowledgments

Dr. Cimino's work on vocabulary development is supported, in part, by contracts with the National Library of Medicine.

References

Cimino, J. J., Clayton, P. D., Hripcsak, G., & Johnson, S. B. (1994). Knowledge-based approaches to the maintenance of a large controlled medical terminology. *Journal of the American Medical Informatics Association, 1*(1), 35–50.

College of American Pathologists. (1971). *Systematized nomenclature of pathology.* Chicago: The College.

Commission on Professional and Hospital Activities. (1978). *International classification of diseases, ninth revision, with clinical modifications* (ICD-9-CM). Ann Arbor.

Côté, R. A., & Rothwell, D. J. (1975). *Systematized nomenclature of medicine.* Skokie, IL: College of American Pathologists.

Côté, R. A., Rothwell, D. J., Palotay, J. L., Beckett, R. S., & Brochu, L. (Eds.). (1993). *The systematized nomenclature of medicine: SNOMED international.* Northfield, IL: College of American Pathologists.

Friedman, C., Huff, S. M., Hersh, W. R., Pattison-Gordon, E., & Cimino, J. J. (1995). The Canon effort: Working toward a merged model. *Journal of the American Medical Informatics Association, 2* (1) 4–18.

3M Health Information Systems (Updated annually). *AP-DRGs: All patient diagnosis related groups.* 3M Health Care, Wallingford, CT.

Kurtzke, J. F. (1979). ICD-9: A regression. *American Journal of Epidemiology, 109*(4), 383–393.

Lambert, H., & Wood, M. (Eds.). (1987). *International classification of primary care.* Oxford: Oxford University Press.

Lindberg, D. A. B., Humphreys, B. L., & McCray, A. T. (1993). The unified medical language system. *Methods of Information in Medicine, 32,* 281–291.

McDonald, C. J., Tierney, W. M., Overhage, J. M., Martin, D. K., & Wilson, G. A. (1992). The Regestrief medical record system: 20 years of experience in hospitals, clinics, and neighborhood health centers. *MD Computing, 9*(24), 206–217.

National Library of Medicine. (Updated annually). Medical subject headings. Bethesda, MD: The Library.

New York Academy of Medicine. (1961). *Standard nomenclature of diseases and operations (fifth edition).* New York: McGraw-Hill.

NHS Centre for Coding and Classification (April, 1994a). *Read codes, version 3.* London: NHS Management Executive, Department of Health.

NHS Centre for Coding and Classification. (April, 1994b). *Read codes and the terms projects: a brief guide.* Leicestershire, UK: NHS Management Executive, Department of Health.

Pryor, T. A., Gardner, R. M., Clayton, P. D., & Warner, H. R. (1983). The HELP system. *Journal of Medical System, 7*(2), 87–102.

Rector, A. L., Glowinski, A. J., Nowlan, W. A., & Rossi-Mori, A. (1995). Medical concept models and medical records: An approach based on GALEN and PEN&PAD. *Journal of the American Medical Informatics Association, 2*(1), 19–35.

Rocha, R. A., Huff, S. M., Haug, P. J., & Warner, H. R. (1994). Designing a controlled medical vocabulary server: The VOSER project. *Computers and Biomedical Research, 27,* 472–507.

Rothwell, D. J., Côté, R. A., Cordeau, J. P., & Boisvert, M. A. (1993). Developing a standard data structure for medical language—the SNOMED proposal. In C. Safran (Ed.), *Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care, Washington, DC* (pp. 695–699). New York: McGraw-Hill.

van der Lei, J., Duisterhout, J. S., Westerhof, H. P., van der Does, E., Cromme, P. V. M., Boon, W. M., & van Bemmel, J. H. (1993). The introduction of computer-based patient records in The Netherlands. *Annals of Internal Medicine, 119,* 1036–1041.

World Health Organization. (1977). *Ninth edition. International classification of diseases index. Manual for the international statistical classification of diseases.* Geneva: The Organization.

World Health Organization. (1992). *International classification of diseases index. Tenth revision. Volume 1: Tabular list.* Geneva: The Organization.