Original **Investigations**

Methods ■

Planned NLM/AHCPR Large-Scale Vocabulary Test: Using UMLS Technology to Determine the Extent to Which Controlled Vocabularies Cover Terminology Needed for Health Care and Public Health

BETSY L. HUMPHREYS, MLS, WILLIAM T. HOLE, MD, ALEXA T. McCRAY, PhD, J. MICHAEL FITZMAURICE, PhD

Abstract   The National Library of Medicine (NLM) and the Agency for Health Care Policy and Research (AHCPR) are sponsoring a test to determine the extent to which a *combination* of existing health-related terminologies covers vocabulary needed in health information systems. The test vocabularies are the 30 that are fully or partially represented in the 1996 edition of the Unified Medical Language System (UMLS) Metathesaurus, plus three planned additions: the portions of SNOMED International not in the 1996 Metathesaurus, the Read Clinical Classification, and the Logical Observations Identifiers, Names, and Codes (LOINC) system. These vocabularies are available to testers through a special interface to the Internet-based UMLS Knowledge Source Server. The test will determine the ability of the test vocabularies to serve as a source of controlled vocabulary for health data systems and applications. It should provide the basis for realistic resource estimates for developing and maintaining a comprehensive "standard" health vocabulary that is based on existing terminologies.

■ **JAMIA.** 1996;3:281–287.

Affiliations of the authors: National Library of Medicine, Bethesda, MD (BH, WH, AM); and Agency for Health Care Policy and Research, Rockville, MD (JMF).

Correspondence and reprints: Betsy L. Humphreys, MLS, National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894. e-mail: blh@nlm.nih.gov

The National Library of Medicine (NLM) and the Agency for Health Care Policy and Research (AHCPR) are sponsoring a large-scale vocabulary test to determine the extent to which a *combination* of existing health-related classifications and vocabularies covers vocabulary needed in information systems supporting health care, public health, and health services research. The test differs from previous studies

of single terminologies[1] or comparisons of single systems[2,3] in its emphasis on determining the aggregate coverage of more than 30 terminology systems. The vocabularies in the test include some that are not yet incorporated into the Unified Medical Language System (UMLS) Metathesaurus,[4] but the test relies heavily on UMLS technology—the Metathesaurus data and format, the SPECIALIST lexicon and lexical programs,[5] and the Internet-based UMLS Knowledge Source Server.[6]

Any organization or person willing and able to adhere to the test procedures and to provide test data by December 15, 1996, may participate. In addition to contributing to a worthwhile cause, participants establish links between their local systems and the UMLS Metathesaurus. These connections can provide synonyms, definitions, hierarchical contexts, pointers to external information sources, and other information useful in health information systems.

Each test participant must have a good Internet connection and a real task for which controlled vocabulary is desired. Examples of such tasks include designing structured data entry screens for a particular clinic or service, developing a lexicon for a problem list, identifying names of all valid values for parameters in a clinical protocol or guideline, and developing templates for retrieving and summarizing patient data. Participation in the vocabulary test should be an extension of a task that will be undertaken anyway. Although NLM and others will be assessing the utility of the test vocabularies in natural language processing, the focus of the test is on the ability of the test set to serve as a source of controlled vocabulary for health data—not necessarily on its ability to represent concepts and terms occurring in free text.

Underlying Assumptions

Unambiguous, sharable, and aggregatable electronic health data will increase the quality, effectiveness, and efficiency of patient care and will facilitate clinical research, public health surveillance, and health services research. The basic premise of the vocabulary test is that controlling the vocabulary used in at least some health data elements is essential, although not sufficient, to achieve electronic health data that are unambiguous, sharable, and aggregatable. Other difficult problems, such as developing a robust structure for explicit, computable definitions of health concepts[7] and of data elements in patient records, must also be solved, but controlled vocabulary is an important step toward comparable health data.

The test assumes that the controlled vocabulary used in United States health data systems can be based on a combination of existing terminologies, thus avoiding the need to start de novo or to replicate terminology already available in an existing system. This position was espoused by AMIA's Board of Directors in 1994.[8]

The combination of vocabularies must include more than statistical and billing codes, such as the International Classification of Diseases, 9th edition, Clinical Modification (ICD9-CM) and Current Procedural Terminology (CPT). There is ample evidence that these systems are not adequate for recording detailed health care and public health data[3] or for outcomes and effectiveness research.[9] This is not surprising, since they were developed to serve statistical reporting and billing objectives. By mapping more specific controlled vocabulary to these systems, we can improve the information content of basic health data and retain the ability to generate statistics and bills at the appropriate level of aggregation.

Whatever its starting point, the controlled health vocabulary must evolve over time to represent the specific terminology and multiple hierarchical arrangements needed in different settings, to keep pace with changes in medicine and health, to accommodate evolving standards for the structure of health care and public health records, and to correct any problems that inhibit efficient electronic processing. The vocabulary test assumes that the UMLS Metathesaurus will provide an appropriate framework for managing the evolution of the controlled health vocabulary. The Metathesaurus provides access to many different health-related vocabularies in a common database format. In a number of cases, the Metathesaurus represents—in an explicit, machine-processable format—information that is only implicit in the original source vocabularies (e.g., relationships implied by indentations in a word-processing file or print tape). Organized by concept or meaning, the Metathesaurus links specific clinical vocabularies to each other, to statistical and billing codes, and to terms used in clinical practice guidelines, knowledge-based expert systems, MEDLINE, and other decision support tools. The Metathesaurus can represent an unlimited number of hierarchical views and can identify a particular concept or name of a concept as belonging to an unlimited number of functional subsets. If system developers design software and systems to use vocabulary data in the Metathesaurus format today, they are likely to have a relatively smooth transition path from the use of current vocabularies to any eventual, standard health vocabulary in the United States.

A significant barrier to the development and mainte-
nance of a United States health vocabulary has been
the lack of an organization with the responsibility, au-
thority, and resources necessary to coordinate a tran-
sition from the current array of independent vocabu-
laries to a coherent, maintainable "standard." Efforts
to address coordination and long-term maintenance
obviously must continue. The new Department of
Health and Human Services Data Council and the re-
vamped National Committee on Vital and Health Sta-
tistics may help this process. In the meantime, the re-
sults of the vocabulary test should help to clarify the
nature and extent of the gaps in existing vocabularies
and to assist in developing realistic estimates of the
resources required to achieve and maintain a compre-
hensive health vocabulary that can be distributed and
linked to statistical, billing, and decision support vo-
cabularies within the UMLS Metathesaurus.

Test Vocabularies

The vocabularies included in the test are the 30 that
are fully or partially represented in the 1996 edition
of the Metathesaurus (Table 1) plus three planned ad-
ditions. These are the portions of SNOMED Interna-
tional[10] not in the 1996 Metathesaurus, the Read Clin-
ical Classification,[11] and the Logical Observations
Identifiers, Names, and Codes (LOINC) system.[12] For
purposes of the test, the three planned additions have
been converted to a Metathesaurus-like format, in-
cluding word and term indexes automatically gener-
ated using the same programs that produce the Me-
tathesaurus indexes. The planned additions are
available to testers through special interfaces to the
UMLS Knowledge Source Server.

The test vocabularies were selected based on data pre-
sented and discussed at a December 5–6, 1994, meet-
ing convened by the NLM and AHCPR to assist in
planning for the test.[13] Attendees included medical in-
formatics researchers, system developers, and repre-
sentatives of standards groups, government agencies,
and interested professional organizations (e.g., AMIA,
American Nurses Association). Presentations at the
meeting summarized the results of a number of pub-
lished studies demonstrating the excellent clinical
coverage of SNOMED International,[2,3] and partici-
pants endorsed its inclusion in the test set. The de-
scription of the LOINC system and its development
from numerous sources, including EUCLIDES,[14]
ASTM E1218-94,[15] and laboratory test names currently
used at sites such as MetPath, Latter Day Saints Hos-
pital, the Regenstrief Institute, and the Department of
Veterans Affairs system, provided a compelling justi-
fication for its inclusion. The Read Clinical Classifi-

Table 1 ■

**Vocabularies Fully or Partially Included in the
1996 UMLS Metathesaurus**

AI/RHEUM diagnoses and findings
Classification of Nursing Diagnoses. NANDA
COSTAR (Computer-Stored Ambulatory Records) Vocabulary
COSTART: coding symbols for thesaurus of adverse reaction
  terms
CRISP (Computer Retrieval of Information on Scientific Proj-
  ects) Thesaurus
Diagnostic and Statistical Manual of Mental Disorders: DSM-
  III-R
Diagnostic and Statistical Manual of Mental Disorders: DSM-IV
DXplain (diagnoses and findings)
Glossary of Methodologic Terms for Clinical Epidemiologic
  Studies of Human Disorders
    (McMaster University)
Home Health Care Classification of Nursing Diagnoses and In-
  terventions
Index for Radiological Diagnoses: including diagnostic ultra-
  sound: revised 3rd ed.
International Classification of Diseases: 9th revision, Clinical
  Modification: ICD-9-CM.
Library of Congress Subject Headings. 12th ed.
Medical Subject Headings. MeSH
    (English, French, German, Portuguese, Spanish)
Neuronames Brain Hierarchy
Nursing Interventions Classification NIC
Omaha System
Online Mendelian Inheritance in Man
Physician Data Query (PDQ) Cancer Terms
Systematized Nomenclature of Human and Veterinary Medi-
  cine: SNOMED International
Systematized Nomenclature of Medicine: 2nd ed.
Thesaurus of Psychological Index Terms. PsycINFO Thesaurus
UltraSTAR. Ultrasound Structured Attribute Reporting
Universal Medical Device Nomenclature System: product cate-
  gory thesaurus
WHO Adverse Reaction Terminology. WHOART

cation was selected for the test set because the results
of a small NLM study of the overlap between SNO-
MED International and preliminary version 3.1 of the
Read Clinical Classification suggested that the two
systems might provide complementary coverage in
some clinical areas.[13]

Other vocabularies discussed at the meeting but not
yet sufficiently developed for inclusion in the test
were the United States version of the Tenth Interna-
tional Classification of Diseases; the in-patient Proce-
dure Coding System, which the Health Care Financ-
ing Administration plans to implement when the
United States begins to use ICD-10 for morbidity sta-
tistics; and the Medical Dictionary for Drug Regula-
tory Affairs (MEDDRA). MEDDRA will replace Cod-
ing Symbols for Thesaurus of Adverse Drug Reaction
Terms (COSTART) and WHO Adverse Drug Reaction
Terminology (WHOART), and it will provide a single
system for reporting adverse drug reactions to the
Food and Drug Administration (FDA) and regulatory
agencies in other countries. Although MEDDRA is not
one of the test vocabularies, the FDA and NLM will

Netscape: Form1

Metathesaurus   Knowledge Source Server

Your term matched the concept "Subdeltoid bursa ".

Your query term is "acromial bursa"

Is this concept equivalent in meaning to your concept?

◇Yes

◇No

If your answer to the above question is no, please provide a brief definition for your concept. Then check the no radio button and click on submit.

Submit

Basic Concept Information

Concept Name	Subdeltoid bursa
UI	C0224794
Definition	no definition found.
Semantic Type	Body Part, Organ, or Organ Component

Concept Variants

Synonym	Acromial bursa
Lexical Variant	None
Concept Name	Subdeltoid bursa
Source	SNM2/PT/T-16170
Source	SNMI95/PT/T-16170

Vocabulary:	SNMI95
SNOMED International	
TOPOGRAPHY	
MUSCULOSKELETAL SYSTEM AND SOFT TISSUES	
BURSAE	
SPECIFIC BURSAE	
Subdeltoid bursa	[T-16170]

Figure 1 Special World Wide Web interface to the UMLS Knowledge Source Server for processing user terminology. Here the user has entered the term "acromial bursa." The term is found in the current Metathesaurus as a synonym for the concept "subdeltoid bursa."

collaborate to search the test set (UMLS Metathesaurus, remainder of SNOMED International, Read Clinical Classification, and LOINC) for concepts and terms that may be needed to augment MEDDRA's coverage.

Test Participants

The core test participants are the following 10 health care organizations that serve as electronic medical record test sites under 8 cooperative agreements funded by NLM and AHCPR: Albert Einstein Medical Center (New York), Beth Israel Hospital (Boston), Children's Hospital (Boston), Columbia–Presbyterian Medical Center (New York), Kaiser-Permanente (Oakland), Massachusetts General Hospital (Boston), Mayo Foundation (Rochester, Minnesota), Oregon Health Sciences University (Portland), Regenstrief Institute (Indianapolis), and Washington University (St. Louis).

Other participants include the University of Pittsburgh (as part of a project funded by an NLM High Performance Computing and Communications contract), the Department of Veterans Affairs, the Department of Defense, Group Health of Puget Sound, the FDA (as previously indicated), commercial developers of clinical software, state public health agencies, drug companies, and biomedical publishers. Additional volunteers are welcome (contact Betsy Humphreys at blh@nlm.nih.gov). NLM and AHCPR will recruit specific participants if they are needed to test coverage of concepts important to key clinical or public health activities.

Test Procedures

In overview, test participants search the test set for controlled vocabulary needed for their specific purposes and send standard electronic messages to the

NLM about what they find and do not find. Testers use a special Web interface (Figs. 1 and 2) to the UMLS Knowledge Source Server for interactive or "batch" processing of groups of terms. In either case, the following steps apply:

■ Each tester is assigned an identifier that will be included in all data submitted to the NLM.

■ The tester describes the purpose for which controlled vocabulary is needed by selecting appropriate categories (Table 2) from a form provided in the Web interface. If testers are looking for different groups of terms for different purposes, each group of terms is categorized separately.

■ The tester searches for controlled vocabulary for particular concepts of interest using a special configuration of the UMLS lexical programs. The same programs that produced the indexes to the Meta-

thesaurus and the additional vocabularies in the test set convert each term sought to a comparable normalized form. Normalization involves breaking a term into its constituent words, converting each word to lower case letters, converting each word to an uninflected canonical form (e.g., for nouns, the singular), and sorting the words in alphabetical order. For searches against the additional vocabularies, potential British spelling variants of the input term will also be generated and normalized to support searching of the Read system. If no exact match for the normalized form of the desired term is found, "looser" lexical searches against the test are made; these involve such techniques as word-level substitution (e.g., "renal" for "kidney") and partial matches (e.g., "obstructive apnea" and "sleep apnea" are returned for a search on "obstructive sleep apnea"). The interface will display enough infor-

Figure 2 Special World Wide Web interface to the UMLS Knowledge Source Server for processing user terminology. Here the user has entered the term "continuous pump driven hemofiltration." This term is not found as an exact match in the Metathesaurus, but further lexical processing has identified some Metathesaurus concepts that are closely related to the meaning of the user's term.
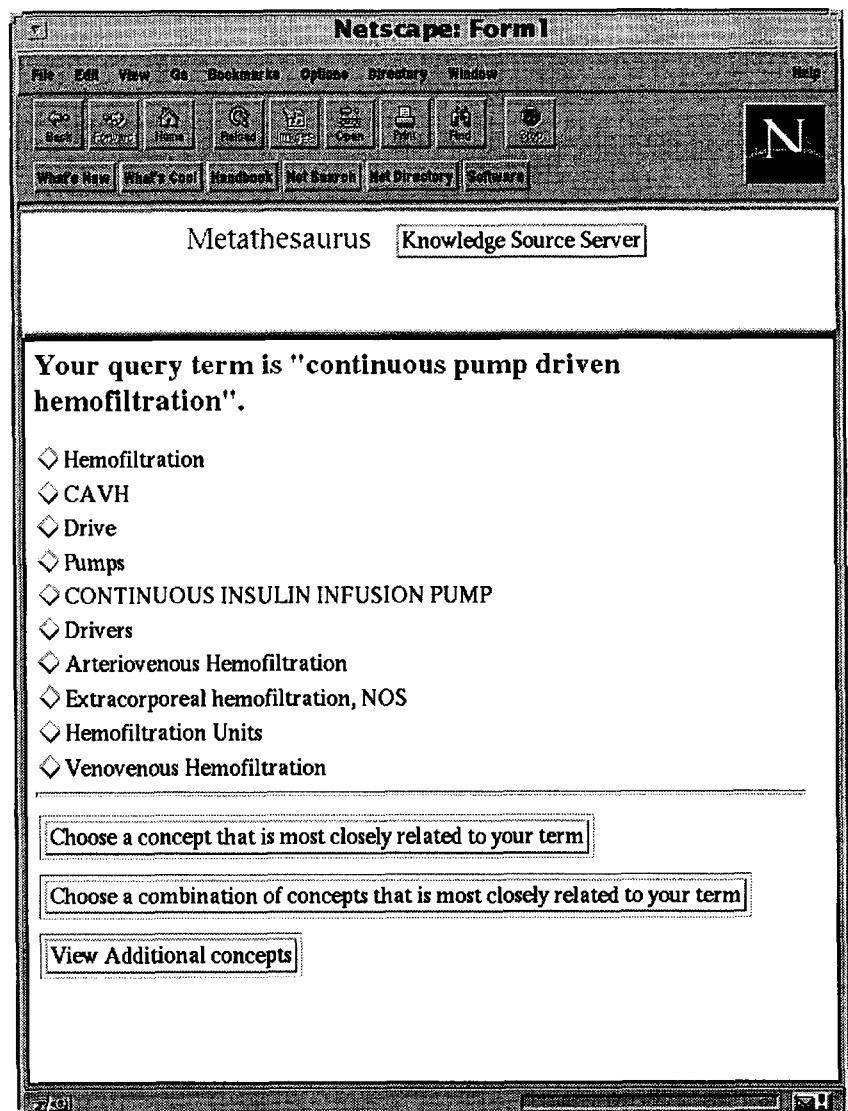
Table 2 ■

Tester's Reason for Searching for Controlled Vocabulary

For each discrete group of terms, select at least one category each from A and B. Select categories from C, D, and E if applicable. In some sections categories may overlap. Select the category that is closest to your view of the task at hand.

A. Data Task (required)
   A1. Record or display data about individuals
   A2. Extract or summarize data about groups of individuals
   A3. Retrieve information from knowledge bases
   A4. Build multi-purpose vocabulary database or tools
   A5. Link natural language to controlled vocabulary
   A6. Other; please specify _____
B. General Purpose of task (required)
   B1. Direct Patient Care
   B2. Decision Support
   B3. Clinical Research
   B4. Public Health Surveillance or Intervention
   B5. Outcomes, Health Services Research
   B6. Development of Guidelines, Pathways, Reminders
   B7. Enhancement of health database or system used for multiple purposes
   B8. Other, please specify _____
C. Care Setting or Facility (required, if applicable)
   C1. Ambulatory Care Office/Clinic
   C2. Inpatient Care Facility
   C3. Longterm Care Facility
   C4. Home Care
   C5. Free-Standing Clinical Laboratory
   C6. Free-Standing Pharmacy
   C7. Other; please specify _____
D. Specific Type of Care or Specialty (required, if applicable)
   D1. Anesthesiology
   D2. Dentistry
   D3. Diagnostic Imaging
   D4. Emergency Medicine
   D5. Family Practice
   D6. Internal Medicine
   D7. Intensive Care/Critical Care
   D8. Neurology
   D9. Nursing
   D10. Obstetrics/Gynecology
   D11. Ophthalmology

   D12. Orthopedics
   D13. Pathology
   D14. Pediatrics
   D15. Pharmacology
   D16. Pharmacy
   D17. Psychiatry/Clinical Psychology
   D18. Social Work
   D19. Surgery
   D20. Urology
   D21. Veterinary Medicine
   D22. Other; please specify _____
E. Specific Segment of Patient Record for which Controlled Terminology is Sought (required, if applicable)
   E1. Chief Complaint
   E2. Problem List
   E3. Discharge Summary
   E4. Medications
   E5. Diagnoses
   E6. Patient History
   E7. Physical Examination
   E8. Review of Systems
   E9. Laboratory Tests
   E10. Procedures
   E11. Progress Notes
   E12. Immunizations
   E13. Family History
   E14. Assessment
   E15. Flowsheet
   E16. Plan
   E17. Intake and Output
   E18. Environmental Exposures
   E19. Demographic Data
   E20. Functional Status
   E21. Consult/Referral
   E22. Patient Education/Teaching Record
   E23. Other; please specify _____

mation to let the tester determine whether the meaning found is identical to the meaning sought.

■ For each term searched, the tester sends an electronic record to the NLM in a standard format. Interactive use of the Web interface generates these records and sends them to the NLM as a by-product of the search process. Testers who employ batch processing produce and send comparable records using locally developed routines. If the exact meaning of the term searched is found as a single concept in either the Metathesaurus or the set of planned additions, the record sent to the NLM includes a variable indicating an exact match. If the exact meaning of the term searched is *not* found, the tester must supply definitional information and a relationship between the concept searched and the most closely related term found. Testers may also suggest useful relationships between two concepts in the test set or identify a combination of two or

more concepts in the test set that is synonymous with a concept for which they were searching.

Data Analysis

After automatic format and preliminary content validation are completed, experts in the vocabularies included in the test set review a sample of records for additional or new concepts. This review determines whether concepts not found by testers are actually present in the test set, although perhaps not retrievable by the search routines used in the test. Based on results of the sampling, additional records may undergo review. Records for any "new" concepts actually found in the test set during this step are modified to reflect the existence of an exact match. Test participants are notified if concepts submitted as new are already in the test set. Those concepts correctly iden-

tified as new are candidates for addition to the Metathesaurus, either through the incorporation of another existing controlled vocabulary that contains them or, failing that, possibly as individual concepts.

When all test data are received, they will be summarized to provide basic statistics—number of testers, number of concepts searched, number of concepts found and not found by testers, number of concepts not found by testers that were located during the review step—both in aggregate and by the purposes for which the concepts were sought. The percentage of terms found for various purposes will serve as one indicator of the relative strength of the test set for various clinical, public health, and research purposes and of any significant gaps in the test set. Test data will also be used to assess tester preferences for precoordinated (e.g., "fractured right leg") versus atomic concepts (e.g., "fracture" and "leg" modified by "right") for various purposes. For concepts not found, some analysis of their frequency of use in health care and public health settings will be undertaken to help determine their relative priority.

Expected Outcomes

We expect the test to have a range of positive effects. Test participants should receive immediate help in solving problems that require controlled health vocabulary. In addition, the NLM will obtain valuable feedback on the utility of its UMLS Knowledge Source Server, the lexical programs used in testing, and priorities for expanding the UMLS Metathesaurus. The test data should increase our understanding of the nature and extent of the gaps in existing vocabularies. Such understanding will assist the developers of individual vocabularies. It will also provide the basis for developing a realistic plan and resource estimates for achieving and maintaining a comprehensive health vocabulary that can be distributed and mapped to statistical, billing, and decision-support vocabularies within the UMLS Metathesaurus. Such a comprehensive vocabulary will help to ensure that the meaning of electronic health data is preserved as the data are used in health care, public health surveillance, and research.

References ■

1. Henry SB, Holzemer WL, Reilly CA, Campbell KE. Terms used by nurses to describe patient problems: can SNOMED III represent nursing concepts in the patient record? J Am Med Inform Assoc. 1994;1:61–74.
2. Campbell JR, Payne TH. A comparison of four schemes for codification of problem lists. SCAMC Proc. 1994;201–5.
3. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR, for Computer-Based Patient Record Institute's Work Group on Codes & Structures. The content coverage of clinical classifications. J Am Med Inform Assoc. 1996;3:224–233.
4. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System Methods Inf Med. 1993;32:281–91.
5. McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. SCAMC Proc. 1994:235–9.
6. McCray AT, Razi A. The UMLS knowledge source server. Med Inf. 1995;8:144–7.
7. Campbell KE, Amar K, Musen MA. A logical foundation for representation of clinical data. J Am Med Inform Assoc. 1994;1:218–32.
8. Board of Directors of the American Medical Association. Standards for medical identifiers, codes, and messages needed to create an efficient computer-stored medical record. J Am Med Inform Assoc. 1994;1:1–7.
9. United States Congress, Office of Technology Assessment. Identifying Health Technologies That Work: Searching for Evidence. Washington, DC: US Congress, Office of Technology Assessment, September 1994. Publication OTA-H-608.
10. Rothwell DJ, Cote RA, Cordeau JP, Boisvert MA. Developing a standard data structure for medical language—the SNOMED proposal. SCAMC Proc. 1993:695–9.
11. Read JD, Sanderson HF, Drennan YM. Terming, coding, and grouping. Med Inf. 1995;8:56–9.
12. Forrey AW, McDonald CJ, DeMoor G, et al. Logical Observation Identifier Names and Code (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. Clin Chem. 1996;42:81–90.
13. Vocabularies for computer-based patient records: identifying candidates for large scale testing. Minutes of a December 5–6 meeting sponsored by the National Library of Medicine and the Agency for Health Care Policy and Research, Bethesda, MD: National Library of Medicine, 1994. Located in http://www.nlm.nih.gov/proceedings.dir/ . . .
14. EUCLIDES. Coding System Version 4.01. Ghent, Belgium: EUCLIDES Foundation International, 1994.
15. ASTM E1238-94. Standard Specification for Transferring Clinical Observations Between Independent Computer Systems. Philadelphia, PA: American Society for Testing Materials, 1994.