J Forensic Sci, January 2009, Vol. 54, No. 1
doi: 10.1111/j.1556-4029.2008.00903.x
Available online at: www.blackwell-synergy.com

TECHNICAL NOTE

Sarah L. Lathrop,[1] *D.V.M., Ph.D.; Wayland L. Davis,*[1] *B.I.S.; and Kurt B. Nolte,*[1] *M.D.*

Medical Terminology Coding Systems and Medicolegal Death Investigation Data: Searching for a Standardized Method of Electronic Coding at a Statewide Medical Examiner's Office*

ABSTRACT: Medical examiner and coroner reports are a rich source of data for epidemiologic research. To maximize the utility of this information, medicolegal death investigation data need to be electronically coded. In order to determine the best option for coding, we evaluated four different options (Current Procedural Terminology [CPT], International Classification of Disease [ICD] coding, Systematized Nomenclature of Medicine Clinical Terms [SNOMED CT], and an in-house system), then conducted internal and external needs assessments to determine which system best met the needs of a centralized, statewide medical examiner's office. Although all four systems offer distinct advantages and disadvantages, SNOMED CT is the most accurate for coding pathologic diagnoses, with ICD-10 the best option for classifying the cause of death. For New Mexico's Office of the Medical Investigator, the most feasible coding option is an upgrade of an in-house coding system, followed by linkage to ICD codes for cause of death from the New Mexico Bureau of Vital Records and Health Statistics, and ideally, SNOMED classification of pathologic diagnoses.

KEYWORDS: forensic science, forensic pathology, medical examiner, information technology, autopsy, public health

Human language is rife with inconsistencies, even among a small subset of similarly trained and educated professionals working within the same field. While conducting autopsies, different forensic pathologists use multiple terms to describe the same condition. Although these synonyms may not present a problem for others reading the reports generated, entering the resulting information into an electronic database and attempting to retrieve it quickly reveal the need for consistent coding of medical terms. The increased use of medical examiner data in public health and epidemiologic studies (1,2) has resulted in an increased need for consistently coded, easily retrievable information. Identifying an adaptable, affordable system with the characteristics needed to successfully achieve this goal can be challenging, and it is one faced by all fields of medicine from research to patient care (3–5).

Background

The New Mexico Office of the Medical Investigator (OMI) is the statewide, centralized medical examiner agency charged with investigating any deaths in the state that are sudden, unexpected, violent, unnatural, or unattended. OMI investigates approximately one-third of all deaths in New Mexico (a state with a 2006 population of 2,010,570) each year (6), and all autopsies are performed at a central facility in Albuquerque. In 2006, OMI investigated 5031 deaths, with 2077 (41%) of these receiving an autopsy. Although OMI does not have jurisdiction over certain federal lands in the state (military installations and tribal domains), the office is frequently contracted to perform autopsies on deaths occurring on these lands. OMI has kept electronic records of death investigations and autopsy data since the mid-1970s, resulting in a large database with over 30 years' worth of death investigation data for the state. Frequently this information proves useful for epidemiologic studies, given the types of deaths investigated (suicides, homicides, drug overdose deaths, deaths of children) and the sheer volume of data collected over the course of medicolegal death investigations (1,2). To maximize the utility of the electronic data for researchers, cases must be readily identifiable by cause and manner, and searchable by more specific parameters, such as demographic variables, geographic location, pathologic diagnoses, and risk factors of interest, such as helmet use in motorcycle fatalities or presence of a note at a suicide scene. With ever-improving technology available, we chose this time to evaluate our coding techniques and analyze the advantages and disadvantages of various clinical terminologies for indexing, storing, and retrieving data generated during 5,000 death investigations per year.

Historically, management of the OMI database had been through the use of MUMPS, the Massachusetts General Hospital Utility Multi Programming System (7,8). Originally developed for use with medical records or any database requiring multiple users to access it simultaneously, MUMPS was specifically designed for searches of text files such as autopsy reports, and had the benefits of minimal hardware requirements and good scalability (9,10). These advantages were offset by its low transaction reliability and its poor integration with other environments, as well as the programming skill required to perform even simple queries (8). Given

[1]Office of the Medical Investigator, University of New Mexico, Albuquerque, NM 87131-0001.

*This study was supported by funding from the Bioterrorism Preparedness and Response Program of the United States Centers for Disease Control and Prevention and the New Mexico Department of Health.

Received 21 Dec. 2007; and in revised form 5 Mar. 2008; accepted 5 Mar. 2008.

the alternative programs now available for searching text files, MUMPS is now little used and widely considered obsolete (9).

To facilitate ease of data queries and accessibility of data, the OMI database was converted to operate on Microsoft's Structured Query Language (SQL) Server 7.0 (a widely used relational database management system) (11) in 2005 in conjunction with the Coroner/Medical Examiner software developed by VertiQ, with a local intranet, active-server web-based interface. Funding at a level of $100,000 was available for this transition from a Centers for Disease Control and Prevention bioterrorism grant administered by the New Mexico Department of Health (NMDOH). Not only was current information entered in this SQL database, but all archival data was transferred as well. With this new system, autopsy reports were indexed and searchable on-line. Autopsy information became more accessible for studies of public health significance than ever before.

One problem remained, however. The methods used at OMI to code cause and manner of death and pathologic diagnoses were developed over the years by various pathologists and staff. As a result, this coding is unique to OMI and not comparable with any other system. The coding is not consistent, varying by pathologist, investigator, and data entry person, all of whom may contribute to the entry of a single record. Unless a researcher knows all possible iterations of a diagnosis, a search for a specific diagnosis is likely to exclude records in which a slightly different term was used, a problem faced by many institutions transitioning from paper records to electronic records (3–5,12). We took the opportunity presented by the migration of the database to a new format to examine available coding systems and determine their usefulness for a statewide medical examiner's office, in an attempt to standardize the coding of pathologic diagnoses and enhance the utility of the tremendous amount of data collected during the course of medicolegal death investigations. In particular, the information collected by medical examiner and coroner's offices can be critical in addressing public health concerns and providing information to local and state public health agencies. The systems under consideration included the current in-house system, Current Procedural Terminology (CPT), International Classification of Disease (ICD-10), and Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT). We present our findings in the hopes of helping provide insight into determining which system would best meet the needs of medical examiner and coroner's offices, as more and more move to electronic storage and retrieval of data and enhanced collaboration with public health agencies and research organizations.

Current OMI Coding System

As one of the first medical examiner's offices to automate their data, there were no standard coding systems available to OMI programmers in 1975 (7,8). A coding system was developed piecemeal, contributed to by various programmers, data entry staff, and pathologists, changing and growing as needed and with minimal expense involved, other than time. While this method worked initially, the increased OMI workload, the increased need for medical examiner data in epidemiologic studies, and the availability of more precise systems call for either integration of a new coding system or an upgrade and overhaul of the current system.

The current OMI computer system requires either investigators or data entry staff to assign an alphanumeric code to cause, manner, and type of death, based on the autopsy and investigation results. As an example, "N43" signifies "blood disorder" as the cause of death, not specifying leukemia, myeloma, or any specific type of blood disorder. Codes starting with "N" are for natural causes of death, "C" is for unnatural, "U" is undetermined, and "X" is other. Manner of death codes are similar, with "A" codes for accidents, "S" for suicides, "H" for homicides and "U" for undetermined, followed by a one- or two-digit number to specify the mechanism or cause, as in S7 encoding suicide by hanging. Deaths are further described by an OMI employee assigning an alphabetic type of death code (TDC). Examples of TDC include "ID" for institutional death, "WR" for work related, and "GNU" for a motor vehicle accident where an air bag was not in use. Codes developed early on were more intuitive, just simple acronyms; but as time went by and more codes were needed they were less likely to reflect the actual text. Numerous people throughout the office assign the codes, leading to the exclusion of certain cases if all possible synonyms are not included in the search. The current system is functional, but suffers from ambiguity in coding and a lack of specificity in information retrieval. Benefits of the current system include the familiarity of OMI staff, doctors, and NMDOH epidemiologists with this system and its proven compatibility with the current electronic database, without the need for new software or a transition period to a new system.

Current Procedural Terminology

The American Medical Association (AMA) developed CPT in 1966 to provide physicians with standard descriptions of clinical procedures (13). Using a four-character code to describe surgical procedures ensured consistency for insurance claims and aided in analyses of procedures. The usefulness of this system prompted the AMA to expand the coding to internal medicine, radiology, and specialties with the second edition in 1970, developing the core five-character code that formed the basis of later editions of CPT (13). Third and fourth editions in the 1970s updated medical terminology. The Health Care Financing Administration (HCFA) implemented CPT as part of its Healthcare Common Procedure Coding System (HCPCS). By the end of the 1980s, HCFA was requiring the use of HCPCS by both Medicare and Medicaid, as well as mandating its use for coding outpatient surgical procedures. With its mandated use in federal agencies and AMA support, CPT is the most commonly used language to communicate clinical, procedural information for administrative and financial purposes (13).

CPT consists of a five-character code, modified by concept attenders for short, medium, or long descriptions of both diagnostic services as well as medical and surgical procedures. By assigning a unique five-character code to a procedure, local and national utilization trends for specific services can be tracked and analyzed, thus allowing CPT to facilitate both insurance billing and health care research (13). The federal mandate and AMA-supported review and updating of CPT ensure it remains a uniform, national, responsive coding system. The system was found to not be readily applicable to medical examiner functions, however, because of its focus on clinical procedures. CPT evolved in the offices and operating rooms of physicians treating living patients, doctors who needed to track procedures for billing purposes. With its emphasis on codifying information for financial and administrative purposes rather than pathologic diagnoses, CPT did not provide the specificity needed for accurate descriptions of autopsy findings. In a 1996 comparison of seven medical classification systems, CPT scored significantly lower than three of the other major systems in the category of procedures, but scored even more poorly in nonprocedure categories (14). The AMA developed CPT to expedite insurance claims and track procedures, neither of which is important for

the field of forensic pathology. CPT is a useful national health care encoding system, but did not meet the needs of OMI.

International Classification of Disease

The World Health Organization (WHO) has long recognized the need for reporting causes of death that are internationally comparable, allowing the collection, classification, and analysis of mortality data from local to global levels (15). This need for consistent, uniform coding of medical conditions was evidenced by the publication of the first ICD system over 100 years ago, in 1900. Over the past century ICD has undergone 10 revisions, reflecting changes in medical terminology and new discoveries. The ninth revision of ICD, in place from 1979 to 1998, was replaced by the larger 10th revision in 1999, changing from strictly numeric categories to alphanumeric categories (15). ICD-10 is used to code causes of death on death certificates, systematically consolidating and ordering reported conditions to produce a single underlying cause of death and multiple nonunderlying, contributory causes.

Within ICD coding, conditions are assigned a three-digit code, followed by a decimal and additional modifying numbers that serve to further describe the specific condition. As an example, pulmonary tuberculosis is assigned the code 011, with 011.1 describing nodular tuberculosis of the lung and 011.2 describing tuberculosis of the lung with cavitation. Four software programs are available to automatically assign these codes to cause-of-death data: SuperMICAR Data Entry, MICAR (Mortality, Medical Indexing, Classification, and Retrieval), ACME, and TRANSAX (15). SuperMICAR automatically codes cause of death into numeric entry codes, while MICAR automates the coding of multiple causes according to standardized ICD rules. ACME automates underlying cause of death coding by applying WHO rules to the MICAR-assigned ICD codes. Data from ACME are converted into a fixed statistical format by TRANSAX, allowing for person-based records (15).

While this system is commonly used in the United States for morbidity and mortality coding, comparisons with other classification systems demonstrates some of it weaknesses. In a 1996 study, Chute et al. (14) found ICD-9, used in conjunction with CPT, failed to capture clinical content abstracted from medical records adequately, and ICD-10 did not perform any better, scoring 1.6 out of 2 for diagnoses, but only 0.62 overall. A 2002 study revealed similar shortcomings of the ICD system, with ICD-9 coded reasons for emergency department visits matching the text entry from the time of the visit in only 40% of included cases (16). However, the ubiquity of ICD codes make them an attractive option for coding cause of death, particularly if the complex coding is performed at an agency which already has the software and expertise needed to correctly and consistently assign the needed codes.

Systematized Nomenclature of Medicine Clinical Terms

As evidenced by the numerous coding systems available and the problems encountered in retrieving and cross-referencing medical examiner data, the consistent coding of medical language is a common problem facing anyone designing studies using data in free text form, whether in medical records or pathology reports (3–5). Programmers, researchers and physicians commissioned by the College of American Pathologists (CAP) to develop Systematized Nomenclature of Medicine (SNOMED), have spent the past 30 years struggling to find solutions to this problem and have created a medical vocabulary which is broad in its scope yet granular in its detail (14,16,17).

Created in 2003 by the convergence of SNOMED RT and the United Kingdom's Clinical Terms Version 3 (Read Codes) SNOMED CT is a structured nomenclature designed to electronically code information in all types of medical documents, from patient records to laboratory results. The basic elements of SNOMED CT are concepts, hierarchies, relationships, and descriptions (18). Procedures, clinical findings, diseases, and therapies can be described by 366,000 concepts, each with a unique name and numeric code, which are then grouped into 19 hierarchies, each of which has sub-hierarchies. Concepts can be linked across these hierarchies through the coded relationships (18). Descriptions provide a preferred term for each concept, as well as one or more synonyms (18). SNOMED relates the multiple terms used by patients and health care providers for one condition to one common concept, linking terms to ideas (17). In addition to creating a final common pathway for synonymous terms, SNOMED provides very precise numeric codes, allowing for granular searches based on anatomic location, severity, and diagnosis.

The consensus of the literature appears to be that SNOMED is the language of choice for electronic coding of medical records (3–5,18). Large health care organizations such as the Veterans Health Administration and insurer Kaiser Permanente are incorporating SNOMED CT in electronic health records (18). Its granularity, compositional structure, and ability to map to other international codes (19,20), give it distinct advantages over ICD and CPT coding for pathologic diagnoses. In the 1996 study cited earlier, investigators parsed 14,247 words from various clinical texts into 3061 concepts, then encoded them using seven different systems and found SNOMED superior to the other six systems based on the degree of matching between the code and the original text (14). McClay and Campbell found that SNOMED coding was significantly more accurate for coding reasons for visits to emergency departments than ICD, providing a lexical match for 93% of the text entries, as compared with 40% for ICD coding (15). In 2003, an agreement between CAP and the United States National Library of Medicine (NLM) provided free access to English and Spanish editions of SNOMED CT through NLM's Unified Medical Language System (UMLS), with annual releases in January, April, July, and October (21). This agreement continues, even after acquisition of SNOMED by the International Health Terminology Standards Organization in 2007 (21).

With proven accuracy and accolades from the medical community, and now free access, why is SNOMED not more widely used? Many medical records are not electronic, making SNOMED impractical (17). Although SNOMED's complexity is one of its advantages, it also makes it difficult to implement, with people preferring to continue using a system with which they are comfortable (15). Even among companies providing fee-for-service SNOMED coding, individual professional SNOMED coders, and physicians, little agreement was found between different users when coding clinical research concepts (22–24), calling into question the system's consistency and reproducibility. With no national incentive to use SNOMED, most health care providers do not see the need to change to a more difficult system, or try to persuade physicians to learn a new system (15,17). In order to implement SNOMED CT at our office, we would need to acquire the NLM UMLS Metathesarus data set (free or at a minimal cost) and then either develop or purchase the software tools needed to access the terminology database and index documents either manually or by using an autocoder. Alternatively, turnkey terminology vendors exist that could provide development, design, and integration services for software systems that index existing documents using SNOMED coding terminology. Their services range from consulting to full-scale implementation. Additionally, vendors can provide a "per piece" indexing service that would take existing documents, scan

TABLE 1—*Comparison of electronic coding system characteristics considered for implementation at the New Mexico OMI.*

System	Utility	Specificity	Generalizability to Other Systems	Cost to Implement/Maintain	Example of Code: Burn
In-house coding	Good for current needs	Adequate for case retrieval	Poor	Minimal	C10 (death because of thermal injuries)
CPT	Minimal for medical examiners/coroners	Good for procedures	Good	Several thousand dollars	16030 (treatment of burns)
ICD-10	Excellent for cause of death	Very good	Excellent	Several thousand (contract with vital records)	T20.0 (burn of head and neck, unspecified degree)
SNOMED CT	Excellent for pathologic diagnoses	Excellent	Excellent	Several hundred thousand if including archival data	284196006 (burn of skin)

OMI, Office of the Medical Investigator; CPT, Current Procedural Terminology; ICD, International Classification of Disease; SNOMED CT, Systematized Nomenclature of Medicine Clinical Terms.

them for terminology, and code them accordingly. Depending on whether the coding is done manually with experienced coders or by automated coding software, the costs will vary substantially. Costs range from $0.25 per code fragment to $7 per electronic document to $15 per hard copy document, scanned, coded, and indexed. Given the large number of legacy records in our agency, the cost of converting them all to SNOMED would be considerable, most likely several hundred thousand dollars or more.

Decision

After studying coding systems and implementation options, with the qualities summarized in Table 1, OMI's information technology (IT) staff conducted both an internal and external needs assessment, to determine if the utility would justify the expense. Pathologists, administrators, and IT personnel within OMI were surveyed to determine opinions regarding current and future coding needs. Questions asked included: (i) What internal OMI functions could be enhanced by using SNOMED to code pathologic diagnoses? (ii) Can these benefits be realized by using existing technology and resources? (iii) How many internal requests for data does OMI receive annually? How are they fulfilled and what is the cost? For the external needs assessment, we asked outside partners (NMDOH, UNM School of Medicine) how often they requested OMI data, if SNOMED coding would be helpful in their data requests, and expected response time for data request fulfillment.

Analyzing the responses to these questions, as well as potential buy-in of OMI's forensic pathologists, utility for in-house searches as well as external agency data requests, IT resource availability, ongoing improvements in computer infrastructure and free-text search capabilities, and funding support for initial costs and annual support, we found that current needs did not justify the expense and system interruption of implementing SNOMED CT at this time. Ideally, the resources needed to transition to SNOMED coding of pathologic diagnoses would be available within the next 5 years.

We then decided to evaluate and upgrade our in-house coding system, working with investigators and pathologists to standardize the assignment of codes, delete outdated codes, and develop a system for assigning new codes as needed. With the office moving toward a "paperless" system, code assignment could be standardized within the new OMI electronic database. An important first step will be to initiate regular auditing of random samples of data, to verify that assigned codes match the text in the file and identify codes where ambiguity is common. Meetings between pathologists, investigators, and data entry staff are underway to clarify code application and improve consistency, as well as identifying additional categories of codes to add. Another option being explored is to hire one person to code all OMI reports, thus ensuring uniformity of coding.

While it would be impractical to code deaths using ICD-10 here at OMI, as it would be for most medical examiner/coroner offices, our state's Bureau of Vital Records and Health Statistics (BVRHS) is already translating OMI causes of death into ICD-10 codes for death certificates. These codes could be transmitted back to OMI and entered in the OMI electronic database, providing the codes without having to implement on-site coding. With the transition to electronic death certificates in New Mexico (initiated in fall of 2007), it may be possible to incorporate multiple cause of death codes for use by both BVRHS and OMI.

Discussion

Many medical examiner and coroner's offices are facing the challenge of converting free-text records into consistent, computer-readable formats, to allow for accurate matches with the original text and easy retrieval of data for research purposes. Considering the options currently available, SNOMED CT is the best, most accurate option available for coding pathologic diagnoses, while ICD codes allow granular classification of cause of death. There are cost considerations and potential system disruptions which may delay implementation of ideal solutions for many offices, as was found to be the case in the experience of OMI. In the meantime, updating the "homegrown" system through more consistent and specific coding will enhance the utility of the OMI database while we work with BVRHS to link our records with theirs, in order to "backfill" ICD codes after BVRHS has assigned them.

With rapid advances in technology, medical examiner and coroner offices will be able to build databases that greatly enhance their utility for research purposes. Several relevant coding tools exist to aid in the retrieval of specific types of deaths, but the availability of financial resources and personnel need to be considered prior to implementation of coding systems, in order to maximize the buy-in of end users and minimize the disruption to the vital daily processes in medical examiner and coroner's offices.

References

1. Shah N, Lathrop S, Landen M. Unintentional methadone-related overdose death in New Mexico (USA) and implications for surveillance, 1998–2002. Addiction 2005;100:176–88.
2. Bennewith O, Hawton K, Simkin S, Sutton L, Kapur N, Turnbull P, et al. The usefulness of coroners' data on suicides for providing information relevant to prevention. Suicide Life Threat Behav 2005;35(6):607–14.
3. Vikstrom A, Nilsson G. SNOMED CT in multidisciplinary clinical practice—evaluation of usefulness for classification and coding of care-planning procedures. Stud Health Technol Inform 2006;124:824–9.
4. Van der Kooij J, Goossen WT, Goossen-Baremans AT, de Jong Fintelman M, van Beek L. Using SNOMED CT codes for coding information

in electronic health records for stroke patients. Stud Health Technol Inform 2006;124:815–23.
5. Richesson RL, Andrews JE, Krischer JP. Use of SNOMED CT to represent clinical research data: a semantic characterization of data items on case report forms in vasculitis research. J Am Med Inform Assoc 2006;13:536–46.
6. http://www.unm.edu/~bber/demo/sts2000s.htm
7. Weston JT. A modern medical examiner facility. J Forensic Sci 1978; 23(1):201–11.
8. Fullilove JP, Weston JT, Reddish CA. NOMIS: a computerized information and communications system for a statewide medicolegal investigative system. J Forensic Sci 1978;23(2):396–412.
9. http://en.wikipedia.org/wiki/MUMPS
10. Schuller G. An open, modular, distributed and redundant computer system for hospitals. Comput Healthc 1991;24–8:31.
11. http://en.wikipedia.org/wiki/Microsoft_SQL
12. Donelly K. SNOMED CT: the advanced terminology and coding system for eHealth. Stud Health Technol Inform 2006;121:279–90.
13. American Medical Association. CPT process. Available at: http://www.ama-assn.org/ama/pub/category/3882.html
14. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. J Am Med Inform Assoc 1996;3:224–33.
15. National Center for Health Statistics. International classification of diseases. Tenth Revision (ICD-10), available at: http://www.cdc.gov/nchs/about/major/dvs/icd10des.htm
16. McClay JC, Campbell J. Improved coding of the primary reason for visit to the emergency department using SNOMED. Proceedings of the Sixth American Medical Informatics Association Annual Symposium, 2002, Nov 9–13, San Antonio, TX. San Antonio, TX: American Medical Informatics Association, 2002.
17. Baldwin G. A flurry of terms, but not many users. New York: Health Data Management, 2002. Available at: http://www.healthdatamanagement.com/HDMSearchResultsDetails.cfm
18. Joch A. A blanket of SNOMED. Falls Church, VA: Government Health IT, 2005.
19. Wasserman H, Wang J. An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. Proceedings of the Seventh Annual American Medical Informatics Association Annual Symposium, 2003, Nov 8–12, Washington DC. Washington DC: American Medical Informatics Association, 2003.
20. Schulz S, Hanser S, Hahn U, Rogers J. The semantics of procedures and diseases in SNOMED CT. Methods Inf Med 2006;45(4):354–8.
21. http://www.nlm.nih.gov/research/umls/Snomed/snomed_faq.html
22. Andrews JE, Patrick TB, Richesson RL, Brown H, Krischer JP. Comparing heterogeneous SNOMED CT coding of clinical research concepts by examining normalized expressions. J Biomed Inform 2008 (Epub ahead of print). DOI: 10.1016/j.jbi.2008.01.010.
23. Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. J Am Med Inform Assoc 2007;14(4):497–506.
24. Chiang MF, Hwang JC, Yu AC, Casper DS, Cimino JJ, Starren JB. Reliability of SNOMED-CT coding by three physicians using two terminology browsers. Proceedings of the Tenth Annual American Medical Informatics Association Annual Symposium, 2006, Nov 11–15, Washington DC. Washington DC: American Medical Informatics Association, 2006.

Additional information and reprint requests:
Sarah L. Lathrop, D.V.M., Ph.D.
Office of the Medical Investigator
MSC11 6030
1 University of New Mexico
Albuquerque, NM 87131-0001
E-mail: slathrop@salud.unm.edu