Clinical Terminologies and Computer-based Patient Records

by Simon P. Cohn, MD, MPH, and Christopher G. Chute, MD, DrPH

Introduction

Words, terms, and codes have had clearly understood functions within patient records for well over a century. Historically, caregivers would describe patient conditions and findings in prose transcriptions, using technical terms for precision and concise communication. Subsequent coding of the encounter, for a variety of purposes, would follow as a distinct and separate process. During the last 30 years, the increasing role of reimbursement classification has formalized and standardized this scenario. Today, medical nosology is a major part of the medical record infrastructure for all healthcare providers in the US, from single practitioners to large, multicampus medical centers.

The American Hospital Association published a major study in 1959 at the beginning of the modern healthcare coding era, comparing the utility of encoding medical records with a highly detailed nomenclature (Standard Nomenclature of Diseases and Operations) versus coding the same information with a high-level classification (an adaptation of ICD-7).[1] Their conclusions shaped the coding world that we know today, pointing out that ICD-style classifications were faster and more accurate. The authors chose, however, to de-emphasize that the breadth of accurate ICD codes belied clinical precision. Nevertheless, the great classification vs. nomenclature debate was waged and raged fiercely for the next decade.

The emergence of computer-based patient records (CPRs) have rekindled this controversy. Kaiser Permanente Medical Care Program, in collaboration with the Mayo Foundation, has allocated resources to develop a common terminology that would be appropriate for our evolving CPR systems. We regard the historical controversy as informative, but essentially irrelevant.

Clinical terms occupy a spectrum from highly detailed pathophysiologic nomenclatures to very broad Aristotelian classifications (animal, vegetable, mineral). We have chosen to focus on elements in between, that retain stated or implied definitions. These interim terms we refer to collectively as the Convergent Medical Terminology (CMT) and are based on the current version of the previously described SNDO (now Standardized Nomenclature of Medicine [SNOMED], developed by the College of American Pathologists). New terms identified by the collaborative effort are being submitted to the College of American Pathologists for possible inclusion in future versions of SNOMED International.

Importance of Terminologies

Those with more detailed, reliable, and comparable data for cost and outcome studies, best practices and guidelines development, and management will be more successful in the marketplace. The intrinsic importance of clinical terminologies is intimately associated with the secondary value that encoded data can provide. The utility of that encoded data will in turn depend on its comparability with similar data, thereby enabling analytic insight into the patterns, profiles, outcomes, and efficiencies of the clinical care process.

Relatively new to healthcare practice are the disciplines of outcomes research, continuous quality improvement, clinical epidemiology, and health services evaluation. All of these activities are similar in their dependence on consistently collected, comparable clinical information. Common terminology, or at least the ability for words to map to a common terminology, underlies comparable clinical data. Put simply, a well-formed clinical terminology enables a clinician to consistently document the healthcare process in the areas of diagnoses, procedures, and perhaps more importantly, symptoms and observable findings. The strategic value of comparable clinical data in turn enables:

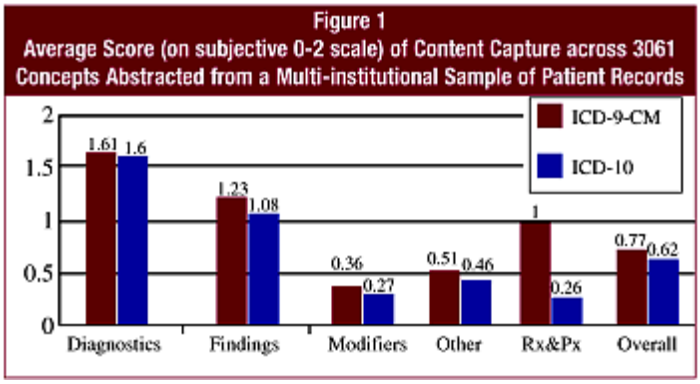
- Cost control
- Improved quality of care
- Improved outcomes
- Improved communication among providers

The advent of CPRs has fundamentally changed the importance of clinical terminologies from an interesting but impractical concern to a central problem within medical record systems. In 1993, the General Accounting Office identified a number of critical technologies and standards that must be in place to make CPRs possible.[2] While many areas such as messaging standards (HL7) or data security were developing or mature, the arena of clinical terminology had not even entered the phase of serious discussion within the US. In the face of this, the Computer-based Patient Record Institute (CPRI) undertook a study to evaluate just how well major terminologies actually worked.

How Well Do Terminologies Work?

CPRI collected samples of medical records from four clinical centers in the Midwest and on the West Coast. Text from inpatient and outpatient notes, histories and physicals, nursing notes, and surgical and radiologic reports were compiled and edited by researchers. From these, the salient concepts and observations were extracted in a way that permitted collaborators to encode these unambiguously identified terms. This encoding was undertaken in ICD-9-CM, ICD-10, CPT, SNOMED International, and the Read codes. The complete report of this work is published in the *Journal of the American Medical Informatics Association*,[3] and the core findings have been presented at many national meetings, including AHIMA's 1996 National Convention.

Figure 1 illustrates the principal conclusion of this study-most major classifications can lose more than half of the underlying information. Represented on the graph are the mean scores as subjectively measured by the authors across 3061 concepts that derived from the source records. These scores were scaled on a 0-2 range, where 0 means that a concept had no acceptable code, 1 implies a code that approximately captured the concept, and 2 means the code accurately represented the clinical concept.



It readily can be seen that ICD-9-CM captures less than half of the clinical data overall (right-hand column). Further, ICD-10 will not substantially improve this situation, performing somewhat less well than the clinical modification of ICD-9. This is not news. Since its second revision in 1909, the introduction of ICD has advised users to not use this classification for clinical purposes.

The tendency of care providers to use ICD classification as a clinical terminology persists. Indeed, many CPR vendors supply ICD as a default terminology for problem lists, order indications, and patient summaries. More importantly, many benchmarks of hospital performance and quality are based on administrative summary data derived from this information. An example is the Health Care Financing Administration (HCFA) Mortality Study, which was issued on an annual basis until recently. It would characterize the quality of a hospital by comparing death rates after discharge within diagnoses. It is readily demonstrated that hospitals with different severities of illness or case mix will fare differently in the mortality analyses. Simply put, hospitals with sicker patients appear worse. This is despite HCFA's reasonable efforts to account for case mix by looking at coexisting conditions; this simply did not directly capture severity of patients' illness.

The physician-oriented CPT classification fared far worse than ICD. Considering its narrow orientation, this is not surprising. It is important to note that CPT did little to augment the content capture of ICD, debunking the widely held perception that administrative data comprised of data from both systems captures most of the clinical information in a patient record. The best performer in the CPRI evaluation was SNOMED International. It captured 87 percent of the clinical content overall, compared to 38 percent captured by ICD-9-CM. Despite its negligible use in the US healthcare industry, it does stand out as a likely candidate upon which to build practical terminologies for use in CPRs.

Developing a Clinical Terminology

Kaiser Permanente recognized the need to generate comparable, detailed patient data as a byproduct of delivering care. The need is driven by the practical requirements of the country's largest managed care organization in the areas of quality improvement, decision support, and resource management. Mayo Foundation also recognized that patient care, education, and research would become vastly more efficient if data was comparable across its group practices and with external patient referrals. Together these organizations agreed to collaborate on refining and extending the CMT initiated by Kaiser Permanente. Research funding to support this effort was secured with a cooperative agreement from the National Institutes of Health (NIH) and the Agency for Health Care Policy and Research (AHCPR).

The CMT project is a participant in the Large-Scale Vocabulary Test[4] of the National Library of Medicine, with emphasis on the best-performing terminology in the CPRI evaluation-SNOMED International. The object of the CMT project is to test the suitability of an enhanced SNOMED to meet our CPR needs. Because the structure and hierarchies of a terminology are important in their use and maintenance, considerable attention is given to how the CMT is put together and cross-linked within itself and to other terminologies. Explicit linkages to ICD-9-CM and CPT are incorporated. To manage these complex relationships, the project uses a sophisticated knowledge representation environment developed by IBM Watson Research labs.[5] This tool is called K-Rep. We have developed K-Rep enhancements to support distributed development with IBM.[6] Thus we are able to support the decentralized development and maintenance of the CMT at multiple sites by identifying inconsistencies such as potentially conflicting additions or changes.

Expansion of CMT's clinical content is occurring at several regions within Kaiser Permanente and with the addition of the Problem Coding terminology of Mayo Foundation. These systems are being implemented at many sites within both organizations, lending real-world evaluations of their completeness, usability, and utility. A major demonstration project of the value of detailed clinical encoding is being designed with academic medical centers within the NIH/AHCPR cooperative agreement and HCFA.

The CMT is intended to become embedded with CPR systems, but may also be used to encode manually created information. Once the clinical detail is captured in the rich representation of the CMT, it can be mapped by computer to required reimbursement or administrative systems. This would diminish the need and expense of duplicative data coding of the medical record, improving accuracy, timeliness, and efficiency.

Implications and Opportunities

Over the next decade, CPR systems will become common at all healthcare locations. It is inevitable that a national standard for a clinical terminology, functionally similar to the CMT, will also emerge. Thus, the implicit encoding of data during the process of care and patient record entry will fundamentally alter the requirements for coding after care into broad classifications, in that such classification can be undertaken algorithmically from the detailed, controlled terminology of the CPR.

Unbounded opportunities exist for involved professionals in health information management. The quality, consistency, and accuracy of the clinical term entry will require extensive quality control infrastructures. The maintenance of local terminologies and their integration into CMT-like national systems will become a major endeavor. The integrity of algorithmic mapping will require periodic validation and audit. Finally, the development and maintenance of the core vocabularies will require substantial resources and input from users.

Healthcare's dependence on detailed, unbiased, precise, and accurate information about the process of care will create enormous demand for trained professionals who understand terminology and classification principles and issues.

Notes

1. Anderson, J.R., ed. "Efficiency in Hospital Indexing of the Coding Systems of the International Statistical Classification and Standard Nomenclature of Diseases and Operations." *Journal of the American Association of Medical Record Librarians* 30, no. 3 (1959): 95-129.

2. General Accounting Office. *Medical ADP Systems: Automated Medical Records: Leadership Needed to Expedite Standards Development.* (GAO/IMTEC-93-17) April 1993:12.

3. Chute, Christopher G., et al. "The Content Coverage of Classical Systems." *Journal of the American Medical Informatics Associatio*n 3, no. 3 (1996): 224-233.

4. Humphreys, Betsy L., et al. "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." *Journal of the American Medical Informatics Association* 3, no. 4 (1996): 281-287.

5. Mays, Eric, et al. "Scalable Expressive Medical Terminologies." *Journal of the American Medical Informatics Association: Symposium Supplement* (1996): 259-263.

6. Campbell, Keith, et al. "Galapagos: Computer-based Support for Evolution of a Convergent Medical Terminology." *Journal of the American Medical Informatics Association: Symposium Supplement* (1996): 269-273.

Simon P. Cohn *is clinical information systems coordinator, Kaiser Permanente Medical Care Program, Oakland, CA.* ***Christopher G. Chute*** *is section head of medical information resources and associate professor of epidemiology, Mayo Medical School, Rochester, MN. Cohn and Chute co-chair the Computer-based Patient Record Institute's Work Group on Codes and Structures.*

Article citation:
Cohn, Simon P., and Christopher G. Chute. "Clinical Terminologies and Computer-based Patient Records." *Journal of AHIMA* 68, no.2 (1997): 41-43.
