Avoiding a War of Words (and Numbers): the Uncertain Future of Terminologies and Classifications

Save to myBoK

by Rosemary Roberts, Kerry Innes, Sue Walker, and Peter Scott

---

With major changes under way worldwide, the evolving roles and relationships of classifications and terminologies are uncertain. But the future lies in collaboration, not competition.

Classifications and clinical terminologies have coexisted for decades. Both use clinical language, but they come from different domains: classifications derive from epidemiology and health information management; terminologies come from health informatics. Both have long histories, as exemplified by the International Classification of Diseases (ICD) from the World Health Organization, SNOMED from the College of American Pathologists, and the Read Codes from the United Kingdom National Health Service.

Until now, the need for and use of health classifications for disciplined reporting of mortality and morbidity statistics were unquestioned, as were the precision and specificity afforded by clinical terminologies. But the future is less clear. Developing clinical terminologies is difficult, and managing the increase in scale and complexity of terminologies is a challenge. Yet they are essential to computer-assisted clinical management and documentation.[1] The adoption of a terminology's patient-centered system can require a significant cultural change from the statistical approach of classifications.[2] Of most relevance in this article is the assumption that clinical terminologies will have to be converted to existing reporting and epidemiological coding systems such as ICD.[3]

Will the uncertain relationship between classifications and terminologies continue? Will one predominate? The degree of synergy and symbiosis between the two systems has yet to emerge. If one believes they are part of a spectrum of granularity of concepts, then there is a better chance of relating one to the other or of describing the class by its component concepts. If, on the other hand, they are regarded as intrinsically different organisms, their relationship is less clear. The emergence of electronic health records will certainly influence the future of data abstraction and reporting.

This article deals primarily with the use of classifications and terminologies in the context of morbidity data. It describes the basis and use of the systems and discusses their possible collaboration. The issue is being discussed against a background of ICD-10 implementation around the world, development of clinical modifications of ICD-10 for morbidity reporting, and the union of SNOMED RT and Clinical Terms Version 3 to form SNOMED CT.

The US and UK country-wide licenses for use of SNOMED CT have provided a catalyst for widespread adoption of SNOMED CT within those countries and examples of implementation for other potential users internationally. The evaluation of clinical terminologies carried out by the National Committee on Vital and Health Statistics has provided important information on the intrinsic value of SNOMED CT and its compliance with criteria for high quality terminologies.[4,5]

Function Follows Form

Traditionally, classifications have been developed, used, and held in a paper environment. Classification manuals and hard-copy updates still predominate in most countries. Until recently, manuals were maintained in word-processing files, making updates and publication extremely resource-intensive procedures. The format has affected classification structure, in that relationships between clinical terms and words have to be relatively simple and straightforward. Diseases with different etiologies and manifestations had to be restricted to one or at the most two places, with conventions and rules to indicate context or interdependence with other diseases or patient characteristics.

In classifications, codes convey meaning, and there are predetermined categories based on body system, etiology, or life phase to which codes are allocated. Classifications are essentially flat and contrived, with conditions usually coded in one place only and labels adapted to cover all terms included in a rubric.

Clinical terminologies, on the other hand, have a much greater agility, are expressed in "natural" language, and are usually maintained electronically. Relationships between terms and concepts and between concepts themselves are indicated by their place in a poly-hierarchy, with interconnections between hierarchies. Terminology codes are meaningless, and terminologies can have many dimensions by virtue of parent-child relationships that constitute their structure. Terminologies aim to represent clinical knowledge through these relationships.

How People and Computers Apply Codes and Terms

The decision on how to classify an episode of care is usually made post-discharge. It involves scrutiny of the record by a health information manager or clinical coder, sometimes in conjunction with a clinician, and allocation of codes for diseases treated and procedures performed during the episode. These codes both record the concepts inherent in the care provided as well as interpret the ordering principles of the classification. Such ordering principles are designed for purposes such as statistical reporting.

Health information managers and clinical coders are trained in the use of classifications, conventions, and rules and in abstracting data from patient records. The process is a manual one, with coders using various resources to determine the correct code or codes to describe the episode and extract meaning in a predictable way. Entry to a class is via a clinical term, which is interpreted for that patient episode by the application of rules. The coding process is carried out by a person who possesses the shared beliefs and business rules of the environment in which he or she works. The clinical term is usually accessed manually through the alphabetic index, although encoders incorporating the logic pathways of ICD, for example, are readily available. The index to ICD is structured according to convention—that is, entry is by disease process rather than site. Many more terms exist in the index than there are valid codes in the tabular list of diseases.

Clinical terming, on the contrary, can be performed by clinicians during the creation of the patient record. Choice of term can inform treatment when the underlying concept is reused as a point of entry to clinical pathways and protocols, prescriptions and allergy information, or to access the medical literature. Similarly, because clinical terms and the concepts for which they may be synonyms may be uniquely identified by code, they can be transmitted electronically in a reasonably unambiguous manner, forming part of messages derived and exchanged in a standard way within the health system.

Computers rely on formal appeals to the reference terminology for their knowledge. Unlike people, computers do not implicitly understand links between the name of a concept and that to which the name refers in reality. Large computerized knowledge bases are built on concepts and the relationships between them as expressed by hundreds of thousands of propositions. Indeed it can be argued that reference terminologies exist expressly to support computers.[6] Issues relating to the context of clinical terms and concepts expressed within the electronic health record are the subject of ongoing work in the terminology world. In reporting, there are as yet few if any predetermined groupings to use for national or international data extraction and interpretation.

The Pros and Cons

The essential predictability and shared understanding of health classifications are their virtues. The long-standing use of health classifications to represent morbidity data nationally has allowed longitudinal studies of health patterns with relative stability between revisions. The down side of classifications is their cumbersome nature, which has resulted from their paper evolution. Maintaining those classifications in databases is now providing revolutionary access to terms within classifications through unique identification of index terms and expression of relationships between terms and code classes. Classifications can be maintained and published electronically with provision for subsets and supersets derived from the core classification. Classification rubrics can be described by the index terms that relate to them. Ostensibly, this could lead to a merger between the tabular list and index with uniquely identified index terms within each class to further flesh out the categories, especially unspecified or other specified diseases of a system or organ.

Because of their electronic format and distribution, clinical terminologies can be updated frequently to reflect clinical knowledge and language. Specific, recognizable terms are connected to clinical information resources so that clinicians can at

the same time compose the patient record and capture terms for treating the patient and for describing that treatment. They exist in real time. The use of clinical terminologies for reporting has not yet been developed, although there has been some discussion about navigational hierarchies that act as templates for data retrieval.

Case-mix groupings are being used around the world to further lump patient episodes of diseases and procedures into clinically and financially homogeneous groups. Such groupings, commonly called diagnosis related groups (DRGs), are based on ICD codes. At present, any connection between clinical terms and DRGs is through a version of a health classification, usually ICD.

Collaboration: Mapping and Joint Use

Countries such as the United Kingdom and the US, where terminologies and classifications have been in joint use for some years, have undertaken disciplined mappings between specific terminologies and classifications. In the UK, the National Health Service has mapped terms in Clinical Terms Version 3 to classes in ICD-10. These have evolved into mappings between SNOMED CT and ICD-10 diseases. In the US, SNOMED International has mapped SNOMED CT to ICD-9-CM diseases. In both countries, these have been large-scale and disciplined efforts. SNOMED International has a working group on mapping to identify standard mapping practices and to agree on definitions of use cases and mapping categories. During the alpha trials of SNOMED CT, Australia's National Centre for Classification in Health undertook preliminary mappings between specific domains of SNOMED CT and ICD-10-AM. Part of this exercise involved capturing the ICD-10-AM index term, so that the map trail included SNOMED CT concepts, ICD-10-AM index terms, and ICD-10-AM codes. This process was facilitated by assigning identification numbers to the ICD-10-AM index terms.

While such mappings allow an automated relationship between a term and a code, many still require human intervention where the context of the code is not straightforward. Two examples are diabetes in a neonate or in pregnancy and hypertension with renal disease.

Classifications and terminologies each have merits. They were designed for similar but different purposes. The adoption of terminologies will both affect the development of electronic health records and be affected by the extent of electronic health record use. The overlap between classifications and terminologies will be governed by how well classifications adopt electronic means of holding and presenting the classification and its relationship with terms and terminologies. The implicit knowledge in classifications needs to be brought out and formalized. Harmonies may well exist between the terminology and the classification as this process happens.

Some criticism of ICD relates to its lack of specificity.[7] This can be remedied in part by more formally merging the expressivity of a large number of index terms with code descriptions and inclusion terms in tabular lists and uniquely identifying terms within classes. More specificity is provided through rules and extension of codes into fifth characters, at the cost of increasing levels of pre-coordination (combination of more than one atomic concept in the one code and associated code title).

The Unified Medical Language System has shown us one way in which terminologies and classifications can be linked electronically.[8] Another example from the US is the National Center for Health Statistics' Mortality Medical Data System, which translates causes of death as documented on the death certificate to ICD code.

Maintaining classifications and terminologies is extremely resource intensive. We must find a way of collaborating between organizations and individuals responsible for classifications and terminologies. We can't afford to compete. We must be clever in our use of information technology to build a triangular or tripartite partnership between clinical terminologies, classifications, and the electronic means of holding and relating the descriptions of disease processes and health interventions. Electronic control, maintenance, and distribution of each system and maps between them are the first steps in expressing the nature of the relationship and in helping make it work.

For Instance . . .

Reporting Queries in SNOMED CT and ICD-10-AM

Classifications and terminologies represent their concepts in different ways. Consider how the ordering principle of "abnormalities of breathing" deals with the clinical concepts of dyspnea (shortness of breath) versus orthopnea (shortness of breath lying flat). In ICD-10 and ICD-10-AM (the Australian modification of ICD-10), both these terms find a home in the statistico-clinical concept R06 Dyspnea. In a reference terminology such as SNOMED CT they possess their own unique

concept IDs; however, the latter is subsumed by the former, so a query on the concept of dyspnea is likely to find cases of orthopnea.*

Where the business of reporting may diverge for ICD-10 and SNOMED CT is where people are needed to both interpret and reconcile (within the complete record) the communicative intent of clinical documenters and apply their tacit knowledge to the reporting task at hand. The rules of the classification need interpreting. Relationships with clinicians need to be built up to further communication, not least to enable clarification of clinical documentation.

As an example, in Australia it has been decided to collect specific information about the connection of diabetes to other medical conditions. If a clinician has documented the phrase "Diabetes, renal failure," the coder seeks to be more specific about the type of diabetes, the presence of proteinuria, and the nature of the renal failure (end stage, acute). The "correct" codes may lie within fifth-character specificity within the ICD-10 parent rubric E10.2 Type 1 diabetes mellitus with renal complication, along with an appropriate code such as N18.0 End-stage renal disease. To do this, the coder uses his or her implicit understanding of the health domain, goes through the whole medical record, and if necessary "corridor chases" hospital clinicians. To achieve similar precision in SNOMED CT, the clinician (who is unaware of coders' business rules), would need to have recorded "Diabetes Type 1 associated with renal failure," by post-coordinating the atomic elements within the phrase. Even with a correct set of terminology concept IDs, we are missing data about the presence or absence of albuminuria and the degree of the renal impairment.

Consider further the ICD-10-AM statistico-clinical concept of "burn." This has informal attributes of site and thickness. Rules exist in ICD-10-AM to assist in the coding process that takes these attributes into account. Similarly, "gastroenteritis" (without further specification) seems to require the age of the patient (above or below 15 years of age) to enable correct code assignment.

Therefore it seems that the classification can be:

- Interpreted with implicit knowledge by people rather than computers, in the main
- Full of rules, notes, and structure that allude at times to complex statistico-clinical concepts
- About the business of aggregating while allowing clinical precision

Thinking of a classification as simply a computerized query on a given set of term data appears overly simplistic.

* Subsumption doesn't help where concepts are expressed as a post coordination of concepts, but where these are equivalent to a given precoordinated concept, mechanisms exist to reliably retrieve them.

Notes

1. Rector, A.L. "Clinical Terminology: Why Is It So Hard?" *Yearbook of Medical Informatics* (2001): 286–99.
2. Ibid.
3. Ibid.
4. National Committee on Vital and Health Statistics Subcommittee on Standards and Security. Draft Report. "Summary and Analysis of Terminology Questionnaires Submitted by Developers of Candidate Terminologies for PMRI Standards." Version 2. April 17, 2003.
5. Cimino, J.J. "Desiderata for Controlled Medical Vocabularies in the Twenty-first Century." *Methods of Information in Medicine* 37 (1998): 394–403.
6. Rector, A.L. "Thesauri and Formal Classifications: Terminologies for People and Machines." *Methods of Information in Medicine* 37 (1998): 501–9.
7. Slee, V. "But Should We Switch to ICD-10-CM?" *Health Care Watch* 1 (2003). Available online at www.imakenews.com/healthcare/e_article000131185.efm.
8. Campbell, K.E., D.E. Oliver, K.A. Spackman, and E.H. Shortliffe. "Representing Thoughts, Words and Things in the UMLS." *Journal of the American Medical Informatics Association* 5 (1998): 421–31.

Rosemary Roberts *(r.roberts@fhs.usyd.edu.au) is director,* ***Kerry Innes*** *is associate director (Sydney),* ***Sue Walker*** *is associate director (Brisbane), and Peter Scott is health informatician at the National Centre for Classification in Health in Australia.*

Article citation:
Roberts, Rosemary, et al. "Avoiding a War of Words (and Numbers): the Uncertain Future of Terminologies and Classifications." *Journal of AHIMA* 75, no.9 (October 2004): 26-32.

Driving the Power of Knowledge

Copyright 2022 by The American Health Information Management Association. All Rights Reserved.