WORLD HEALTH ORGANIZATION

WHO/GPE/CAS/C/02.83
Distr.: LIMITED
ENGLISH ONLY

MEETING OF HEADS OF WHO COLLABORATING CENTRES FOR THE FAMILY OF INTERNATIONAL CLASSIFICATIONS

Brisbane, Queensland, Australia
14-19th October 2002

Title: The Who Family Of International Classifications And Its Relationship To Clinical Vocabularies

Authors: David J. Berglund, M.D., M.P.H.
Medical Officer, National Center for Health Statistics
Centers for Disease Control and Prevention (CDC)
Hyattsville, Maryland, USA

Willem Hirs, M.D.
Department for Public Health Forecasting (VTM), RIVM
The Netherlands

André L'Hours
Classification, Assessment, Surveys and Terminology
Health Financing and Stewardship
World Health Organization
Geneva, Switzerland

Catherine Sykes
Australian Institute of Health and Welfare
Canberra, Australia

Purpose: For information

Abstract:

This paper addresses differences between detailed clinical vocabularies and aggregate classifications, issues related to mapping between vocabularies and classifications, and potential for machine-readable definitions of classification codes. Specific vocabularies and aggregate classifications have many differences, related to their different purposes. Vocabularies ideally provide specific codes for each concept, incorporate discoveries quickly, and have multiple hierarchies. In contrast, classifications aggregate for a balanced number of exhaustive, mutually exclusive, stable categories.

Specific vocabularies usually can be mapped to aggregate classifications. However, issues arise. Complex definitions, exclusions, and other factors complicate mappings. Automated mappings may be desired but not fully achievable. Recognizing correspondence of vocabulary representations to classification codes requires complicated rules, or human review. Map creation might be partially automated, but requires manual validation.

Using a reference terminology, machine-readable definitions of classification codes could be created, enabling development of complex mappings, and sophisticated computerized analyses. However, again issues arise. Combining codes consistently requires a grammar. Definitions could be considered at different levels. To be practical, definitions should use general concepts, not excessive specificity. With electronic medical records, such code definitions have significant potential in morbidity applications, but will require people with detailed understanding of both the vocabulary and classification for use.

This paper will address some differences between detailed clinical vocabularies and aggregate classifications, issues related to mapping between vocabularies and classifications, and potential for machine-readable definitions of codes in a classification.

Differences Between Classifications and Vocabularies

There are a number of differences between specific clinical vocabularies and aggregate classifications, related to the different purposes these have. For purposes of this paper, "terminology" will be used as an umbrella term, "vocabulary" will mean a clinically specific terminology, and "classification" will mean an aggregating terminology. From the practical standpoint, there may not be a sharp demarcation between classifications and vocabularies. Some terminologies may have some characteristics of both. Classifications can have definitions with the similar and distinguishing characteristics of each category, such as in vocabularies. Vocabularies can be presented in the form of a hierarchical classification.

The difference in principle would be the difference in purpose, of aggregation for classifications, and discrimination for vocabularies. Aggregation is achieved by skipping non-essential characteristics to group almost similar objects. In discrimination, the purpose is to find new characteristics in order to identify every separate object.

The optimum level of detail for a terminology will depend on its potential applications. Vocabularies should ideally provide a specific code for each specific concept. In contrast, classifications should aggregate to provide a balanced number of categories. For clinical use, more detail is needed than for epidemiological or administrative uses. Clinical subspecialty use of a terminology may require even more detail than other clinical use, for finer discrimination.

With increasing medical knowledge, more detail is added to classifications, for greater discrimination. While this can lead classifications and vocabularies to look more similar, the differences in principle and purpose remain.

While different descriptions may be given of the ideal characteristics of a clinically specific vocabulary, one frequently cited set of criteria is the desiderata described by Cimino in a 1998 update to an earlier 1989 description. (Cimino, 1998; Cimino, 1989) However, it should be noted that these desiderata are intended to be applicable for a vocabulary that is multiple purpose, not necessarily for those which are for specific purposes. As such, these criteria should not be applied broadly to classifications, although some of them may also be desirable for classifications. Certain of these desiderata will be discussed here, and a more complete listing and discussion of these will be provided as an appendix.

Classifications have the purpose of aggregation, with multiple concepts sometimes being represented by a single code. Cimino's desiderata would call this ambiguity, and consider it undesirable. However, this would be an example of where the different purposes for classifications and vocabularies lead to different interpretations of what is best.

Graceful evolution is one of Cimino's desiderata, which would be desirable for both classifications and vocabularies. However, what makes for graceful evolution may differ depending on the purpose. The ideal timeframe for updates differs, depending on purpose. For vocabularies, shorter time between updates is better, to quickly incorporate new knowledge. In contrast, for classifications a longer time is reasonable, to allow for more stability, for ease of tracking over time, and for ease of use.

There are some difficulties with placing all medical entities in a single-variable axis classification such as the ICD. However, this does have benefits for certain uses, including accurate enumeration. Cimino suggests that polyhierarchy or multiple classifications are the best approach, because no single hierarchy will be universally acceptable. This allows a concept to be a descendant from more than one parent, and an entity may be found in more than one location or more than one way in the system. (Cimino, 1998) However, this adds complexity which is not necessary for a classification, which needs mutually exclusive categories. A classification may enable finding an entity related to another part of the hierarchy by use of an excludes note.

For classifications, in order for categories to be exhaustive, an "Other, not elsewhere classified," category is generally provided. This is specifically considered undesirable in Cimino's desiderata, so other approaches are recommended which may be implementation dependent.

Mapping Between Vocabularies and Classifications

Cross-mappings between terminologies are desirable, to enable comparison of data collected using one system with data in another system. An ideal expressed by some is that detailed clinical data could be collected using a clinically specific vocabulary, and could then both be used clinically, and also mapped to other systems such as classifications for administrative and research purposes. The intent is to automate such mapping as much as possible. For example, data could be collected using a system such as SNOMED or Clinical Terms Version 3, and then mapped to a classification such as ICD-9-CM or ICD-10.

It is usually possible to map from more specific vocabularies to more aggregate classifications. However, issues can arise. In theory, a mapping from a vocabulary to a classification should consist mainly of n-to-1 relations, with some 1 to 1 relations possible. However, 1-to-m relations could appear to indicate inconsistency in the vocabulary or classification. In practice, vocabularies and classifications may be imperfect, and attempts to map from one to the other may generate a number of 1-to-m relations, which require review and resolution.

One explanation for some 1-to-m relations in mapping from a vocabulary to a classification is that the classification may include more complicated diagnoses with multiple findings, and consequently with multiple concepts. Assigning a single code to these more complicated concepts gives a pre-coordinated representation. In some cases a pre-coordinated

representation may be given in a classification, but not in a vocabulary. For such cases, the vocabulary may then require multiple codes for a post-coordinated representation. Such complex definitions and exclusions, among other factors, can make mapping from vocabularies to classifications extremely complicated.

Automation of mapping might work for some content, but it would not be practical for all cases. Similar issues arise as those found in automating mortality coding, where a certain percentage of cases must be coded by hand. Not all needed information may be represented by the vocabulary, and even if all the information is represented, properly recognizing the correspondence to a specific code or set of codes in the classification can require complicated rules and human review. Creation of such mappings also might be partially automated, but would require significant manual review and testing for validation.

Work on mapping between terminologies has been done in a number of places. Some of the most extensive work on mapping between terminologies has been done with the Unified Medical Language System (UMLS), in the Metathesaurus. It provides concepts, and relationships between concepts, from a number of source terminologies. (Bodenreider 1998) It is possible to use the relationships in the UMLS to map between different terminologies. For example, Cimino et al developed a method of automated mapping of ICD-9-CM preferred terms to UMLS Metathesaurus concepts, which were then linked to Medical Subject Headings (MeSH), which are used for indexing medical literature. (Cimino, 1994) A similar approach was used to link non-MeSH terms to MeSH terms in 1998. (Bodenreider, 1998) Since that time, MeSH has been changed to have a more concept-based underpinning. (Savage, 2000) This would allow more detailed representation of semantics, and might improve results, were this to be repeated.

Machine-Readable Formal Definitions for a Classification

A reference terminology could be used to create machine-readable formal definitions of terms or codes in a classification. This would enable development of more complex mappings, and sophisticated types of analyses using computers. Such definitions could be considered at different levels. One potential application of such definitions is in the electronic medical record, with the idea of encoding the entire record using a vocabulary, and then automating assignment of ICD codes. While certain types of uses such as in an expert system for assisted diagnosis could benefit from more detailed clinical definitions, that detail would not be needed for many applications.

It would not be feasible to attempt very specific clinical definitions for purposes of the classification, such as defining hypertension in terms of specific blood pressure levels. Creating definitions would still require significant input from clinical experts. Additional issues would be raised in attempting to create such definitions, since different clinical experts may use the same natural language term but with different meanings. This could make creating such definitions in some cases extremely problematic, or even impossible.

Formal definitions involve a concept being defined using a symbolic form (able to be manipulated by computer), to express a collection of relationships with other concepts in the vocabulary. Cimino describes these as being widely desired for vocabularies. (Cimino, 1998) It could be argued that these should not be used for classifications, but rather terms should be used with accepted medical meanings. On the other hand, it is possible these could also be useful for classifications. One benefit could be to avoid semantic drift, where term meanings may change over time.

For a vocabulary, other concepts within the vocabulary could be used for formal definitions. The lower level of detail in a classification would keep it from being feasible to define codes in the classification only in terms of other codes. It would be possible to create such definitions for a classification, using concepts in a vocabulary (or from multiple vocabularies, for specific domains). However, this would require work to enumerate all terms and concepts actually associated with each code.

It should be noted that combining multiple codes in a post-coordinated fashion is complicated. A grammar is needed to specify how the codes combine consistently, and interpret the meaning of the combined codes. Existing vocabularies might be able to express definitions of concepts in a classification directly, or might require modification to properly capture all such definitions.

One system with a well defined method for combining codes is the Generalised Architecture for Languages Encyclopaedias & Nomenclatures (GALEN), with the GALEN Representation And Integration Language (GRAIL). However, GALEN does not have the depth of content of SNOMED CT. Also, making use of GRAIL can be complex, and could require significant effort. On the other hand, SNOMED CT does not have such detailed methods for combining codes, even though it is intended to function as a reference terminology. The SNOMED CT structure for cross-maps could possibly suffice, but a SNOMED CT extension and subset for ICD might be needed. Another potential system that might be considered for such use would be the Medical Subject Headings (MeSH), with its new concept basis. However, MeSH was designed for other purposes, so it is not clear how well it would work for this purpose.

As an interim approach, it would be possible to use a concept clustering approach similar to that adopted for MeSH. (Nelson, 2001) Since the change to a concept basis for MeSH structure, a MeSH descriptor (used for indexing) can be considered a class of concepts. In turn, a concept is a class of synonymous terms. In order to partition concepts for organizing information in MeSH, concepts must be members of exactly one descriptor class.

For the ICD, this approach would involve defining relationships between different terms associated with a code. It would then be possible to identify where one term “is-a” more specific example of another term. Synonyms for terms could be identified, as well as related but non-synonymous terms. For such an approach to be applied, it could be practical to evaluate using the UMLS and MeSH. A related issue would be identifying terms; full code titles would be needed. Further, index entries would need to be normalized to create terms for comparison with other systems (thus listing the text string with words in natural order). The entry terms and code titles for ICD-9-CM which have been created for the UMLS could be reviewed. This would require significant time to accomplish, but less than recreating these entirely.

Discussion

One of the issues to consider in evaluating the potential benefit of developing a system of formal relationships or definitions for a classification using a clinically specific vocabulary would be the licensing available for the vocabulary. If a time-limited license were used, which had to subsequently be renewed or renegotiated, then the cost of using the data could potentially increase substantially, or the data could even become unavailable for use in the future. It would be important if a proprietary vocabulary were to be considered, that pricing be reasonable and non-discriminatory. This would also require transparency of pricing.

MeSH is freely available in English, in the United States. Derivative translations to other languages are available, which may be proprietary. While the UMLS Metathesaurus is freely available, it includes various proprietary terminologies, and thus requires signature of a license agreement prior to usage. This limits uses of the data without further separate agreements with terminology copyright holders. GALEN is freely available from OpenGALEN, and its license appears to require that derivations remain freely available. SNOMED is proprietary, and unfortunately, the licensing has not been transparent in the past, although that could change.

A classification serves the purpose of counting diseases well, and thus meets epidemiological needs. This is not a purpose of a vocabulary, and how well it could serve for counting diseases could be implementation dependent. Vocabularies can be used to describe diseases, or specific cases of disease. The increased detail of a vocabulary lends itself well to clinical applications. However, detailed classifications can also be useful for certain clinical applications. The potential for similar uses of classifications and vocabularies supports the idea that these are part of a continuum of terminologies, ranging from detailed to aggregate.

Some vocabularies can enable different views or levels of detail to be considered. Likewise, a classification can be used with different levels of detail. For example, category level ICD data could be used for some research purposes.

Fundamental challenges remain in how to make classifications and vocabularies more compatible. Support for mappings remains a significant challenge, both from the classification side (e.g., for the ICD in normalizing terms), and from the vocabulary side (e.g., in providing a grammar or way to combine codes, or a structured approach to mappings). Normalized terms could be produced, associated with ICD codes (to include all terms from the index and other places). This could support mappings from vocabularies. However, producing such a list would be resource intensive. Some method of at least partially automating production of such terms would be needed, for this to be feasible. Work done in the UMLS Metathesaurus could be reviewed, for potential application.

With development of electronic medical record systems, machine-readable definitions of codes have significant potential use in morbidity applications. However, there will be a need for people with detailed understanding of both the vocabulary and classification, to properly set up analyses and to interpret the results.

Bibliography

Bodenreider O, Nelson SJ, Hole WT, Chang HF.

Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies.

J Am Med Informatics Assoc (Symposium Suppl) 1998;:815-9.

http://www.nlm.nih.gov/mesh/beyond.html

Cimino J.
Desiderata for Controlled Medical Vocabularies in the Twenty-First Century.
Meth Inform Med 1998; 37:394-403.

Cimino J, Johnson S, Peng P, Aguirre A.

From ICD9-CM to MeSH using the UMLS: a how-to guide.

Proc Annu Symp Comput Appl Med Care 1993:730-4.

Cimino J, Hripcsak G, Johnson SB, Clayton PD.

Designing an Introspective, Multipurpose, Controlled Medical Vocabulary.

In: Kingsland LC, ed. Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care. New York: IEEE Computer Society Press, 1989:513-518.

Nelson SJ.; Johnston D, Humphreys BL.

Relationships in Medical Subject Headings.

In: Bean, Carol A.; Green, Rebecca, editors. Relationships in the organization of knowledge. New York: Kluwer Academic Publishers; 2001. p.171-184.

http://www.nlm.nih.gov/mesh/meshrels.html

Savage, Allan.

Changes in MeSH Data Structure.

NLM Tech Bull 2000;:March-April.

http://www.nlm.nih.gov/pubs/techbull/ma00/ma00_mesh.html

Madden RC, Üstün TB, Sykes CR, Ashley J, Hirs WM, Schiøler G

The WHO Family of International Classifications

Appendix – Desiderata of Cimino

This gives a brief overview of the desiderata for controlled medical vocabularies described by Cimino (see Cimino, 1998), and comments on how these should apply for classifications. The primary intent for these desiderata is to describe characteristics which would enable a terminology to be used for multiple purposes. Thus, it is important to note that not all of these should be applied to classifications intended for specific purposes.

Content

To meet multiple purposes, addition of detailed content is one of the most needed areas for vocabularies. More detailed content is needed for a vocabulary than a classification. Both vocabularies and classifications need a formal methodology for expanding, but the optimal timeframes for updates differ.

Concept orientation

Concept orientation is identified as nonvagueness, nonredundancy, and nonambiguity. The ideas of nonvagueness (terms must correspond to at least one meaning) and nonredundancy (meanings correspond to no more than one term) are important for both vocabularies and classifications. However, the idea of nonambiguity (terms have no more than one meaning) is not consistent with the purpose of aggregation with codes in a classification.

Concept permanence

For stability, meanings of codes cannot change, and concepts cannot be deleted. This is important for both vocabularies and classifications. However, some caveats are that indexing of new terms, and expansion of an "Other" category should not be considered a change of meaning, as may be the strict interpretation of Cimino.

Nonsemantic concept identifiers

Identifiers should not convey hierarchical information, to avoid running out of room to expand in specific areas, and also more readily handle polyhierarchy and relocation of codes within the hierarchy if needed. On the other hand, hierarchical codes have the advantage of being more understandable to humans. While Cimino suggests that humans will not need to see the codes, it appears likely that this will be necessary for the immediate future. These may be beneficial for vocabularies, but classifications have a goal of limiting categories to a manageable number. Thus, this should not be applied to classifications.

Polyhierarchy

Cimino suggests that polyhierarchy or multiple classifications are the best approach, because no single hierarchy will be universally acceptable. This allows a concept to be a descendant from more than one parent. This adds complexity which is not necessary for a classification.

Formal definitions

A concept can be defined using a symbolic form (able to be manipulated by computer), to express a collection of relationships with other concepts in the vocabulary. Cimino describes these as being widely desired for vocabularies. It could be argued that these should not be used for classifications, but rather terms should be used with accepted medical meanings. On the other hand, these potentially could also be useful for classifications, as explored in more detail in the final section of this paper.

Reject "Not Elsewhere Classified"

It may be feasible for clinically specific controlled vocabularies not to use "Not Elsewhere Classified" categories, but such categories are necessary for classifications. This allows easy distinction from "Not Otherwise Specified" terms, which otherwise could not be distinguished, unless systems had other changes which would be implementation dependent to allow capturing the information.

Multiple Granularities

Different levels of detail could be provided for different uses, both for a clinically specific controlled vocabulary, and for a classification system. However, for clinical use, rules requiring that the highest known level of detail be coded in a classification are helpful to ensure that the data be of most use. The use of multiple granularities in a clinically specific vocabulary may be more complex, allowing different levels for clinical use by different specialties, for example.

Multiple Consistent Views

In systems with polyhierarchy, multiple consistent views allow looking at information in different ways, and ensure that the same information is available regardless of the approach used to find it. As a negative example, the Medical Subject Headings (MeSH) provide different (or inconsistent) views for some parts of the hierarchies, depending on the specific way the hierarchy was found within the MeSH polyhierarchy. This allows searches to be more specific based on the requested information. Part of the reason for this is the MeSH use of descriptors which include multiple concepts. In any case, multiple consistent views would be more applicable for clinically specific vocabularies, but not for classifications.

Beyond Medical Concepts: Representing Context

The idea of providing formal, explicit information about how concepts are used is intended for clinically specific vocabularies. Classifications such as the ICD may represent context in certain codes, and have guidelines which specify how the codes should be applied in various contexts. However, classifications do not have the same issues in this respect as clinically specific vocabularies.

Evolve Gracefully

This is needed for all terminologies. However, such graceful evolution may be defined differently for a classification and a clinically specific vocabulary, with the classification being more stable, and the vocabulary being updated more promptly.

Recognize Redundancy

This is important for both classifications and clinically specific vocabularies. For classifications, this involves care when setting up the categories. The ICD uses excludes notes to assure that related disorders which belong elsewhere are properly excluded from a code or category. Redundant representations have been an issue for practical use of vocabularies which allow for complex combinatorial representations of concepts. Description logic can also be used to resolve this.