A Clinical Terminology in the Post Modern Era: Pragmatic Problem List Development

CG Chute, MD, PL Elkin, MD, SH Fenton, GE Atkin
Mayo Foundation, Rochester, MN

ABSTRACT

A brief review of the rich heritage of classifications and terminologies is the background for a description of the Mayo Clinic's clinical terminology development. Vender specific system constraints prompted the scope and style of an interim problem list vocabulary. We describe the sources and review process which led to a working terminology for use in a Computer-based Patient Record (CPR).

Because terminology development is often subjective and metrics against which to measure the quality of individual human judgements are few, we decided to compare the selection of preferred terms made by general internists with those made by sub-specialists. A significant difference between a sub-specialist's assignment of preferred terms and a general internist's (948 vs. 2271, $P<0.001$) was observed. Sub-specialists were less than half as likely as a generalist to designate a term as a preferred form. These results emphasize the need for sub-specialty editing when assigning preferred terms to concepts.

INTRODUCTION

Clinical Terminology has evolved over the past four hundred years, dating from the London Bills of Mortality[1] to the elegant knowledge relationships in Galen[2]. Chronicling this intellectual development could require volumes. However, an art period metaphor may bring simplicity to a tumultuous evolution.

If our tradition began with Graunt and his writings on the London Bills of Mortality, than it's direct descendents are the international classifications of disease which have paraded in various versions and revisions since William Farr and the First Statistical Congress of Brussels in 1853[3]. Here the nascent "ICD" was born, generating the Classical tradition of terminology which focused upon systematic classification of mortality. The rise of health care costs and demands for outcomes information commanded a surrender to the needs of morbidity; quite different from death statistics. These neo-Classical modifications, dominant today in the ICD-9-CM and ICD-10-CM[4], share cause of death underpinnings in base rubrics and anatomic structuring.

A central characteristic of Classical clinical terminologies is the hierarchical list. This system of ordering facilitates code finding and higher level aggregations. Traditional terminologies carry this system forward, but permit the co-existence of multiple hierarchies within a terminology simultaneously. First successful in the creation of the Systemized Nomenclature of Diseases and Operations (SNDO) at the New York Academy of Medicine in 1928[5], it is perhaps more familiar today as the underlying model for older versions of SNOMED[6] and the UK based Read Codes. Indeed, SNOMED is a lineal descendent of SNDO, by way of the Systemized Nomenclature of Pathology (SNOP)[7]. These traditional nomenclatures exhibit enormous expressive power, as clinical events and conditions can be composed by combining terms from different hierarchies into a descriptive "sentence." Thus, an introduction of complex structure into clinical terminologies begins with the traditional period.

If structure is the key distinguishing feature of traditional terminologies, structure taken to its limit characterizes Modern terminologies. Formal theories of knowledge and information representation guide the evolution of these systems. An approachable summary of these features appears in the series of articles from the Canon group[8], and in works by several authors about specific systems. Present day examples include Galen, the new NHS Clinical Terms (formerly Read V3), and Kaiser-Permanente's Convergent Medical Terminology (CMT).

Why are Modern era's followed by post-Modern? Aesthetics probably has less to do with this transition for clinical terminologies than tractability. Fully object-class structured terminologies are difficult to build and are therefore usually small. Those that attain any appreciable size tend to trade off purely modern structures for practical efficiencies; after all, these terminologies are intended to be used in healthcare and health policy. Information that could be structured as attributes of a concept may be represented as non-functioning "facets" about the term instead.[9] Indeed, the new SNOMED-RT (Reference Terminology) and the

1091—8280/98/$5.00

CMT incorporate some traditional tools to facilitate content editing.[10]

Table 1

Clinical Term Sources

Term Source	Total Terms Entered	Unique Terms
ICD-9-CM Title Terms	15205	15205
ICD-9-CM Index Terms	18214	16121
"Friendly" Terms from Decision Support System	5782	5782
"Friendly" Terms from Service Recognition Sheets	5449	4059
High Level Branch Titles from Service Recognition Forms	779	744
Mayo Abbreviations	337	337
Headword Inverted Mayo Abbreviations	71	71
Tailword Inverted Mayo Abbreviations	27	26
Mayo Problem List Terms	5010	4797
Headword Inverted Mayo Terms	1517	1428
Tailword Inverted Mayo Terms	1175	1032
Totals	53564	49602

This manuscript will not attempt to establish a rigid classification of clinical classifications. We do hope to demonstrate how traditional and classical features can be rapidly harnessed to create serviceable terminologies for problem list entry. Mayo Clinic generated a customized version of a multi-hierarchical terminology system using ordinary tools, including 3x5 cards. The discussion elaborates on how Modern behavior can be elicited from traditional structures and terminology development methodologies.

METHODS

Context

Mayo Foundation has a long heritage of indexing, retrieving, and analyzing the treatment and outcomes of its patients[11]. Mayo will implement the IDX LastWord system throughout its Rochester based clinic and hospitals over the next three years. Present versions of this product constrain the structure and expressiveness of a central problem list. This list is the basis for a) Indication for Order; b) Clinical Problems; c) Billing Codes; and d) Summary Diagnoses. In view of our commitment to education and research, this problem list must be optimally structured within LastWord constraints to support detailed indexing and retrieval. Thus, we needed to synthesize a clinical terminology that fulfilled our needs and conformed to a Classical structure.

Terminology Sources

Subscribing strictly to the creed of not inventing that which already is, we assembled several sources of clinical terminology as the raw clay from which we would shape our interim problem terminology. Because the auditing requirements of billing and reimbursement are inflexible, we chose ICD-9-CM as the primary backbone for our model. Our goal was to identify terms clinicians could relate to and accept as a clinical description of a patient problem; most ICD-9-CM titles do not illustrate this property. Thus, a major task was identifying synonyms and clinically specific terms for and within ICD-9-CM rubrics.

We used the ICD-97 titles, which are contextually expanded by the NLM and Lexical Technology to be fully specified or semantically self contained. This is required when terms inherit context from their hierarchical parent, but this is not evident in the terms string alone. An example is:

001 Cholera
001.0 Due to Vibrio cholerae

which is changed to

001.0 Cholera due to Vibrio cholerae

Terms which had their meaning specified by a table expansion for the 4th or 5th digit, were de-normalized to incorporate each permutation as fully specified string for each expanded code.

The index to the ICD-9-CM is rich with synonyms and more specific conditions. Machine readable sources for this are not widely available, although the UMLS contains index terms derived from the typeset tapes for the 1991 version of ICD-9-CM. These were mapped to the present meaning of the ICD term, which had changed in a few instances over the 6 year period.

Clinicians may expect to find terms beginning with the root form of a concept, e.g. *Myocardial Infarction, Acute* rather than *Acute Myocardial Infraction.* This is pertinent in the LastWord product, which does not navigate in a word order independent way. For this reason, we algorithmically constructed terms which were tail-word and head-word inversions. Inversion boundaries were based on the

scope of modifiers or qualifiers[12], such as *acute*, that appear in a term.

Mayo has long attempted to buffer clinicians from the ICD by presenting multiple check box service recognition sheets labeled with "friendly" terms in lieu of formal ICD titles. The corpus of such terms was harvested, as were the mini-hierarchies created by their ordering and nesting on the printed service recognition forms. Similarly, we have accepted sanctioned abbreviations overtime, such as *CHF* for *Congestive Heart Failure*. Some common abbreviations, such as *MI*, are not sanctioned due to the ambiguity arising between *Myocardial Infarction* and *Mitral Insufficiency*.

Finally, we had compiled a corpus of Mayo Problem Terms, which arose from the most frequent problem statements among the many millions, which have accrued, on our Master Sheet and as Assessment statements in machine-readable clinical notes. A description of their development was previously presented[12]. Table 1 summarizes term sources and relative contributions.

Compilation and Sifting

All terms, including redundant ones, were entered into an INGRES database. A Mayo unique identifier (MUI) was assigned to each term. When terms derived from UMLS sources, we carried CUI and SUI identifiers as well. Redundant term-code pairs were marked as obsolete, and excluded from further processing. Integrity checks to evaluate whether identical lexically normalized terms were assigned to different ICD codes were made. A sub-study was done to establish whether term-code discordance could be explained by word order dependent contexts, using a variant of the Specialist Lexicon's LVG which maintained word order within strings.

From the 49,602 source terms, the Specialist Lexicon NORM utility[13] yielded 38,897 unique normalized terms in the corpus. These spanned 12,628 non-procedural codes. Because part of our task was to ask clinicians to identify preferred terms where many existed, and to evaluate whether multiple terms within a code were synonymous, we put aside 7,342 codes for which we had only a single term. An additional 215 terms spanning 75 "E-codes" were also put aside. This left 30,352 terms spanning 5,203 codes.

Choosing and Ordering

To facilitate the task of human review, the corpus of terms was printed onto 3x5 cards. The term was in 24pt Times centered on the card. ICD codes were printed in a smaller font in the upper right corner. To expedite subsequent data input, a machine-readable bar code was printed in the lower left corner from the Mayo unique identifier (MUI). Cards were sorted by ICD code, with blue colored separator cards between codes. Blue cards had a distinguishing bar code. Cards were distributed to clinical sub-specialists, corresponding to the ICD domain.

The primary task of this exercise was to identify terms acceptable to clinicians with respect to word content and clinical specificity. We sought to have clinicians evaluate multiple terms within an ICD-9-CM rubric to establish if more than one clinically significant concept was lumped into that code, as represented by their terms. If so, the clinician was sort cards into as many concept piles he or she thought appropriate, and insert additional blue cards between these concepts (large numbers of spare blue cards were provided).

Within each concept pile, or ICD code if there was but a single clinically relevant concept, they were asked to order terms so that the most clinically preferable term appeared first after the blue card. If no term was deemed acceptable, they were allowed to write in an appropriate preferred term on yellow cards. Yellow cards were assigned MUIs at the time of scanning from pre-printed labels, the write-in text and MUI were keyed after scanning in order. Terms judged inappropriate, mis-leading, or simply awkward, were placed after a Red card, which signified that these should be marked obsolete and not considered further.

Table 2

Distribution of Concepts per ICD Code

# ICD Codes	# Concepts
3174	1
301	2
134	3
66	4
31	5
31	6
26	7
13	8
5	9
5	10
4	11
2	12
2	13
2	14
2	21
1	29
1	31
1	34

Table 3

		Large List	
		PT	OT
ProblemList		5132	15858
PT	3087	948	1323
OT	2556	405	1423

Distribution of 5,643 "Problem List" terms, and 20,990 "Large List" terms, classified into Preferred Terms (PT) or Other Terms (OT) by Generalists (Problem List) and Sub-specialists (Large List). Not all terms cross-tabulate, due to an arbitrary selection of identical strings from multiple sources into the Large List.

Processing and Tabulation

Card decks were collected and bound, to preserve card order. Cards were unbound at an optical scanning station, where the sequential order of MUIs and blue separator cards was captured. Concept relationships and assignments were added to the INGRES database, as were the keyed input from yellow write-in cards. Separate scanning passes were made for cards in the Red-delete decks.

Results tabulate the number of concepts discovered within ICD codes, the frequency of new preferred terms, and deleted terms.

Initial Review

During the creation of our original Problem List terms[12] ten clinicians, all generalists, were given terms that were algorithmically thought to be similar. They reviewed the terms and classified them as a Preferred Term (PT), an Other Term (OT) or not a term (NT). Two clinicians reviewed each term. A term was considered a Primary Term if either clinician classified it a PT. A term was considered not to be a valid term if both clinicians thought it was an NT. All other terms were OT or roughly equivalent to entry terms in ICD.

This subsequent study, using the colored index cards, was distributed to sub-specialists in each ICD rubric. The selection of terms by specialists were then cross-tabulated with the original decisions of the generalists in our earlier study, for terms which over-lapped (by MUI designation) in both studies (Table 3).

Data Analysis

Terms from the initial review were compared with the second sub-specialty review to determine if there were significant differences in preferred term assignments. An assessment of inter-reviewer variability was performed, by comparing PT and OT rates between reviewers who reviewed the same using the McNemar test.

RESULTS

The sorting and separation process yielded 5,611 concepts, each with a preferred term which spanned 3806 ICD codes. An additional 19,568 synonyms were judged useful and mapped to these concepts. Deleted terms numbered 3,481.

There were 528 write-ins, all of them of course preferred forms. These also demonstrated a highly skewed distribution: 331 ICD codes had one write-in; 42 codes had two; 9 had three; 7 had four; and 1 code had thirteen write-ins.

The distribution of concepts per ICD code, created by separating stacks with blue cards appears in Table 2. Note that these include the write-ins.

The sub-specialty review yielded significantly different results than the generalist review (Table 3; $P<0.001$, McNemar Test). In general sub-specialists were almost half as likely to call a term a primary term than were generalists. Additionally, sub-specialists were able to remove terms from the list, which were ambiguous or not clinically relevant; these numbered fully one quarter of the terms from the initial data set.

DISCUSSION

This work combines the practical approaches of traditional terminology development with a "good-enough" relational model to support structure. It assembles off-the-shelve terms to make them usable in a clinical problem list. The fundamental enhancements to the final corpus are an identification of clinically acceptable preferred or canonical forms for each concept, and the discrimination of different concepts within an ICD-9-CM code. A secondary benefit is an authoritative identification of clinically reasonable synonyms for many concepts.

The comparative analysis of the Initial Review with the second Sub-Specialty review showed profound differences in the assignment of preferred terms and other terms. Sub-specialists were about half as likely to call a term preferred than generalists. Additionally, many terms could be filtered out as either inappropriate or incorrectly assigned synonyms. This simple example illustrates the need to have the most knowledgeable groups or

individuals involved in the authoring process. There may be nothing good enough about "good-enough" editing.

Mayo intends to use these terms for its initial implementation of the LastWord products. Over time, this lexicon will become more compliant with SNOMED-RT content and style, particularly as more sophisticated terminology server objects are modularly introduced into the software. Meanwhile, we have demonstrated a practical,. neo-classical method for rapid clinical terminology development and customization.

ACKNOWLEDGEMENTS

Thanks to Jim Buntrock and Doug Crowson for database maintenance, Lesa Rohde for scanning, Donna Ihrke for ICD expertise, and Karen Elias for administrative assistance, and Beth Atkinson for statistical support. This work was supported in part by grants HS/LM 08751, LM 05416, and AR 30582.

REFERENCES

[1] Graunt J. *Natural and Political Observations Made Upon the Bills of Mortality.* London, 1662. (Republished) Baltimore, MD: The Johns Hopkins Press, 1939.

[2] Rector AL, Glowinski AJ, Nowlan WA, Rossi-Mori A. Medical concept models and medical records: An approach based on GALEN and PEN&PAD. *JAMIA* 1995;2(1):19-35.

[3] Sixteenth Annual Report of Registrar-General of England, 1853, Appendix, p. 73.

[4] ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/ICD10

[5] *Standard Nomenclature of Diseases and Operations (SNDO).* Chicago, IL: American Medical Association, 1933

[6] Cote Ra, Rothwell DJ, Palotay JL, Beckett RS. *SNOMED International.* Northfield, IL: College of American Pathologists, 1994.

[7] *Systematized Nomenclature of Pathology (SNOP).* Chicago, IL: College of American Pathologists, 1965.

[8] Evans DA, Cimino JJ, Hersh WR, Huff SM, Bell DS, for the CANON Group. Position Statement: Toward a Medical Classification Language. *JAMIA* 1994;1:207-217.

[9] Mays E, Weida R, Dionne R, Laker M, White B, Liang C, Oles FJ. Scalable and expressive medical terminologies. *JAMIA* 1996:259-63.

[10] Dolin RH, Huff SM, Rocha RA, Spackman KA, Campbell KE. Evaluation of a "Lexically Assign, Logically Refine" Strategy for Semi-automated Integration of Overlapping Terminologies. *JAMIA* 1998;5:203-213.

[11] Kurland LT, Molgaard CA. The patient record in epidemiology. *Scientific American* 1981;245(4):54-63.

[12] Chute CG, Elkin PL. A clinically derived terminology: Qualification to reduction. *JAMIA* 1997;Symp Suppl:570-74.

[13] McCray at, Srinivasan S, Browne AC. Lexical methods for managing variation in Biomedical Terminologies. *JAMIA* 1997;Symp Suppl:193-201.