A Clinically Derived Terminology: Qualification to Reduction

Christopher G. Chute, MD, DrPH, Peter L. Elkin, MD
Mayo Foundation, Rochester, MN

Mayo Foundation is developing synonym rich entry points for the recording of patient problems by clinicians, which will map to the KP-Mayo Convergent Medical Terminology. We describe the empirical sources for these terminology components, and how the number and complexity of the terms could be substantially reduced by the introduction of a Qualifier axis. The expressive power of these entry points is dramatically enhanced by this axis. This work is being integrated into terminology navigation modules being jointly developed with Lexical Technology, which leverages UMLS content. It will form the basis for structured problem entry into Mayo's Computer-based Electronic Record.

INTRODUCTION

Computer-based Patient Records are penetrating the American healthcare environment with an accelerating rapidity. Less rapid in appearance are standard terminologies that can capture the clinical detail of patient problems, diagnoses, events, or interventions with sufficient detail to support evidence based medicine. While some terminologies approach this functional behavior, most do not[1] and arguably none presently seem natural to clinicians using them for the routine documentation of healthcare.

The Mayo Foundation has indexed patient problems, diagnoses, and procedures for a century. These indices form the basis of outcomes research, disease natural history studies, treatment response evaluations, continuous quality improvement efforts, and clinical epidemiology. During more recent decades, reimbursement and capitation have been computed from their derivative mappings against statistical classifications such as the ICD-9-CM. The process of indexing has become increasingly computerized. It is approaching the logical final step of being integrated with the Computer-based Patient Record as a structured entry using controlled medical terminology.

This report characterizes some preliminary work at Mayo by our Clinical Terminology Committee to make the presentation of such a controlled terminology familiar and relevant to Mayo practitioners. This terminology development is a module within several larger projects, including Mayo's Electronic Medical Record initiative, the Kaiser-Permanente and Mayo Foundation partnership to evolve toward a Convergent Medical Terminology[2], and the Lexical Terminology Inc. initiatives to produce a multi-functional terminology navigation and browsing resource.

METHODS

Clinical Sources

Two clinical applications served as the empirical source of problem statements entered by clinicians, the Mayo Master Sheet and Assessment statements within electronic Clinical Notes.

The Mayo Master Sheet has captured summary diagnoses, problems, and dismissal impressions since its introduction at the Clinic in 1907. The full text of these entries has been electronically available since 1993. This evaluation sorted the 2,548,119 strings captured over a 15 month period into frequency order, yielding 811,060 unique strings. The most frequent string appeared 30,092 times, the 5000th rank ordered string was present 37 times. From these, the 5000 most frequent strings formed the basis of the Master Sheet input.

The Clinical Notes application was introduced at Mayo in 1994, and is being used by an increasing number of Mayo's 1,200 staff clinicians in the outpatient setting. The 1,389,781 strings which appeared in the Assessment or Impression part of this repository during 1995 were sorted into 547,979 unique terms. The most frequent string was recorded 38,550 times, and the 5000th rank ordered string appeared 18 times. These strings were then merged with the Master Sheet input.

1091—8280/97/$5.00 © 1997 AMIA, Inc.

Clinical Review

The unique reduction of the merged Master Sheet and Clinical Notes terms were reviewed subjectively. Two conclusions were immediately evident:

1. Logistic and operational terms were embedded in these lists, which have little clinical relevance.
2. Many terms were combinations of diagnoses and operational qualifiers.

We carefully pruned the list of logistical elements, and recorded the reason for reductions. In the course of this process, a preliminary pattern of embedded qualifiers was created based on their appearance in the clinical sources.

At this point, we established the utility and feasibility of distinguishing operational qualifiers from clinical modifiers. Qualifiers we defined as words or phrases that operationally or administratively qualify the meaning of a diagnosis or problem, e.g. *history of* a condition, *status post* a procedure, or *rule out* a condition (Mnemonic: Q for Quality Assurance Studies). These are opposed to a different but equally important group of variations which we class as Modifiers. Modifiers are words or terms which modify the severity, location, acuity, or other intrinsic clinical detail of a diagnosis or condition, e.g. *Stage I, acute, antero-lateral,* etc. (Mnemonic: M for Medical).

These distinctions are practical in part because we deemed Qualifier enumeration and recognition to be straightforward. Modifiers are a vastly more complex problem[3], which invoke semantic and linguistic issues we chose to defer.

Qualifiers

Having distinguished the role of Qualifiers, and establishing an empirical starting point for a Qualifier ontology, we completed a preliminary thesaurus and ordering for Qualifiers. These concepts and their synonyms could then be identified within our merged clinical terms corpus, and removed. This process was facilitated by invoking the lexical normalization tools[4] developed by Alexa McCray at the NLM and distributed with the UMLS.

The underlying Mayo Terminology was reinforced by the re-addition of terms that had been present only in combination with a Qualifier. Similarly, some phrases had Qualifiers intrinsically embedded within a composite term, where the meaning would be distorted by the excision of the Qualifier; these were retained in the terminology.

RESULTS

Terminology edits which resulted from combination, review and reduction of the Master Sheet and Clinical Notes sources appear in the table. Overlap between these sources approximated only 20%. An additional 20% were deleted due to superfluous combinations of terms already present with our Qualifier lexicon.

Process or Code	Process Step or Code Meaning	Count of Terms for Each Step		
		Added	Deleted	Remaining
Master Sheet	Most Frequent MS Terms over 3 years	5,000		5,000
Clin. Notes	Most Frequent from Clinical Notes	5,000		10,000
Unique Merge	Drop Terms Redundant Across Sources		2,047	7,953
D	Qualifier marked for Deletion		2,032	5,921
E	Type of Examination		691	5,230
A	Abbreviation		461	4,769
S	Site of Diagnoses or Examination Including anatomy.		4	4,765
L	Location of Visit		105	4,660
Pr	Prescription drugs or therapy		12	4,648
Q	A new term added for a Qualifier	129		4,777
T	A new Term added for an Abbreviation	387		5,164

Table: "Balance register" of term volume, by review "transaction."

Measurable deletions took place for administrative or provincial reasons, accumulating another 10% reduction in aggregate. Thus, from our original source of 10,000 terms, just 5,164 unique terms remained after review and reduction.

The Appendix enumerates the 269 strings that form our preliminary ontology of Qualifiers. These include 104 Preferred Terms, or unique concepts; the remainder are variants, synonyms, or abbreviations. The two character prefix to the term identifiers, borrowed from the UMLS, indicates:

PT	Preferred Term
SY	Synonym
LV	Lexical Variant
AB	Abbreviation

The three digit term code conveys a superficial hierarchy and notation of synonymy. Terms with identical numbers are functionally synonymous (even if not exactly synonymous – they are so interpreted for purposes of Problem List entry).

DISCUSSION

This report begins a suite of presentations that address multiple steps in the empirical creation of a clinical terminology. While the desirable attributes of clinical vocabularies have become relatively well understood[5,6], their actualization into widely used or useful terminologies for computer-based patient records has lagged.

The practical consequences of introducing a Qualifier axis are almost self-evident. Whereas our raw lexicon had an ability to recognize about 8,000 terms as potential problem statements, the simple combinatorial of 270 qualifiers across some 5,000 disease specific terms parleys this by three orders of magnitude – a considerable advantage. Invoking the lexical mapping utilities associated with the Specialist Lexicon[4] would further multiply the number of "strings" which could match our preliminary terminology.

The recognition that an independent axis can geometrically expand the expressive power of a terminology is hardly novel. Multi-axial coding dates at least to the development of the SND[7] by the New York Academy of Medicine in 1928; this of course is the ancient forerunner to SNOMED[8]. Nevertheless, the development of a well-structured "General" axis, capturing Qualifiers and Modifiers lags behind in the evolution of most terminologies. This has practical manifestations, since reimbursement codes presently cannot discriminate *Rule-Out* conditions from final diagnoses, nor have they any reasonable mechanism to recognize disease severity.

Our empirical modality has prompted a series of Usability Laboratory evaluations[9] which focus on the practical ability of clinicians to navigate or use our terminology in prototypes of computer-based patient records. The ability to qualify existing terms to fit the clinical scenario is warmly received in these subjective studies. Indeed, a moderate clamor to implement parallel functionality for clinical modifiers affirms a known limitation.

The recognized limitations of this present work are well understood. Future research must address an ontology of Modifiers, analogous to that created for Qualifiers. Whether these should be inheritable within classes of diagnoses to guard against non-sense term coordination and to inform term composition, bears scientific cost-benefit consideration. This work also does not provide an ontology for the residual 5,164 Mayo Clinical Terms, although we are approximating that in our pre-alpha evaluations by mapping to UMLS hierarchies. Finally, while an empirically derived starting point affords familiarity and relevance, it suffers incompleteness and non-comprehensiveness. This is being addressed by Mayo specialist reviews of our terminology content for completeness.

ACKNOWLEDGEMENTS

We thank Jim Buntrock, Geoff Atkin and Donna Ihrke who worked diligently with the primary clinical sources in preparation for clinical review. Lexical Technology and members of the Mayo Terminology Committee respectively critiqued the structure and content of our preliminary qualifier axis. Karen Elias facilitated this manuscript and logistics. This work was funded in part by grants HS/LM 08751, LM 05416, and AR 30582.

REFERENCES

[1] Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. *JAMIA* 1996;3:224-233.

[2] Campbell KE, Cohn SP, Chute CG, Rennels G, Shortliffe EH. Gâlapagos: Computer-Based Support for Evolution of a Convergent Medical Terminology. *JAMIA* 1996;Suppl.:269-273.

[3] Blois MS. *Information and Medicine. The Nature of Medical Description.* Berkley, CA: University of California Press, 1984.

[4] McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. *JAMIA* 1994;Suppl.: 235-239.

[5] Cimino JJ. Desiderata for Controlled Medical Vocabularies in the Twentieth Century. In: Chute CG (ed), Proceedings of the 1997 International Medical Informatics Association Working Group 6 on Natural Language (preprint) 257-267.

[6] Chute CG, Cohn SP, Campbell KE. A Framework for Comprehensive Health Terminology Systems in the United States: Development Guidelines, Criteria for Selection, and Public Policy Implications. Submitted.

[7] Logie HB (Ed.). *Standard Classified Nomenclature of Disease.* New York, NY: The Commonwealth Fund, 1932.

[8] Cote RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L (Eds.). *SNOMED-Snomed International. The Systematized Nomenclature for Human and Veterinary Medicine.* Northfield, IL: College of American Pathologists, 1993.

[9] Elkin PL, Mohr DN, Tuttle MS, et al. Standardized Problem List Generation, Utilizing the Mayo Canonical Vocabulary embedded within the Unified Medical Language System. Submitted to 1997 Proceedings of AMIA Fall Symposium.

No	Qualifier
PT100	Indeterminate
SY100	Unknown
PT101	Indeterminate etiology
LV101	of indeterminate etiology
SY101	of undetermined etiology
SY101	Unknown cause
SY101	Unknown etiology
LV101	of unknown etiology
PT110	Uncertain
PT111	Uncertain etiology
LV111	of uncertain etiology
PT120	Unspecified
SY120	Not Otherwise Specified
AB120	NOS
SY120	Nonspecific
AB120	NS
LV120	Non-specific
PT121	Unidentified
SY121	Not identified
PT200	History
AB200	Hx of
AB200	HO
AB200	H/O
LV200	History of
AB200	Hx
PT210	by history
AB210	by Hx
PT220	Previous
PT230	Status Post
AB230	SP
AB230	S/P
PT231	Prior
SY231	Prior history of
AB231	Prior Hx of
PT240	Past
SY240	Past history
LV240	Past history of
PT241	Remote
SY241	Remote History
LV241	Remote History of
AB241	Remote Hx of
PT250	Recent
PT260	Recurrent
PT261	Relapsing
SY261	Repeating
SY261	Re-occurring
PT290	No history
AB290	No Hx
SY290	Negative history
AB290	Neg history
AB290	Negative Hx
AB290	Neg Hx
SY290	No past history
AB290	No past Hx
PT291	Never
PT300	Family History
AB300	Family Hx
AB300	FH
LV300	Family history of
AB300	Family Hx of
SY300	Positive family history
AB300	Positive family Hx
LV300	Positive family history of
AB300	Positive family Hx of
PT301	Strong family history
LV301	Strong family history of
AB301	Strong family Hx
PT302	First degree relative with
PT390	No family history
AB390	No family Hx
LV390	No family history of
AB390	No family Hx of
SY390	Negative family history
AB390	Negative family Hx
LV390	Negative family history of
AB390	Negative family Hx of
PT400	Follow Up
AB400	FU
AB400	F/U
PT401	Follow Up Care
AB401	F/U Care
AB401	FU Care
PT410	Check
SY410	Recheck
AB410	ReCK
PT500	Normal
AB500	NI
SY500	Within normal limits
AB500	WNL
PT510	Negative examination
AB510	Neg exam
AB510	Neg examination
PT511	Normal examination
AB511	Normal exam
PT520	No diagnosis
AB520	No Dx
PT530	Satisfactory
AB530	Satis
PT540	Negative
AB540	Neg
AB540	-
LV540	Negative for
AB540	Neg for
SY540	No
SY540	Not
PT541	No evidence
SY541	No evidence of
LV541	No evidence for
SY541	Not seen
SY541	Not observed
PT542	No evidence of disease
AB542	NED
PT543	Ruled out
AB543	R'd/O
AB543	Rd/O
PT544	without
AB544	W/O
AB544	w/o
SY544	Free from
LV544	Free of
SY544	sine
AB544	s
AB544	s/
PT550	Resolved
AB550	Resolv
PT560	Remission
AB560	Remiss
LV560	in remission
SY560	Remitted
PT561	in continued remission
PT562	No evidence of recurrent disease
PT570	No Additional Disease
AB570	NAD
PT600	Treated
AB600	Rxed
PT610	on therapy
LV610	under therapy
SY610	on treatment
AB610	on Rx
LV610	under treatment
PT611	on medical therapy
PT612	on replacement
SY612	on replacement therapy
PT620	Controlled
SY620	under control
SY620	under adequate control
PT621	under good control
SY621	well controlled
SY621	with good control
PT630	Controlled on therapy
SY630	Controlled on treatment
PT631	Controlled on medication
PT640	Stable
PT690	Uncontrolled
SY690	Not controlled
SY690	Out of control
PT691	Unstable
SY691	Not stable
PT692	Labile
PT800	Possible
PT801	Suspected
SY801	Suspect
AB801	Susp
PT803	Questionable
SY803	Question
AB803	Quest
LV803	Question of
PT804	Preliminary
AB804	Prelim
PT810	Rule Out
AB810	RO
AB810	R/O
PT820	Differential
AB820	Diff
PT821	Working
PT830	Include
SY830	Includes
PT831	Cannot exclude
PT832	Cannot rule out
AB832	Cannot RO
AB832	Cannot R/O
PT840	High Risk
PT841	Likely
PT850	Probable
SY850	Most likely
PT851	Suspicious for
AB851	Susp
PT852	Evidence of
LV852	Evidence for
SY852	Evidence
AB852	Evid
PT853	Compatible
LV853	Compatible with
SY853	Consistant
LV853	Consistant with
AB853	CW
AB853	C/W
AB853	cw
AB853	c/w
PT860	Unconfirmed
PT861	Observed
SY861	Seen
SY861	Witnessed
SY861	Watched
PT862	Reported
PT870	Clinical
SY870	Clinically
PT871	Surgical
SY871	Surgically
SY871	by surgery
PT872	Radiologic
AB872	Radiol
SY872	X-Ray
SY872	Radiologically
SY872	Radiological
PT873	Laboratory
AB873	Lab
SY873	by laboratory
AB873	by lab
PT874	Pathologic
AB874	Path
SY874	Pathologically
SY874	by pathology
AB874	by path
PT875	Cytologic
SY875	Cytology
SY875	by cytology
AB875	by cyto
PT890	Positive
AB890	Pos
AB890	+
LV890	Positive for
AB890	Pos for
PT891	Present
PT892	Known
PT893	Specific
LV893	Specific for
PT894	Identified
SY894	Identifiable
PT895	With
AB895	W/
AB895	W
AB895	w/
AB895	w
SY895	cum
AB895	c
AB895	c/
PT896	Absolute
PT899	Certain
PT900	Principal
PT901	Primary Diagnosis
PT910	Secondary Diagnosis
PT911	Incidental
PT912	Innocuous
PT920	Administrative
PT921	Reimbursement
SY921	Billing
PT923	Management
PT924	Accounting
PT925	Accreditation
PT926	Review
PT930	Research
PT931	Protocol
PT932	Study
PT933	Statistical
PT940	Reportable
PT941	Mandatory Public Health
PT942	Recommended Public Health
PT950	Screening

Appendix: Tabular list of Qualifiers derived from clinical sources and used in reduction.