

Direct Comparison of MEDCIN[®] and SNOMED CT[®] for Representation of a General Medical Evaluation Template

Steven H. Brown MS MD^{1,2}, S. Trent Rosenbloom MD MPH², Brent A. Bauer MD³, Dietlind Wahner-Roedler MD³, David A. Froehling, MD, Kent R. Bailey PhD, Michael J Lincoln MD, Diane Montella MD¹, Elliot M. Fielstein PhD^{1,2}, Peter L. Elkin MD³

1. Department of Veterans Affairs 2. Vanderbilt University, Nashville TN 3. Mayo Clinic, Rochester MN

Background: Two candidate terminologies to support entry of general medical data are SNOMED CT and MEDCIN. We compare the ability of SNOMED CT and MEDCIN to represent concepts and interface terms from a VA general medical examination template.

Methods: We parsed the VA general medical evaluation template and mapped the resulting expressions into SNOMED CT and MEDCIN. Internists conducted double independent reviews on 864 expressions. Exact concept level matches were used to evaluate reference coverage. Exact term level matches were required for interface terms.

Results: Sensitivity of SNOMED CT as a reference terminology was 83% vs. 25% for MEDCIN ($p < 0.001$). The sensitivity of SNOMED CT as an interface terminology was 53% vs. 7% for MEDCIN ($P < 0.001$).

Discussion: The content coverage of SNOMED CT as a reference terminology and as an interface terminology outperformed MEDCIN. We did not evaluate other aspects of interface terminologies such as richness of clinical linkages.

Background

Computerized structured data entry systems have been used in a variety of environments to improve clinical documentation quality and timeliness, quality of care, practice guideline compliance, research data collection and other aspects of patient care.(1-9) Structured data entry permits the implementation of reminders, alerts, data-driven monitors, and other types of decision support in addition to monitoring of documentation completeness

To realize the potential of structured data entry, systems should be built using standard terminologies for underlying knowledge representation.(10-13) This strategy facilitates sharing of patient-specific data and decision support rules and content.(14) Without standards for knowledge representation, sharing requires time consuming and often challenging mapping

efforts. Usable and functionally complete standard terminologies need to be available to systems designers and architects. Two candidate terminologies to support entry of general medical data are SNOMED CT and MEDCIN.

SNOMED CT is a reference terminology that has been recommended for various components of patient medical record information by the Consolidated Health Informatics Council and the National Committee on Vital and Health Statistics.(12) SNOMED CT, licensed for US-wide use by the National Library of Medicine in 2003, was evaluated in 15 Medline indexed studies in 2006. According to the College of American Pathologists, SNOMED CT is "the universal health care terminology that makes health care knowledge usable and accessible wherever and whenever it is needed".(15) SNOMED CT includes approximately 370,000 concepts and over 1 million synonyms. Despite SNOMED CT's relatively good performance in content coverage studies, the Department of Defense elected to use MEDCIN as its point of care terminology for CHCS II (now AHLTA).

MEDCIN is a clinical terminology designed to support medical documentation entry into electronic health record systems. MEDCIN was initially developed as "an intelligent clinical database for documentation at the time of care."(16) MEDCIN's producer, Medicomp, states that their software "makes capture of the encounter information fast enough, sufficiently comprehensive and rewarding to overcome physician reluctance." MEDCIN has evolved to include more than 250,000 concepts since 1978, and has been installed in several EHR systems as an interface terminology for clinical documentation including AHLTA, the EHR system developed for the US Department of Defense. MEDCIN covers concepts commonly used in medical histories, physical examination, tests, and some diagnoses and therapies. Unlike SNOMED CT, MEDCIN evaluations are few.(17) In general, MEDCIN concepts are designed to be pre-coordinated to a level that

allows them to contain adequate clinical meaning to be useful for documentation. MEDCIN has been mapped to other terminologies, including CPT-4, ICD-9, ICD-10 and DSM-IV.

We previously evaluated SNOMED CT's ability to represent concepts and interface terms(18) needed for the US Department of Veterans Affairs (VA) general medical evaluation template for compensation and pension (disability) examinations.(19) In the current study we evaluate MEDCIN's ability to represent the VA general medical examination template and compared its performance to SNOMED CT's performance on the same task.

Methods

We utilized a portion of the same data set, and the same review methodology, reviewers, and review assignments to assess MEDCIN that we used previously to assess SNOMED CT.

The general medical evaluation template examined in the current study of MEDCIN and SNOMED CT was created using the Progress Note Construction Set (PNCS) template design environment.(1) Three definition files fully define a PNCS template: a form definition file, a script definition file, and a report definition file. The form definition file details each data object on the template. The script definition file controls form behavior at run-time. Report definitions specify how captions are merged with patient data elements to create a free-text report for upload into the Veterans Health Information Systems and Technology Architecture (VistA), an integrated outpatient and inpatient clinical information system.(20)

We extracted all objects from the three PNCS template definition files including captions, labels, object names and object contents (e.g. list box choices). We mapped the extracted data to MEDCIN (3rd quarter 2005 R2) via automated and manual processes. Two independent reviewers from our previous study of SNOMED-CT (released July 2006) examined a convenience sample of half of the template objects based on their previous review of SNOMED CT. The two internist reviewers assessed MEDCIN as a reference terminology (i.e., concepts present or absent) and as an interface terminology (i.e., normalized terms present or absent). The reviewers were required to search MEDCIN manually using the Multi-threaded Clinical Vocabulary Server's Browser(21) before categorizing a term as a "non match". We used Boolean operators to construct post-coordinated compositions of MEDCIN terms when needed to model complex input data from the template. Mapping examples are given in table 1. Linking semantics were not required to be present for post-coordinations to be considered a match. A third independent review was conducted when needed to establish a consensus rating for subsequent analysis. Agreement statistics (e.g. kappa) are not required for consensus ratings and are not reported.

Results were classified according to the schema in figures 1 and 2. We report sensitivity and positive predictive value (PPV) as a reference terminology and as an interface terminology. We compared the sensitivity and PPV of MEDCIN and SNOMED using uncorrected Pearson Chi-square test for comparing two independent binomial proportions.

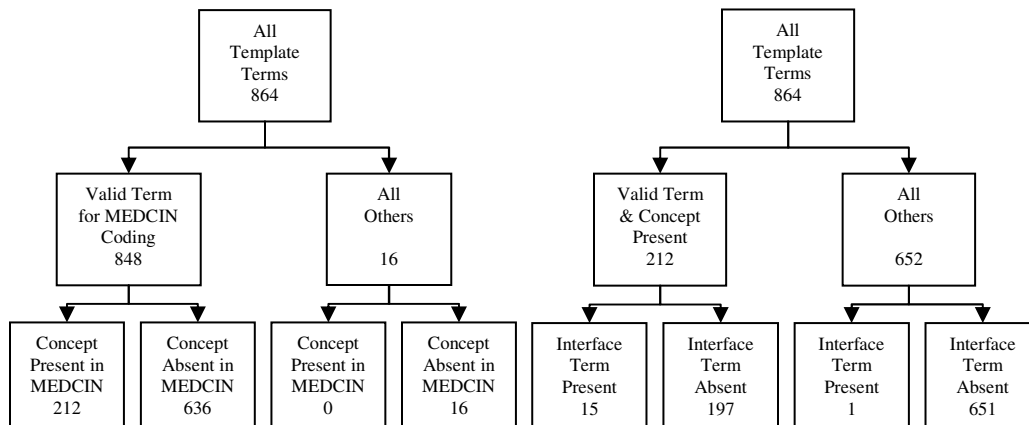


Figure 1. Reference terminology classification

Figure 2. Interface terminology classification

Template Term	CT Concept	Medcin Concept
Hx of Congestive Heart Failure	- Congestive heart failure (disorder) [42343007] [K] - [is Qualified By] . History of (present illness) (contextual qualifier) (qualifier value) [51042001] [Q]	History [5141] [K] Congestive Heart Failure [33268] [K]
Neurologic Exam Deep Tendon Reflexes	- Neurological assessment (procedure) [225398001] [K] - [is Modified By] . Deep (qualifier value) [795002] [M] - [is Modified By] . Tendon reflex (observable entity) [38299000] [M]	Physical Examination [6000] [K] Deep Tendon Reflexes (DTR) [9051] [K]]
Location of enlarged lymph nodes	. Location (attribute) [246267002] [M] . Enlarged (qualifier value) [260376009] [M] . Entire lymph node (body structure) [181756000] [M]	Lymph Nodes Enlarged [9325] [K]
Calf Tenderness	- Tenderness (finding) [247348008] [K] - [has Finding Site] . Entire calf of leg (body structure) [244015008] [M]	(Lower) Leg Tenderness On Palpation Gastrocnemius [8118] [K]

Table 1. Examples of terms and MCVS mappings.

Group	n	Reference Terminology						Interface Terminology					
		Sensitivity (Recall)			PPV (Precision)			Sensitivity (Recall)			PPV (Precision)		
		MED	CT	p	MED	CT	p	MED	CT	p	MED	CT	P
Full group	864	0.25	0.83	<0.001	1.00	0.98	0.034	0.07	0.53	<0.001	0.94	0.96	0.634
Excluding disability	853	0.25	0.83	<0.001	1.00	0.99	0.100	0.07	0.53	<0.001	0.94	0.98	0.331
Only disability	11	0.00	0.20	0.292	n/a	0.14		n/a	0.00		n/a	0.00	
Excluding navigation	829	0.03	0.83	<0.001	1.00	0.98	0.044	0.07	0.52	<0.001	1.00	0.96	0.456
Navigation only	35	0.26	0.82	<0.001	1.00	0.93	0.786	0.00	0.78	0.018	0.00	0.91	0.007
Excluding labels only	437	0.09	0.86	<0.001	1.00	0.98	0.385	0.10	0.34	0.002	1.00	0.95	0.635
Labels only	427	0.41	0.79	<0.001	1.00	0.98	0.043	0.06	0.75	<0.001	0.92	0.97	0.328
no branch points	738	0.28	0.81	<0.001	1.00	0.98	0.033	0.07	0.59	<0.001	0.94	0.96	0.585
branch points only	126	0.05	0.93	<0.001	1.00	0.98	0.747	0.00	0.22	0.200	n/a	0.93	
history only	221	0.30	0.41	0.013	1.00	1.00		0.02	0.23	<0.001	1.00	1.00	
physical only	623	0.23	0.98	<0.001	1.00	0.99	0.489	0.10	0.57	<0.001	0.93	0.99	0.011
assessment only	13	0.00	0.00		n/a	0.00		n/a	n/a		n/a	0.00	

Table 2. Comparison of the performance of SNOMED CT and MEDCIN as reference and interface terminologies. MED = MEDCIN

Results

We evaluated 864 unique General Medical Template objects. Of these, 853 were appropriate for any general medical evaluation, and 11 were concepts specific to disability

exams. Items specific to history numbered 221. Physical exam objects numbered 623 and assessment objects numbered 13. We categorized 16 phrases, such as “Do you believe the veteran is capable of managing personal financial affairs?” as not valid to attempt to map into MEDCIN.

MEDCIN's sensitivity as a reference terminology for the general medical evaluation template was 25%. MEDCIN's sensitivity as an interface terminology was 7%. SNOMED CT's sensitivity as a reference terminology was 83% and its sensitivity as an interface terminology to be 53%. Positive predictive values (precision) for the full group exceeded 90% for each determination. Detailed results of MEDCIN and SNOMED CT performance are presented in table 2.

Discussion

In the current study, MEDCIN's overall sensitivity for concepts used in a general medical evaluation template was 25% and its sensitivity as an interface terminology was 7% when measured against the set of covered concepts. SNOMED CT performed significantly better than MEDCIN in representing underlying concepts (83% vs 25% $p < 0.001$) and terms needed at the human computer interface (53% vs 7% $p < 0.001$).

We offer three possible explanations for this difference in performance. First, SNOMED is a larger terminology than MEDCIN. Second, approximately 75% of the general medical evaluation template terms required compositional expressions of clinical language.(19) The complexity of the target terms made this a difficult mapping task. We found many partial matches. SNOMED contains numerous atomic terms and a set of linking semantics that can be used to create post-coordinated expressions. MEDCIN contains a high percentage of pre-coordinated terms and does not offer formal mechanisms for post-coordination. In the current study, we use Boolean operators to attempt to use MEDCIN for post-coordinations. In neither case did we count the absence of a linking semantic as a terminology failure. We believe that the performance difference, in part, reflects the power and flexibility of post-coordination for content representation.(22, 23) Third, the reviewers noted that MEDCIN often contained more granular terms than were required for the VA general medical evaluation template and at times failed to have appropriate organizing concepts. For example, MEDCIN lacked the organizing concept of "New York Heart Association Class" for congestive heart failure

but did have the more granular concepts of class I thru class IV. If the template had required this level of structured data entry, MEDCIN coverage would have been found to be acceptable. This suggests that the addition of hierarchical, organizing concepts may improve MEDCIN performance for tasks similar to ours. Our reviewers were board certified internists, not terminologists. It is possible that highly trained experts might have found somewhat different results. It is important to note that the current study evaluated only content coverage. We did not attempt to evaluate the richness of clinical links within the terminology - a parameter we have previously found to be important at the human-computer interface terminologies.(13)

The substrate for this study was derived from a single general medical evaluation template used for the purpose of VA disability examinations. It is possible that the results may not generalize to other exam types or to primarily treatment oriented (vs. disability) exams. We believe that the general medical evaluation is similar to other exam types for two reasons. First, only 3.8% of objects were disability-specific. Second, 98% of data elements (i.e. not labels) from the general medical template are also used on at least one other VA template exam type (such as hypertension, diabetes, GI, GU exams). Despite these observations, the general medical template is broad in scope and does not call for subspecialty depth evaluation. Thus, the current study results may not apply in circumstances when subspecialty concepts are in heavy use. We believe additional evaluation of subspecialty content should be conducted to better answer this important question.

In conclusion, we remain optimistic about SNOMED CT's potential role in representing concepts utilized on structured healthcare templates. We look forward to future versions with even better concept coverage. Neither SNOMED-CT nor MEDCIN sufficiently expressed the precise terms needed at the human-computer interface for the studied template. Controlling such near-natural language expressivity is considerably more difficult than concept representation, and may not be easily achievable in the general case. Additional study is required to better understand which aspects of interface terminologies, such as clinically relevant linkages between terms, are needed for these terminologies to facilitate the Human-

Computer interaction with controlled languages in future health informatics solutions.

Acknowledgement: Supported by US National Library of Medicine grant LM06918-04 & 1K22 LM08576-02, and grants from the Centers for Disease Control and Prevention PH000022 and HK00014 and Vanderbilt University Medical Center Funds.

References

1. Brown SH, Hardenbrook S, Herrick L, St Onge J, Bailey K, Elkin PL. Usability evaluation of the progress note construction set. Proc AMIA Symp 2001:76-80.
2. Brown SH, Lincoln MJ, Groen PJ, Kolodner RM. VistA--U.S. Department of Veterans Affairs national-scale HIS. Int J Med Inf 2003;69(2-3):135-56.
3. Johnson KB, Ravich WJ, Cowan JA, Jr. Brainstorming about next-generation computer-based documentation: an AMIA clinical working group survey. Int J Med Inform 2004;73(9-10):665-74.
4. McDonald CJ. The barriers to electronic medical record systems and how to overcome them. J Am Med Inform Assoc 1997;4(3):213-21.
5. Musen MA, Carlson RW, Fagan LM, Deresinski SC, Shortliffe EH. T-HELPER: automated support for community-based clinical research. Proc Annu Symp Comput Appl Med Care 1992:719-23.
6. Payne TH, Hirschmann JV, Helbig S. The elements of electronic note style. J Ahima 2003;74(2):68, 70.
7. Rosenbloom ST, Kiepek W, Belletti J, Adams P, Shuxteau K, Johnson KB, Elkin PL, Shultz EK. Generating complex clinical documents using structured entry and reporting. Medinfo 2004;2004:683-7.
8. Slack WV, Hicks GP, Reed CE, Van Cura LJ. A computer-based medical-history system. N Engl J Med 1966;274(4):194-8.
9. Stead WW, Brame RG, Hammond WE, Jelovsek FR, Estes EH, Parker RT. A computerized obstetric medical record. Obstet Gynecol 1977;49(4):502-9.
10. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37(4-5):394-403.
11. Elkin PL, Brown SH, Carter J, Bauer BA, Wahner-Roedler D, Bergstrom L, Pittelkow M, Rosse C. Guideline and quality indicators for development, purchase and use of controlled health vocabularies. Int J Med Inform 2002;68(1-3):175-86.
12. Lumpkin JR, Richards MS. Transforming the public health information infrastructure. Health Aff (Millwood) 2002;21(6):45-56.
13. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface Terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc 2006.
14. Hammond WE. The role of standards in creating a health information infrastructure. Int J Biomed Comput 1994;34(1-4):29-44.
15. SNOMED Clinical Terms. [cited 2007; Available from: <http://www.snomed.com>
16. Medicom Systems. [cited; Available from: <http://www.medicomp.com/>
17. Warnekar P, Carter J. HIV terms coverage by a commercial nomenclature. AMIA Annu Symp Proc 2003:1046.
18. Campbell KE, Oliver DE, Spackman KA, Shortliffe EH. Representing thoughts, words, and things in the UMLS. J Am Med Inform Assoc 1998;5(5):421-31.
19. Brown SH, Elkin PL, Bauer BA, Wahner-Roedler D, Husser CS, Temesgen Z, Hardenbrook SP, Fielstein EM, Rosenbloom ST. SNOMED CT(R): Utility for a General Medical Evaluation Template. AMIA Annu Symp Proc 2006:101-5.
20. Brown SH, Lincoln MJ, Groen P, Kolodner RM. VistA: The U.S. Department of Veterans Affairs National Scale Hospital Information System. International Journal of Medical Informatics 2003;vol 69(2-3):135 - 156.
21. Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, Speroff T. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. Mayo Clin Proc 2006;81(6):741-8.
22. Elkin PL, Bailey KR, Chute CG. A randomized controlled trial of automated term composition. Proc AMIA Symp 1998:765-9.
23. Elkin PL, Bailey KR, Ogren PV, Bauer BA, Chute CG. A randomized double-blind controlled trial of automated term dissection. Proc AMIA Symp 1999:62-6.