MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press

VA National Drug File Reference Terminology: A Cross-Institutional Content Coverage Study

Steven H. Brown[a,b], Peter L. Elkin[c], S. Trent Rosenbloom[b], Casey Husser[c], Brent A. Bauer[c], Michael J. Lincoln[a,e], John Carter[d], Mark Erlbaum[d], Mark S. Tuttle[d].

[a]U.S. Department of Veterans Affairs; [b]Vanderbilt University, Nashville Tennessee U.S.A; [c]Mayo Clinic, Rochester Minnesota U.S.A; [d] Apelon Inc Ridgefield Connecticut U.S.A; [e]University of Utah, Salt Lake City Utah U.S.A.

Abstract

***Background**: Content coverage studies provide valuable information to potential users of terminologies. We detail the VA National Drug File Reference Terminology's (NDF-RT) ability to represent dictated medication list phrases from the Mayo Clinic. NDF-RT is a description logic-based resource created to support clinical operations at one of the largest healthcare providers in the US.*

***Methods**: Medication list phrases were extracted from dictated patient notes from the Mayo Clinic. Algorithmic mappings to NDF-RT using the SmartAccess Vocabulary Server (SAVS) were presented to two non-VA physicians. The physicians used a terminology browser to determine the accuracy of the algorithmic mapping and the content coverage of NDF-RT*

***Results**: The 509 extracted documents on 300 patients contained 847 medication concepts in medication lists. NDF-RT covered 97.8% of concepts. Of the 18 phrases that NDF-RT did not represent, 10 were for OTC's and food supplements, 5 were for prescription medications, and 3 were missing synonyms. The SAVS engine properly mapped 773 of 810 phrases with an overall sensitivity (precision) was 95.4% and positive predictive value (recall) of 99.9%.*

***Conclusions**: This study demonstrates that NDF-RT has more general utility than its initial design parameters dictated*

Keywords:

Vocabulary, Controlled; Information Storage and Retrieval; Information Theory

Background

In fiscal year 2001, the Department of Veterans Affairs (VA) Veterans Health Administration (VHA) provided health care to 4.1 million veterans and dependents in the form of 43 million outpatient visits, 573,000 inpatient admissions, and 167 million prescriptions (as 30-day equivalents). VHA has developed and deployed a variety of electronic tools to assist clinicians, including VISTA (Veterans Integrated Service and Technology Architecture) [1,2], CPRS (Computerized Patient Record System) [3,4], BCMA (Bar Code Medication Administration) [5,6], and others. VHA is continually looking for ways to use information technology (and other tools) to improve care quality, promote patient safety, and reduce costs. Reference terminologies and terminology services that permit retrospective and real-time aggregation, and sophisticated decision support are one such area under investigation. Formal terminologies are also being evaluated as a way to reduce maintenance and mapping effort.[7,8] VHA's initial reference terminology project is NDF-RT [9,10], a formalization of the National Drug File. Other reference terminologies will be deployed under the VHA Enterprise Reference Terminology project.

NDF-RT uses a Description Logic-based reference model which includes: a defined set of abstractions denoting levels of description for drug products (based on work performed within the Health Level 7 (HL7) Vocabulary Technical Committee)[11]; a set of hierarchical and definitional relationships; and sets of non-definitional properties used at each hierarchical level to capture associated details. The model includes hierarchies for chemical structure, mechanism of action, physiologic effect [10], and therapeutic intent [9]. As of September 2003, NDF-RT is the final phases of expert review by doctors of clinical pharmacology. The most recent version of NDF-RT includes 4202 active ingredients (including salt forms) and 108,112 National Drug Code (NDC) level products. Role definition counts (including inferred roles) are 118,504 mechanism_of_action roles, 119,095 physiologic_effect roles, 123, 379 may_treat roles, 52,827 may_prevent roles, and 5522 may_diagnose roles.

A number of papers detailing desirable characteristics of terminologies have been published in the past five years. For example, in 1998 Cimino [12] described 12 "desiderata" synthesized from the literature of medical vocabulary research. That same year, Chute et al documented 11 desirable characteristics of terminologies to meet health care needs [13]. Two additional publications [14,15] by Elkin et al advance our understanding of terminology quality indicators even further. In the Standards world, ASTM E 2087-00 [16] and ISO TS17117 [17] make significant contributions. Each of these works acknowledges the importance of content coverage. One of the keys of terminology is (to paraphrase an old adage) "Content, content, content" [12].

Obviously, content is also important for clinical records. One type of clinical record "content" that remains challenging to take advantage of is information found in free-text clinical notes [18]. Even at sites such as VA that have extensive electronic medical record systems, virtually all clinical notes are stored as free text.

Table 1: Specific scoring instructions provided to reviewers

Evaluation NDF-RT	
Case I: SAVS and NDF-RT terms matched & the reviewer agrees:	**TP**
Case II: SAVS and NDF-RT terms matched, and reviewer disagrees. Search for accurate NDF-RT term(s) in the SAVS terminology browser	If the term is well-represented: **TP** If the term is not in NDF-RT: **FN**
Case III: Partial match: Search for accurate NDF-RT term(s) in terminology browser	If the term is well-represented: **TP** If the term is not in NDF-RT: **FN**
Case IV: Misspelled or Non-drug term leads to a non-match:	If the misspelled term matched: **FP** If the misspelled term did not match: **TN**
Evaluation of SAVS Engine	
Case I: NDF-RT represents the term and SAVS maps it correctly:	**TP**
Case II: NDF-RT finds a term, and SAVS does not map it correctly:	**FN**
Case III: NDF-RT doesn't have coverage for a term, and SAVS does:	**FP**
Case IV: NDF-RT doesn't find a term, and SAVS does not find a term	**TN**

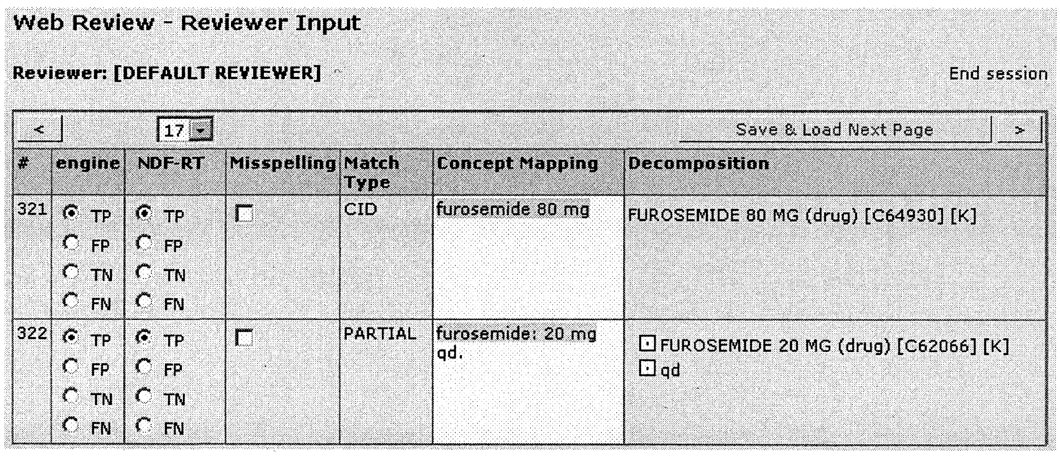

Figure 1 - SAVS web-based reviewer interface

Free-text notes may have structure that is obvious to a human reader, with distinct sections such as "history of present illness", "review of systems", "current medications", "physical exam", "study results", and "assessment". Unfortunately, the data locked in free text notes is inaccessible for algorithmic processing even when sophisticated decision support systems are available (as is often the case with medications).

The purpose of this paper is to evaluate the use of NDF-RT to cover concepts present in free-text medication lists generated at the Mayo Clinic. We examine the utility of NDF-RT in two important ways: 1) We evaluate NDF-RT in a setting beyond its institution of origin (VA) and 2) We evaluate NDF-RT in a documentation role not within its original design parameters (direct clinical care, provider order entry, and pharmacy operations).

Methods

Document Sample

The medication lists used in this study were derived from Mayo Clinic records in the following manner. Mayo Clinic personnel, not otherwise associated with this evaluation, selected a convenience sample of 300 patients. We extracted all transcribed documents about these patients entered within a 2-week time frame into the Mayo EMR (n = 509). The documents were stored in the Mayo Electronic Medical Record (EMR) in a Health Level 7 (HL7) Clinical Document Architecture (CDA) [19] level 1 format and sent as HL7 v2.3.1 messages to our lab. A Java-based CDA object parser created by the Laboratory of Biomedical Informatics at the Mayo Clinic was used to transform the HL7 messages into a series of Java objects containing CDA information (e.g. history, medication list, physical exam). For this study, the data contained within the medication list objects was saved into a relational database.

Automated Mapping

The SmartAccess Vocabulary Server (SAVS; pronounced "saves") is a set of tools developed by the Laboratory of Biomedical Informatics in the Mayo Clinic's Department of Medicine, that facilitate health vocabulary indexing. SAVS uses a four-tier architecture which provides a set of (terminology independent) terminology services designed to expose a flexible web-based application programmer interface. The tools are a combination of client-side and server-side programs that support the mapping of free text into coded data. On the server side, SAVS performs preprocessing of underlying source vocabularies, building complex coded expressions (using automated compositional expressions [20] and automated concept dissection methods [21]), and maintaining a personal list of commonly used coded expressions. On the client side, there are Web-based GUI building blocks for searching and navigating the source vocabularies.

Study Design

We used the SAVS engine to preprocess the medication list phrases and to algorithmically map them to NDF-RT. The preliminary mappings were used to populate evaluation forms for human expert review. The review forms were presented in a web-based interface (see Figure 1) linked to an Access Database™.

Two non-VA staff physicians were asked to review the medications extracted from Mayo clinical records and to make two judgments for each mapping between medication list phrases and NDF-RT concepts. First, the physicians independently determined if NDF-RT concepts could represent the concept. Second the physicians judged how well the SAVS engine automatically mapped the medication list phrase to NDF-RT. Only clinically exact conceptual matches were accepted. Non-matches and "partial" matches were evaluated via a failure analysis. In cases where the two independent reviewers disagreed, a third non-VA clinician was employed to make a tie breaking judgment. The specific scoring instructions are detailed in table 1.

Analysis

Descriptive statistics about the phrases used in transcribed Mayo medication lists are provided. We report the sensitivity, specificity, positive predictive value, negative predictive value and positive likelihood ratios for both NDF-RT's coverage of medication list phrases and the SAVS engine's ability to automatically create the mappings. Table 2 contains definitions of true positive, false positive, true negative and false negative results for NDF-RT We also report the McNemar's test for correct judgments between the two raters. We did not use the Cohen's Kappa statistic because the high proportion of true positive judgments produced a base-rate bias.

Table 2. Definitions of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) for NDF-RT's coverage of Mayo medication list concepts.

Mayo Concept is:	Mayo Concept is: Medication	Not a Medication
In NDF-RT	True Positive	False Positive
Not In NDF-RT	False Negative	True Negative

Results

The medical record extraction process produced 509 documents on 300 patients. These 509 documents contained 847 medication concepts in medication list document sections. Each of these medication concepts contained a reference to a "common" name (e.g. "aspirin", n=335) or a trade name (e.g. "avandia", n=512). Other components of concepts listed in the medication list included: doses (n=557), routes (n=79), frequency of administration (n=237), and units (n=453).

Table 3. Results of algorithmic and subject matter expert mappings of transcribed phrases to NDF-RT.

Results	NDF-RT	SAVS
True positives	814	773
False positives	0	1
True negatives	15	36
False negatives	18	37
Sensitivity	0.978	0.954
Specificity	1	0.973
Positive Likelihood Ratio	∞	35.333
Positive Predictive Value	1	0.999

The coverage of medication list phrases by NDF-RT and the SAVS engines mappings are presented in table 3. Overall NDF-RT covered 97.8% of concepts. 18 phrases could not be mapped to NDF-RT by the SAVS engine or human experts. The most common type of medication found in the transcribed phrases that could not be mapped to NDF-RT were specific trade names, specific over-the-counter products or food supplements (n=7; 38% of total misses) such as "Juice plus", "Mineral Blast" and "Herbalife vitamin". NDF-RT missed 3 generic OTC's (oil of oregano (n=1) and coral calcium (n=2)). Three misses were due to lacking synonymy ("baby aspirin" as a synonym for 81 mg aspirin tablet, and "childrens multivitamin"). There were two instances in which NDF-RT did not cover the underlying substance (motilium (domperidone) and lexotan (Bromazepam)), one instance of a missed ester form (paba present but not as paba dmg), and one instance of a missing locally compounded item (Wilson's solution).

The SAVS engine properly mapped 809 of 847 phrases (TP = 773; TN = 36). Of 37 false negative mappings, the most common reason (n=15) was a failure to map a set of words from the transcriptions to an NDF-RT concept containing those words (e.g. mapping "pulmicort" to the "Pulmicort Turbuhaler"). In 8 cases the SAVS engine mapped vitamin types improperly (e.g. mapping "Vitamin C" to "Vitamin A". Complete mismappings numbered 4. Proper mappings of one ingredient of a two-ingredient medication combined with failure to map the second ingredient (e.g. "calcium 1200 mg with vitamin d") occurred 4 times.

The raters agreed on 92.8% of NDF-RT judgments (786/847). The inter-rater correlation on their assessments of NDF-RT's coverage was significant by McNemar's test ($p < 0.001$). The raters agreed on 92.4% of SAVS engine judgments (783/847). The correlation between the two raters on their assessments of the SAVS engine's ability to create the mappings trended towards significance by McNemar's test ($p = 0.072$).

Discussion

Overall NDF-RT covered 97.8% of medication list phrases generated via transcription at the Mayo Clinic. Furthermore, the study demonstrated that the two raters made independent judgments of the Engine assessment that were well correlated, despite the high baseline rate of true positive results. The raters' judgments of NDF-RT's coverage for the concepts were definitely significantly correlated. The high rate of true positive and true negative judgments made congruently by both raters in these two analyses indicates that the engine produces reliable mappings and that the coverage of NDF-RT for the content is very complete. The significant agreement between the raters indicates that their judgments are likely to be reliable. The high rate of true judgments indicates that the engine produces highly accurate mappings and that the content coverage of NDF-RT for the domain is very complete.

This is an important and useful result for at least three reasons. First, the US Department of Veterans Affairs developed NDF-RT as a formalization of its National Drug File (NDF). The NDF is the sole terminology used for mail-out prescriptions in VA; it has been used to issue hundred of millions of prescriptions over more than a decade. NDF content is functionally complete for VA purposes despite the fact that formal content coverage studies have never been performed. Because each medication in NDF is present in NDF-RT, it is only a small leap of faith that NDF-RT is also functionally complete for VA purposes. The current study's finding that NDF-RT covers medication list phrases from the Mayo Clinic demonstrates that NDF-RT may be used beyond the US Department of Veterans Affairs. Second, the task of representing transcribed phrases from the medical record is a different task than the usual application of NDF, and was not a task represented in the use cases that guided NDF-RT development. The representation of concepts from transcribed records places different demands on a terminology than does structured data entry supporting provider order entry, and pharmacy operations. For example, the transcription representation task requires elements typically associated with an interface terminology, such as extensive synonymy. The third reason this study's results are useful is that they point out opportunities for improvement in NDF-RT. This study demonstrates that NDF-RT has more general utility than its initial design parameters (supporting VA medication prescribing and dispensing) dictated. This finding is further corroborated by the use of NDF-RT as a data management aid by the U.S. Pharmacogenomics Research Network.

Mapping phrases found in free text to concepts in a reference terminology is a difficult task with great potential. The SAVS engine performed well in mapping phrases from transcribed medication lists to the NDF-RT. For this task, its sensitivity (precision) was 95.4%, its positive predictive value (recall) was 99.9% and its positive likelihood ratio was 35.3. Further studies need to be performed to replicate and extend this finding.

The current study has several limitations. First, the source data came from only one institution. It is possible that other institutions may have different data representations. Second, the transcriptions came from the Mayo Clinic's Department of Adult Medicine. VA's healthcare beneficiaries are exclusively adults. Thus, the current study did not evaluate NDF-RT's coverage of pediatric medications. It is possible that the different dosage forms and strengths that exist for the pediatric population may not be represented in NDF-RT as well as those intended for adults. Third, the current study only evaluated one possible use of a drug terminology. Doubtlessly, there are many others. Despite these limitations, we are encouraged by the findings of the present study.

Content coverage studies are an important component of terminology evaluation. They provide data to help systems designers, implementers and users make informed decisions. By necessity, individual content coverage studies cover specific uses of a terminology. It is the authors' opinion that potential users should evaluate terminologies by reviewing a number of content coverage studies in order to formulate a broad and fully representative view. From this perspective, the authors of this first content coverage study of NDF-RT look forward to additional studies by different research groups, covering different uses.

References

[1] 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;69(2/3):135-156.

[2] Kolodner RM, editor. Computerizing large integrated health networks: the VA success. New York: Springer-Verlag; 1997.

[3] Meldrum K, Volpp B, Vertigan R. Department of Veterans Affairs' Computerized Patient Record System. Proc AMIA Symp 1999(1-2):1214.

[4] Brown SH. No free lunch: institutional preparations for computer-based patient records. Proc AMIA Symp 1999:486-90.

[5] Johnson CL, Carlson RA, Tucker CL, Willette C. Using BCMA software to improve patient safety in Veterans Administration Medical Centers. J Healthc Inf Manag 2002;16(1):46-51.

[6] Patterson ES, Cook RI, Render ML. Improving patient safety by identifying side effects from introducing bar coding in medication administration. J Am Med Inform Assoc 2002;9(5):540-53.

[7] Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. Journal of the American Medical Informatics Association 1994;1(1):35-50.

[8] Lau LM, Johnson K, Monson K, Lam SH, Huff SM. A method for the automated mapping of laboratory results to LOINC. Proc AMIA Symp 2000:472-6.

[9] Carter JS, Brown SH, Erlbaum MS, Gregg W, Elkin PL, Speroff T, et al. Initializing the VA Medication Reference Terminology Using UMLS Metathesaurus Co-Occurrences. Proc AMIA Symp 2002:116-20.

[10]Rosenbloom ST, Awad J, Speroff T, Elkin PL, Rothman R, Spickard IA, et al. Adequacy of representation of the national drug file reference terminology physiologic effects reference hierarchy for commonly prescribed medications. Proc AMIA Symp 2003:569-78.

[11]Cimino JJ, McNamara TJ, Meredith T, Broverman CA, Eckert KC, Moore M, et al. Evaluation of a proposed method for representing drug terminology. Proc AMIA Symp 1999:47-51.

[12]Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37(4-5):394-403.

[13]Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. ANSI Healthcare Informatics Standards Board Vocabulary Working Group and the Computer-Based Patient Records Institute Working Group on Codes and Structures. J Am Med Inform Assoc 1998;5(6):503-10.

[14]Elkin PL, Brown SH, Chute CG. Guideline for health informatics: controlled health vocabularies--vocabulary structure and high-level indicators. Medinfo 2001;10(Pt 1):191-5.

[15]Elkin PL, Brown SH, Carter J, Bauer BA, Wahner-Roedler D, Bergstrom L, et al. Guideline and quality indicators for development, purchase and use of controlled health vocabularies. Int J Med Inf 2002;68(1-3):175-86.

[16]ASTM-E31. E2087-00 Standard specification for quality indicators for controlled health vocabularies. Conshohocken, PA: American Society for Testing and Materials; 2000.

[17]ISO/TC 215 Health informatics. ISO 17117 Health informatics — Controlled health terminology — Structure and high-level indicators. ISO Technical Specification 2002;2002-02-15:32.

[18]Rassinoux AM, Michel PA, Wagner J, Baud R. Current trends with natural language processing. Medinfo 1995;8(Pt 2):1657.

[19]Dolin RH, Alschuler L, Beebe C, Biron PV, Boyer SL, Essin D, et al. The HL7 Clinical Document Architecture. J Am Med Inform Assoc 2001;8(6):552-69.

[20]Elkin PL, Bailey KR, Chute CG. A randomized controlled trial of automated term composition. Proc AMIA Symp 1998:765-9.

[21]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.

Address for Correspondence

Steven Brown MD

1310 24th Avenue South, Nashville Tennessee USA. 37212

steven.brown@med.va.gov