Initializing the VA Medication Reference Terminology Using UMLS Metathesaurus Co-Occurrences

John S. Carter, MBA[1,2], Steven H. Brown, MD, MS[3,4], Mark S. Erlbaum, MD, MS[2], William Gregg, MD[4], Peter L. Elkin, MD[5], Ted Speroff, PhD[4], Mark S. Tuttle, FACMI[2]

[1]University of Utah [2]Apelon, Inc. [3]Department of Veterans Affairs [4]Vanderbilt University [5]Mayo Foundation for Medical Education and Research

ABSTRACT

We developed and evaluated a UMLS Metathesaurus Co-occurrence mining algorithm to connect medications and diseases they may treat. Based on 16 years of co-occurrence data, we created 977 candidate drug-disease pairs for a sample of 100 ingredients (50 commonly prescribed and 50 selected at random). Our evaluation showed that more than 80% of the candidate drug-disease pairs were rated "APPROPRIATE" by physician raters. Additionally, there was a highly significant correlation between the overall frequency of citation and the likelihood that the connection was rated "APPROPRIATE." The drug-disease pairs were used to initialize term definitions in an ongoing effort to build a medication reference terminology for the Veterans Health Administration. Co-occurrence mining is a valuable technique for initializing term definitions in a large-scale reference terminology creation project.

BACKGROUND AND INTRODUCTION

The Department of Veterans Affairs (VA) Veterans Health Administration (VHA) provides health care to more than four million veterans and dependants. VHA has developed and deployed a variety of electronic tools to assist clinicians, including VISTA[1] (Veterans Integrated Service and Technology Architecture), CPRS[2] (Computerized Patient Record System), BCMA[3] (Bar Code Medication Administration), and others.

The VHA National Drug File (NDF)[4] is a nationally maintained medication terminology used to support VHA clinical applications. NDF is used at each of VHA's 172 medical centers for order entry, decision support and to send outpatient prescriptions to regional automated mail-out pharmacies (57 million prescriptions in 2001). It includes information about drug costs, ingredients, and inventory management.

NDF is implemented as a single-inheritance hierarchy of approximately 400 Drug Classes, approximately 11,000 Drug Products, and approximately 80,000 National Drug Codes (Figure 1). A cross-reference file lists the ingredients of each product.

VHA is continually looking for ways to improve care quality, promote patient safety, and reduce costs through a variety of means, including information technology. One area under investigation is the use

```
National Drug File

   AH000 : Antihistamines
   AM000 : Antimicrobials
   .  .  .
   CV000 : Cardiovascular Medications
      CV050 : Digitalis Glycosides
      CV100 : Beta Blockers/Related
      .  .  .
      CV700 : Diuretics
         CV702 : Loop Diuretics
            FUROSEMIDE 10MG/ML INJ
            FUROSEMIDE 20MG TAB
               NDC : 00005370823
               NDC : 00005370831
               .  .  .
            .  .  .
         .  .  .
```

Figure 1: Veterans Health Administration National Drug File Drug Class Hierarchy Sample.

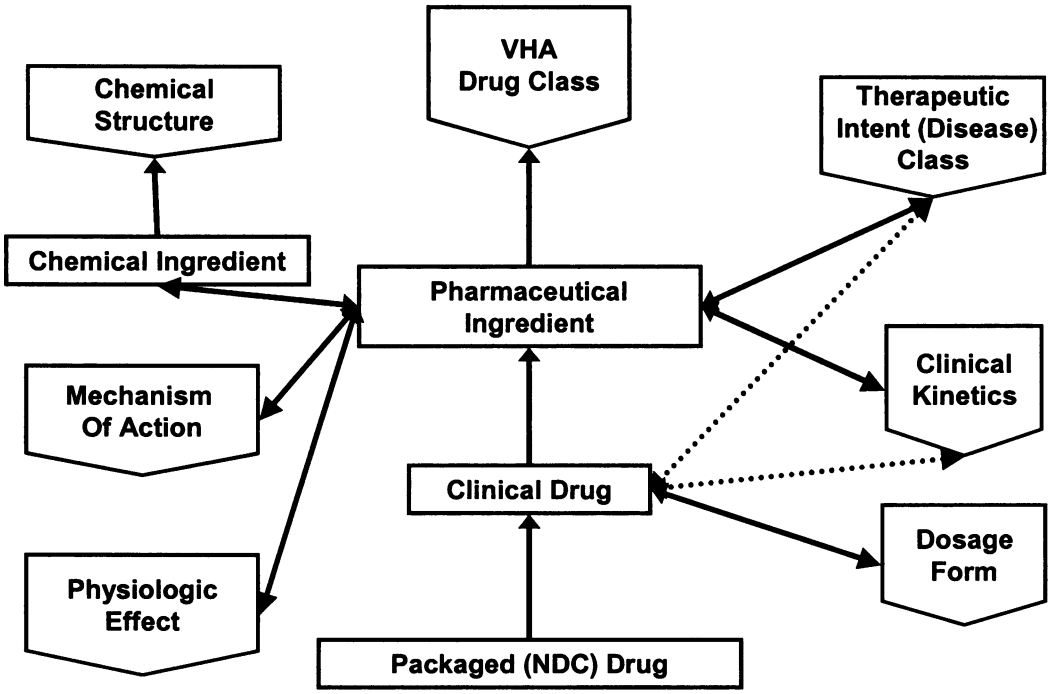

Figure 2: The NDF-RT Model. Shield-shapes represent multiple-inheritance reference hierarchies. Rectangles are named sets of concepts each representing a level of abstraction used to describe medications. Single-headed arrows are IS-A hierarchies, and double-headed arrows represent other semantic relationships. Solid arrows are the most commonly instantiated connections, while dotted arrows are used to describe unusual or problematic cases. NDC=National Drug Code. VHA=Veterans Health Administration.

of reference terminologies and terminology services to permit retrospective and real-time aggregation, comparison, and sophisticated decision support.

Reference terminologies are a "third generation terminology" as described by Rossi Mori.[5] Reference terminologies have a formal definition for each term, and are designed for data aggregation and retrieval. Formal definitions can be represented using symbolic logic and can be manipulated by computer. Reference terminologies should have other characteristics[6, 7] including concept orientation, domain completeness, the use of meaningless identifiers, and the ability to support multiple consistent views. Formal terminologies have been found to aid the terminology mapping process[8] and to reduce term maintenance costs.[9]

VHA's initial reference terminology project is NDF-RT, a formalization of the National Drug File. Medications were selected for several reasons. Along with their many benefits, medications pose a variety of risks including adverse reactions and interactions with other substances. Secondly, medications are one of the fastest-growing components of health care costs. Finally, a recent review of existing medication terminology products found several areas for improvement.[10] NDF-RT will show how drug products are different (which could be accomplished using simply a unique identifier such as the National Drug Code), and also how they are similar. The ability to browse, retrieve and aggregate drugs along multiple axes of similarity is necessary to support functions such as decision support and retrospective analysis. Maintaining the familiar Drug Class structure is also necessary for users to have a familiar and clinically relevant drug terminology. The three basic steps being followed to create NDF-RT are model development, reference taxonomy development, and term definition.

NDF-RT uses a Description Logic-based reference model (Figure 2) adapted from the Government Computer-Based Patient Record (GCPR) Project's Reference Terminology Model Demonstration Project.[11] The model includes orthogonal hierarchies for chemical structure, mechanism of action, physiologic effect, clinical kinetics, and therapeutic intent (disease) class, while preserving the existing VHA Drug Classes.

We initialized the Mechanism of Action, Physiologic Effect, and Chemical Structure axes by matching VHA ingredient names to the National Library of Medicine's Medical Subject Headings (MeSH)[12] terms. The MeSH "D – Chemicals and Drugs" tree

provides the chemical structure hierarchy, and selected Pharmacologic Action links were used to initialize the mechanism of action and physiologic effect trees. However, MeSH does not categorize drugs by the diseases or manifestations they treat. With more than 3,000 ingredients, each treating multiple diseases, we sought a way to algorithmically initialize the "therapeutic intent" axis as an efficiency measure.

Medline indexing is based on assigned keywords from the MeSH vocabulary. An article's keywords also can be further specified using a controlled set of qualifiers. The UMLS[13] (Unified Medical Language System) includes a file (MRCOC) listing the frequency of all qualified index term heading pairs. Zeng and Cimino[14] stated that the co-occurrence pairs show good sensitivity for drug-disease relationships. Burgun and Bodenreider[15] found that drug-disease co-occurrences were among the most frequent in the UMLS. Based on these findings, we hypothesized that the MeSH co-occurrences would provide useful drug-disease links and that an algorithm to usefully initialize NDF-RT based on mining the co-occurrence data could be created. This hypothesis is evaluated in the remainder of this manuscript.

METHODS

We combined the co-occurrence files from 1986 to 2001 (total 16 years) and searched as follows. First, we mapped ingredients from the VHA NDF to MeSH Main Headings using a combination of lexical matching and human review. Then, we collected all co-occurrence pairs pointing from one of these drug ingredients to a MeSH disease heading. We kept drug-disease pairs where the TU (therapeutic use) qualifier frequency was greater than the maximum frequency of either the AE (adverse effects), PO (poisoning), or TO (toxicity) qualifiers. We then excluded pairs where the number of drug-disease co-occurrences identified by the TU qualifier was less than 60% of the total co-occurrences (including the AE, PO, TO, TU and all other qualifiers) for that drug-disease pair.

The resulting data produced a prohibitively large number of drug-disease pairs for review, primarily because if the drug-disease co-occurrence only occurred once with the TU qualifier, our algorithm included it. To make the review task more manageable, we limited our review to those pairs that occurred five or more times in the 16 years. For example, our raw data included 59 rows for furosemide, a commonly prescribed ingredient in diuretic medications, of which only the eight in Figure 3 occurred five or more times.

```
Ingredient:
        Furosemide

Disease Categories:
        Ascites
        Bronchopulmonary Dysplasia
        Edema
        Heart Failure, Congestive
        Hemorrhage
        Liver Cirrhosis
        Nephrotic Syndrome
        Pulmonary Edema
```

Figure 3: Sample Candidate Ingredient-Disease Pairs.

For our experiment, we sought to determine whether the algorithmic initialization of the terminology was clinically relevant. We selected 100 ingredients from the 3,300 total ingredients. First, we included the 50 most commonly prescribed drugs from the Nashville VAMC. These represent two-thirds of total prescriptions and more than half of total prescription costs. Therefore, we expected that the reviewers would easily be able to review the drug-disease connections for these ingredients, and that the review of these ingredients would prove of immediate value to VHA. We then selected 50 additional ingredients at random from the remaining ingredients. We prepared a spreadsheet with one line for each drug-disease candidate pair.

Three of the authors (SB, PE, WG) used the 9-point Rand Appropriateness Scale[16] to rate the appropriateness of the candidate drug-disease connection according to the following criteria:

> "The use of medication X for the prophylaxis of, treatment of, or diagnosis of the disease noted, or its associated symptoms, or closely associated diseases (e.g. specific opportunistic infections in AIDS) is APPROPRIATE (score 7-9) / AMBIGUOUS (score 4-6) / NOT APPROPRIATE (score 1-3) given the usual course of the disease being treated, the usual risks of the medication and the usual benefits derived from that medication."

Within the larger groupings of APPROPRIATE, AMBIGUOUS and NOT APPROPRIATE, the reviewers considered the likelihood and sensibleness of the disease-drug association based on their clinical

experience. The reviewers worked independently and did not discuss their results until all ratings were turned in. The review task took between four and eight hours per reviewer.

RESULTS

Of the 50 most common VHA ingredients, 15 were not initialized either because they are not MeSH main headings or because our mapping algorithm failed to identify them.

Initial examination of the results revealed a systematic discrepancy between one of the raters' interpretation of the scoring task and the interpretation of the other two. Based on a follow-up discussion among the raters, the outlying rater agreed that he used the extreme ends and not the full 9-point scale, resulting in dichotomizing the scale rather than rating according to the written instructions. That rater's data are not included in the results presented.

Each of the two remaining reviewers examined a total of 498 drug-disease co-occurrence pairs for the commonly prescribed ingredients and 479 co-occurrence pairs for the randomly selected ingredients. There was no statistically significant difference between the raters for either the common ingredients (McNemar's test p = .411 with an intraclass correlation coefficient = .47) or the random ingredients (McNemar's test p = .08 with an intraclass correlation coefficient = .21). Therefore, the raters' scores were averaged into a pooled score. Based on pooled scores, 414 of 498 (83.1%) of the common ingredient co-occurrences and 378 of 479 (78.9%) of the random ingredient co-occurrences were rated as "APPROPRIATE." There was a highly significant ($p < .001$) correlation between the citation frequency of a drug-disease pair and its appropriateness rating, as shown in Figure 4.

Of the remaining candidates, 14.9% of the common and 19.6% of the random ingredient pairs were rated "AMBIGUOUS" and the remainder were rated "NOT APPROPRIATE" (2% of common and 1.5% of random).

Examination of the cases where the raters disagreed the most (greatest absolute difference between scores) revealed that a small number of ingredients caused the majority of these disagreements. For example, the ingredient "cilastatin sodium" was rated "APPROPRIATE" against a wide range of infectious diseases by one rater and "NOT APPROPRIATE" by the other. Discussion revealed that one rater considered the ingredient as if it were combined with

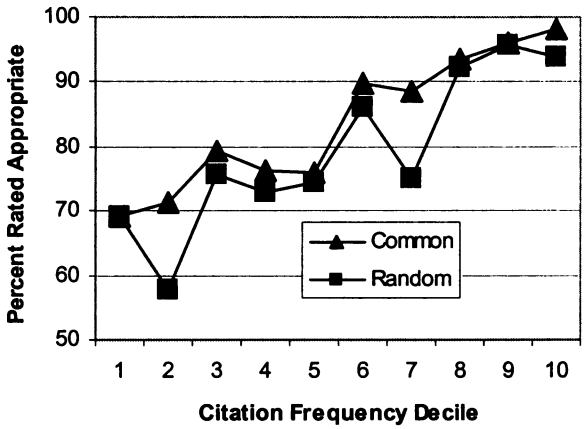

Figure 4: Highly significant correlation between citation frequency and "APPROPRIATE" rating ($p < .001$).

the antibiotic imipenem, which is the usual form in which it is administered.

DISCUSSION

Our results indicate that mining UMLS Metathesaurus MeSH co-occurrences to initialize term definitions as part of the reference terminology creation process can result in the creation of useful data for human editors to review. Even though up to 30% of the NDF ingredients were not successfully initialized in this experiment, the mining algorithm uncovered more than 25,000 candidate drug-disease connections. This study suggests that more than three fourths of those may be valid and retained in the terminology.

As NDF-RT development continues, the task of the human editors will be to eliminate inappropriate diseases from the list generated in this experiment rather than to create the entire disease list from scratch. We anticipate that this will result in a substantial productivity boost. The fact that only 50 ingredients account for a majority of prescriptions filled and total costs suggests that incremental success can be achieved without an extensive long-term modeling process.

One limitation of this evaluation is that we do not measure the false negatives, that is, the diseases that are in fact treated by a drug but are not included in the candidate pairs. Mining data from another source, such as a corpus of patient records, will be required to capture any missing "real-world" uses of medication ingredients.

The decision to include drug-disease links in the terminology can itself be challenged. However, as Rector[17] points out, the test of a terminology is how well it supports software for key functions, including data entry, information retrieval, mediation, indexing and authoring. In the VHA's clinical environment, data entry, information retrieval and authoring for medications most often takes place in the context of a patient's disease. In addition, drugs are only approved for sale by the FDA in the context of a specific, limited set of diseases, manifestations or diagnostic situations in which they have been found to be "safe and effective." Therefore, while a complete list of diseases, contraindications, interactions and the like are indeed outside the scope of the terminology, we argue that a limited list of clinically important diseases appropriately treated by a given drug do form a part of the drug's clinical definition. Further exploration of the boundaries between terminologies and knowledge bases is needed for medications and other subject domains.

Initializing the medication reference terminology using UMLS Metathesaurus MeSH co-occurrences provides a way to jump-start the terminology creation and maintenance process while taking advantage of work already done. Our results show that this method produces a large quantity of clinically relevant information. This suggests that other terminology and knowledge base initialization efforts could benefit from a similar method.

Acknowledgements

The authors would like to acknowledge the support of Jim Demetriades of the VHA Enterprise Architects Group and Don Lees, RPh, of the VHA Pharmacy Benefits Management Group. Nels Olson and Munn Maung provided programming support.

REFERENCES

1. Kolodner R, editor. Computerizing large integrated health networks: the VA success. New york: Springer; 1997.
2. Meldrum K, Volpp B, Vertigan R. Department of Veterans Affairs' Computerized Patient Record System. Proceedings of the AMIA Annual Symposium 1999(1-2):1214.
3. Johnson CL, Carlson RA, Tucker CL, Willette C. Using BCMA software to improve patient safety in Veterans Administration Medical Centers. Journal of Healthcare Information Management 2002;16(1):46-51.
4. http://www.vapbm.org/PBM/natform.htm
5. Rossi Mori A, Consorti F, Galeazzi E. Standards to support development of terminological systems for healthcare telematics. Methods of Information in Medicine 1998;37(4-5):551-63.
6. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of Information in Medicine 1998;37(4-5):394-403.
7. Subcommittee E. Standard Specification for Quality Indicators for Controlled Health Vocabularies. West Conshohocken: American Society for Testing and Materials; 2001.
8. Lau LM, Johnson K, Monson K, Lam SH, Huff SM. A method for the automated mapping of laboratory results to LOINC. Proceedings of the AMIA Annual Symposium 2000:472-6.
9. 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.
10. Lau LM, Lam SH. Applying the desiderata for controlled medical vocabularies to drug information databases. Proceedings of the AMIA Annual Symposium 1999:97-101.
11. Carter JS, Brown SH, Nelson SJ, Lincoln MJ, Tuttle MS. The Creation and Use of a Reference Terminology for Inter-Agency Computer-based Patient Records: The GCPR RTM Demonstration Project. Proceedings of the AMIA Annual Symposium 2001:809.
12. Lipscomb CE. Medical Subject Headings (MeSH). Bull Med Libr Assoc 2000;88(3):265-6.
13. http://www.nlm.nih.gov/research/umls/UMLSDOC.HTML
14. Zeng Q, Cimino JJ. Automated knowledge extraction from the UMLS. Proceedings of the AMIA Annual Symposium 1998:568-72.
15. Burgun A, Bodenreider O. Methods for exploring the semantics of the relationships between co-occurring UMLS concepts. Medinfo 2001;10(Pt 1):171-5.
16. Merrick NJ, Fink A, Park RE, Brook RH, Kosecoff J, Chassin MR, et al. Derivation of clinical indications for carotid endarterectomy by an expert panel. American Journal of Public Health 1987;77(2):187-90.
17. Rector AL. Thesauri and formal classifications: terminologies for people and machines. Methods of Information in Medicine 1998;37(4-5):501-9.