Evaluation of a Proposed Method for Representing Drug Terminology

James J. Cimino, M.D.,[1] Timothy J. McNamara, M.D., M.P.H.&T.M.,[2]
Terri Meredith, R.Ph.,[2] Carol A. Broverman, Ph.D.,[3] Karen C. Eckert, M.P.M., R.Ph.,[3]
Michael Moore, M.D.,[4] David J. Tyree, Pharm.D.,[4]
[1] Columbia University Department of Medical Informatics, New York, New York
[2] Multum Information Services, Inc., Denver, Colorado
[3] First DataBank, San Bruno, California
[4] Micromedex, Inc., Englewood, Colorado

In the absence of a single, standard, multipurpose terminology for representing medications, the HL7 Vocabulary Technical Committee has sought to develop a model for such terms in a way that will provide a unified method for representing them and supporting interoperability among various terminology systems. We evaluated the preliminary model by obtaining terms, represented in our model, from three leading vendors of pharmacy system knowledge bases. A total of 2303 terms were obtained, and 3982 pair-wise comparisons were possible. We found that the components of the term descriptions matched 68-87% of the time and that the overall descriptions matched 53% of the time. The evaluation has identified a number of areas in the model where more rigorous definitions will be needed in order to improve the matching rate. This paper discusses the implications of these results.

INTRODUCTION

Many clinical information tasks can benefit from the use of a standard terminology for representing drug information, including electronic medical records, automated decision support, quality assurance, health care research, reimbursement, and mandatory reporting.[1] The Food and Drug Administration's National Drug Codes (NDC) is a tempting choice for such uses. It has the advantage of being widely available and in the public domain. However, these codes are created by individual drug suppliers, rather than by a central authority, with very few rules regarding editorial management. As a result, the NDC codes lack many of the widely-recognized desirable characteristics for controlled terminologies.[2,3] Ample evidence of these shortcomings is demonstrated by the fact that several pharmacy system knowledge base vendors (PSKBVs) expend considerable effort to construct their own drug terminologies, independent from (although inclusive of) the NDC codes. Unfortunately, these terminologies are not interchangeable in their current forms, contributing to the prevalent Tower of Babel state found in medical information systems.[4]

Recent efforts in terminology development and maintenance have turned toward the use of knowledge representation techniques to capture the meaning of individual terms in order to promote better understanding of terms,[5] support integration of terms from multiple sources into a single terminology,[6] and exchange coded patient data.[7] In this approach, the meanings of terms are made explicit through the use of definitions that adhere to a formal structure, using descriptive terms that are themselves drawn from controlled terminologies. For example, in the Logical Observations, Identifiers, Names and Codes (LOINC) system, laboratory test terms are represented by stating the substance measured by the test, the property measured, the specimen, the method, and so on.[8]

Knowledge-based representation of terms not only supports use of the terminology, but evaluation of it as well. Multiple parties can create standardized definitions of the terms; subsequent comparison of the definitions can help identify ambiguity in the knowledge representation scheme or in the terminology itself. For example, Campbell et al. have studied discrepancies among definitions created by multiple vocabulary editors of the SNOMED RT terminology.[9] Baorto et al. have compared LOINC encodings across multiple institutions to identify and rectify problems in the LOINC scheme.[10]

We have applied similar methods to evaluate a drug term representation scheme being developed by the HL7 organization.[11] Specifically, we are comparing definitions for entities called "clinical drugs" (defined below) to determine if the current knowledge representation scheme will support development of a way to unify or translate terms from individual PSKBVs.*

BACKGROUND

The HL7 organization provides a standard for electronic health messages for exchange between computer systems. The standard includes advice on how to use controlled terminologies to represent patient data in messages. In order to represent drug information, the standard allows for the use of NDC codes and codes used by pharmacy systems (that are, for the most part, supplied by one of a small number of PSKBVs). There is no single standard, however,

* This project was not intended to compare the quality or content of individual data sets. In order to avoid this, these sets are not specifically identified in such comparisons this paper.

Name	Dosage Form	Ingredient #1	Ingredient #2
Valium 5mg Tablet	Tablet	Diazepam^5^mg	
Tylenol #3	Tablet	Acetaminophen^325^mg	Codeine^30^mg
Chloral Hydrate Syrup	Syrup	Chloral hydrate^100.00000^mg^1.00000^ml	

Table 1: Formal representation of some clinical drugs. Each clinical drug has one dosage form and one or more active ingredients. The ingredients are composed of 3 required parts (active substance, ingredient strength, and ingredient units). Many drugs also include two optional fields to represent volume and volume units, when the ingredient is in some infinitely divisible form (such as a syrup) expressed as a concentration.

for supporting "plug and play" integration, nor is there a mechanism for translating reliably among these choices, should the sender choose one terminology and the receiver desire another. HL7 created a Vocabulary Technical Committee (VTC) to study this problem and recommend solutions.

Several PSKBVs are members of the VTC and are also interested in finding ways to support the interchange of data coded in their various proprietary terminologies. A variety of mechanisms are being explored, including the notion of placing the terminologies in the public domain. However, simply making the terminologies generally available does not solve the translation problem, unless one terminology were to predominate and become, by itself, a *de facto* standard. For a variety of reasons, such an outcome is not likely to happen soon.

The VTC therefore created a subcommittee to explore ways that HL7 might exploit the willingness of PSKBVs to share their terminologies. This subcommittee has convened several meetings and conducted lively on-line discussions. Based on this work, the subcommittee has created a hierarchical model for representing drug terms that includes a specification for creating formal definitions.[12] This model is preliminary in nature, and has not been subjected to formal HL7 ballot. Nevertheless, it provides a useful framework for discussing ongoing activities.

One of the key concepts in the model is the notion of a *clinical drug*.[13] Informally, a clinical drug is roughly equivalent to the concept embodied in a drug order, such as "diazepam 5 mg tablet". Formally, it requires explicit representation of two attributes: *active ingredients* and *dosage form*. Active ingredients are further defined to consist of one or more *ingredients* (usually chemicals) and an *ingredient strength* for each ingredient. Ingredient strength is further defined as having an *ingredient amount* and *ingredient units*; it may also include an *ingredient volume amount* and *ingredient volume units*. Table 1 shows some typical definitions.

METHODS

We obtained the September, 1998 NDC Code tables from the FDA's Web site (www.fda.gov) and extracted 1000 terms (1.4%) at random. Information regarding these 1000 terms, as represented in the FDA database, was supplied to all interested parties in the VTC. In addition to the NDC code, we supplied the FDA's unique identifier, the drug name, strength, units, form, dose size and package size.

We asked each PSKBV to return to us a set of clinical drug descriptions for each of the terms in the test set, if a match could be found in their respective systems based on NDC code. We examined the sets returned by the PSKBVs to determine the degree of overlap among the sets. Where the same drug was represented in two or more sets, we examined the similarities and differences of the various representations.

When differences between like terms occurred, we analyzed the differences, looking for systematic reasons (for example, tablets being called “TAB” in one data set and “TABLET” in another data set). In cases where the differences appeared to be solely due to a difference in naming (as with tablets), we normalized the data sets to eliminate the differences and then reanalyzed the pair-wise comparisons.

RESULTS

Three PSKBVs responded to the request for sample data by providing a total of 5 sets of drug descriptions. Table 2 shows the number of terms supplied in each set and the overlap with terms supplied from each of the other sets. In general, there was good overlap between sets. Overall, 367 terms were not represented in any set, 71 appeared in only one set, 77 appeared in exactly two sets, 83 appeared in three sets, 91 appeared in four, and 311 terms appeared in all five terminologies. Failure to map NDC terms was attributed to the presence of discontinued medications – that is, old FDA terms that the PSKBVs did not have in their databases.

Pair-wise comparisons were done based on the overlap between sets: terms covered by two sets have one pair-wise comparison, terms covered by three sets have three pair-wise comparisons, terms covered by four sets have six pair-wise comparisons, and terms covered by all sets provided ten pair-wise comparisons. Thus, a total of 3982 pair-wise comparisons (77∗1 + 83∗3 + 91∗6 + 311∗10) were available for review.

Comparison of Form

When identifying the dosage forms in each of the 3982 pair-wise comparisons, we found that only 236 cases (6%) had exact matches. After converting the form names to all upper case, exact matches increased to 645, or 16%. Simple equivalence was established based on lexical similarity. Thus "TAB" was considered equivalent of "TABLET", but "LIQUID" was not assumed to be synonymous with "ORAL LIQUID". After establishing 111 dosage form synonyms, equivalence mappings were accomplished for 2859 (72%) of the pair-wise comparisons

Comparison of Ingredients

Since many drugs had multiple ingredients, we examined the pair-wise match of ingredients within each of the 3982 pair-wise drug description comparison. Each ingredient from the drug with the most ingredients was compared to the ingredient list of the other drug in the pair. With this algorithm, there were a total of 5507 comparisons.* Initially, only 1273 matches could be made but, after converting all drug names to upper case, we found 3607 (65%) exact matches.

When we examined the nonmatches, we found many examples of synonyms, the most common being terms containing "HYDROCHLORIDE" versus "HCL". In some cases, failure to match was due to differences in punctuation or word order. In other cases, the differences were due to inappropriate inclusion of dose route (e.g., "ORAL") and strength information (e.g., "0.9%") in the ingredient name. In all, 54 transformation rules were created to handle these simple cases of synonyms. In other cases, however, synonymy was not assumed. In most of these, the differences were due to inconsistent mention of a salt form of the ingredient (such as "HEPARIN" versus "HEPARIN SODIUM") and the animal source (e.g., "PORCINE HEPARIN SODIUM" and "HUMAN RECOMBINANT INTERFERON ALPHA 2A"). After applying the 54 transformation rules, we found exact matches in 4337 (79%) of the comparisons. The remaining 1170 included 502 (9%) failures to match due to a discrepancy between the number of ingredients listed in the drug descriptions. Thus, only 663 (12%) of the comparisons failed due to differences in ingredient names.

* This maximizes the denominator in the comparisons, such that when one term has more ingredients than another, the extras will be counted as nonmatches.

Comparison of Ingredient Strength and Units

When a match was found between ingredients (4337 instances), an examination was made of the ingredient strength and units. Initially, the format of the numeric strength values from each vendor differed widely (e.g., "500", "500.00", and "00500.00"), and the number of matches was only 374 (9%). After converting these values to real numbers, however, the number of matches rose to 3262 (75%).

The initial comparison of units showed 1845 matches (43%). This was due, in large part, to differences in abbreviations (e.g., "G", "GM", and "GRAM(S)"). After conversion of units to a standard form, 2964 matches (68%) were found. Of the remaining 1373 mismatches, 655 (15%) were due to missing values (nulls) in the descriptions and 450 (10%) were due to the inappropriate inclusion of concentration information (e.g., "MG/5ML") in the strength units field. In 56 (1.3%), the units differed by a factor 1000 (e.g., "GM" versus "MG") or even 1,000,000 (e.g., "GM" versus "MCG") with a corresponding difference in the strength (suggesting that additional normalization could correct for this). The remaining differences were due to the use of "%" in one of the pairs and "MG" or "GM" in the other.

Comparison of Ingredient Volume and Units

As with the strength and strength units, when an ingredient comparison was made (4337), we examined the volume and volume units field. Since these fields are optional, we found that some data sets the default value for the volume was null, while in other sets the default was 1 or 0. Different numeric formats (similar to the strength field) were also used. A simple comparison of the volume fields showed 1054 (24%) matches. After removing 0's and 1's, 3754 matched (87%).

Similarly, the default volume units varied from null to "EACH" and legitimate values were often in

Set	Mapped	A	B	C	D	E	None
A	358	-	340	358	326	357	642
B	459	340	-	452	361	449	541
C	605	358	452	-	393	554	395
D	405	326	361	393	-	392	595
E	566	357	449	554	392	-	434

Table 2: Number of terms mapped in each set and degree of overlap across any two data sets. Right-most column refers to terms in each set that were not mapped by the PSKBV.

different forms (e.g., "HOUR(S)" versus "HR"). The initial comparison showed 2319 matches (53%), which improved to 3486 (80%) after applying conversion rules.

Overall Matching

Table 3 summarizes the findings from each of the comparisons. Once a method was developed for comparing description components, it was possible to see which of the pair-wise comparisons matched on *all* components and the overall comparisons. With no conversion at all, the comparisons yielded *no* matches, mainly due to the different formats used for the strength and volume fields. With simple conversion of the fields to numeric form, the number of matches was 1122 of 3982 (28%). After applying all conversion rules, the matches rose to 2128 (53%).

DISCUSSION

The drug terminology modeling work being carried out by HL7 is still in its preliminary stages. Indeed, there remains great uncertainty over whether the proposed approach will prove to be valuable for achieving conversion of terms across different terminologies. The experiment presented in this paper is a first step in examining this hypothesis.

The results presented here are replete with lower-than-desired values. However, careful analysis is required in order to understand the implications of each of the steps in the matching process. First, the data sets from the vendors fell far short of the 1000 terms in the NDC sample set. As previously stated, this is due to the historical nature of the NDC code file. It is unlikely that older drugs are any more difficult to model than current ones and are, in any case, unlikely to appear in messages from pharmacy systems; the ability to adequately represent these terms is therefore of relatively little importance.

Analysis of the drug form matching shows that a simple standardized terminology can go a long way toward supporting automated term matching. In fact, the VTC is working on a second project to unify the drug form terminologies of all the PSKBVs. It is beyond the scope of the current project to comment on the ultimate feasibility of unifying drug form terminologies, but preliminary results are encouraging (Shalaby J, in presentation to HL7).

	Total	Before Conversion	After Conversion
Components:			
Ingredients	5507	3607 (65%)	4337 (79%)
Strength	4337	374 (9%)	3262 (75%)
Units	4337	1845 (43%)	2964 (68%)
Volume	4337	1054 (24%)	3754 (87%)
Volume Units	4337	2319 (53%)	3486 (80%)
Overall:			
Each Ingred	4337	0 (0%)	2773 (64%)
All Ingred	3982	0 (0%)	2519 (63%)
Form	3982	645 (16%)	2859 (72%)
Complete Drug	3982	0 (0%)	2128 (53%)

Table 3: Summary of comparison statistics.

Matching on ingredients falls short in two ways: name and count. Although there is no single standard for chemical names, agreeing on one (such as Chemical Abstracts, or SNOMED) would probably go a long way toward resolving the types of discrepancies found in this study. Resolving the discrepancies in the number of ingredients is more problematic, and suggests vendor-specific differences in editorial policy. On-going discussions by the VTC will be needed to determine if these differences are resolvable.

The differences in strength and concentration are probably the easiest to resolve. The present guidelines developed by the VTC have provided a great deal of latitude in interpretation and, clearly, the vendors have exercised this latitude. For example, there is no rule about how to express concentration of liquids. Thus, the concentration of ingredients in a cough syrup might be expressed as quantity per milliliter or per teaspoon (5 milliliters). We are hopeful that a modest amount of additional attention paid to this part of the model will yield favorable results.

Finally, we are left with the question of how to interpret the overall result of a 53% match rate. On the one hand, this is much better than would be expected from the simple product of the match rates of each of the components (.79*.68*.87*.80*.80*.72, or 22%, from Table 3). This suggests that the majority of drug terms are well-behaved and easy to model, while problems mapping separate components tend to accrue in a smaller percentage of "troublesome" terms. However, in order for automated translation to be of real value, a match rate of 53% falls far short of what will be deemed adequate.

The work described in this paper represents only the first step in a difficult task. We believe there is much here that is encouraging, given that the rules for generating the data sets left much room for interpretation. The participation of the three PSKBVs is significant: their products represent the terminologies used in the vast majority of pharmacy systems in the United States. If the preliminary method described here can be expanded (through the use of a few additional rules and naming conventions) to provide consistent, accurate translations across the term sets used by the participating PKSBVs, it will establish an enormous potential for sharing clinical drug data among health information systems,

including those used for patient care and clinical research.

The ability to translate medication data as "clinical drugs" (as opposed to specific products or packages), at an almost-100% level may be a powerful tool for some applications, such as clinical research. It may be woefully inadequate for others, such as automated drug dispensing. The volunteer workforce of HL7 will have to decide whether the results presented here are *sufficiently* encouraging to justify investment in a full-scale, rigorous attempt to generate full tests sets (with tens of thousands of terms in each). The answer to this question will depend on whether one wishes to interpret a simple result of 53% as being a glass approximately half empty or half full.

CONCLUSION

We have developed a model for representing clinical drug terms and applied that model to random samples of five commercial products. With relatively little effort to standardize or normalize the resulting representations, the individual components matched fairly well, and the overall matching of drug terms was encouraging. This study lays the groundwork for developing the model and representation rules further, in order to achieve better mapping between existing proprietary drug terminologies.

Acknowledgments

The authors readily acknowledge that credit for the model presented here, and the impetus to evaluate it, go to the too-many-to-mention members of the HL7 Vocabulary Technical Committee who have participated in meetings and e-mail discussions over the past two years, especially co-chairs Ed Hammond and Stan Huff. Thanks also go to the HL7 Steering Committee and Board of Directors, for their spiritual and financial support, and the HL7 Staff, led by Mark McDougal and Karen van Hentenryk, for their technical and logistical support.

References

1. Board of Directors of the American Medical Informatics Association. Standards for medical identifiers, codes, and messages needed to create an efficient computer-stored medical record. JAMIA. 1994;1:1-7.
2. 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. JAMIA. 1998;5:503-510.
3. Cimino, JJ. Desiderata for controlled medical Vocabularies in the Twenty-First Century. Methods of Information in Medicine; 1998;37:394-403.
4. McDonald CJ, Overhage JM, Dexter P, Takesue B, Suico JG. What is done, what is needed and what is realistic to expect from medical informatics standards. Int J Med Inf 1998;48(1-3):5-12.
5. Spackman KA, Campbell KE. Compositional concept representation using SNOMED: towards further convergence of clinical terminologies. JAMIA 1998;5(suppl):740-744.
6. Cimino JJ, Johnson SB, Hripcsak G, Hill CL, Clayton PD. Managing vocabulary for a centralized clinical system. In Kaihara S, Greenes RA, eds. Proceedings of the World Congress on Medical Informatics - Medinfo '95; Vancouver, Canada; Healthcare Computing and Communications Canada, Edmonton, Alberta, 1995: 117-120.
7. Huff SM, Rocha RA, McDonald CJ, et al. Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary. JAMIA 1998;5(3):276-92.
8. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leaville D, Leland D, Fiers T, Chalrse L, Griffin B, Stalling F, Tullis A, Hutchins K, Baenziger J. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. Clinical Chemistry. 1996:42:81-90.
9. Campbell KE, Cohn SP, Chute CG, Shortliffe EH, Rennels G. Scalable methodologies for distributed development of logic-based convergent medical terminology. Methods Inf Med 1998;37(4-5):426-39.
10. Baorto DM, Cimino JJ, Parvin CA, Kahn MG. Using Logical Observation Identifier Names and Codes (LOINC) to exchange laboratory data among three academic hospitals. JAMIA; 1997;4 (Suppl):96-100.
11. Health Level 7: http://www.hl7.org
12. Cimino JJ, Huff SM, Broverman CA, McNamara T, Nelson SJ. Panel: Development of a standard terminology to support medication messages. Presented at the 1998 AMIA Annual Fall Symposium, Orlando, Florida.
13. Sperzel WD, Broverman CA, Kapusnik JE, Schlesinger JM. The need for a concept-based medication vocabulary as an enabling infrastructure in health informatics. JAMIA, 1998;5(suppl):865-869.