Comparing Pharmacologic Classes in NDF-RT and SNOMED CT

Jonathan Mortensen[1], Olivier Bodenreider*[2]

[1]Department of Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio, USA
[2]National Library of Medicine, Bethesda, Maryland, USA

Email: Jonathan Mortensen - Jonathan.Mortensen@case.edu; Olivier Bodenreider*- olivier@nlm.nih.gov;

*Corresponding author

Abstract

Background: Clinical decision support systems and semantic mining require interoperable representations of pharmacologic classes across reference terminological systems. We explore two such systems: NDF-RT and SNOMED CT.

Methods: We evaluate the overlap of pharmacologic classes in NDF-RT (VA Classes) and SNOMED CT. We compare classes based on the set of their members (drugs) across systems, using the Jaccard coefficient as a measure of overlap between two classes.

Results: There is a limited overlap among the two systems. The average Jaccard value is 0.293. Only 11.5% of the VA classes have a Jaccard value of 0.75 or above.

Conclusions: The analysis of discrepancies between pharmacologic classes across systems offers a strategy for identifying classes in need of critical review. Due to the heterogeneity of the representation of pharmacologic classes in various terminologies, we recommend that drugs, not classes, be annotated in text for semantic mining purposes.

Introduction

Pharmacologic classes are typically established in reference to some of the properties of the active moiety, with respect to chemistry, physiology, metabolism and therapeutic intent [1]. For example, the classes *platelet aggregation inhibitors* and *anticoagulants* refer to the physiologic effect of drugs decreasing platelet aggregation and coagulation, respectively. In contrast, the class *cardiac glycoside* refers to the chemical structure of drugs such as *digoxin*, while the class *antianginal* refers to the therapeutic properties of some drugs on angina pectoris. Some classes are also defined in reference to several properties, e.g., *nitrate vasodilator*, referring to both the chemical structure of nitrates and their relaxing action on the musculature of blood vessels (physiologic effect). In other words, pharmacologic classes provide an abstract representation of drug properties, useful in the context of clinical decision support and for the annotation of biomedical resources, including clinical text and the biomedical literature.

While interoperability among terminologies is a requirement for clinical decision support, in which decision support rules are defined in reference to concepts in various terminologies (e.g., concepts for drug classes), it is also important that annotations to biomedical entities such as drug classes be consistent within and across datasets when such datasets are exchanged and integrated, as these annotations form the basis for knowledge discovery through semantic mining.

The National Drug File - Reference Terminology (NDF-RT) is a drug terminology produced by the Department of Veterans Affairs in the United States and is recommended as the standard in e-prescribing systems [2]. Other clinical terminologies such as SNOMED CT also include pharmacologic information.

The objective of this work is to evaluate the degree to which annotations to drug classes in various terminological systems are interoperable, with a focus on pharmacologic classes from NDF-RT. More specifically, we evaluate the overlap of VA classes to those in SNOMED CT. The analysis of the classes reveals discrepancies between the two systems and offers a strategy for identifying classes in need of critical review.

Background

In this section, we give a brief presentation of NDF-RT and SNOMED CT and present some related work on NDF-RT.

The **National Drug File - Reference Terminology (NDF-RT)** is based upon the National Drug File, a listing of medications produced by the Department of Veteran Affairs [3]. It serves as a reference standard for a variety of medical situations related to drugs and medications. NDF-RT is a description logic-based model available in OWL and XML formats. It includes 9 "Kinds" of information: Cellular or Molecular Interactions, Clinical Kinetics, Diseases Manifestations or Physiologic States, Pharmaceutical Preparations, Physiological Effects, RxNorm Dose Forms, Therapeutic Categories, and VA Drug Interactions. The Pharmaceutical Preparations hierarchy organizes drugs into three categories: Products by Generic Ingredient Combination, Products by VA Class and External Pharmacologic Classes (EPC).

There are 485 VA Drug Classes organized into a basic hierarchy. A drug generally belongs to only one class. Examples of VA classes include *ANTIMALARIALS*, of which the clinical drug *QUININE SO4 162.5MG TAB* is a member. Its parent class is *ANTIPROTOZOALS*. In addition, there are 425 EPCs (not used in this work). Differing from the VA Classes, the EPCs have a nearly flat hierarchy and are defined in reference to various properties, such as physiologic effect, therapeutic intent, ingredient and mechanism of action.

The July 11th 2010 Version of NDF-RT was used in the evaluation.

SNOMED CT is currently the largest clinical terminology. It is developed and maintained by the International Health Terminology Standard Development Organization (IHTSDO) [4]. In SNOMED CT, the drugs are simply related to pharmacologic classes through the ISA relationship. For example, there is an ISA relationship between the drug *Quinine* and the class *Cinchona antimalarial*. The January 31, 2010 Version of SNOMED CT was used in the study.

Related Work. Others have examined many aspects of NDF-RT. [5] investigated the coverage of the Physiologic Effects hierarchy in NDF-RT. It was found that the physiologic effects category was sufficient for classifying medications. [6] investigated the addition of pharmacogenomics into the hierarchy. [7] applied NDF-RT to mapping text from medication lists at the Mayo Clinic using the SmartAccess Vocabulary Server. NDF-RT covered 97.8% of the concepts found in the medication lists, indicating NDF-RT can be used in a clinical setting for medication purposes. [8] compared NDF-RT to the National Drug File, Medicare Part D and a proprietary knowledge base. It was determined that 76% of the classes from the three original terminologies were contained in NDF-RT. In recent work, [9] evaluated the correspondence of NDF-RT drugs and classes to RxNorm drugs and classes. As of October 2009, approximately 50% of the drugs did not correspond between terminologies. [10] mapped medications to diseases, showing a clear example of how NDF-RT can be applied in clinical decision support situations. As another example of clinical applications [11] integrated NDF-RT into the process of generating structured product labeling. Finally, [12] used the NDF-RT drug classes to determine the anti-coagulation status of patients based on their medication list, demonstrating a first step in clinical decision support.

Our study focuses not on content coverage, but rather on interoperability among systems of drug classes in various terminologies, including NDF-RT and SNOMED CT. These terminologies were chosen as a reference because they contain drug hierarchies, are mature, and are widely used. More specifically, we want to assess whether similar sets of drugs are linked to the same classes in different systems.

As part of the evaluation, we use a concept alignment technique described by [13]. NDF-RT was loaded into a Virtuoso endpoint [14] for SPARQL querying, which allowed for evaluation of the drug classes.

Methods

To evaluate the drug classes in NDF-RT, we developed an extensional method of evaluation, comparing between VA classes and SNOMED CT drug classes. Instead of comparing pharmacologic classes based on lexical resemblance of their names, we compare the extensions of these classes.

The extension of a pharmacologic class is the set of drugs a class has as members. The degree to which any two drug classes are similar was determined by the overlap of their extensions. This is measured by the Jaccard Coefficient,

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|},$$

where the intersection is the number of drugs which are the same between any two classes and the union is the total number of drugs between any two drug classes [15]. An example of extension is presented in Figure 1a. Here, the VA class *THROMBOLYTICS* and the SNOMED CT class *Thrombolytic* share 6 drugs, including *streptokinase*, while the drug *drotrecogin* is specific to the SNOMED CT class. The Jaccard value is computed as the cardinality of the intersection (6) over that of the union of the classes (7), i.e., 0.86. (In actuality, the classes are compared, not based on the ingredients, but based on the clinical drugs they have as members. The corresponding ingredients are shown in Figure 1a for brevity.)

The extension of each VA class is compared to that of every class in SNOMED CT. For a given VA class, the SNOMED CT class for which the highest Jaccard value is found is selected as the best match. The average Jaccard of the pairwise comparisons between NDF-RT and SNOMED CT is used to summarize the external comparison and determine the overall similarity between the two class systems.

To obtain an extension, the clinical drug members of a drug class were obtained. As opposed to VA (where drug classes are linked directly to clinical drugs), in SNOMED CT, the ingredients of a class were first obtained, then the clinical drugs for those ingredients were obtained using relations in NDF-RT, thus keeping the domain of clinical drugs limited to only NDF-RT. In addition, the drug members of a class included its drugs and all drugs which were members of any subclasses. For example, the clinical drug *QUININE SO4 260MG TAB* is linked directly to the VA class *ANTIMALARIALS*, but is also considered a member of the its parent class *ANTIPROTOZOALS*. Using the drug members, the drug member intersection was found, comparing the extension of the VA classes to the SNOMED CT classes.

Results

There are 485 VA and 722 SNOMED CT classes. To reduce comparisons (and noise), classes which did not have any drug members were removed. There were 95 VA (20%) and 195 SNOMED CT (27%) classes without members. Examples of classes with no members include *INVESTIGATIONAL ANTI-TUBERCULAR DRUGS* (VA), *ANTIFUNGALS,TOPICAL OTIC* (VA), *Antineoplastic alkaloid* (SNOMED CT) and *Corticosteroids used in the treatment of asthma* (SNOMED CT).

Among the 15,027 clinical drugs in NDF-RT, 8414 correspond to ingredients also present in SNOMED CT. Examples of clinical drugs specific to NDF-RT include medicinal products from classes such as *HERBS/ALTERNATIVE THERAPIES* (e.g., *WILD CHERRY BARK PWDR*).

The extensional comparison was obtained by calculating the overlap between the sets of drug members of class pairs and can be summarized by the average Jaccard coefficient for all class pairs between NDF-RT and SNOMED CT. Through their average Jaccard value, pairs of pharmacologic class systems can be compared for their overall similarity. The average Jaccard value is 0.293, indicating limited overlap overall between drug extensions across the two class systems.

The distribution of the average highest Jaccard

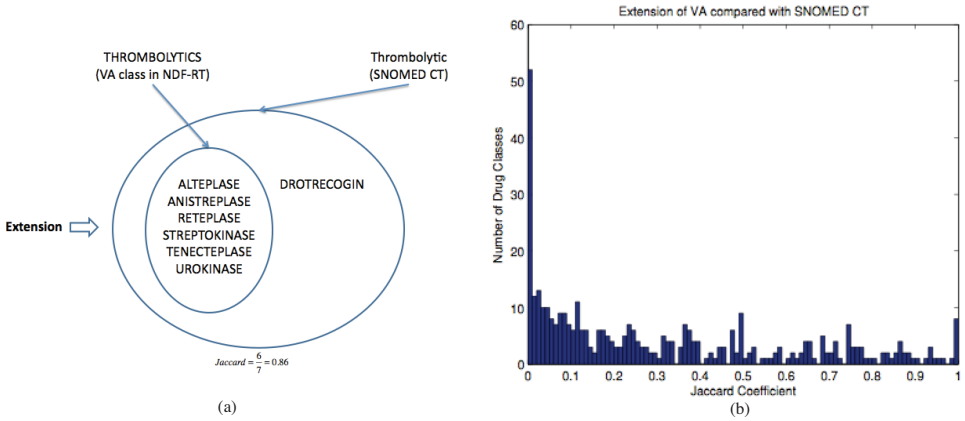

Figure 1: (a) Comparison of the extensions of the VA class *THROMBOLYTICS* and the SNOMED CT class *Thrombolytic*, (b) Distribution of Jaccard (highest per class)

value for the VA classes is shown in Figure 1b. Very few classes exhibit complete overlap (Jaccard = 1.0). Examples include the VA class *DIRECT RENIN INHIBITOR* and the SNOMED CT class *Renin Inhibitor*. This particular class contains the clinical drugs corresponding to only one ingredient, *aliskiren*. The proportion of VA classes with a Jaccard value of 0.75 or above is 11.5%. For example, the Jaccard value for the overlap between the VA class *THROMBOLYTICS* and the SNOMED CT class *thrombolytic* is 0.86. As shown in Figure 1a, the clinical drugs corresponding to six ingredients are common to both the VA class and the SNOMED CT class. These ingredients are *alteplase*, *anistreplase*, *reteplase*, *streptokinase*, *tenecteplase* and *urokinase*. Additionally, SNOMED CT also lists *drotrecogin* as a member of the class *thrombolytic*, although the indications for this drug seem to be limited to severe sepsis.

Finally, 75.6% of the VA classes have a Jaccard value lower than 0.5. For example, the Jaccard value for the overlap between the VA class *HYPEROSMOTIC LAXATIVES* and the SNOMED CT class *osmotic laxatives* is only 0.16. While clinical drugs corresponding to the ingredients *lactulose* and *magnesium sulfate* are common to both classes, many clinical drugs found in the VA class are not in the SNOMED CT class (e.g., other magnesium salts such as *magnesium biphosphate* and *magnesium hydroxide*). Interestingly, clinical drugs corresponding to *magnesium hydroxide* are part of a different SNOMED CT class, *saline hydroxide*. Conversely, solutions of *glycerol* are classified as *osmotic laxatives* in SNOMED CT, but as *LAXATIVES, RECTAL* in NDF-RT.

Discussion

Overlap among Class Systems

The similarity among the two pharmacologic class systems under investigation (VA and SNOMED CT) is relatively limited. The average Jaccard values among classes based on shared drugs is 0.293. The following reasons can be proposed as an explanation, in addition to sheer differences in classification and discrepancies illustrated in the section above. In some cases, there is no equivalent class in SNOMED CT for a given VA class, especially for high-level aggregation classes (e.g., *BLOODPRODUCTS / MODIFIERS / VOLUME EXPANDERS*), residual classes (e.g., *CARDIOVASCULAR AGENTS, OTHERS*) and classes specific to topical forms (e.g., *BETA-BLOCKERS, TOPICAL OPHTALMIC*). Another reason is that partially overlapping classes are defined using different classificatory criteria. For example, ophtalmic forms of beta-blockers such as *TIMOLOL MALEATE 0.5% GEL, OPH* are classified as *BETA-BLOCKERS, TOPICAL OPHTALMIC* in NDF-RT and as *anti glaucoma agent* in SNOMED CT. While the former class only contains beta-blockers, the latter includes a wider range of products (e.g., *apraclonidine*). Another difference between the two class systems is that

the pharmacologic class is a property of the clinical drug for the VA classes, whereas it is inherited through the ingredient for SNOMED CT classes. For example, injectable forms of *acetylcysteine* are classified as *ANTIDOTES/DETERRENTS, OTHERS* in NDF-RT, while topical solutions (e.g., for inhalation) are classified as *MUCOLYTICS*. In contrast, all forms of this drug are classified as both *drugs used in the treatment of paracetamol poisoning* and *mucolytic agent* in SNOMED CT. Finally, we also found a limited number of errors, such as the classification of the antibacterial drug *NORFLOXACIN 0.3% SOLN, OPH* as *BETA-BLOCKERS, TOPICAL OPHTALMIC* in NDF-RT.

Classes without Drug Members

One particular difference between the two pharmacologic class systems is the number of classes for which there is no drug in NDF-RT. (These classes were omitted from our statistics). These differences have different causes in different systems. For VA classes, most classes with no drugs correspond to investigational drugs. In contrast, in SNOMED CT, such classes correspond essentially to classes for which the corresponding medicinal products are out of the scope of NDF-RT, including blood products (e.g., *Red cells - irradiated*), dietary products (e.g., *Gluten free food product*) and various prescribable entities (e.g., *Sterile maggots*).

Consequences for Semantic Mining

As the sets of drugs available in terminological systems vary considerably across systems, with minimal overlap among them, annotation of the literature directly with classes from a given system is likely to result in annotated datasets that will not be interoperable, and whose annotations will be difficult to reconcile. Even if some terminologies such as SNOMED CT and NDF-RT tend to provide good coverage of clinical drugs, their overlap with other terminologies in terms of pharmacologic classes remains limited.

In practice, a better option for semantic mining is to annotate drugs rather than pharmacologic classes. Drugs names are relatively standard (at least at the ingredient level) and integration resources such as RxNorm are already available. Once resources have been annotated at the ingredient level, the corresponding classes can be added automatically in reference to the most useful pharmacologic class system in a particular context. Annotations to another pharmacologic class system can be recomputed from the ingredients in case of reuse of these resources for a different purpose.

Limitations and Future Work

There are a few limitations to this work. The evaluation was only a quantitative evaluation, comparing the two terminologies. It was not an evaluation of the clinical quality or the use of NDF-RT in a clinical situation. In addition, the domain of drugs used in the comparison was only clinical drugs in NDF-RT, as we assumed the clinical drugs to be complete. No comparisons were done at the ingredient level. Because of this, we obtained all the NDF-RT clinical drugs of an ingredient from the terminology. Some classes that have no drugs members may have had ingredients; however, these ingredients were either not present in NDF-RT or they did not have clinical drugs associated with them, resulting in the class not having drug members.

For a complete evaluation of NDF-RT, the external pharmacologic classes (EPCs) will be included in future work. To leverage these classes, they first must be enriched with drugs. A technique to do such an operation has been piloted by [12], which utilizes a description logics based classifier to classify drugs into EPCs.

In addition to the extensional approach used in this study, we would like to explore an intensional approach to comparing the classes, leveraging synonymy relations in the Unified Medical Language System (UMLS). In practice, classes across systems could be mapped through the UMLS and the extensions of equivalent classes could be compared.

Finally, this work may be considered a class centric approach, focused around drugs associated with classes. Future work will include a drug centric approach, which focuses on classes associated with drugs. More specifically, we will study the set of pharmacologic classes associated with a given drug in different pharmacologic class systems.

Conclusions

By using an automated method of comparing classes using drug class extensions, inconsistencies between terminologies were discovered. These inconsistencies serve as an indicator for possible review. The automated method of pairwise class member comparison complements standard lexical matching and can serve as an additional quality assurance tool for terminologies. This methodology sets a framework for pairwise comparison of drug classes between terminological systems using only their drug members. Finally, due to the heterogeneity of the representation of pharmacologic classes in various terminologies, we recommend that drugs, not classes, be annotated in text for semantic mining purposes.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Jonathan Mortensen and Olivier Bodenreider conceived and designed the study. Jonathan Mortensen acquired the data and performed the analysis and interpretation of the data. Both authors contributed to the redaction of the manuscript and approved its final version.

Acknowledgements

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM) in addition to the Choose Ohio First Scholarship Program.

References

1. Carter JS, Brown SH, Erlbaum MS, Gregg W, Elkin PL, Speroff T, Tuttle MS: **Initializing the VA medication reference terminology using UMLS metathesaurus co-occurrences.** In *Proceedings of the AMIA Symposium* 2002:116–20.
2. Miller RA: **Clinical Decision Support and Electronic Prescribing Systems: A Time for Responsible Thought and Action**. *Journal of the American Medical Informatics Association* 2005, **12**(4):403–409.
3. Lincoln MJ, Brown SH, Nguyen V, Cromwell T, Carter J, Erlbaum M, Tuttle M: **US Department of Veterans Affairs enterprise reference terminology strategic overview**. *Medinfo* 2004, **11**(Pt 1):391–395.
4. **SNOMED CT (Systematized Nomenclature of Medicine-Clinical Terms)**[http://www.ihtsdo.org/snomed-ct/].
5. Rosenbloom ST, Awad J, Speroff T, Elkin PL, Rothman R, III AS, Peterson J, Bauer BA, Wahner-Roedler DL, Lee M, et al.: **Adequacy of representation of the National Drug File Reference Terminology Physiologic Effects reference hierarchy for commonly prescribed medications**. In *AMIA Annu Symp Proc* 2003:569–78.
6. Chute CG, Carter JS, Tuttle MS, Haber M, Brown SH: **Integrating pharmacokinetics knowledge into a drug ontology as an extension to support pharmacogenomics**. In *AMIA Annu Symp Proc* 2003:170–4.
7. Brown SH, Elkin PL, Rosenbloom ST, Husser C, Bauer BA, Lincoln MJ, Carter J, Erlbaum M, Tuttle MS: **VA National Drug File Reference Terminology: a cross-institutional content coverage study**. *Medinfo* 2004, **11**(Pt 1):477–81.
8. Carter JS, Brown SH, Bauer BA, Elkin PL, Erlbaum MS, Froehling DA, Lincoln MJ, Rosenbloom ST, Wahner-Roedler DL, Tuttle MS: **Categorical information in pharmaceutical terminologies**. In *AMIA Annu Symp Proc* 2006:116–20.
9. Pathak J, Chute CG: **Analyzing categorical information in two publicly available drug terminologies: RxNorm and NDF-RT**. *Journal of the American Medical Informatics Association* 2010, **17**(4):432–439.
10. Burton MM, Simonaitis L, Schadow G: **Medication and Indication Linkage: A Practical Therapy for the Problem List?** In *AMIA Annual Symposium Proceedings* 2008:86–90.
11. Schadow G: **Structured product labeling improves detection of drug-intolerance issues**. *Journal of the American Medical Informatics Association* 2009, **16**(2):211–219.
12. Bodenreider O, Mougin F, Burgun A: **Automatic determination of anticoagulation status with NDF-RT**. In *Proceedings of the 13th ISMB'2010 SIG meeting "Bio-ontologies"* 2010:140–143.
13. Bodenreider O, Burgun A: **Aligning knowledge sources in the UMLS: methods, quantitative results, and applications**. In *Stud Health Technol Inform, Volume 107(Pt 1)* 2004:327–31.
14. **Virtuoso Universal Server**[http://virtuoso.openlinksw.com/].
15. Jaccard P: **Étude comparative de la distribution florale dans une portion des Alpes et des Jura**. *Bulletin de la Société Vaudoise des Sciences Naturelles* 1901, (37):547–579.