Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (IHI 2012) 2012:437-443.

Comparing Drug-Class Membership in ATC and NDF-RT

Fleur Mougin
Univ. Bordeaux, ISPED, LESIM
F-33000, Bordeaux, France
0033 557 574 885
fleur.mougin@isped.u-bordeaux2.fr

Anita Burgun
INSERM U936, EA3888
School of Medicine
Univ. Rennes 1, IFR 140, France
0033 299 284 215
anita.burgun@univ-rennes1.fr

Olivier Bodenreider
National Library of Medicine
Bethesda, Maryland, USA
001 301 435 3246
olivier@nlm.nih.gov

ABSTRACT
The Anatomical Therapeutic Chemical (ATC) classification system is widely used in Europe for the classification and coding of drugs. However, ATC is not well integrated with other medication terminologies (e.g., NDF-RT – the National Drug File-Reference Terminology), which hinders the integration of data coded to these two systems. In this work, we propose to map ATC to NDF-RT, via the Unified Medical Language System (UMLS), in which several medication terminologies are integrated, including NDF-RT but not ATC. Only half of ATC terms were successfully mapped to the UMLS using automatic lexical techniques, resulting in very few overlapping drug-class pairs between ATC and NDF-RT. To improve these results, we performed a manual mapping of cardiovascular ATC and NDF-RT classes, which increased the number of common drug-class pairs from 39 to 128. We believe that the discovery of mappings between ATC and NDF-RT classes could be further automated and made more effective by identifying mappings between the drugs in these classes.

Categories and Subject Descriptors
I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods; J.3 [**Computer Applications**]: Life and Medical science.

General Terms
Standardization.

Keywords
Mapping of terminologies, ATC, NDF-RT, UMLS, coding of drugs, data integration.

1. INTRODUCTION
Terminological drug information sources specify the names and relations of drug entities, in particular, relations between drugs and pharmacological classes. While ingredient names generally exhibit minimal variation across sources, the names of pharmacological classes tend to be poorly standardized. For example, a cursory examination reveals that the names of pharmacological classes differ largely between the Anatomical Therapeutic Chemical (ATC) classification system and the National Drug File-Reference Terminology (NDF-RT). As a consequence, applications relying on the names of pharmacological classes in these two systems cannot be expected to be interoperable. Such applications include clinical decision support and integration of data about adverse events.

Clinical decision support generally relies on specialized knowledge bases. This knowledge tends to be expressed at the highest level possible. Instead of specifying all drug-drug interactions pairwise, the interactions can be expressed between a drug and a pharmacological class. An inference engine can then compute all interactions between the first drug and all members of the pharmacological class. For such an inference to yield similar results in two different systems, a given pharmacological class must be represented in both systems in a comparable way and have the same list of drug members.

For example, *Phenelzine* is a *monoamine oxidase (MAO) inhibitor* and *Citalopram* is a *Selective Serotonin Reuptake Inhibitor (SSRI)*. Both are antidepressant drugs. The granularity of information about antidepressants varies depending on the application. Drug prescription is made at the drug level (e.g., *Phenelzine*, *Citalopram*), while interactions, adverse reactions and pharmacogenomics information are often represented at the class level (e.g., 'Response to selective serotonin reuptake inhibitors is associated with a functional 5-HT1A receptor gene polymorphism').

ATC is widely used in Europe in applications related to clinical decision, as well as clinical data integration and mining. The DebugIT project (Detecting and Eliminating Bacteria UsinG Information Technology [6]) focuses on antibiotics resistance, analyzes practices and outcomes in the domain of antibiotics treatments. EU-ADR [5] aims to develop an innovative computerized system exploiting clinical data from electronic healthcare records to detect adverse drug reactions. Both projects use ATC for representing drugs. It is noteworthy that ATC is not only used in Europe, but also in Northern America. PharmGKB, an integrated resource about how variation in human genetics leads to variation in response to drugs [9], uses the ATC classification for categorizing drugs according to its therapeutic classes. DrugBank, the bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information [18], provides ATC codes for each drug entry (as does Wikipedia). Despite its popularity, ATC is not well integrated with other medication terminologies (e.g., NDF-RT), which hinders the integration of data coded to these two systems.

The objective of this work is to compare the ATC classification system with a reference clinical drug terminology, namely NDF-RT. We focus our analysis on the relations between drugs and

IHI'12, January 28–30, 2012, Miami, Florida, USA.

pharmacological classes (i.e., drug-class membership relations) in these two systems.

2. RESOURCES

The **Anatomical Therapeutic Chemical** (**ATC**) classification system was primarily developed to support drug utilization research [7]. It is controlled by the WHO Collaborating Centre for Drug Statistics Methodology (WHOCC) and was first published in 1976. In the ATC classification system, the drugs are divided into different groups according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties. As a result, one drug can be assigned more than one category, and consequently more than one code: for example, *acetylsalicylic acid (aspirin)* is coded A01AD05 as a drug for local oral treatment, B01AC06 as a platelet inhibitor, and N02BA01 as an analgesic and an antipyretic. More precisely, the drugs are classified according to five different levels: the first level is formed by fourteen anatomical main groups; the second level consists of one pharmacological/therapeutic subgroup (94 in total); the third and fourth levels are chemical/pharmacological/therapeutic subgroups (267 and 877, respectively); and the fifth level contains the drugs themselves (4,406). In Table 1, the structure of ATC codes is illustrated by the complete ATC classification of *digoxin* (C01AA05).

Table 1. Hierarchy of the drug *digoxin* (C01AA05) in the ATC system

Code	Label
C	Cardiovascular system *(1st level, anatomical main group)*
C01	Cardiac therapy *(2nd level, therapeutic subgroup)*
C01A	Cardiac glycosides *(3rd level, pharmacological subgroup)*
C01AA	Digitalis glycosides *(4th level, chemical subgroup)*
C01AA05	Digoxin *(5th level, drug)*

The **National Drug File Reference Terminology** (**NDF-RT**) is a resource developed by the Department of Veterans Affairs (VA) Veterans Health Administration, as an extension of the VA National Drug File [11]. This version covers 6,960 active moieties (level = "ingredient") and 15,313 clinical drugs (level = "VA product"). Two independent kinds of drug classes are represented in NDF-RT: legacy VA classes and "external pharmacologic classes" (EPC). Legacy VA classes are simply listed as parents of clinical drugs. For example, the drug *DIGOXIN 0.5MG TAB* is a subclass of *[CV050] DIGITALIS GLYCOSIDES*. There are 485 such VA classes, organized in a shallow hierarchy. The set of VA classes forms a classification system, i.e., accommodates virtually any drug through residual classes (e.g., *[BL900] BLOOD PRODUCTS, OTHER*). The EPC classes are not used in this investigation. The version of NDF-RT used in this study is dated February 7, 2011.

The **Unified Medical Language System**® (**UMLS**) [3] includes three sources of semantic information: the Metathesaurus®, the Semantic Network and the SPECIALIST Lexicon. The UMLS Metathesaurus is assembled by integrating over 150 source vocabularies, including NDF-RT, but not ATC. It contains more than two million concepts, which correspond to clusters of terms coming from the different vocabularies. Nearly 46 million relations exist among these concepts. The Semantic Network is a much smaller network of 133 semantic types organized in a tree structure. The semantic types have been aggregated into fifteen coarser semantic groups [4], which represent subdomains of biomedicine (e.g., **Anatomy**, **Disorders**). Each Metathesaurus concept has a unique identifier (CUI) and is assigned at least one semantic type. Additionally, the MetaMap Transfer (MMTx) program allows the mapping of text to concepts in the Metathesaurus [1]. Finally, the SPECIALIST Lexicon is a general English lexicon that includes many biomedical terms [12]. It consists of a set of lexical entries with one entry for each spelling or set of spelling variants in a particular part of speech. Lexical items may be multi-word terms made up of other words if the multi-word term is determined to be a lexical item by its presence as a term in general English or medical dictionaries, or in medical thesauri. The 2010AA version of the UMLS is used in this study.

3. METHODS

Our method for comparing drug-class pairs between ATC and NDF-RT can be summarized as follows. First, we acquire drug-class pairs from ATC and map the corresponding drugs and classes to UMLS concepts for comparison purposes. Then we acquire drug-class pairs from NDF-RT for those drugs present in ATC (NDF-RT provides cross-references to UMLS concepts for its drugs and classes). Finally, we compare the UMLS concepts for the drug-class pairs in the two systems.

3.1 Acquiring drug-class pairs from ATC

3.1.1 Mapping ATC to the UMLS

In order to increase the chances of mapping ATC terms to the UMLS, we expanded abbreviations in ATC names with the complete corresponding word. In practice, we transformed "excl." into "excluding", "incl." into "including", "comb." into "combinations", "adm." into "administration", "gr." into "group", "I.V" into "Intravenous". We then mapped each pre-processed ATC term (i.e., groups, and drugs) to the UMLS through the MMTx program with the following parameters: strict model, term processing and a restriction to semantic types belonging to the semantic group **Chemicals and Drugs**. In practice, for each ATC term, variants are generated by MMTx using the knowledge in the SPECIALIST lexicon and a supplementary database of synonyms. UMLS concepts having at least one synonym, which matches exactly one of these variants are selected. Finally, a score is attributed to each candidate concept according to a weighted average of four metrics: centrality (involvement of the head), variation (an average of inverse distance scores), coverage (how much of a candidate matches the term) and cohesiveness (how many synonyms match the term). Only mappings for which an exact match is found are kept, i.e., when an ATC term is mapped to a unique UMLS concept and whose mapping score is 100%.

3.1.2 Constituting ATC drug-class pairs

In ATC, the drugs are situated at the lowest level. The four other levels represent anatomical, therapeutic, pharmacological, and chemical groups in which a given drug is involved. We thus considered that the classes correspond to the groups described in these four upper levels. Thus, we computed the ATC drug-class pairs by associating each code of the fifth level with each of its upper levels. Only those pairs for which a UMLS CUI was obtained via MMTx for the drug and the class are kept. As an illustration, the ATC drug-class pairs generated for the drug *digoxin*

(C01AA05) are (Table 1): *digoxin-Cardiovascular system*, *digoxin-Cardiac therapy*, *digoxin-Cardiac glycosides*, and *digoxin-Digitalis glycosides*, resulting in only the two last drug-class pairs, which are resolved to UMLS CUIs and become respectively: C0012265-C0012253 and C0012265-C0007158. The two first pairs are eliminated because *Cardiovascular system* is categorized by the semantic type "Body System", which does not belong to the semantic group **Chemicals and Drugs** while *Cardiac therapy* could not be found as such in the UMLS.

3.2 Acquiring drug-class pairs from NDF-RT

NDF-RT drugs (at the ingredient level) can be linked to their corresponding pharmacological class ("VA class") through the corresponding products ("VA Product"). For example, the NDF-RT ingredient *DIGOXIN* from the drug hierarchy (N0000146388) is linked to several products, including *DIGOXIN 0.25MG TAB* (N0000151459), whose pharmacologic class is *[CV050] DIGITALIS GLYCOSIDES* (N0000029117). All the other products linked to *DIGOXIN* have the same pharmacologic class. *[CV050] DIGITALIS GLYCOSIDES* is child of the pharmacologic class *[CV000] CARDIOVASCULAR MEDICATIONS* (N0000029116). Therefore, the classes associated in NDF-RT with the corresponding drug *digoxin* are *[CV050] DIGITALIS GLYCOSIDES* and *[CV000] CARDIOVASCULAR MEDICATIONS* (Table 2).

Table 2. Relata of the drug *DIGOXIN* (N0000146388) in NDF-RT

Code	Label
N0000029116	[CV000] CARDIOVASCULAR MEDICATIONS *(VA class)*
N0000029117	[CV050] DIGITALIS GLYCOSIDES *(VA class)*
N0000146388	DIGOXIN *(drug)*
N0000151459	DIGOXIN 0.25MG TAB *(VA Product)*

Moreover, NDF-RT provides cross-references to UMLS concepts for each drug and class. The drug-class pairs *digoxin-[CV050] DIGITALIS GLYCOSIDES* and *digoxin-[CV000] CARDIOVASCULAR MEDICATIONS* are resolved to UMLS concept identifiers through these cross-references and become C0012265-C0012253 and C0012265-C1874729, respectively. In practice, we used applications programming interfaces to access NDF-RT [14] and traverse the appropriate relations between ingredients, drugs and pharmacologic classes. We also leveraged RxNorm, a rich source of synonyms for drug entities, through the RxNorm API [15].

3.3 Comparing drug-class pairs between ATC and NDF-RT

3.3.1 Identifying drug-class pairs common to ATC and NDF-RT

The algorithm summarizing the acquisition of drug-class pairs from ATC and NDF-RT and their comparison is displayed in Figure 1. We computed the pairs which are common to both terminologies, i.e., where both the CUI of the drug and the CUI of the class are common in ATC and NDF-RT. When multiple classes were common for a given drug, the pair involving the most specific class was chosen and the other pairs were eliminated. For example, *hydroxychloroquine* (C0020336) belongs to the classes *ANTIPROTOZOALS* (C0003416) and *ANTIMALARIALS* (C0003374) (both present in ATC and NDF-RT). The latter class being the more specific, only the pair (C0020336, C0003374) was kept (Figure 2).

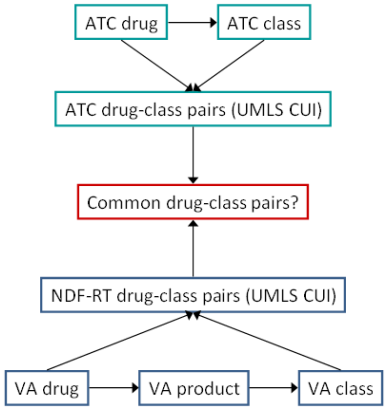

Figure 1. Algorithm for the acquisition and comparison of ATC and NDF-RT drug-class pairs

3.3.2 Exploring the Cardiovascular System group

As shown in the next section, the number of common drug-class pairs between ATC and NDF-RT is small, in particular because the mapping of ATC class terms to the UMLS (via MMTx) was not very productive. To complete this mapping and to be able to perform a more realistic comparison of ATC and NDF-RT, one of the author (AB) performed a manual mapping of the 22 NDF-RT classes of the Cardiovascular branch to the ATC classes of the Cardiovascular System group. We then recovered the ATC drugs and classes from this group, excluding combinations. Toward this end, we eliminated every ATC terms, which includes the word "combination" or "and". In addition, when a class was eliminated, all its subclasses and their drugs were also ignored. An exception was made for the ATC class *ANTIARRHYTHMICS, CLASS I AND III* (C01B) because it was the best candidate to be mapped to the NDF-RT class *[CV300] ANTIARRHYTHMICS*. We finally recomputed the common drug-class pairs between ATC and NDF-RT.

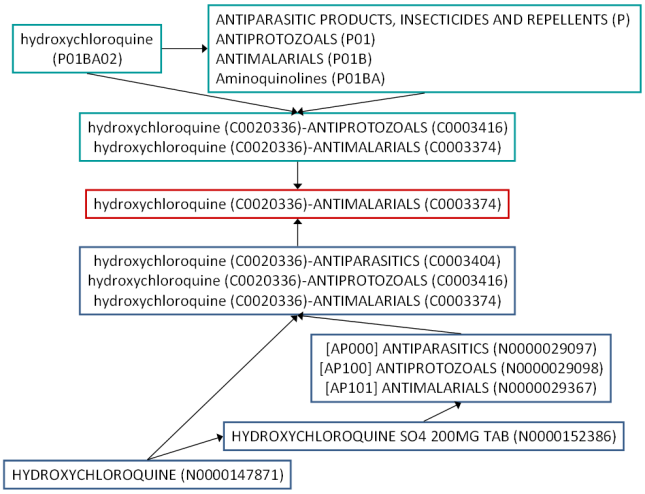

Figure 2. Detailed example for the acquisition of common ATC and NDF-RT drug-class pairs for the drug *hydrochloroquine*

4. RESULTS

4.1 Acquiring drug-class pairs from ATC

Overall, 198 ATC classes (15.8%) and 2,644 drugs (60.0%) are mapped exactly to a UMLS concept of the semantic group **Chemicals and Drugs**. It is noteworthy that only 148 of these mapped ATC classes and 1,760 drugs are present in NDF-RT. Among them, the class *Prostaglandins* (C01EA) and the drug *digoxin* (C01AA05) were mapped to the UMLS CUIs C0033554 and C0012265, respectively.

From the 17,624 ATC drug-class pairs, only 2,704 pairs were resolved to UMLS CUIs (15.3%). An example is the pair *alprostadil* (C01EA01)-*Prostaglandins*, corresponding to the UMLS CUIs pair C0002335-C0033554.

4.2 Acquiring drug-class pairs from NDF-RT

Starting from the 4,406 ATC drugs:

- 2,707 correspond to a drug entity in RxNorm;
- 1,945 correspond to a NDF-RT ingredient;
- 1,794 correspond to some VA product;
- 6,512 NDF-RT drug-class pairs were generated.

One of these drug-class pair is *rifapentine* (N0000148581)-*[AM500] ANTITUBERCULARS* (N0000029084), corresponding to the CUIs pair C0073372-C0003448.

Different explanations can be proposed for justifying the missed mappings at the different steps of the algorithm:

- not in UMLS: this is mainly due to the complexity of ATC terms, to which MMTx can not find an exact mapping to a UMLS concept. An example is the ATC drug *inulin and other polyfructosans* (V04CH01);
- not in RxNorm or not an ingredient in RxNorm: RxNorm focuses on clinical drugs and may not have all experimental drugs. For example, *gitoformate* (C01AA09) comes from the supplementary concepts in MeSH (probably a substance referred to in the literature). More rarely, a drug can be in RxNorm but not as an ingredient (i.e., which has no RxNorm properties, because it simply comes from an external source). For instance, *sparteine* (C01BA04) has a CUI in RxNorm but is in this category because it is not an ingredient;
- not in NDF-RT or not an ingredient in NDF-RT: some drugs are "recognized" by RxNorm, but are not true RxNorm concepts (because no clinical drugs are attached to them). An example is *acetyldigitoxin* (C01AA01), which is in FirstDataBank and thus in RxNorm but not in NDF-RT. Sometimes, the drug exists in NDF-RT but is not an ingredient, such as *acetyldigoxin* (C01AA02), and is thus not expected to be present in ATC;
- no VA class: some drugs are ingredients in NDF-RT but they have no associated clinical drugs ("VA Product") to which VA classes are assigned. One such ATC drug is *pinacidil* (C02DG01);
- not the same class as ATC: this categorization would require further investigation to be correctly explained. An example is the drug *rutoside* (C05CA01) which is associated with the ATC class *Bioflavonoids* (C05CA) while it is mapped to *HERBS/ALTERNATIVE THERAPIES* (HA000) in NDF-RT.

4.3 Comparing drug-class pairs between ATC and NDF-RT

Overall, only 333 drug-class pairs are common to ATC and NDF-RT. One of these common pairs is *chlorpromazine* (C0008286)-*ANTIPSYCHOTICS* (C0040615).

The manual mapping between the 22 NDF-RT classes of the Cardiovascular branch to the ATC classes of the Cardiovascular System group is displayed in Table 3 (see at the end of the paper). Only the two following NDF-RT classes could not be mapped to the ATC classes of the Cardiovascular system group:

- *[CV703] CARBONIC ANHYDRASE INHIBITOR DIURETICS* because the corresponding class is described in the *SENSORY ORGANS* (S) group in ATC. This is explained by the fact that the carbonic anhydrase inhibitors (e.g., *acetazolamide*) are used primarily for glaucoma (and, incidentally, for intracranial hypertension and high altitude sickness);
- *[CV900] CARDIOVASCULAR AGENTS, OTHER* because its semantics is vague and related to the extension of the class.

This manual mapping was particularly useful to acquire new ATC drug-class pairs. The Cardiovascular system group is composed of 554 drugs and 169 classes (Table 4). After eliminating those corresponding to combinations, 319 drugs and 115 classes remained. Originally, only 202 drug-class pairs were generated for the Cardiovascular System group. Thanks to the manual mapping, 508 additional drug-class pairs were obtained. For example, the pair *quinidine* (C0034414)-*ANTIARRHYTHMICS, CLASS I AND III* (C0003195) was found through the mapping of *ANTIARRHYTHMICS, CLASS I AND III* to *[CV300] ANTIARRHYTHMICS*. The resulting effect of these additional pairs is the increase of the number of drug-class pairs in common between ATC and NDF-RT from 39 to 128. In particular, the previous C0034414-C0003195 pair is also described in NDF-RT.

Table 4. Number of drugs in the Cardiovascular system group before and after the elimination of combination terms, which were mapped to a UMLS concept, to a RxNorm concept, to an ingredient in NDF-RT, to some VA class. Finally, the number of drug-class pairs common to ATC and NDF-RT is displayed

	Original Cardiovascular system group	Cardiovascular system group without combinations
Number of drugs	554	319
Map to UMLS	359	306
Map to RxNorm	286	240
Map to an ingredient in NDF-RT	185	150
Map to some VA class	179	144
Common drug-class pairs	39	128

5. DISCUSSION

5.1 Findings

Overall, the mapping of ATC to the UMLS is disappointing because only 50% of all ATC terms were mapped to a UMLS CUI. One should however notice that we required a mapping score of 100% with MMTx, which was very restrictive. In particular, we

ignored 1-1 mappings when they were not complete but which could be correct and could have thus been useful. For example, the mapping of *trandolapril and verapamil* (C09BB10) to *Trandolapril/Verapamil* (C0718096) has a score of 91.3% and is total. In contrast, the mapping of *dihydroxialumini sodium carbonate* (A02AB04) to *sodium carbonate* (C0074732) has also a high score (90.1%) although it is not complete. This second example illustrates the reason why we decided to keep only mappings with a score of 100%.

The mapping of ATC was particularly poor for classes as only 15.8% could be mapped to UMLS CUIs. This is due to the fact that many ATC classes are combinations, have complex names or have names specific to ATC. Examples are *OTHER COLD COMBINATION PREPARATIONS* (R05X) and *Adrenergics and other drugs for obstructive airway diseases* (R03AK). This makes it difficult to find a corresponding UMLS concept as such. It is noteworthy that some combinations were correctly mapped to multiple UMLS concepts. For example, *Vitamin B-complex with anabolic steroids* (A11ED) was correctly mapped to both *Vitamin B Complex* (C0042849) and *Anabolic steroids* (C0002744). However, this combination necessitates further processing to be efficiently exploited.

The manual mapping performed between the NDF-RT classes of the Cardiovascular branch and the ATC classes of the Cardiovascular System group substantially improved the number of drug-class pairs generated and thus the number of common pairs between ATC and NDF-RT. More practically, these numbers increased by nearly 3.5 times and 3 times, respectively. This more detailed study of a given group of ATC showed that the overlap between ATC and NDF-RT could be largely better. A possible solution for detecting mappings between ATC and NDF-RT classes more automatically is discussed in the perspectives.

5.2 Limitations

This study presents some limitations, which result from different causes: ATC as illustrated above and further here, the UMLS, and NDF-RT. For the mapping of ATC to the UMLS, we opted for using MMTx, which is a linguistically-motivated mapping approach. Sometimes the structural knowledge can be exploited in combination with lexical information in order to enhance the mapping results [17]. However, the ATC description of drugs being limited to their label, only a lexical method could be used.

As mentioned earlier, the overlap of drug-class pairs between ATC and NDF-RT is limited and an explanation for this poor performance is the well-known problem of missed synonymy in the UMLS [8]. For example, no mapping could be found between ATC and NDF-RT ophthalmological drugs because the ATC class *OPHTHALMOLOGICALS* (S01) is mapped to the UMLS concept *Ophthalmological agents* (C2013096) whereas the NDF-RT class is [*OP000] OPHTHALMIC AGENTS*, which has a different CUI (C0973585). These two distinct UMLS concepts, which should be clustered into a unique UMLS concept, obviously result in missing common pairs between ATC and NDF-RT.

The VA classes of NDF-RT are organized in a shallow hierarchy, which is not complete. Indeed, some hierarchical relations between VA classes are clearly missing. For example, *[OP230] ANTIVIRALS, TOPICAL OPHTHALMIC* is listed as a child of *[OP200] ANTI-INFECTIVE, TOPICAL OPHTHALMIC*, but not as a child of *[AM800] ANTIVIRALS*. When introducing NDF-RT, we mentioned the existence of another kind of classes: the "external pharmacologic classes" (EPC), which are defined in reference to some of the properties of the active moiety. There are 408 such external pharmacologic classes, with no hierarchical organization. In a previous work [2], we proposed a method for inferring relations between these "external pharmacologic classes" and the drugs. It would be interesting to investigate these classes for comparison with ATC classes and study whether they provide a viable alternative to the legacy VA classes.

5.3 Perspectives

When performing the manual mapping of cardiovascular classes between ATC and NDF-RT, we made decisions on the basis of the set of drugs present in a given class. For example, *[CV702] LOOP DIURETICS* was mapped to the ATC class *HIGH-CEILING DIURETICS* (C03C). High ceiling diuretics are diuretics that may cause a substantial diuresis – up to 20% of the filtered load of NaCl and water. Loop diuretics have this ability, and are therefore often synonymous with high ceiling diuretics. Both *[CV702] LOOP DIURETICS* and *HIGH-CEILING DIURETICS* (C03C) contain, among others, *furosemide*. We thus believe that an instance-based mapping, i.e., exploiting the overlap of ATC and NDF-RT drugs, could be useful for discovering or for checking automatically correspondences between the classes to which these drugs belong. Such an instance-based mapping has already been performed to map NDF-RT and SNOMED CT classes [13]. In our case, this approach would however necessitate manual validation, in particular because some drugs can be part of distinct ATC classes of a same upper class. An example of such drug is *sodium phosphate* which belongs to the class *Osmotically acting laxatives* (A06AD) and the class *Enemas* (A06AG), which are both in the *LAXATIVES* (A06A) class.

Finally, conceptual models of drugs have been designed for specific purposes, e.g., pharmacogenomics [10] or computerized physician order entry [16]. In the first case, it may be used to search for drugs that share the same mechanism of action, or the same target. In the second case, the model takes into account several properties of medications that can be useful for drug substitution for example.

6. CONCLUSION

We present a comparison of the ATC classification system, which is widely used to code drugs in Europe, to NDF-RT, a reference drug terminology used in clinical applications. We showed that only 50% of ATC terms were mapped automatically to the UMLS, which resulted in a very poor overlap with NDF-RT. By performing a manual mapping of the NDF-RT classes of the Cardiovascular branch to the ATC classes of the Cardiovascular System group, we increased the number of common cardiovascular drug-class pairs from 39 to 128.

7. ACKNOWLEDGMENTS

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM).

8. REFERENCES

[1] Aronson, A.R. and Lang, F.-M. 2010. An overview of MetaMap: historical perspective and recent advances. *Journal of the American Medical Informatics Association: JAMIA*. 17, 3 (may. 2010), 229-236.

[2] Bodenreider, O. 2004. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*. 32, Database issue (jan. 2004), D267-270.

[3] Bodenreider, O. and McCray, A.T. 2003. Exploring semantic groups through visual approaches. *Journal of Biomedical Informatics*. 36, 6 (dec. 2003), 414-432.

[4] Bodenreider, O., Mougin, F., and Burgun, A. 2010. Automatic determination of anticoagulation status with NDF-RT. *Proceedings of the 13th ISMB'2010 SIG meeting « Bio-ontologies »* (2010), 140-143.

[5] Coloma, P.M. et al. 2011. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiology and Drug Safety*. 20, 1 (jan. 2011), 1-11.

[6] DebugIT: 2011. http://www.debugit.eu/.

[7] Guidelines for ATC classification and DDD assignment, Oslo: 2010.
http://www.whocc.no/filearchive/publications/2011guidelines.pdf.

[8] Hole, W.T. and Srinivasan, S. 2000. Discovering missed synonymy in a large concept-oriented Metathesaurus. *Proceedings of the AMIA Annual Symposium* (2000), 354-358.

[9] Klein, T.E. et al. 2001. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. *The Pharmacogenomics Journal*. 1, 3 (2001), 167-170.

[10] Knox, C. et al. 2011. DrugBank 3.0: a comprehensive resource for « omics » research on drugs. *Nucleic Acids Research*. 39, Database issue (jan. 2011), D1035-1041.

[11] Lincoln, M.J. et al. 2004. U.S. Department of Veterans Affairs Enterprise Reference Terminology strategic overview. *Studies in Health Technology and Informatics*. 107, Pt 1 (2004), 391-395.

[12] McCray, A.T., Srinivasan, S., and Browne, A.C. 1994. Lexical methods for managing variation in biomedical terminologies. *Proceedings of the AMIA Annual Symposium* (1994), 235-239.

[13] Mortensen, J. and Bodenreider, O. 2010. Evaluating pharmacologic classes in NDF-RT. *Proceedings of the Fourth International Symposium on Semantic Mining in Biomedicine* (2010), 119-124.

[14] NLM. NDF-RT API: 2011.
http://rxnav.nlm.nih.gov/NdfrtAPI.html.

[15] NLM. RxNorm API: 2011.
http://rxnav.nlm.nih.gov/RxNormAPI.html.

[16] Senger, C. et al. 2011. Design and evaluation of an ontology-based drug application database. *Methods of Information in Medicine*. 50, 3 (may 2011), 273-284.

[17] Sun, Y. 2004. Methods for automated concept mapping between medical databases. *Journal of Biomedical Informatics*. 37, 3 (jun. 2004), 162-178.

[18] Wishart, D.S. et al. 2008. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research*. 36, Database issue (jan. 2008), D901-906.

Table 3. Mapping of Cardiovascular NDF-RT classes to the Cardiovascular system ATC classes

ATC class	NDF-RT class	UMLS CUI
C - CARDIOVASCULAR SYSTEM	[CV000] CARDIOVASCULAR MEDICATIONS	C1874729
C01AA - Digitalis glycosides	[CV050] DIGITALIS GLYCOSIDES	C0012253
C01B - ANTIARRHYTHMICS, CLASS I AND III	[CV300] ANTIARRHYTHMICS	C0003195
C01D - VASODILATORS USED IN CARDIAC DISEASES	[CV250] ANTIANGINALS	C1874267
C02 - ANTIHYPERTENSIVES	[CV400] ANTIHYPERTENSIVE COMBINATIONS [CV490] ANTIHYPERTENSIVES,OTHER	C1874305 C0350167
C02C - ANTIADRENERGIC AGENTS, PERIPHERALLY ACTING	[CV150] ALPHA BLOCKERS/RELATED	C1874153
C03 - DIURETICS	[CV700] DIURETICS	C0012798
C03A - LOW-CEILING DIURETICS, THIAZIDES	[CV701] THIAZIDES/RELATED DIURETICS	C0012802
C03C - HIGH-CEILING DIURETICS	[CV702] LOOP DIURETICS	C0354100
C03D - POTASSIUM-SPARING AGENTS	[CV704] POTASSIUM SPARING/COMBINATIONS DIURETICS	C1875688
C03X - OTHER DIURETICS	[CV709] DIURETICS,OTHER	C1875040
C04 - PERIPHERAL VASODILATORS	[CV500] PERIPHERAL VASODILATORS	C0724804
C05BB - Sclerosing agents for local injection	[CV600] SCLEROSING AGENTS	C0036426
C05BX - Other sclerosing agents	[CV600] SCLEROSING AGENTS	C0036426
C07 - BETA BLOCKING AGENTS	[CV100] BETA BLOCKERS/RELATED	C1874540
C08 - CALCIUM CHANNEL BLOCKERS	[CV200] CALCIUM CHANNEL BLOCKERS	C0006684
C09A - ACE INHIBITORS, PLAIN	[CV800] ACE INHIBITORS	C0003015
C09C - ANGIOTENSIN II ANTAGONISTS, PLAIN	[CV805] ANGIOTENSIN II INHIBITOR	C1874242
C09XA - Renin-inhibitors	[CV806] DIRECT RENIN INHIBITOR	C1950687
C10 - LIPID MODIFYING AGENTS	[CV350] ANTILIPEMIC AGENTS	C0003367