

# Analyzing U.S. prescription lists with RxNorm and the ATC/DDD Index

Olivier Bodenreider, MD, PhD, Laritza M. Rodriguez, MD, PhD

Lister Hill National Center for Biomedical Communications, National Library of Medicine,  
National Institutes of Health, Bethesda, MD

## Abstract

**Objectives:** To evaluate the suitability of the ATC/DDD Index (Anatomical Therapeutic Chemical (ATC) Classification System/Defined Daily Dose) for analyzing prescription lists in the U.S. **Methods:** We mapped RxNorm clinical drugs to ATC. We used this mapping to classify a large set of prescription drugs with ATC and compared the prescribed daily dose to the defined daily dose (DDD) in ATC. **Results:** 64% of the 11,422 clinical drugs could be precisely mapped to ATC. 97% of the 87,001 RxNorm codes from the prescription dataset could be classified with ATC, and 97% of the prescribed daily doses could be assessed. **Conclusions:** Although the mapping of RxNorm ingredients to ATC appears to be largely incomplete, the most frequently prescribed drugs in the prescription dataset we analyzed were covered. This study demonstrates the feasibility of using ATC in conjunction with RxNorm for analyzing U.S. prescription datasets for drug classification and assessment of the prescribed daily doses.

## 1. Introduction

Medication errors have been identified as a significant cause of mortality in hospitalized patients [1] and medication safety remains an important issue today [2]. Medication dose errors are a specific category of medication errors [3]. Large variations can be observed in prescribed doses, some of which correspond to medication dose errors, including tenfold medication dose errors [4]. One strategy for reducing medication errors, including dose errors, is to use Computerized Physician Order Entry (CPOE) systems offering clinical decision support [5]. The information used for clinical decision support in CPOEs generally comes from proprietary drug knowledge bases.

The Anatomical Therapeutic Chemical (ATC) classification of drugs is widely available and provides basic information such as drug classification and defined daily doses. This information may be insufficient to fully assist prescription, but can be used for analyzing a prescription dataset retrospectively. One typical use of ATC is to measure drug utilization for pharmaco-epidemiology purposes. However, most of the published studies leveraging the classification and defined daily doses features of ATC have been performed in Europe (e.g., [6-8]).

ATC was recently integrated in RxNorm. While RxNorm only integrates the terminological features of ATC, and not its defined daily doses and routes of administration, this integration already facilitates the analysis of prescription lists indexed with RxNorm identifiers, by providing a reliable entry point into ATC.

The objective of this study is to assess the suitability of the ATC/DDD Index (Anatomical Therapeutic Chemical (ATC) Classification System/Defined Daily Dose) for analyzing prescription lists in the U.S. More specifically, we propose to analyze drug classification based on ATC groupings and to compare the prescribed daily dose to the defined daily dose in ATC for a large prescription dataset from Surescripts. To our knowledge, this study is the first application of ATC to the analysis of a U.S. prescription dataset.

## 2. Background

This investigation leverages ATC, RxNorm and a large prescription list obtained from Surescripts.

### 2.1. ATC

The Anatomical Therapeutic Chemical (ATC) classification [[http://www.whocc.no/atc\\_ddd\\_index/](http://www.whocc.no/atc_ddd_index/)], a system developed by the World Health Organization (WHO) Collaborating Centre for Drug Statistics Methodology, is recommended for worldwide use to compile drug utilization statistics. The system includes drug classifications at 5 levels; anatomical, therapeutic, pharmacological, chemical and drugs or ingredients. Also included are defined daily doses (DDDs) and administration routes assigned to most drugs in accordance to the therapeutic and pharmacological groups.

For example, as shown in Figure 1, drugs from the “Digitalis glycosides” 4<sup>th</sup>-level group are included in the anatomical group “Cardiovascular system”. The route of administration (Adm.R) and the defined daily dose (DDD) are listed for each of the four 5<sup>th</sup>-level drugs.

C <b>CARDIOVASCULAR SYSTEM</b>						
C01 <b>CARDIAC THERAPY</b>						
C01A <b>CARDIAC GLYCOSIDES</b>						
C01AA <b>Digitalis glycosides</b>						
ATC code	Name	DDD	U	Adm.R	Note	
C01AA01	<a href="#">acetyldigitoxin</a>	0.2	mg	O		
C01AA02	<a href="#">acetyldigoxin</a>	0.5	mg	O		
C01AA03	<a href="#">digitalis leaves</a>	0.1	g	O		
C01AA04	<a href="#">digitoxin</a>	0.1	mg	O		
		0.1	mg	P		
C01AA05	<a href="#">digoxin</a>	0.25	mg	O		
		0.25	mg	P		

**Figure 1. Drugs from the Digitalis glycosides 4<sup>th</sup>-level group in ATC (partial screenshot from the ATC website)**

The active ingredients in the classification include a wide range of chemical entities used in a variety of countries. New ingredients are not included in the ATC system until they are approved for pharmaceutical use in at least one country. Only herbal medicinal products approved by regulatory authorities are included in the classification.

The active moieties are classified according to the main therapeutic use of the main ingredient. Since an ingredient can have therapeutic applications on different anatomical sections, ATC assigns a different code to the same ingredient in different anatomical sections. For example, the beta-blocker *timolol* has different codes when used as a cardiovascular drug (*C07AA06*) and as a treatment for glaucoma (*S01ED01*).

The defined daily dose (DDD) is the assumed average maintenance dose per day for a drug used for its main indication in adults. The DDD is calculated based on adult weight of 70 kg. The DDD can be an average of doses from different countries and might be the reflection of the more commonly used strengths. The DDD is not necessarily the prescribed daily dose, as the latter depends on individual patient characteristics such as age, weight and pharmacokinetic considerations. Topical products, sera, vaccines, antineoplastic agents, allergen extracts, anesthetics and contrast media are not assigned a DDD.

The 2014 edition of ATC used in this study contains 4580 5<sup>th</sup>-level ATC drugs, of which 3904 correspond to single-active moieties (as opposed to combinations).

## 2.2. RxNorm

RxNorm is a standardized nomenclature for clinical drugs compiled from 12 drug source vocabularies, and maintained by the National Library of Medicine (NLM). A clinical drug is defined as a pharmaceutical product with therapeutic or diagnostic properties available to patients. A clinical drug includes the ingredient(s), strength or concentration, and dose form appropriate for the intended administration route (e.g., *Thyroglobulin 32 MG Oral Tablet*). The February 2014 edition of RxNorm is used in this study.

Base ingredients are the active moieties of clinical drugs (e.g., *amoxicillin*). RxNorm also covers their various salts, esters and complexes (e.g., *amoxicillin trihydrate*), referred to as “precise ingredients” in RxNorm parlance. Unlike RxNorm, ATC represents mostly base ingredients and does not distinguish between base and precise ingredients. Single-ingredient drugs have a unique chemical component, and multiple-ingredient drugs have two or more (e.g., *Amoxicillin 250 MG / Clavulanate 125 MG Oral Tablet*). While drug combinations are precisely defined in RxNorm, they are often unspecified in ATC (e.g., *meprobamate, combinations*).

Dose forms are administration vehicles, such as pills, tablets, syringes and lotions. Dose form groups (DFGs) are grouping of dose forms (DFs). For example, the DFG *Oral Product* includes the DFs *Oral Tablet*, *Oral Capsule*, *Chewable Tablet*, etc. RxNorm DFGs roughly correspond to administration routes in ATC.

RxNorm identifies a subset of drugs intended to be an approximation of the prescription drugs currently marketed in the U.S. We refer to this subset as the “prescribable subset” of RxNorm drugs, and use it as to restrict our analysis to the most clinically significant drugs.

**Ingredient-level mapping between RxNorm and ATC.** Since August 2013, ATC is a source vocabulary in RxNorm, which provides a mapping between RxNorm ingredients and 5<sup>th</sup>-level ATC drugs. Of the 3166 RxNorm single ingredients (base and precise), 1552 (49%) mapped to a 5<sup>th</sup>-level ATC drug, corresponding to 1991 ATC codes and 1554 distinct ATC drug names. These ingredients in common represent 51% of the 3904 5<sup>th</sup>-level ATC drugs, ignoring drug combinations.

Since the scopes of RxNorm and ATC are slightly different, the mapping is not expected to be complete. For example, RxNorm includes several hundred allergenic extracts (e.g., *papaya allergenic extract 50 MG/ML Injectable Solution*) that are out of the scope of ATC. Conversely, diagnostic and therapeutic radiopharmaceuticals (e.g., *technetium (<sup>99m</sup>Tc) bismate*) are present in ATC (under *V09* and *V10*), but out of the scope of the prescribable subset of RxNorm.

In contrast to RxNorm, in which each ingredient is represented only once, ATC can have multiple codes for the same active moiety, depending on the anatomical system or therapeutic domain in which it is used. As a consequence, there will often be multiple ATC mappings for a given RxNorm ingredient. For example, the RxNorm ingredient *Ketoconazole (6135)* maps to the following ATC codes for this drug: *D01AC08* (from the *ANTIFUNGALS FOR DERMATOLOGICAL USE* group), *G01AF11* (from the *GYNECOLOGICAL ANTIINFECTIVES AND ANTISEPTICS* group) and *J02AB02* (from the *ANTIMYCOTICS FOR SYSTEMIC USE* group).

### 2.3. Surescripts dataset

The prescription drug list is a de-identified list comprised of 102,709 clinical drugs dispensed to emergency room patients over a period of three months in 2011 at Suburban Hospital in Bethesda, Maryland. Each drug includes an anonymized prescription identifier, clinical drug name, drug form, strength, prescribed amount, and the intake duration. This prescription list was annotated with RxNorm identifiers for clinical drugs. When updated against the February 2014 version of RxNorm, 99,576 drugs were valid (or could be mapped to a valid code), while 3133 were obsolete. Of these, we only investigate the 87,001 drug codes corresponding to single-ingredient drugs from the prescribable subset of RxNorm.

### 2.4. Related work

Many studies have been published reporting on drug utilization based on ATC for various classes of drugs, including antibiotics [9, 10], cardiovascular drugs [8], and anti-depressants [11], or across classes [7]. Some studies specifically compare prescribed daily doses to defined daily doses in ATC for anti-epileptic drugs [6] and for several classes of anti-hypertensive drugs [12]. One characteristic of most of these studies is that they were performed in Europe, where ATC is more widely used than in the U.S.

More recently, ATC has also been used as a terminological reference for drugs. For example, ATC has been used to support the detection of adverse events in the EU-ADR project [13]. Additionally, ATC has been used as a reference in research projects where drug classes were predicted by integrating chemical-chemical interactions and similarities [14] or through text mining [15]. In earlier work, we compared and contrasted ATC with the National Drug File-Reference Terminology (NDF-RT) developed by the U.S. Department of Veterans Affairs (VA) Veterans Health Administration [16].

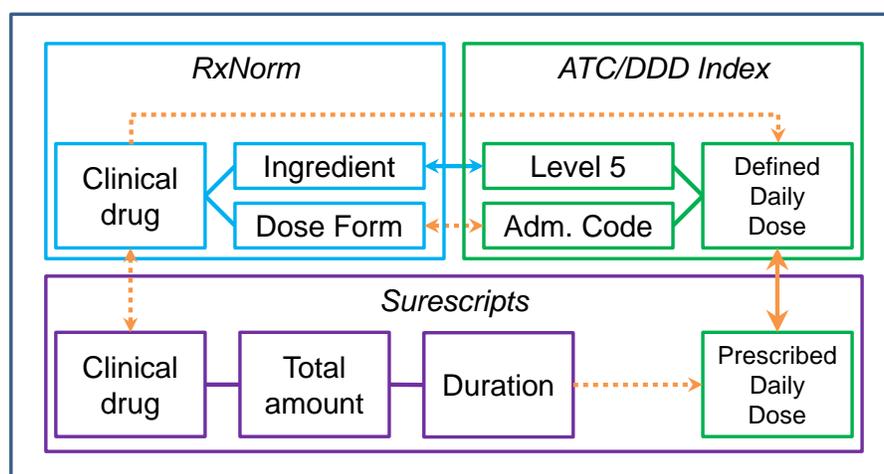
The specific contribution of our work is the application of ATC in combination with RxNorm, the standard drug vocabulary in the U.S. While many pharmaco-epidemiology studies leveraging ATC have been published in Europe, to the best of our knowledge, this study is the first analysis of a prescription dataset in the U.S. with ATC and RxNorm.

## 3. Methods

In ATC, a 5<sup>th</sup>-level code is assigned not to an ingredient, but to an ingredient for a specific therapeutic intent. For example, the beta-blocker *timolol* can be used orally or parenterally as a cardiovascular drug (*C07AA06*) and in eye drops as a treatment for glaucoma (*S01ED01*). Moreover, a defined daily dose (DDD) is assigned not to an ingredient, but to an ingredient with a specific route of administration. For example, the DDD for *acetylsalicylic*

*acid* is 3 g for oral forms, but 1 g when administered parenterally. As a consequence, for the purpose of finding the DDD, the mapping of RxNorm clinical drugs to ATC requires a match for both the ingredient and the route of administration. Our approach to comparing the prescribed daily dose to the defined daily dose is depicted in Figure 2. While the ingredient-level mapping is provided by RxNorm, we had to create a mapping for the routes of administration in order to relate RxNorm clinical drugs to their appropriate ATC 5<sup>th</sup>-level code. We then mapped clinical drugs from the prescription dataset to RxNorm and computed the prescribed daily dose for comparison to the corresponding defined daily dose in ATC.

In this investigation, we restrict the scope of the mapping to single-ingredient clinical drugs in RxNorm and ATC, because combination drugs are often underspecified in ATC. For example, the ATC 5<sup>th</sup>-level code *N05BC51* corresponds to *meprobamate*, *combinations*, and is distinct from the single-ingredient category for *meprobamate* (*N05BC01*), but without specifying which ingredients can be associated with *meprobamate* or what the DDD for *meprobamate* is in this case. Moreover, since our goal is to analyze prescription lists, we restrict the mapping to clinical drugs from the prescribable subset of RxNorm. Finally, since the prescribed daily dose is not explicitly mentioned in the Surescripts data, we further restrict the comparison of daily doses to oral solid dose forms of clinical drugs, for which we can rely on RxNorm to extract the quantity per prescription dose (i.e., pill).



**Figure 2. Overview of the methods for comparing the prescribed daily dose to the defined daily dose**

### 3.1. Mapping RxNorm clinical drugs to ATC

In order to support our use cases of drug classification and assessment of prescribed daily doses, the mapping of RxNorm clinical drugs to ATC must account for both the ingredient and the route of administration. More specifically, we define a mapping between a clinical drug in RxNorm and an ATC 5<sup>th</sup>-level drug when the following two conditions are met.

1. The ingredient (active moiety or salt ingredient) of the clinical drug in RxNorm maps a 5<sup>th</sup>-level drug in ATC. We use the ingredient-level mapping provided in RxNorm.
2. The dose form group for the clinical drug in RxNorm and the administration code (or one of the administration codes, if multiple) of the 5<sup>th</sup>-level drug in ATC are compatible. (i.e., are associated through the same administration route), as defined below.

For example, as illustrated in Figure 3, the RxNorm clinical drug *Amoxicillin 25 MG/ML Oral Suspension* (313797) maps to the ATC code *J01CA04* (*amoxicillin*), because the ingredient of the RxNorm clinical drug, *amoxicillin*, maps to this ATC code, and the dose form group of the RxNorm clinical drug, *Oral Product*, matches one of the routes of administration for the ATC code *J01CA04*, *O*, through the administration route *oral*. In contrast, despite the fact that both drugs have the same ingredient, *butoconazole*, we failed to map *5000 MG Butoconazole nitrate 20 MG/ML Prefilled Applicator* (890780) to *G01AF15*, because the dose form group of RxNorm drug, *Prefilled Applicator Product*, is not listed as compatible with the vaginal route, *V*, listed for this drug in ATC. Finally, some

RxNorm drugs have no mapping to ATC because their ingredient is simply not present in ATC (e.g., *oregano allergenic extract 50 MG/ML Injectable Solution*).

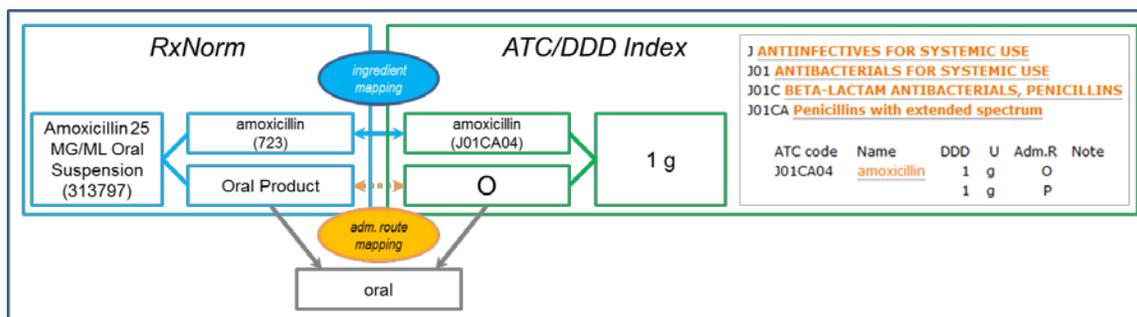


Figure 3. Mapping between RxNorm and ATC through both the ingredient and the route of administration

**Mapping routes of administration between RxNorm and ATC.** In order to map routes of administration between RxNorm and ATC, we harmonized the dose form groups in RxNorm and the administration codes in ATC. We also assigned administration codes to ATC drugs when they were missing.

**Harmonization of administration codes between RxNorm and ATC.** RxNorm and ATC have different ways of representing routes of administration. In RxNorm, the route is expressed through the dose form group (DFG), but RxNorm DFGs actually represent the dose form (e.g., *Pill*), the route (e.g., *Ophthalmic Product*) or a mix of both (e.g., *Oral Gel Product*). Many clinical drugs are associated with multiple DFGs, typically one for the dose form and one for the route. For example, *Ketoconazole 200 MG Oral Tablet (197853)* is associated with both *Pill* and *Oral Product*. Of the 45 DFGs in RxNorm, 22 represent dose forms exclusively, but some of them are indicative of topical products nonetheless (e.g. *Shampoo Product*).

ATC assigns administration codes to the drugs in scope for the defined daily dose (e.g., *O* for oral, *N* for nasal, etc.). In addition to the 22 administration codes, ATC defines 10 coarser administration routes. Although ATC does not provide a correspondence between administration codes and administration routes, this correspondence is usually trivial to establish. Missing from the list of ATC administration routes are entries for the routes of ophthalmologic, otic, stomatologic and other topical products, for which ATC typically does not provide DDDs.

We extended the list of 10 administration routes from ATC with *ophthalmologic*, *otic*, *stomatologic* and *topical*, adding *urethral*, as it exists as an administration code. We mapped all relevant DFGs from RxNorm to the extended list of 15 administration routes derived from ATC.

**Assignment of missing administration codes in ATC.** As mentioned earlier, one issue for mapping drugs between RxNorm and ATC is that ATC assigns administration codes only to a subset of its drug entities, as required for the DDD. In practice, drugs for which no DDD is asserted are also missing an administration code. These drugs typically include topical products and systemic drugs for which there are large inter-individual dose variations (sera, vaccines, antineoplastic agents, allergen extracts, general and local anesthetics and contrast media).

For these drugs, we used various strategies to semi-automatically infer an administration code when it was missing. More specifically, we manually created rules to assign administration codes to sets of drugs based on various characteristics, so that missing administration codes could be automatically inferred at the level of individual drugs. These rules were applied in the following order.

1. The authors assigned one or more administration code manually, not to individual drugs, but to specific ATC groups (at various levels), based on clinical knowledge (e.g., the group *Enemas (A06AG)* was associated with the administration route *rectal*).
2. The authors assigned one or more administration code manually based on specific expressions found in the labels of ATC groups (e.g., ATC drugs from groups, whose label contain the expression “systemic” were associated with the administration routes *oral* and *parenteral*).

3. All administration codes found in the drugs for a given 4<sup>th</sup>-level ATC group were propagated to the drugs with missing administration codes in the same group (e.g., the drug *alogliptin (A10BH04)* “inherits” the administration code *oral* from the other drugs in the group *Dipeptidyl peptidase 4 (DPP-4) inhibitors (A10BH)*, namely *sitagliptin*, *vildagliptin*, *saxagliptin*, and *linagliptin*).
4. The administration code *oral* was assigned by default to any digestive drug (e.g., the digestive drug *tilactase (A09AA04)* was assigned the administration code *oral*).
5. The administration code *topical* was assigned by default to any drug that has not been assigned one in the previous steps (e.g., the drug *tetracycline (D06AA04)* was assigned the administration code *topical*).

### 3.2. Analysis of the Surescripts prescription dataset

The Surescripts prescription dataset is coded to RxNorm clinical drugs and we use the mapping to ATC in order to be able to classify the prescription drugs with ATC groups and to compare the prescribed daily doses to the defined daily doses listed in ATC.

**Assessing coverage.** The proportion of RxNorm clinical drugs from the Surescripts dataset to which we can associate a specific ATC 5<sup>th</sup>-level code (i.e., accounting for both the ingredient and the route of administration) assesses the coverage of RxNorm clinical drugs in ATC.

**Classifying prescription drugs.** Through the mapping to ATC we extract the ATC classification of the drugs for characterizing the prescription list. For example, the RxNorm clinical drug *sitagliptin 50 MG Oral Tablet (665042)* maps to the 5<sup>th</sup>-level ATC drug *A10BH01*, classified under the diabetes drugs in the ATC level-1 group A.

**Assessing prescribed daily doses.** We also compare the prescribed daily doses to the defined daily doses listed in ATC. In the Surescripts dataset, the prescribed daily dose is not explicitly provided. From the total number of prescription doses and duration of the prescription, we can calculate the number of prescription doses for a day. We then use RxNorm to get the quantity in each clinical drug. For example, a prescription of 45 doses *Clonazepam 0.5 MG Oral Tablet* for 30 days yields a prescribed daily dose of .75 mg ( $.5 * 45 / 30$ ). Of note, for comparability between drugs, RxNorm normalizes the strength of solutions per milliliter, of inhalers per “puff” for metered-dose inhalers, and of topical creams and gels to mg/mg. As a consequence, the normalized quantity reflected in RxNorm often does not correspond to the prescribed dose. For this reason, we restrict the analysis of prescribed daily dose to oral solid drug form drugs from the Surescripts dataset. We also ignore from the dataset RxNorm drugs for which ATC provides more than one DDD for a given route of administration.

### 3.3. Implementation

From a technical perspective, this investigation can be thought of as a data integration project. The datasets to be integrated include RxNorm, ATC, the ingredient mapping between RxNorm and ATC, and the mapping of both RxNorm dose form groups and ATC administration codes to administration routes, as well as the prescription dataset. Semantic Web technologies are known to provide support for data integration. Here we converted all the datasets to the Resource Description Format (RDF triples) and loaded them into the triple store Virtuoso. The query language for RDF, SPARQL, also provides support for writing production rules (of the “if ... then” type). We created production rules in order to infer the missing administration codes. We also created rules to infer the mapping of clinical drugs to ATC. Finally, we queried the integrated dataset in order to export the prescribed and defined daily doses for each prescription for statistical analysis.

## 4. Results

### 4.1. Mapping RxNorm clinical drugs to ATC

Of the 11,422 single-ingredient clinical drugs from the prescribable subset of RxNorm, 7748 (68%) had an ingredient mapping to ATC, and 7260 (64%) had both an ingredient and an administration route mapping. In other words, a mapping between a clinical drug in RxNorm and a drug in ATC (at the 5<sup>th</sup> level) for a particular administration route was found for 64% of the clinical drugs in RxNorm.

These RxNorm clinical drugs mapped to 1912 unique ATC codes (96% of the 1991 ATC codes to which an ingredient mapping was found) and 1479 unique drug names (95% of the 1554 ATC drug names to which an

ingredient mapping was found), corresponding to 49% of the 3904 ATC codes for single active moieties (and 44% of the drug names).

**Harmonization of administration codes between RxNorm and ATC.** The correspondence between RxNorm DFGs, ATC administration codes and the extended administration routes is shown in Table 1. Each of the 22 dose form groups from RxNorm and each of the 24 administration codes from ATC (including the four codes we created) is mapped to one of the 15 administration routes (extended list). As a result, each dose form group from RxNorm can be associated with at least one administration code from ATC.

**Table 1.** Correspondence between RxNorm dose form groups and ATC administration codes through then extended list of administration routes derived from ATC.

<b>RxNorm Dose Form Group</b>	<b>Route of administration</b>	<b>ATC Administration Code</b>
Drug Implant Product	<i>implant</i>	implant s.c. implant
Inhalant Product	<i>inhalation</i>	Inhal Inhal.solution Inhal.powder Inh.aerosol Inhal.aerosol Instill.sol.
Nasal Product	<i>nasal</i>	N
Oral Product	<i>oral</i>	O Chewing gum oral aerosol
Ophthalmic Product	<i>ophthalmic</i>	lamella [ophthalmic] *
Otic Product	<i>otic</i>	[otic] *
Injectable Product	<i>parenteral</i>	P
Rectal Product	<i>rectal</i>	R
Sublingual Product	<i>sublingual/buccal</i>	SL
Buccal Product Dental Product Oral Cream Product Oral Foam Product Oral Gel Product Oral Ointment Product Oral Paste Product	<i>stomatologic</i>	[stomatologic] *
Transdermal Product	<i>transdermal</i>	TD TD patch
Intraperitoneal Product Irrigation Product Mucosal Product Prefilled Applicator Product Shampoo Product Soap Product Topical Product	<i>topical</i>	intravesical ointment [topical] *
Urethral Product	<i>urethral</i>	urethral
Vaginal Product	<i>vaginal</i>	V

\* added to the original ATC administration codes for mapping purposes

### Assignment of missing administration codes in ATC

Of the 3904 5<sup>th</sup>-level codes in ATC for single active moieties, 2059 (53%) are missing an administration code. The distribution of the number of ATC codes for which administration codes were generated is listed in Table 2, by type of technique. Since the rules for ophthalmic, otic, stomatologic and rectal products were allowed to generate administration codes even when one had been asserted by ATC, the total number of ATC drugs for which administration codes were generated is slightly higher than the number of ATC codes with missing administration codes.

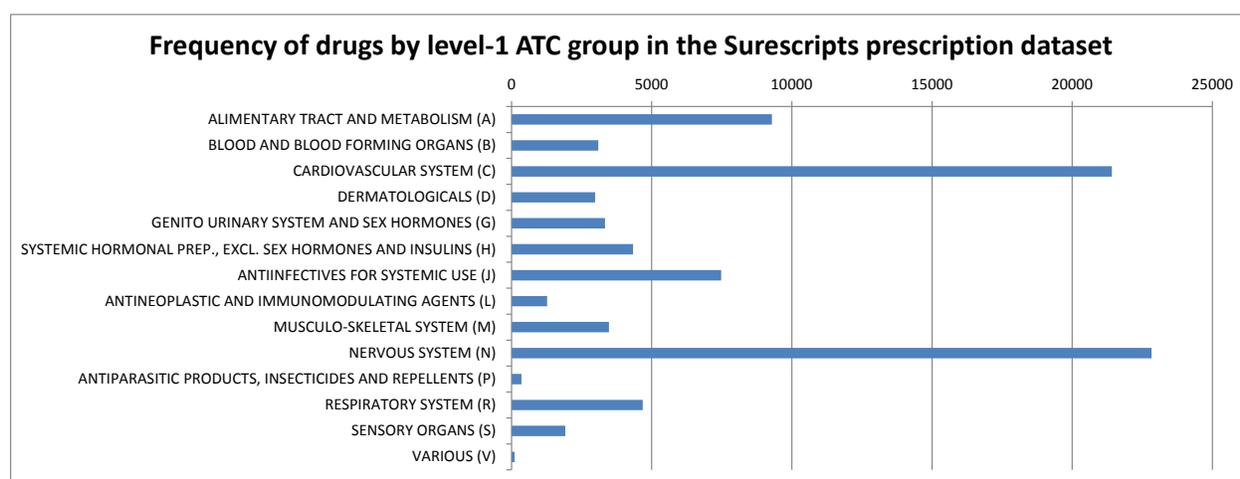
**Table 2.** Number of ATC codes for which administration codes were generated automatically, by type of technique.

Expression in ATC group label	#
Administration code inferred from ATC group	725
Administration code inferred from expressions found in the labels of ATC	232
Administration code inferred from drugs from the same ATC group	492
Oral administration code inferred by default (digestive drugs)	23
Topical administration code inferred by default (remaining drugs)	643
Total	2115

#### 4.2. Analysis of the Surescripts prescription dataset

**Assessing coverage.** Of the 87,001 RxNorm codes from this Surescripts dataset (restricted to single-ingredient drugs from the prescribable subset of RxNorm), 84,380 (97%) mapped to at least one code in ATC (through both the ingredient and the route). Moreover, of the 1695 distinct RxNorm clinical drugs found in the Surescripts dataset, 1606 (95%) were found in ATC.

**Classifying prescription drugs.** Using the mapping to ATC, we classified the 84,380 prescriptions from the Surescripts set against the top-level categories in ATC, resulting into 86,578 ATC codes. The distribution of Surescripts drugs by top-level ATC groups is shown in Figure 4. The top categories are cardiovascular and nervous system drugs. Of note, some RxNorm clinical drugs map to more than one code in ATC (e.g., drugs with multiple therapeutic uses for the same route of administration, such as *clonidine hydrochloride 0.3 MG Oral Tablet*, used orally as both as an antihypertensive drug (*C02AC01*) and an antimigraine agent (*N02CX02*)).



**Figure 4.** Distribution of Surescripts drugs by top-level ATC groups

**Assessing prescribed daily doses.** Of the 72,360 RxNorm clinical drugs corresponding to oral solid dose forms in the Surescripts dataset, 70,394 (97%) could be associated with a defined daily dose in ATC, of which 1932 were associated with more than one DDD (and were ignored from the comparison). For the remaining 68,462 drugs, we compared the prescribed and defined daily doses. The distribution of the ratios of the prescribed daily doses (PDDs) to the defined daily doses (DDDs) is plotted in Figure 5 (using a logarithmic scale, because of the amplitude of the variation among the ratios). Overall, the PDD exactly matches the DDD in 28.6% of the prescriptions. The ratio is in a 66%-150% range for 49.5% of the prescriptions, in a 50%-200% range for 76.1%, and in a 33%-300% range for 86.1%. Only 3.4% of the PDDs are beyond 300% of the DDD and 10.4% below 33% of the DDD. The proportions covered by each range are shown in Figure 5.

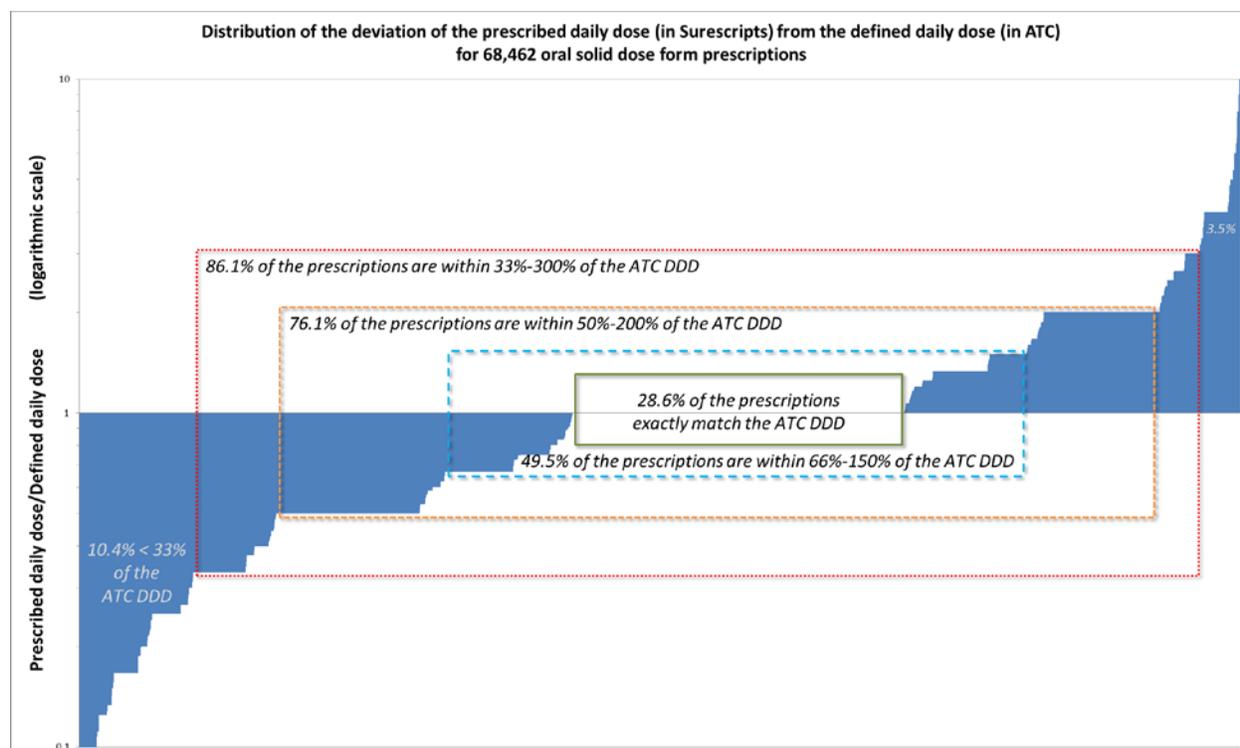


Figure 5. Distribution of the deviation of the prescribed daily dose from the defined daily dose

## 5. Discussion

**Significance.** Although the overall coverage for ingredient mapping is limited, we were able to demonstrate that prescription drugs in current use in the U.S. are mapped reliably and in a considerable proportion. This study confirms that the use of ATC in conjunction of RxNorm is a valid strategy for analyzing prescription datasets in the U.S., both from the perspective of classifying drugs and for the comparison of prescribed and defined daily doses. While ATC is routinely used in Europe for pharmaco-epidemiology, our study is the first application of ATC to prescription data in the U.S.

**Limitations.** The main limitation of our work is the limited size and scope of the prescription dataset, in which the variation of drug ingredients is necessarily limited, even more so in the case of drugs from emergency room patients only. While the proportion of clinical drugs mapped to ATC may be smaller in other datasets, the method of mapping to ATC through RxNorm should be generally applicable, including when drugs are represented with codes from the National Drug Code (NDC) or not coded at all. Another limitation is that the analysis of the prescribed daily doses was restricted to oral solid dose forms, because the prescribed dose was not explicitly mentioned in the Surescripts dataset and could not be reliably extracted from RxNorm. In fact, this issue is being addressed in RxNorm by creating different entities for solution with identical (normalized) concentrations, but different quantities per volume (e.g., 1 mg/ml and 5mg/5ml).

**Future work.** In addition to the exploration of larger and more diverse prescription datasets, the focus of future work is to refine the administration route assignment for missing routes in the ATC, allowing for more precise mapping. We also would like to combine two aspects of the current work, i.e., drug classification and deviation from the defined daily dose, in order to investigate whether certain classes of drugs tend to be prescribed at higher or lower doses compared to the defined daily dose.

## 6. Conclusions

Although the mapping of RxNorm ingredients to ATC appears to be largely incomplete, the most frequently prescribed drugs in the prescription dataset we analyzed were covered. This study demonstrates the feasibility of using ATC in conjunction with RxNorm for analyzing U.S. prescription datasets for drug classification and assessment of the prescribed daily doses.

## Acknowledgments

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.

## References

1. Medicine Io. To Err is Human: Building a Safer Health System. Washington, DC: National Academies Press; 2000
2. Clancy CM. Ten years after To Err is Human. *Am J Med Qual* 2009;24(6):525-8.
3. Keers RN, Williams SD, Cooke J, Ashcroft DM. Prevalence and nature of medication administration errors in health care settings: a systematic review of direct observational evidence. *Ann Pharmacother* 2013;47(2):237-56.
4. Lesar TS. Tenfold medication dose prescribing errors. *Ann Pharmacother* 2002;36(12):1833-9.
5. Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Arch Intern Med* 2003;163(12):1409-16.
6. Koristkova B, Grundmann M, Brozmanova H. Differences between prescribed daily doses and defined daily doses of antiepileptics--therapeutic drug monitoring as a marker of the quality of the treatment. *Int J Clin Pharmacol Ther* 2006;44(9):438-42.
7. Ravera S, Hummel SA, Stolk P, Heerdink RE, de Jong-van den Berg LT, de Gier JJ. The use of driving impairing medicines: a European survey. *Eur J Clin Pharmacol* 2009;65(11):1139-47.
8. Stimac D, Polic-Vizintin M, Skes M, Cattunar A, Cerovic R, Stojanovic D. Utilization of cardiovascular drugs in Zagreb 2001-2005. *Acta Cardiol* 2010;65(2):193-201.
9. Bozkurt F, Kaya S, Tekin R, Gulsun S, Deveci O, Dayan S, et al. Analysis of antimicrobial consumption and cost in a teaching hospital. *J Infect Public Health* 2013
10. Xu J, Duan X, Wu H, Zhou Q. Surveillance and correlation of antimicrobial usage and resistance of *Pseudomonas aeruginosa*: a hospital population-based study. *PLoS One* 2013;8(11):e78604.
11. Viola R, Benko R, Nagy G, Soos G. National trend of antidepressant consumption and its impact on suicide rate in Hungary. *Pharmacoepidemiol Drug Saf* 2008;17(4):401-5.
12. Grimmsmann T, Himmel W. Discrepancies between prescribed and defined daily doses: a matter of patients or drug classes? *Eur J Clin Pharmacol* 2011;67(8):847-54.
13. Avillach P, Dufour JC, Diallo G, Salvo F, Joubert M, Thiessard F, et al. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: a contribution from the EU-ADR project. *Journal of the American Medical Informatics Association : JAMIA* 2012
14. Chen L, Zeng WM, Cai YD, Feng KY, Chou KC. Predicting Anatomical Therapeutic Chemical (ATC) classification of drugs by integrating chemical-chemical interactions and similarities. *PLoS one* 2012;7(4):e35254.
15. Gurulingappa H, Kolarik C, Hofmann-Apitius M, Fluck J. Concept-based semi-automatic classification of drugs. *Journal of chemical information and modeling* 2009;49(8):1986-92.
16. Winnenburger R, Bodenreider O. Exploring pharmacoepidemiologic groupings of drugs from a clinical perspective. *Stud Health Technol Inform (Proc Medinfo)* 2013;192:827-831.