Mapping NDC codes to ATC classification

Ximena Oyarzun-Gonzalez (Mentor: Olivier Bodenreider)
National Library of Medicine, National Institutes of Health, Bethesda, MD

Introduction

Medication classifications have been widely used in medical research, especially when studying medication use. However, there are several classification systems available, and they vary by country, hospital systems, access (not all of them are free), etc. Some of these classification systems, are, for example, the ATC classification system, VA classification system, American health system pharmacist therapeutic classification system, etc.

The **ATC classification** is free and maintained by the WHO. ATC is a hierarchy classification with five levels, where the first one is anatomical or pharmacological level, the second is pharmacological or therapeutic, and the third and fourth are chemical, pharmacological, or therapeutic groups. Finally, the 5th level generally corresponds to the specific chemical substance(s). The classification takes into consideration active ingredient, use, and administration route. Thus, there is a 5th-level code for a formulation comprising an active ingredient administered orally and a different code for a formulation with the same active ingredient administered ophthalmically. In practice, ATC links the active ingredient to all relevant classes, leaving it as an exercise for the user to determine which class is appropriate given the specific dose form or route of administration.

The **VA drug classification** also associates medications with drug classes. It is a 2- to 3- levels hierarchical system where each RxNorm product is assigned a VA class. Unlike ATC, however, the VA classification associates drug products (not just ingredients) with drug classes. The advantage is that a specific formulation is explicitly and unambiguously associated with a drug class regardless of whether several classes are relevant to the active ingredient. While the VA and ATC classifications rely on different classificatory principles, there is an equivalent class in ATC for most classes from the VA classification.

Medications in the USA are assigned an **NDC** (National Drug Code), which provides the formulation with a unique identifier used by the FDA. The code is a three-segment number, where the first segment is a number representing the manufacturer or labeler; the second segment identifies the specific strength, dosage form, and formulation of the product; finally, the third segment identifies package size and type.

RxNorm also identifies drug products. Unlike NDC, RxNorm focuses on clinically useful information and defines drug products based on ingredient(s), strength, and dose form. For all practical purposes, mapping NDC products to ATC can be decomposed into two steps: 1) NDC to RxNorm product, and 2) RxNorm product to ATC. The RxNorm API supports the mapping of NDC codes with the corresponding RxNorm product. Therefore, we focus on mapping RxNorm products to ATC.

The goal of this project is to develop automatic methods for mapping RxNorm products to their corresponding ATC class(es).

Methods

To evaluate the mapping of RxNorm drug products to ATC, we first developed a ground truth dataset. We then created rules for mapping VA classes to ATC classes at the highest possible level for single-ingredient formulations so that we could leverage the existing mapping to the VA classification to derive

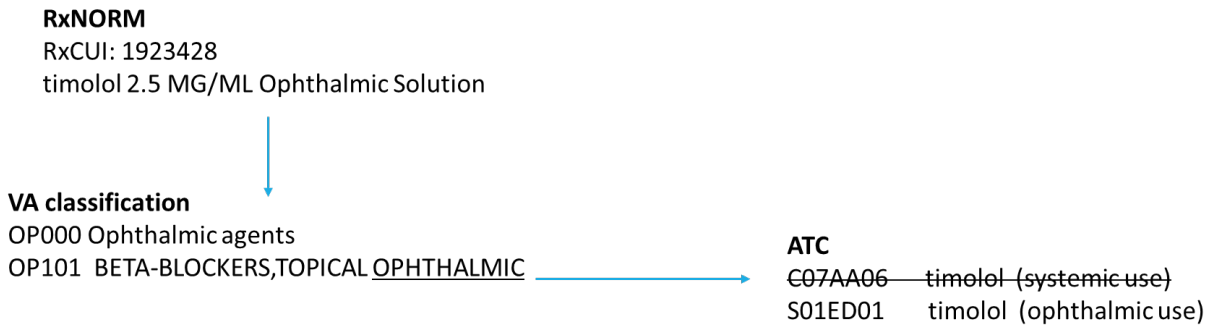
mappings to equivalent classes in ATC. Finally, we developed methods for resolving class names found in underspecified ATC 5th-level terms for multi-ingredient formulations to specific substances.

Establishing a "Ground truth" dataset

To support the evaluation of methods for mapping RxNorm products to ATC codes, we established a "ground truth" dataset that covers the 10,028 RxNorm products (SCDs). This work requires subject matter expertise and familiarity with the ATC coding guidelines. To facilitate this manual work, we used the VA classification as an intermediary between RxNorm products and ATC. The grouping of RxNorm products by VA class made it easier to review these products with ATC classes. For example, it made it easier to identify when RxNorm products associated with a given VA class were not also associated with the corresponding ATC class(es).

Creating VA-ATC class pairs for the disambiguation and validation of single-ingredient formulations

ATC may have more than one code for the same active ingredient, and this will vary depending on the use and formulation. For example, Timolol, as an active ingredient, can be used ophthalmically or systemically depending on the intended pharmacological effect. The VA classification system has RxCUI 1923428 (timolol 2.5 MG/ML Ophthalmic Solution) classified as OP101 (BETA-BLOCKERS, TOPICAL OPHTHALMIC). ATC has timolol classified as C07AA06 (Beta blocking agents, non-selective, for systemic use) and S01ED01 (Beta blocking agents for ophthalmic use). However, we can easily establish that the VA class, OP101, corresponds to the ATC class, S01 (OPHTHALMOLOGICALS), rather than C (CARDIOVASCULAR SYSTEM). This high-level correspondence between a VA class and an ATC class allows us to resolve the ambiguity not only for this Timolol formulation but for all ophthalmic beta-blocker formulations. In cases where a given ingredient is unambiguously associated with a single ATC class, the correspondence established between VA and ATC classes can be used for validating the ATC class of the ingredient. Since the classification of multi-ingredient formulations ("combinations") involves specific classificatory principles in ATC, the method described above is restricted to single-ingredient formulations in RxNorm.



We collected pairs (*V*, *A*) of equivalent VA and ATC classes in a spreadsheet. In practice, these pairs will be transformed into production rules of the sort "IF drug product *D* is coded with (a descendant of) VA class *V*, THEN it should be coded with (a descendant) of ATC class *A*". This information was used to expedite the manual curation of the ATC codes for each RxNorm product and will be used to implement an automated curation system.

Resolving class names in multi-ingredient formulations

ATC has specific principles for the classification of multi-ingredient formulations compared to single-ingredient formulations. Namely, to avoid naming all possible combinations of ingredients, ATC often refers to combinations of one (or more) specific ingredient and a class (e.g., "C07BA02-oxprenolol and thiazides") or simply to the underspecified notion of one (or more) specific ingredient in combination with others (e.g., "R01BA52-pseudoephedrine, combinations"). Less often, ATC creates residual classes simply labeled "combinations" to represent the combination of several active ingredients classified one level above (i.e., at the 4th level) in the ATC classification (e.g., "J01EB20-combinations" represents the class for formulations of more than one short acting sulfonamide listed under "J01EB-Short-acting sulfonamides"). To appropriately map multi-ingredient formulations to ATC requires 1) to expand classes to the list of their ingredients and 2) understand prioritization guidelines provided as part of the ATC documentation in case more than one multi-ingredient class is applicable.

Expanding classes to the list of their ingredients. To support the mapping of multi-ingredient formulations an expansion dataset was created. This dataset consists of the classes that appear in 5th-level ATC names and the corresponding ATC class to which they correspond. For example, in "C07BA02-oxprenolol and thiazides", thiazides refer to the active ingredients from the ATC class "C03AA-Thiazides, plain". The list of specific ingredients for the class thiazides can automatically be extracted by leveraging the hierarchical structure of ATC, and specific combinations of ingredients can automatically be mapped to C07BA02. For example, the RxNorm product "cyclopenthiazide 0.25 MG / oxprenolol 160 MG Extended-Release Oral Tablet" [250553] can be mapped to "C07BA02-oxprenolol and thiazides", because it contains the ingredients "oxprenolol" and "cyclopenthiazide", knowing that "thiazides" corresponds to the ATC class "Thiazides, plain" (C03AA) and that the ATC substance "cyclopenthiazide" (C03AA07) is a descendant of "Thiazides, plain" (C03AA).

More generally, frequently observed patterns in ATC combination classes include

- Multiple drug names
- Drug name AND drug class
- Multiple drug classes

In some cases, the class listed in a combination is not defined in the ATC classification. When this happened, a list was created with the most common active ingredients present in that group. For example, "potassium salts" is a term that does not have its own class under the ATC classification system, but it can easily be expanded to "potassium chloride, potassium lactate, and potassium acetate" based on common pharmacological knowledge.

Creating prioritization rules among multi-ingredient classes. We manually extracted such prioritization rules from the ATC guidelines. Our contribution was in formulating actionable rules that can be integrated into a computer program. These rules include general rules and medication-specific rules.

General rules serve to codify the priority given to more specific classes over less specific classes. For example, drug formulation including enalapril and a thiazide drug (e.g., hydrochlorothiazide) should be classified as "C09BA02 enalapril and diuretics" rather than "C03AX01 hydrochlorothiazide, combinations" because the former is more specific.

Medication-specific rules provide coding guidance for medications that would otherwise have similar levels of priority. For example, the ATC guidelines state that "*Combined preparations which contain more than one analgesic, should be classified by using the following ranking:*
1. Phenacetin
2. Bucetin
3. Dipyrocetyl
4. Paracetamol
5. Acetylsalicylic acid
6. Phenazone
7. Salicylamide
8. Propyphenazone
This means that a product containing paracetamol and phenazone should be classified in N02BE51 - paracetamol, combinations excl. psycholeptics and not in N02BB51 - phenazone, combinations excl. psycholeptics."

Results

Establishing a "Ground truth" dataset

A dataset with 10,028 RxCUI and its corresponding VA code was manually curated to assign the correct ATC code to each formulation. In total, 9,941 ATC codes were assigned (99.1%). To complete the missing codes, emails with the list of formulations that do not currently have an ATC code were sent to the WHO Collaborating Centre for Drug Statistics Methodology, and new codes will be assigned for some of them (the next meeting for this process will be held in October 2022). Because some formulations are not marketed anymore, WHO cannot get a new code assigned, but they suggested classifying them at the 4th ATC level when corresponding.

Creating VA-ATC class pairs for the disambiguation and validation of single-ingredient formulations

From 340 VA classes comprising 10,028 RxCUI, 282 (7,924 RxCUI, about 79%) were associated with one or more ATC classes at level 2 or below. From all the associations created, 88% were associated with only one level of the ATC classes. From these, 2% were associations with a 2nd-level ATC class, 31% with a 3rd-level ATC class, and 68% with a 4th-level ATC class, covering about 56% of the total RxCUIs.

The 282 association pairs created include 6,422 RxCUIs for single ingredient formulations, which represent 80% of the total single ingredient formulations.

Resolving class names in multi-ingredient formulations

Expanding classes to the list of their ingredients. The expansion dataset is comprised of mainly 4 groups of 5th-level ATC codes: 1) a drug and a class of drugs, 2) combinations of drugs from one class of drugs, 3) "drug, combinations", and 4) those ATC 5th-level classes that are coded as "combinations". The multi-ingredient formulations where the ATC explicitly lists all the components are not included in this dataset, as expansion is not needed.

For the first two cases, the following terms were associated with a particular ATC class of drugs:

1. ACE inhibitors
2. acid preparations
3. adrenergics (for systemic use)
4. analgesics
5. other non-opioid analgesics
6. antacids
7. antibacterials (for systemic use)
8. antibiotics (for topical/dermatological use)

9. anticholinergics
10. antiinfectives (for ophthalmic use, for otic use, and for ophthalmic and otic use)
11. antiseptics (topical use)
12. antispasmodics
13. barbiturates
14. belladonna alkaloids
15. beta-lactamase inhibitor
16. Calcium Channel Blockers
17. COMT inhibitor
18. Corticosteroids
19. cough suppressants
20. decarboxylase inhibitor
21. different antibiotics diuretics
22. expectorants
23. mucolytics
24. mydriatics
25. omega-3 fatty acids
26. other diuretics
27. other potassium salts
28. potassium-sparing agents
29. proton-pump inhibitors
30. psycholeptics
31. sulfonamides (as diuretics)
32. sulfonylureas
33. tetracyclines
34. thiazides
35. potassium (different salts in combination)
36. magnesium (different salts in combination)
37. calcium (different salts in combination)
38. contact laxatives
39. other drugs for obstructive airway diseases

Moreover, the following terms were not included in the ATC classification, so independent lists were created for them: toxoids, omega-3-triglycerides incl. other esters and acids, mineral salts in combination, antiflatulents, and lactic acid-producing organisms.

Regarding those ATC codes that fall under number 3) above, the datasets list all the current codes in that form, so when developing the rules for mapping, it is clear what combinations of drugs are allowed.

Finally, for those drugs where the 5th level only indicates “combinations”, the dataset comprises a list of the ATC codes where the only indication is “combinations” (see below), and its corresponding group where the active ingredients should be looked up. For example, ATC R05CA10 represents combinations of different expectorants, thus any formulation comprising 2 or more active ingredients present in the ATC class R05CA Expectorants should be classified as R05CA10. The same applies to the following ATC codes and their corresponding groups for search.

1. A01AA30-combinations →search in A01AA
2. A02AA10-combinations →search in A02AA
3. A02AB10-combinations →search in A02AB
4. A02AC10-combinations →search in A02AC
5. A11CC20-combinations →search in A11CC
6. B01AC30-combinations →search in B01AC
7. B02BC30-combinations →search in B02BC
8. B05BA10-combinations →search in B05BA
9. B05CA10-combinations →search in B05CA
10. B05CB10-combinations →search in B05CB
11. B05CX10-combinations →search in B05CX
12. C01CA30-combinations →search in C01CA
13. D01AA20-combinations →search in D01AA
14. D01AE20-combinations →search in D01AE
15. G03GA30-combinations→search in G03GA
16. G04BE30-combinations →search in G04BE
17. J01CA20-combinations →search in J01CA
18. J01CE30-combinations →search in J01CE
19. J01EB20-combinations →search in J01EB
20. J01EC20-combinations →search in J01EC
21. J01ED20-combinations →search in J01ED
22. J06BB30-combinations →search in J06BB
23. J07BC20-combinations →search in J07BC
24. N06DX30-combinations→search in N06D
25. R01AX30-combinations →search in R01AX
26. R05CA10-combinations →search in R05CA
27. R05CB10-combinations →search in R05CB
28. R05DA20 -combinations→search in R05DA
29. R05DB20-combinations →search in R05DB
30. R07AA30-combinations →search in R07AA
31. S01HA30-combinations →search in S01HA
32. S02DA30-combinations →search in S02DA
33. A06AG20-combinations→search in A06AG
34. A07BC30-combinations→search in A07BC

35. A10AB30-combinations→search in A10AB
36. A10AC30-combinations→search in A10AC
37. A10AD30-combinations→search in A10AD
38. A10AE30-combinations→search in A10AE

Creating prioritization rules among multi-ingredient classes. The list of rules is extracted from the 2022 ATC guidelines for code assignment, with a focus on those rules that establish a hierarchy among different active ingredients. Currently, 75 rules have been extracted from the 2022 ATC guidelines, and 45 of those are regarding the hierarchy of the ingredients. However, this list is not finished, and more information needs to be extracted from the guidelines.

Discussion

Findings. The VA classification is a very helpful tool to map RxNorm products (and NDC codes) to ATC classes, as its higher classification levels are similar to those in the 3rd and 4th levels of ATC classes. The expansion and rules datasets should also support the process of mapping to the correct ATC code.

Significance. Since RxNorm and the ATC classification evolve over time, it is important to automate the ongoing curation of the mapping of RxNorm to ATC. By codifying high-level VA-ATC class pairs, this work is a contribution to the disambiguation and validation of the ATC classification of RxNorm products for single-ingredient drugs. Additionally, by providing the expansion of classes in ATC names and codifying the prioritization among multiple possible classes for multi-ingredient formulations, this work supports the automated mapping of multi-ingredient drugs to ATC. Finally, by establishing an exhaustive ground truth mapping between RxNorm products and ATC, this work is a contribution to the evaluation of tools developed to support the automated curation of this mapping.

Limitations. In our work, we attempted to cover the largest number possible of cases to facilitate the development of tooling support for automated mapping between RxNorm products and ATC classes. However, it will likely be impossible to assign an ATC code to all the RxNorm products based on the information we extracted. Therefore, there will always be a need for the knowledge of a pharmacist or trained professional to adjudicate edge cases, as well as important changes made to the ATC classification.

Acknowledgments

I'd like to express my deepest thanks to Olivier Bodenreider and Lee Peters for giving me the opportunity to provide my support in the developing of this project, their endless patience, and introducing me to people working in a field different from mine, which forced me to view things with a different perspective. I would also like to thank the NLM and ORISE for the support and additional opportunities for professional development provided.