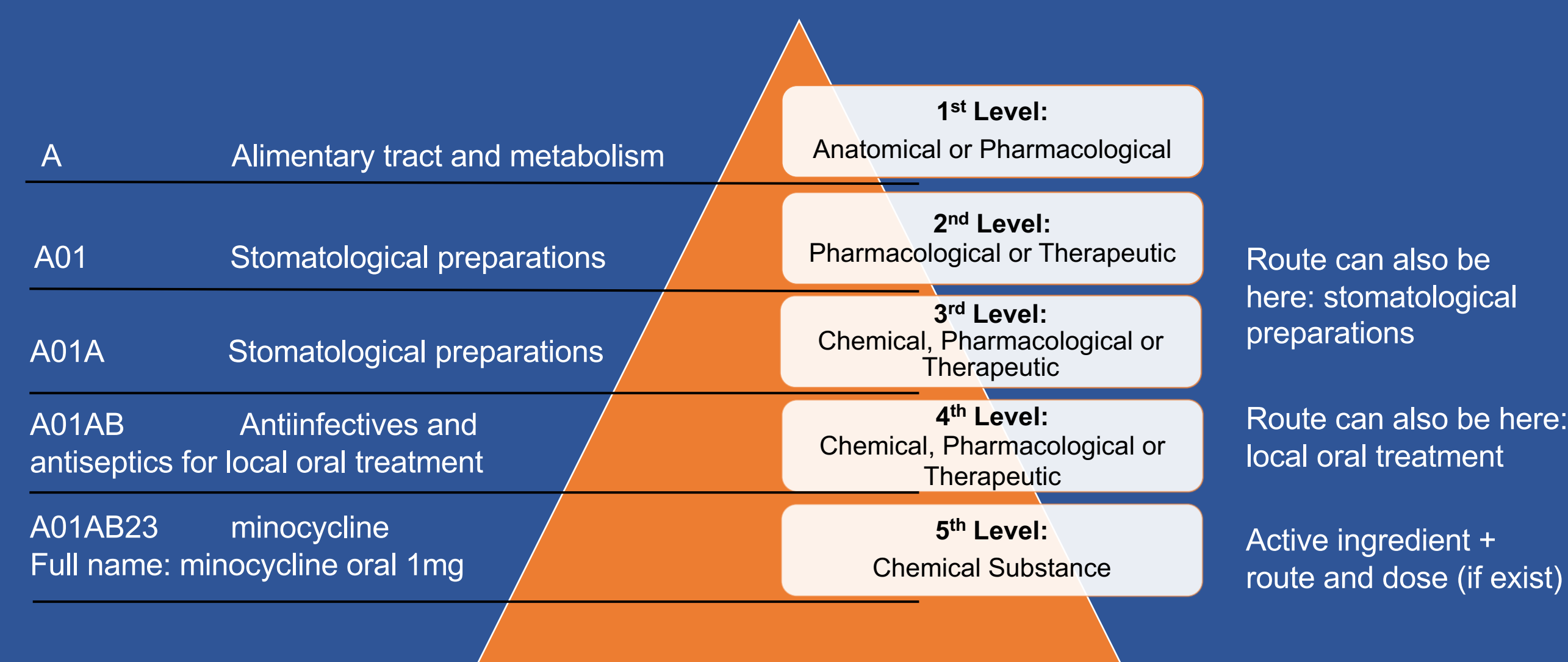


Why is it important?

ATC, developed by the WHO, is the most commonly standard in pharmacoepidemiology, pharmacovigilance, pharmacoecconomy, and observational in the world. OMOP uses RxNorm as a standard for marketed drug products, and a crosswalk between ATC and RxNorm is essential for observational studies if ATC drug classes used to define exposure. Currently, the link happens between ATC level 5 concepts and RxNorm Ingredients. But this link is “greedy” and can result in gross misclassification, because route, dose and indication information implied in the ATC5 Concepts is lost in this crosswalk. Moreover, only simple mono-ingredient ATC 5th concepts are matched leaving combinations and multicomponent drugs uncovered.

ATC Structure and Challenges

Figure 1. ATC Classification structure.



ATC drugs are classified into multiple levels (Figure 1) according to the main therapeutic use of the main ingredient. However, the same ingredient can be used in different indications, often requiring different drug forms and/or drug strengths. For example, the various therapeutic purposes of prednisolone is reflected in differences in the formulation:

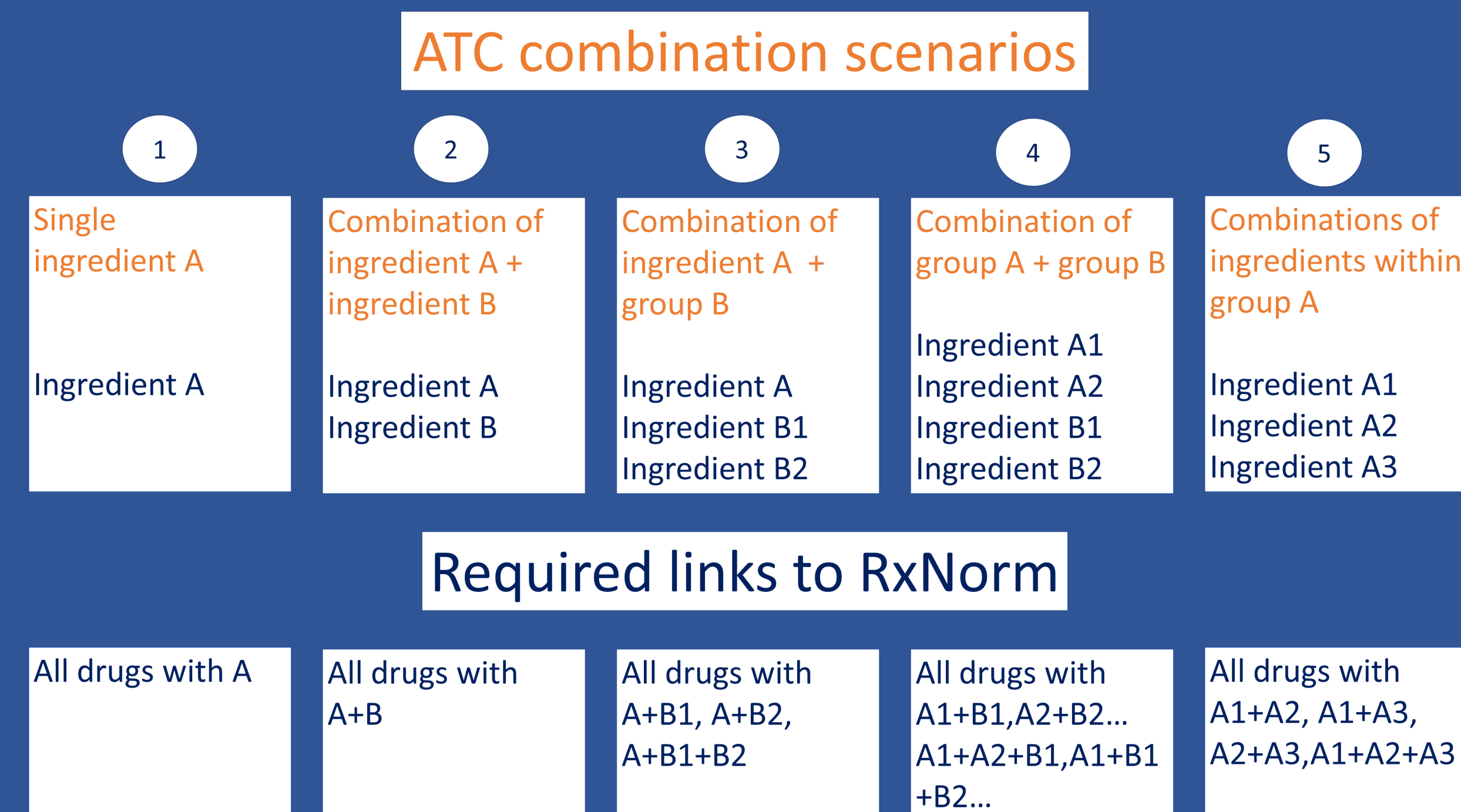
A07EA01 - enemas and foams; S02BA03 - otic forms.
C05AA04 - suppositories;
D07AA03 - topical forms;
H02AB06 - oral and injectable forms;
R01AD02 - nasal forms;
S01BA04 - ophthalmic forms;

The associated drug forms often are not explicitly defined at the 5th level, but need to be inferred from the ATC levels 2-4, likely causing confusion.

Ibuprofen is an example of division by high and low strength for different indications:
M01AE01 ANTIINFLAMMATORY AND ANTIRHEUMATIC 1200 mg Parenteral
C01EB16 CARDIAC THERAPY 5 mg Parenteral

ATC also covers fixed combinations, which need to be linked to the correct drugs. The combinations can be defined between ingredients, and groups of ingredients :

Figure 2. Fixed combinations in ATC



Combining the ATC Drug Classification System with the RxNorm Drug Nomenclature into a comprehensive Drug Ontology: Challenges and Achievements

Anna Ostropolets, MD¹, Hamed Abedtash, PharmD, PhD², Peter Rijnbeek, PhD³, Marcel de Wilde³, Alexander Davydov⁶, Olha Marushchak⁴, Christian Reich, MD⁵

¹ Columbia University, New York City, NY, USA, ² Eli Lilly & Company, Indianapolis, IN, USA,

³ Erasmus University Medical Center, Rotterdam, The Netherlands, ⁴ Danylo Halytsky Lviv National Medical University, Lviv, Ukraine, ⁵ IQVIA, Cambridge, MA, USA, ⁶ Odysseus Data Services, Cambridge, MA, USA

Process

From the original five-level ATC hierarchy, we sought to create two types of connections:

- From the lowest level ATC5 system to RxNorm Drug Products
- From any level ATC to the relevant drug ingredients.

Connecting these ATC codes requires to not only match the active ingredient, but also the other attributes such as indication, therapeutic usage, route of administration (which is closely related to dose forms) or drug strength

Figure 3. Mapping of individual ATC 5th level concepts

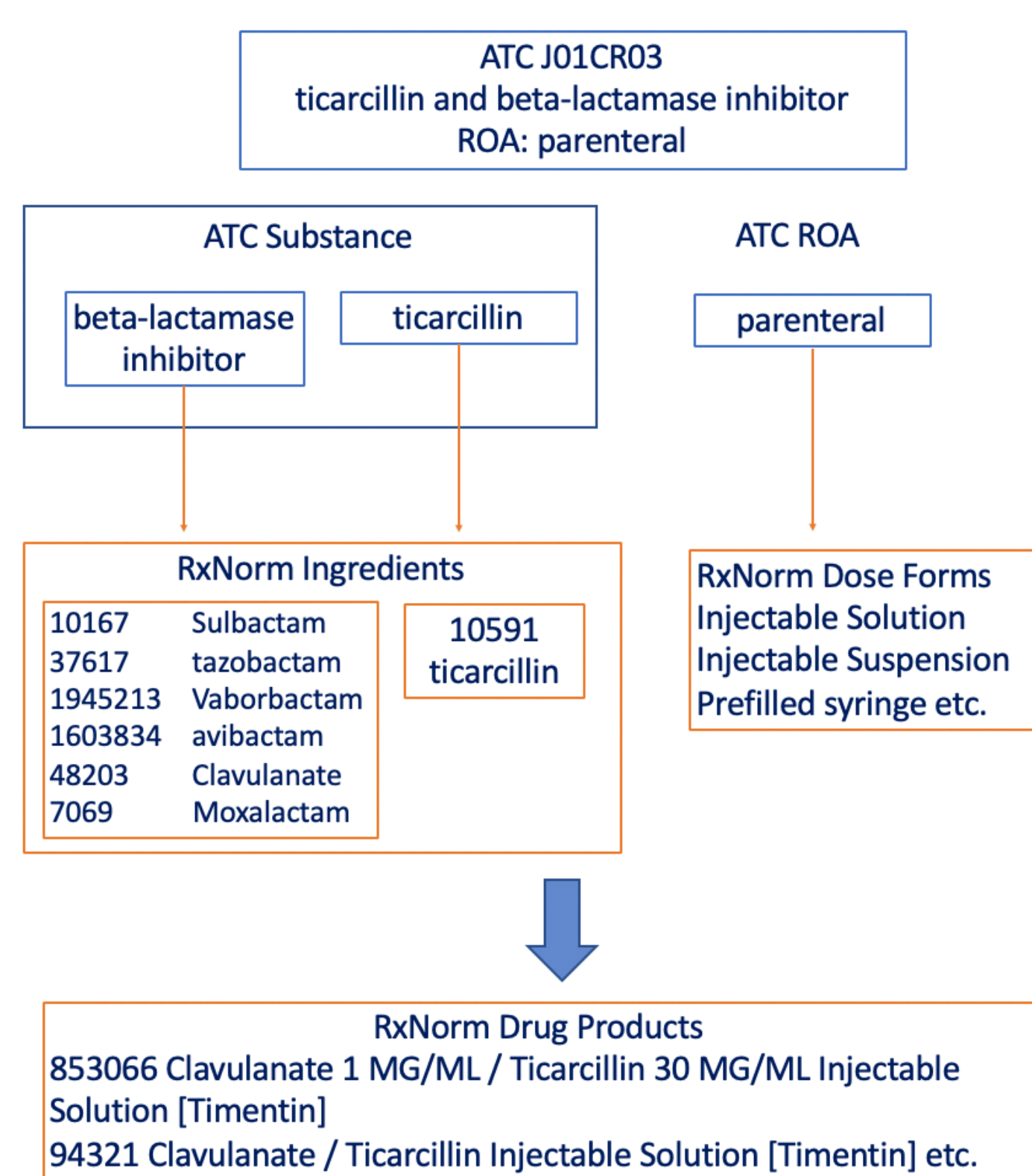


Figure 4. Mapping and ancestry relationships for ATC 1st – 5th level concepts

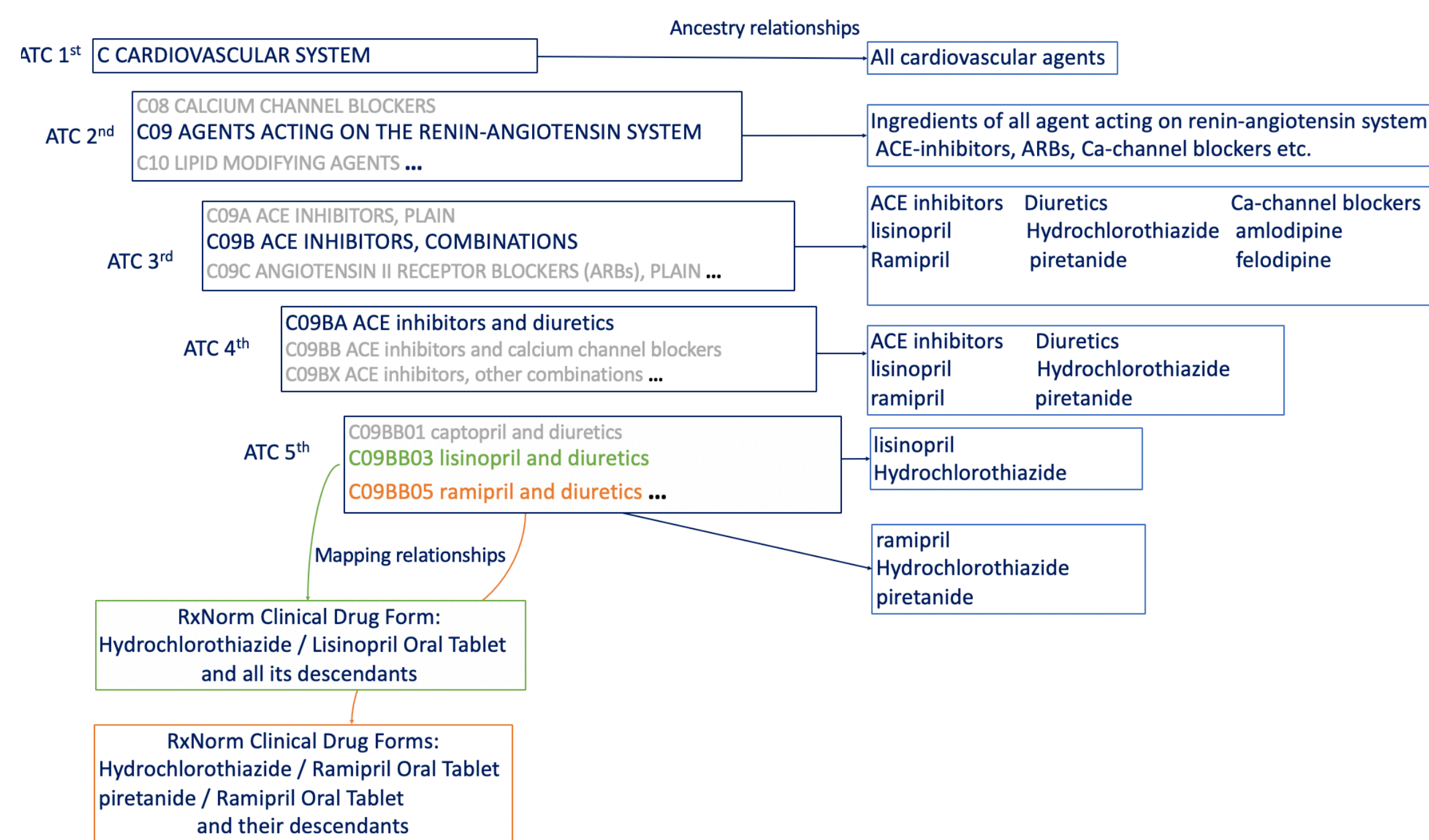
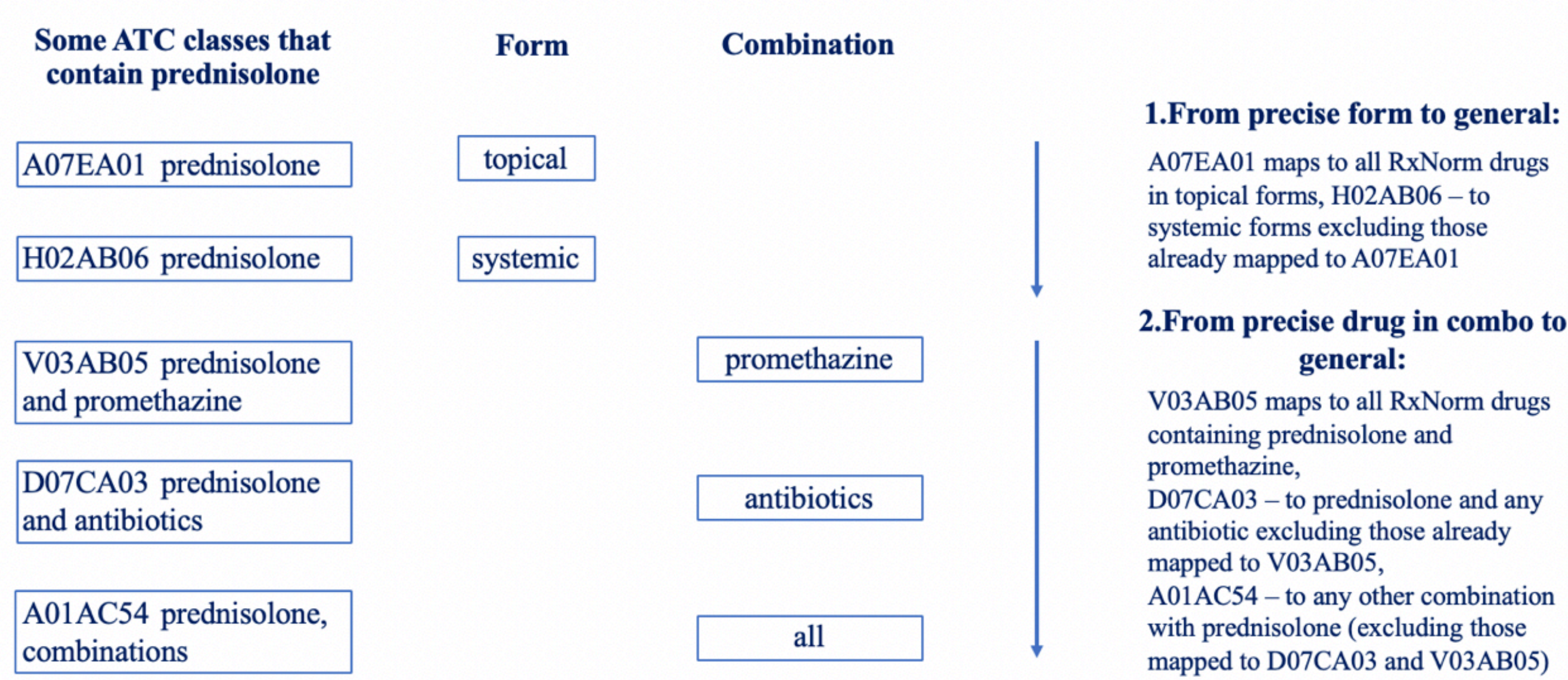


Figure 5. Elimination process

After deriving combination and dose form attributes, the highest and most appropriate RxNorm concepts are selected based on the similarity of attributes using an automated script. Then the RxNorm hierarchy is followed down to expand the mapping to all possible descendants. If descendants overlap, we established a ranking system to prioritize matching based on ATC attribute complexity (Figure 4).



Briefly, defined component combinations come first (e.g., N02AJ13 tramadol and paracetamol); combinations with broader groups second (e.g. N02BE71 paracetamol, combinations with psycholeptics), and single-substance ATC 5th concepts last (N02BE01 paracetamol). In this way, we eliminated pseudo-duplicates that might have been assigned other ATC 5th-level codes.

On the form level, we first defined all the Clinical Drug Forms that are applicable to an ATC with a certain ingredient and ROA (e.g. C02DA01 diazoxide, parenteral form) and mapped them to corresponding RxNorm Drug Products. Second, we mapped the ATC concepts with the same ingredient but without defined ROA to the remaining RxNorm Drug Products eliminating ones from the first step (e.g. V03AH01 diazoxide, no ROA defined). In this example, C02DA01 was mapped to 374410 Diazoxide Injectable Solution and its descendants, and V03AH01 to 371788 Diazoxide Oral Capsule, 371789 Diazoxide Oral Suspension and 378824 Diazoxide Oral Tablet.

Results and Validation

Of all 4,964 ATC 5th-level concepts **3,809 (77%)** are mapped to RxNorm concepts. Among the mapped ATC codes, **518 (10.4% of the total) were unambiguous** and covered any RxNorm concept containing that ingredient with no additional attribute constraints, while the others underwent the above heuristic. Unmapped codes include unapproved ingredients in the US (e.g. A08AA06 etilamfetamine), combinations of ingredients not marketed in a supported region (e.g. A10BD12 pioglitazone and sitagliptin), and ingredients with non-typical ROA (e.g. V10AA03 yttrium (90Y) silicate colloid).

NLM crosswalks comparison

Figure 6. ATC-to-RxNorm mapping performance overlap.

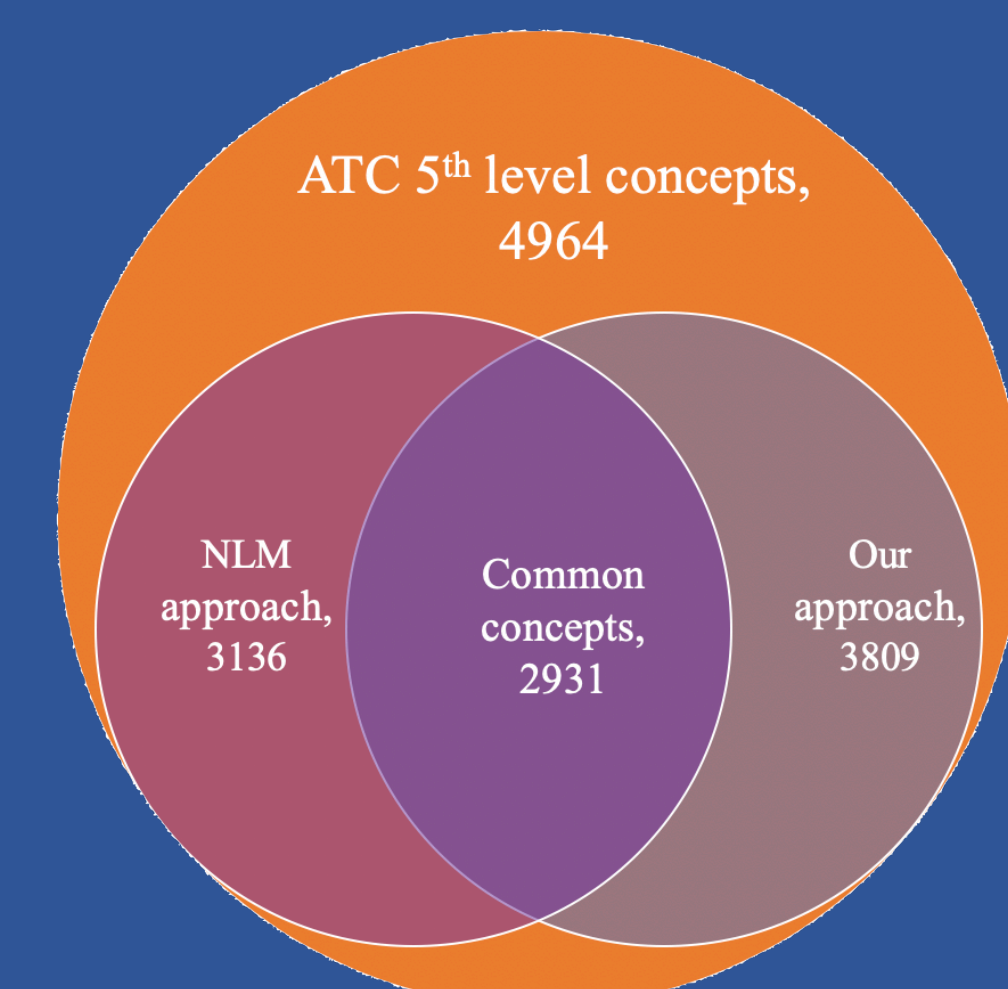


Table 1. Our and NLP approach comparison	Our approach	NLM approach
Multi-component drugs	Yes	No
Complicated groups: insulins, vaccines	Yes	No
Adjusting for ROA	Yes	No
RxNorm Clinical Drug coverage	71%	28%

23,931 RxNorm concepts out of a total of 33,872 (71%) have an ATC ancestor of Clinical Drugs, compared to the mapping provided by the NLM (27.7%). It includes not only monoingredient drugs, but also combinations and complicated and ambiguous groups (Table 1).

Since the ATC hierarchy is not comprehensive, we do not expect to cover all drugs, but our approach allows to automatically expand the mappings once ATC is updated. Compared to the NLM-provided cross-maps, our approach yielded 93% overlap (Figure 5), and covered 673 additional concepts as we also processed multi-component drugs and combinations.

Validation on patient data

We validated our approach on the IPCI database by comparing the number of records coded under an ATC code when using our approach versus G-Standaard RxNorm codes assignment (Figure 7).

Our analysis included Clinical Drug Forms and gained 91% matching (3990 concepts out of 4358) between our approach and the manual mapping. We missed 5% of ATC-RxNorm crosswalks, mainly because of deprecated RxNorm concepts. We had discrepancies in ATC 5th level code assignment for 4% of records. Table 2 presents the examples of discrepancies and their analysis.

Figure 7. Validation plot. X axis represent the number of records per an ATC code (a dot) using source data RxNorm codes assignment and Y axis represent the number of records gained using our approach.

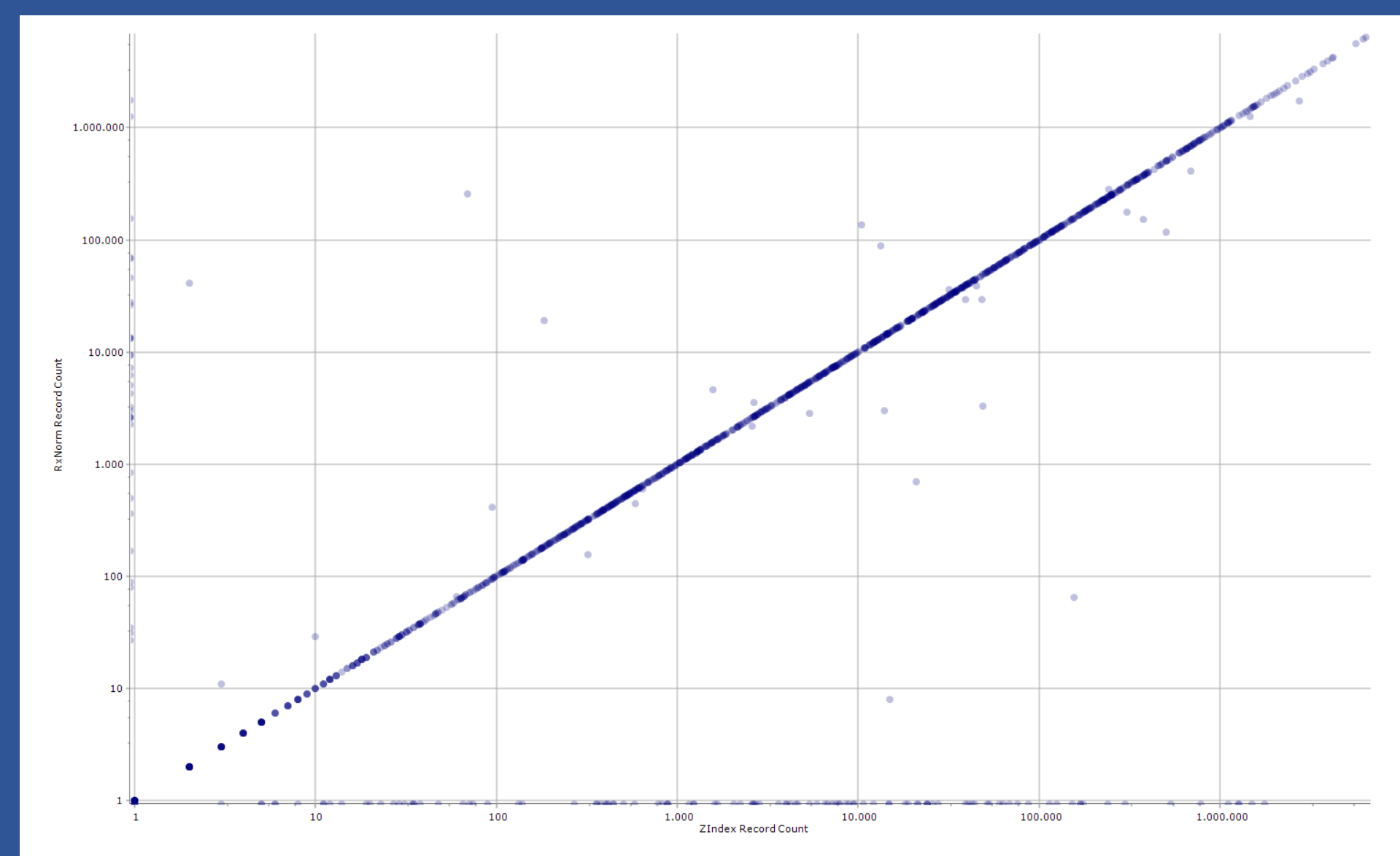


Table 2. Analysis of mapping discrepancies.

Source RxNorm concept	Our ATC assignment	G-Standaard ATC assignment	Comment
Diclofenac Sodium 0.03 MG/MG Topical Gel	D11AX18 diclofenac	M02AA15 diclofenac	Impossible to distinguish
medroxyprogesterone acetate 200 MG Oral Tablet	G03DA02 medroxyprogesterone	L02AB02 medroxyprogesterone	We do not account for different daily dosages, yet
Sodium Hyaluronate 23 MG/ML Injectable Solution	M09AX0 hyaluronic acid	S01KA01 hyaluronic acid	The source does not account for ATC ROA (M09AX01 –ophth., S01KA01-parent.)
Mitomycin 0.2 MG/ML Ophthalmic Solution	L01DC03 mitomycin	S01XA Other ophthalmologicals	Our mapping is more precise

Conclusion

Our semi-automated mechanism preserves the **correct** mapping between ATC and drugs, but also maintains a semantically **correct assignment of ingredients**. This will allow OMOP vocabulary users to use ATC as the standard classification system for drug products. We are currently working on the mapping to extend the mapping to remaining ATC codes that will mostly cover indications and non-therapeutic agents (e.g., media contrasts).