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

Anna Astropolots, MD, Hans Abadhezan, PharmD, PhD, Peter Birnbeek, PhD, Marcel de Wilde, Alexander Davydov, Olha Maruschak, Christian Reich, MD

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 Halytskyi Lviv National Medical University, Lviv, Ukraine, 5 OVIDIA, Cambridge, MA, USA, 6 Odysseyx Data Services, Cambridge, MA, USA

Why is it important?
ATC, developed by the WHO, is the most commonly used standard in pharmacoeconomics, pharmacovigilance, pharmacoeconomy, and observation 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. The 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 ingredients. But this link is "greedy" and can result in gross observational in the world. OMOP uses RxNorm as a standard for marketed pharmacoepidemiology, pharmacovigilance, pharmacoeconomy, and indication.

ATC Structure and Challenges

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 strength. For example, the various therapeutic purposes of prednisone is reflected in differences in the formulation: an anti-inflammatory, a corticosteroid, and an immunosuppressant. This is why it is important to have a comprehensive drug classification system to accurately reflect the different uses of the same drug.

Process

From the original five-level ATC hierarchy, we sought to create two types of connections:
1. From the lowest level ATC5 system to RxNorm Drug Products
2. From any level ATC to the relevant drug ingredients.

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

Figure 3. Mapping of individual ATC 5th level concepts

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

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, an established ranking system to prioritize matching based on ATC attribute complexity (Figure 4).

Brieﬂy, deﬁned combination comes ﬁrst (e.g., N02A13 tramadol and paracetamol); combinations with broader groups second (e.g. N02BE1 paracetamol, combi-150nations with psychostimulants); and single-ingredient ATC 5th concepts last (N02B51 paracetamol). In this way, we eliminated pseudo-duplicates that might have been assigned another ATC 5th level code.

The process involves the following steps:

1. Identify active ingredients
2. Identify dose forms
3. Prioritize matching based on ATC attribute complexity
4. Map to RxNorm concepts

Figure 5. Elimination process

Results and Validation

Of all 4,964 ATC 5th level concepts, 3,890 (77%) are mapped to RxNorm concepts. Among the mapped ATC codes, 518 (10.4%) of the total were unambiguous and covered the RxNorm concept containing that ingredient with no additional attribute constraints, while the others underlie the above heuristic. Unmapped codes include unapproved ingredients in the US (e.g. A05BA06 etilamfetamine), combinations of ingredients marketed in a supported region (e.g. A10B12 ibuprofen and sitaglipin), and ingredients with non-typical ROA (e.g. V10A03 yttium (90Y) silicate colloid). 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 with RxNorm in five ATC levels.

NLM crosswalks comparison

Table 1. Our and NLP approach comparison

<table>
<thead>
<tr>
<th>Components</th>
<th>Ours</th>
<th>NLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-component drugs</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Combinations: insulins, vaccines</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Adjusting for ROA</td>
<td>71%</td>
<td>28%</td>
</tr>
</tbody>
</table>

Validation on patient data

We validated our approach on the IPCI database by comparing the number of records codes with an ATC code under our approach versus G-Standard 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 6% of ATC-Five-level crosswalks, mainly because of deprecated RxNorm concepts. We had discrepancies in ATC 5th level code assignment for 4% of records, Table 2 presents the discrepancies of their performance overlap.

Table 2. Analysis of mapping discrepancies.

<table>
<thead>
<tr>
<th>Source RxNorm concept</th>
<th>Target RxNorm concept</th>
<th>Our approach</th>
<th>G-Standard RxNorm</th>
<th>Discrepancies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diclofenac Sodium 0.03</td>
<td>M22AX03 Glucosaminic acid</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Diclofenac Sodium 0.1</td>
<td>M22AX03 Glucosaminic acid</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Diclofenac Sodium 0.5</td>
<td>M22AX03 Glucosaminic acid</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Diclofenac Sodium 1</td>
<td>M22AX03 Glucosaminic acid</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

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 expand the existing remaining ATC codes that will most cover indications and non-therapeutic agents (e.g., media contrasts).