Automatic Mapping of Drug Concepts to the RxNorm Vocabulary

Maxim Moinat* [1], Lars Pedersen [2], Jolanda Strubel [1], Marinel Cavelaars [1], Kees van Bochove [1], Peter Rijnbeek [3], Michel van Speybroeck [4], Martijn Schuemie [4]

[1] The Hyve, Utrecht, The Netherlands
The Hyve, Cambridge, United States
[2] Aarhus University Hospital, Aarhus, Denmark
[3] Erasmus MC, Rotterdam, The Netherlands
[4] Janssen Pharmaceuticals, Inc.

*E-mail: maxim@thehyve.nl.

the hyve
We empower scientists by building on open source software

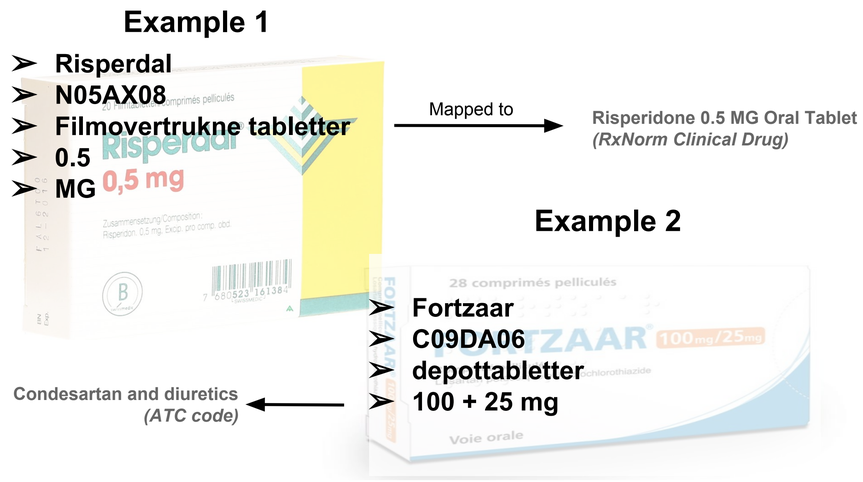
Aarhus University Hospital

1. Background

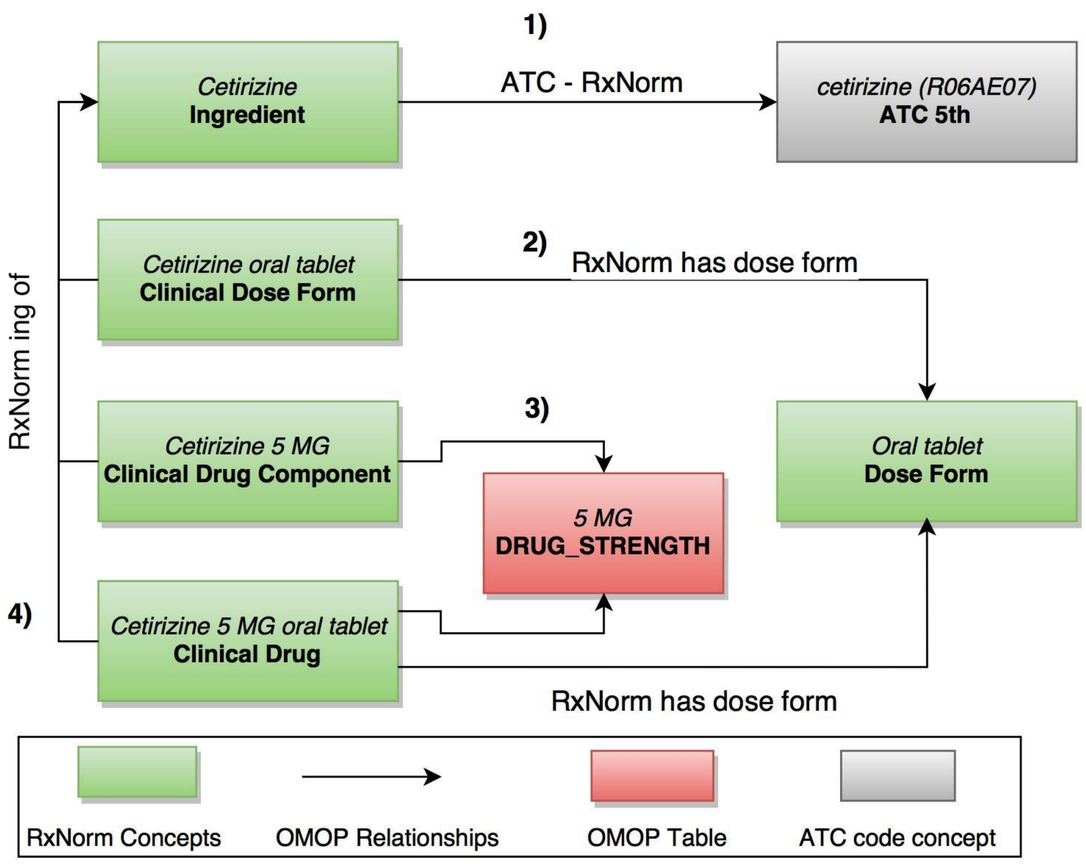
Mapping source concepts to the standard concepts in the OMOP vocabularies is one of the most time-consuming tasks during the transformation to the **OMOP Common Data Model**. Drug mapping is in particular challenging, because different components have to be mapped: **ingredient, dose form and strength.**

As part of the **European Medical Information Framework** (EMIF) project, Danish population health data are mapped to the OMOP CDM, including the local drug codes. The Hyve assists in creating a script to **automatically map** a set of **4754 drugs to the RxNorm vocabulary**. The input data contains ATC codes, dosage forms, numerical strengths and strength units. Two examples are shown in Figure 1.

The mapping procedure presented here is based on the drug mapping for the Japan Medical Data Center Claims Database[I].



▲ ***Figure 1: Examples of input data.*** *Example 1 is successfully mapped automatically. Example 2 consists of two ingredients and has an ATC concept that could not be mapped to a RxNorm concept.*



▲ ***Figure 2: Mapping concepts and relationships*** *with an example for each concept.*

2. Mapping Procedure

The mapping uses the **RxNorm hierarchy** and consists of four steps (see Figure 2).

1. Drugs are mapped to RxNorm **Ingredient** via the 5th level ATC code. The OMOP relationship 'ATC - RxNorm' is used for this purpose.
2. Dose form is added to the ingredient level, to map to **Clinical Drug Form** level.
3. The information on drug strength (including unit) is added to map to **Clinical Drug Component**. The strength is rounded to two decimals.
4. The above three mappings are combined to map to a **Clinical Drug** concept.

Manual mappings are added for a small number of frequently prescribed drugs. The Danish dose forms and units are also manually mapped to OMOP concepts.

3. RxNorm extension

A major limitation of the mapping is the **incomplete RxNorm vocabulary**. Multiple drugs in this Danish dataset do not have a counterpart in RxNorm. An example is Litarex 6 mMol, where RxNorm only contains strengths in mg. To be able to map these drugs, the OHDSI community has proposed to use an extension on RxNorm, called **Pseudo-RxNorm**. Work is in progress to add as much drugs as possible to Pseudo-RxNorm for a complete mapping.

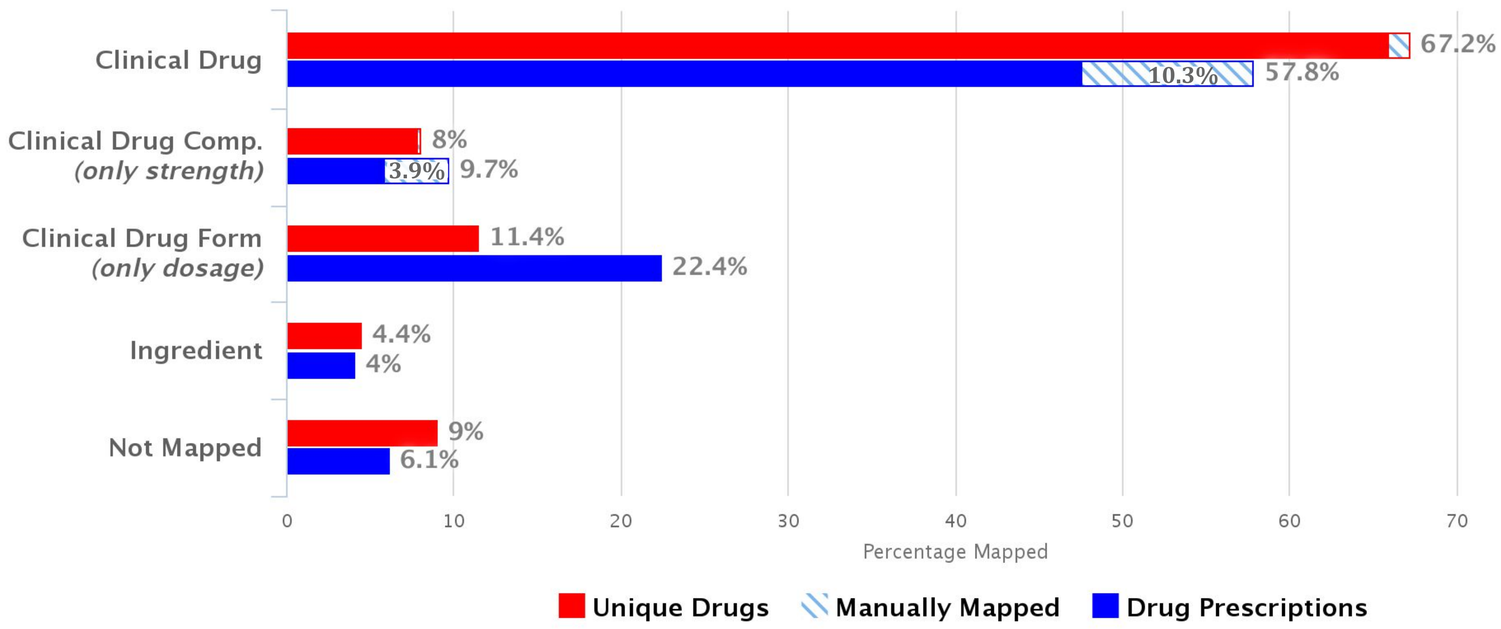
4. Challenges

Many unmappable drugs are drugs consisting of **multiple ingredients** and cannot be automatically mapped to one RxNorm ingredient. The ATC concept is often too general, see Example 2 in Figure 1. The automatic mapping from ATC to RxNorm ingredient should be revisioned to accommodate for mapping of these drugs.

Other challenges include:

- Synonymous dose forms (e.g. 'Cream' and 'Topical Cream')
- Numerator and denominator unit (e.g. 'GL' to 'gram' and 'liter')
- Strength derivation (e.g. 8 gram to 8000 milligram)
- Duplicate mappings (e.g. one drug to multiple Drug Forms)

5. Results



▲ ***Figure 3: Mapping results.*** *Percentages are based on the count of unique drugs (red bars, n= 4754) or on the number of prescriptions (blue bars, n= 1,093,056). The striped bars show the percentage of manually mapped drugs and prescriptions. Each drug is mapped to only one of the concept classes. If a drug could be mapped to Clinical Drug, then it is not included in the percentages of Clinical Drug Component or Clinical Drug Form. It can be seen that 91% of the drugs could be mapped and 67.2% without any loss of information (red bars).*

6. Conclusions

The majority of the Danish drug concepts have been mapped automatically to the RxNorm vocabulary (red solid bars in Figure 3). Further improvements can be made by extending the manual mappings and **supporting multi-ingredient drugs**.

However, a 100% mapping is not achievable with this method. Our work clearly demonstrates the need for the addition of a RxNorm extension to enable the mapping of currently missing drugs, forms and strengths in the standard OMOP vocabulary.

References: [I] Schuemie M, Kubota K, "JMDC drug to OMOP Vocabulary mapping", 2014 August 26

the hyve janssen EMIF Aarhus University Hospital PostgreSQL