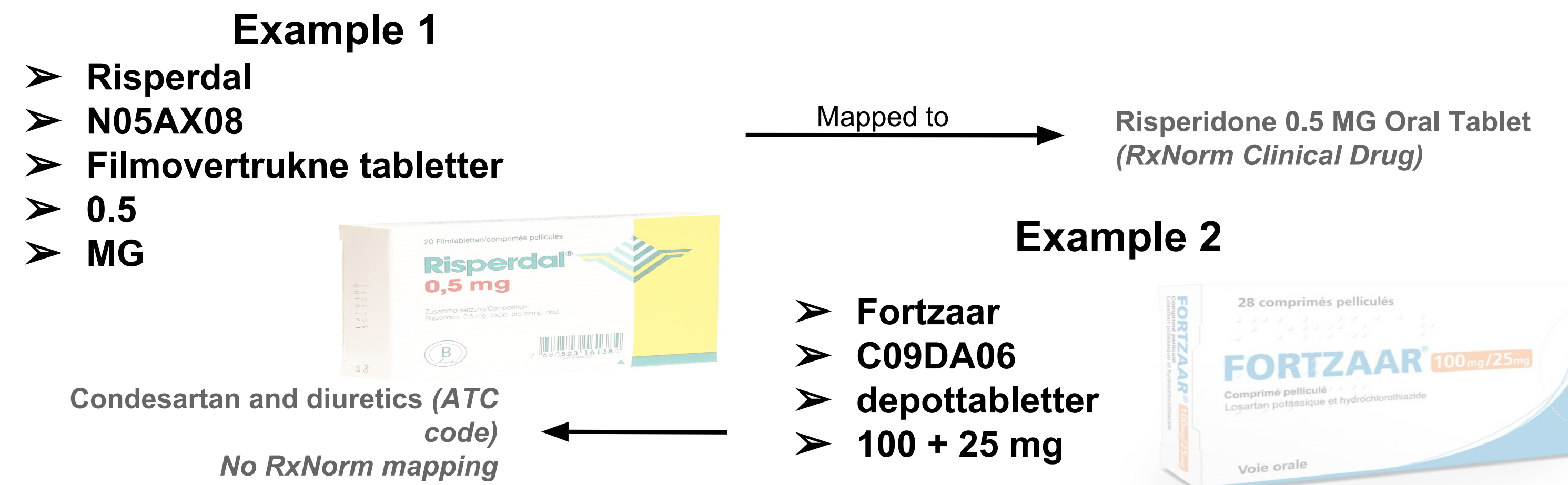


Mapping Danish drug concepts via ATC to the RxNorm vocabulary

Maxim Moinat [1] (maxim@thehyve.nl), Lars Pedersen [2], Jolanda Strubel [1], Marinel Cavelaars [1], Kees van Bochove [1], Michel van Speybroeck [3], Martijn Schuemie [3]
[1] The Hyve, Utrecht, The Netherlands
[2] Aarhus University Hospital, Aarhus, Denmark
[3] Janssen R&D, Inc.



▲ **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 an RxNorm concept.

Method

The mapping uses the **RxNorm hierarchy** and consists of four main steps (Figure 2). This method requires that the Danish dose forms and units are already mapped to RxNorm dose forms and UCUM unit concepts, respectively.

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.

Lastly, **manual mappings** are added for a small number of frequently prescribed drugs that could not be mapped automatically.

Challenges

There are two major points where information is lost, either there is **no ATC to RxNorm ingredient mapping** or the specific drug strength does not exist. An example of the latter is 'Primcillin 800 mg', which can only be mapped to 'Penicillin V Oral Tablet'.

The main cause of missing ATC to RxNorm ingredient mapping are **multi-ingredient drugs**, e.g. 'Bendroflumethiazide and potassium, 2,5+573 mg'. The multi-component RxNorm ingredient does not exist and the source data is inferior; it does not specify the strengths of the individual components. A different approach is necessary to allow mapping of these drug concepts.

Other challenges include:

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

Conclusions

The majority of the Danish drug concepts have been mapped automatically to the RxNorm vocabulary (Figure 3), although many challenges are remaining. Further improvements can be made by **supporting multi-ingredient drugs**, enriching the source data with detailed drug strength information and harmonizing strength units. This work is a starting point for subsequent mappings of local European drug ontologies to the OMOP standard vocabulary.

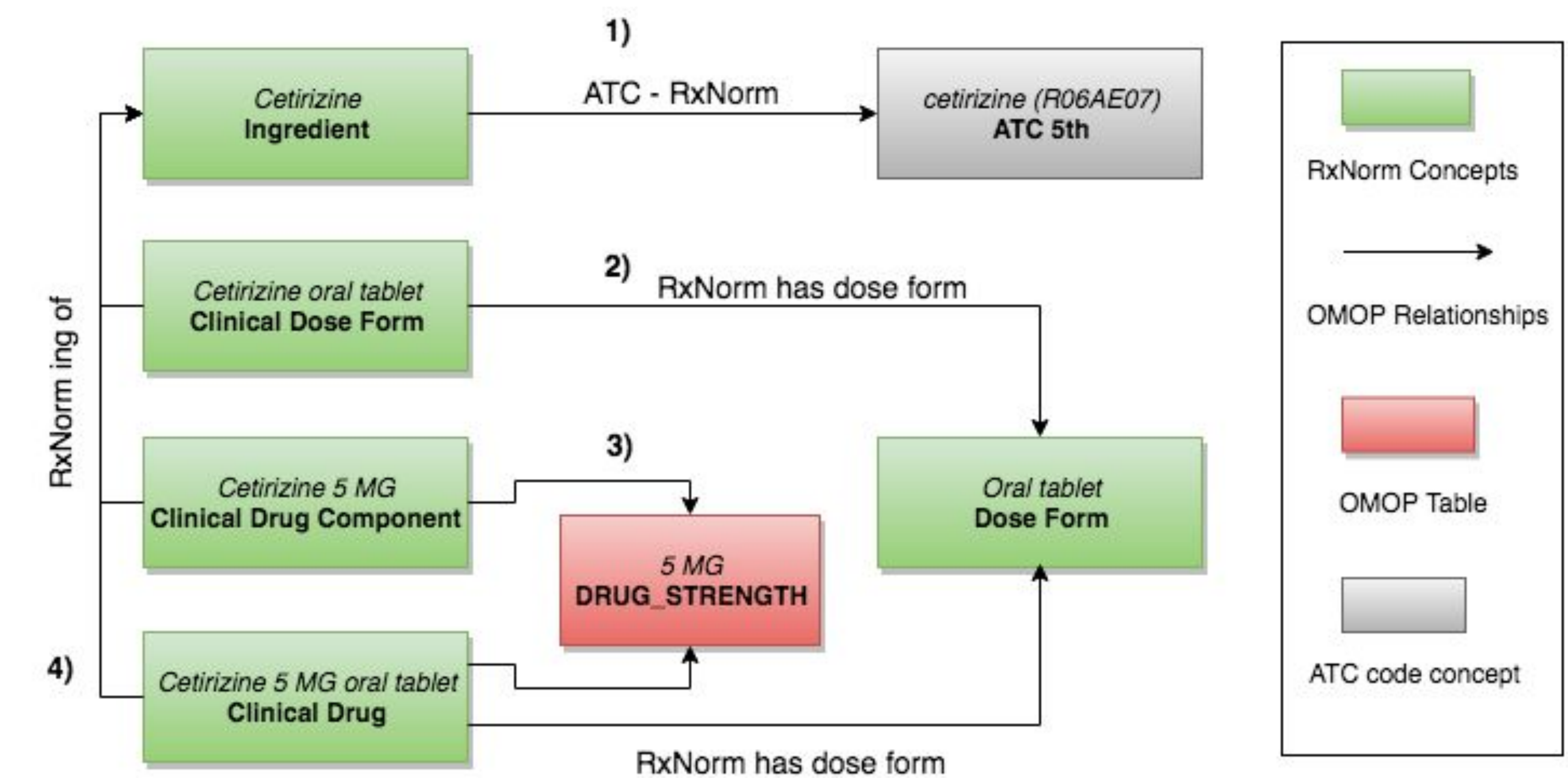
References: ¹ Schuemie M, Kubota K, "JMDC drug to OMOP Vocabulary mapping", 2014 August 26

Background

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

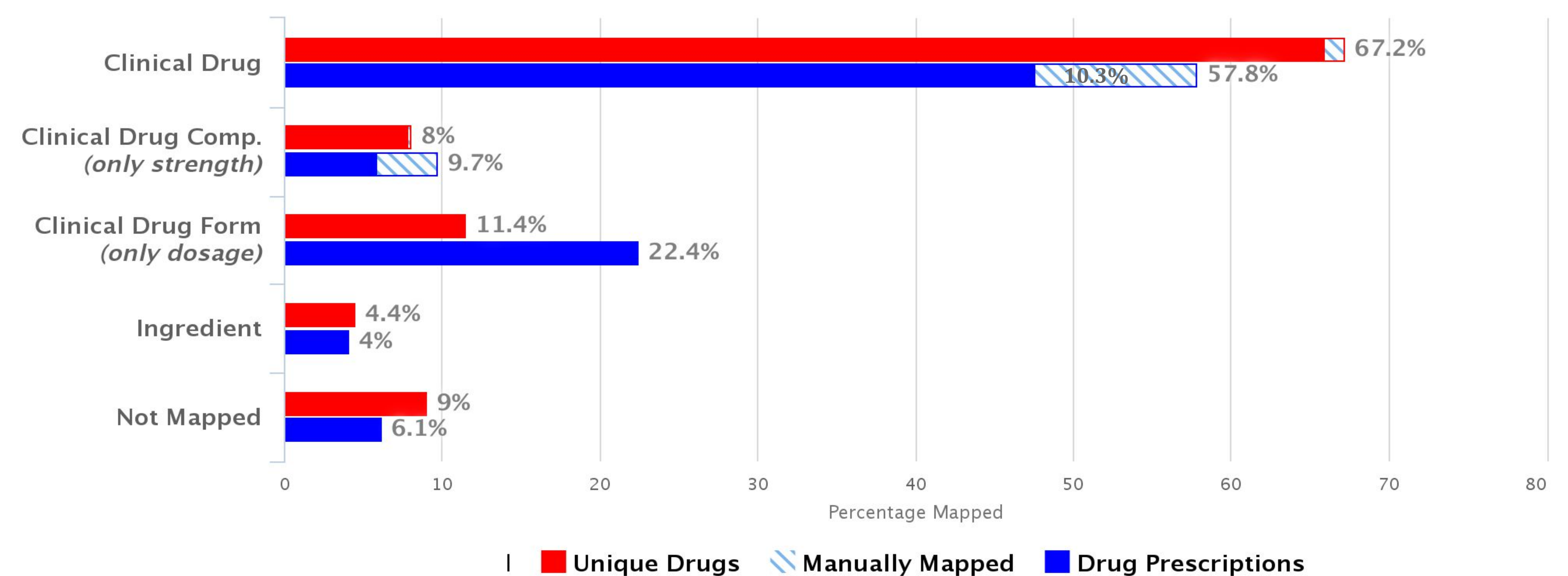
As part of the **European Medical Information Framework (EMIF)** project, Danish population health data are mapped to the OMOP CDM, including the local drug codes. The Hyve assisted in creating a script to **automatically map a set of 4754 drugs to the RxNorm vocabulary**. The input data contains ATC 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 Database¹.



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

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).