Name:	Maxim Moinat
Affiliation:	The Hyve
Email:	maxim@thehyve.nl
Presentation type (select one):	Poster

Mapping Danish drug concepts to the RxNorm vocabulary

Maxim Moinat[1], Lars Pedersen[2], Jolanda Strubel[1], Marinel Cavelaars[1], Kees van Bochove[1], Peter Rijnbeek[4], Michel van Speybroeck[3], Martijn Schuemie[3]
[1]The Hyve, Utrecht, Netherlands; [2]Aarhus University, Aarhus, Denmark; [3]Janssen Pharmaceuticals, Inc.; [4]Erasmus MC, Rotterdam, The Netherlands

Abstract *(100 - 200 words)*

Mapping drugs to the RxNorm vocabulary is a time-consuming and difficult task, which has to be specially designed for each source dataset. We developed an automated script that maps Danish drug concepts to RxNorm, the OMOP standard vocabulary for drugs. The mapping is performed on four levels: Ingredient, Drug Form, Drug Component and Clinical Drug. We managed to map a majority (67%) of the drugs to the RxNorm vocabulary. Improvements are needed, especially to map multi-ingredient drugs. Our work also demonstrates the need for addition of concepts to the standard vocabulary, which will be supported in the 'Pseudo-RxNorm' vocabulary.

Introduction

Extract, Transform and Load (ETL) data from a source database to the OMOP CDM[1] is a challenging task. Mapping the source concepts to the standard concepts in the OMOP vocabularies can be a time-consuming and difficult task. Drug mapping is in particular challenging, because it has different components that have to be mapped: ingredient, dose form and strength.

As part of the European Medical Information Framework (EMIF)[2] project, Danish data was mapped to the OMOP CDM, including the local drug codes. A standard mapping to RxNorm[3] was not available. The Hyve assisted in creating an automatic script to map a set of 4754 Danish drugs to the RxNorm vocabulary.

An example of the data of one local drug code is given below.

vnr	Pname	Dose form	Strength	packtext	strnum	strunut	PACKSIZE	ATC	Frequency
148	Loratadin "HEXAL"	tabletter	10 mg	10 stk. (blister)	10	MG	10	R06AX13	152

Method

The mapping process was based on the approach developed by Martijn Schuemie and previously successfully implemented for mapping the Dutch Z-codes and Japanese JMDC drugs. All scripts were are written in SQL and developed for a PostgreSQL database. The mapping is performed in mainly four steps, Each mapping the drug to a higher level in the RxNorm vocabulary.

In the first step, drugs are mapped to the RxNorm *Ingredient* via the supplied 5th level ATC code. The OMOP relationship 'ATC - RxNorm' was used for this purpose.

ATC codes that consist of two or more ingredients cannot be mapped, because there is no build in mapping from ATC to multiple RxNorm ingredients.

> Example of more than one ingredient: ATC=C03AB01, "bendroflumethiazide and potassium". Has no mapping to a RxNorm ingredient level. The ingredients "bendroflumethiazide" and "potassium" do exist separately. At the Clinical Drug level they could be combined to the RxNorm concept "Bendroflumethiazide 2.5 MG /

In the second step, dose form is added to the ingredient level, to map to *Clinical Drug Form* level. The Danish dose forms (e.g. 'injektionsvæske') were each mapped to a standard dose form in the OMOP vocabulary. There were 490 different dose forms in the Danish dataset. Only the 90 most frequently occurring were mapped.

One particular thing we encountered were the synonymous dose forms in the OMOP vocabulary, e.g. 'Cream' and 'Topical Cream'. To solve this, we looked at how often each dose form was used by RxNorm *Clinical Drug*

Form. We found that 'Cream' was never used, but 'Topical Cream' was used 4177 times. Therefore, we chose to map the Danish dose form 'Creme' to 'Topical Cream' instead.

In the third step, the information on drug strength (including unit) is added to map to *Clinical Drug Component*. All the single ingredient drugs are retrieved from the DRUG_STRENGTH table and the source data units are mapped to standard unit concepts.

An extra challenge were the units with a numerator and denominator. These units, e.g. 'mg/ml', were mapped to two standard unit concepts: milligram as numerator and milliliter as denominator. Additionally, the strength values were rounded to two decimals to increase the number of matching drugs (e.g. 0.132 mg is the same as 0.13 mg).

In the last step, all three mappings (ingredient, dose form and strength) are combined to map to a *Clinical Drug* concept.

Finally an additional manual mapping to *Clinical Drug* was performed for the most frequent drugs that could not be mapped automatically. This is a labor intensive procedure and therefore only a small number of drugs were mapped in this way.

Results

Table 1 shows the result of the mapping procedure. More than half of the drugs could be mapped to a Clinical Drug. A total of 19.3% of the drugs could be mapped to ingredient, but had either the Dose Form or Strength information missing (Clinical Drug Comp and Clinical Drug Form). 9% of the unique drugs, constituting just 6% of the prescriptions, could not be mapped at all to a RxNorm concept **(no ATC mapping)**. **Why?** Please note that the automatic mappings are not verified, there could be false positive mappings among them.

Table 1. *Mapping statistics. Percentages based on the count of unique drugs (n= 4754) and based on the number of prescriptions (n= 1,093,056). Each of the four levels of mappings contain information on one or more of the following domains: Ingredient (Ingr.), Dose Form (Form) or Strength (Str.). Drugs are mapped to only one of the concept classes. For instance, 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. Ingr=ingredient; str=strength*

Concept Class in RxNorm	Information on			Percentage based on Drug Name count	Percentage based on Prescriptions
	Ingr.	*Form*	*Str.*		
Clinical Drug	x	x	x	67.2%	57.8%
Clinical Drug Comp	x		x	8.0%	9.7%
Clinical Drug Form	x	x		11.4%	22.4%
Ingredient	x			4.4%	4.0%
Not Mapped				9.0%	6.1%

Missing vocabulary concepts

Only the drugs available to the US market are present in the RxNorm vocabulary, so not all dose forms and strengths are available for each drugs. This makes that some Danish drugs do not have a counterpart in RxNorm. This is a widely recognized problem and efforts are made by the OHDSI community to alleviate this problem. by introducing a 'Pseudo-RxNorm', where the missing drug concepts can be added.

Conclusion

We have shown good progress in the automatic drug mapping using ATC codes. The script is reproducible on any PostgreSQL OMOP database and the manual mappings can be added.

If new drugs are added to the Danish market, this script will be able to create a new mapping to the RxNorm vocabulary. However, the new drugs have to exist in the RxNorm vocabulary, which currently only contains US based drugs. Our work clearly demonstrates the need for the addition of a 'Pseudo-RxNorm' to enable the mapping of currently missing drugs, forms and strengths in the standard OMOP vocabulary.

References

1. Observational Medical Outcomes Partnership [Internet]. OMOP. 2016 [cited 22 June 2016]. Available from: http://omop.org/
2. European Medical Information Framework [Internet]. EMIF. 2016 [cited 22 June 2016]. Available from: http://emif.eu/
3. US National Library of Medicin [Internet]. RxNorm. 2016 [cited 20 June 2016]. Available from: https://www.nlm.nih.gov/research/umls/rxnorm/