Mapping the Article 57 Database from the European Medicines Agency containing all medicines authorized in the European Economic Area

Marcel de Wilde[1], Mees Mosseveld[1], Steven Le Meur[2], Sofia Zastavnik[2], Gianmario Candore[2], Peter Rijnbeek[1]

[1]Erasmus Medical Center, Rotterdam, The Netherlands

[2]Business Data Department, European Medicines Agency, Amsterdam, The Netherlands

Abstract

The Article 57 database from the European Medicines Agency (EMA) contains all the authorized medicines in the European Economic Area. In this study we made a first attempt in mapping this database to RxNorm and RxNorm Extension. This work provides insights in the necessary additions in the Observational Medical Outcomes Partnership (OMOP) Standardized Vocabularies to cover the European Market. In total 97.68% of all drugs can be mapped to the ingredient level or more granular, while 86,96 % of all the drugs can be mapped to the Clinical Drug Level (ingredient, dose form, and strength). Our next steps will be to perform additional quality control steps together with the EMA and involve the vocabulary team of the Observational Health Data Sciences and Informatics (OHDSI) initiative to extend the vocabularies and obtain full European coverage.

Introduction

Today, in Europe, we are challenged to generate insights and evidence from real world clinical data at scale, to support patients, clinicians, payers, regulators, governments, and the pharmaceutical industry in understanding wellbeing, disease, treatments, outcomes and new therapeutics and devices. Such data is difficult to use at scale, in a myriad of languages, systems and structures, with challenging policy restrictions and technology considerations.

In response to this, the European Health data and Evidence Network (EHDEN) project was launched as a public private partnership under the framework of the Innovative Medicines Initiative (IMI) running from November 2018 until April 2024**.**

The EHDEN project aspires to be the trusted observational research ecosystem to enable better health decisions, outcomes and care. Its mission is to provide a new paradigm for the discovery and analysis of health data in Europe, by building a large-scale, federated network of data sources standardized to the OMOP common data model.

To enable large-scale analytics, we need to improve the syntactic interoperability, i.e. to standardize the structure of the data. Furthermore, we need to improve the semantic interoperability, in other words we need to speak the same language. This is challenging in Europe because there are many local coding systems used for all data domains, including the drugs.

Article 57 database

The Article 57 database is a comprehensive repository of structured information on all medicines authorized in the European Economic Area. Marketing authorization holders must submit and maintain this information in accordance with European Union legislation**.** A public version of this database is made available by the EMA (link) and contains information such as the product name, active substance, route of administration, country of authorization, marketing authorization holder (company), etc.

This database is of great value to the EHDEN project because it allows us to extend the vocabulary to cover all the authorized drugs in Europe and the mapping of all local drug coding systems without loss of information. The EMA made available a recent copy of the database that contains additional fields and will provide frequent updates in the future.

Methods

1. To execute the mapping, we developed an application that takes as input the datafiles provided by the EMA and generates a file containing the mappings to the RxNorm or RxNorm Extension code at the most granular level possible: Clinical Drug (ingredient, strength, form), Clinical Form (ingredient and form), Clinical Comp (ingredient and strength), or ingredient only level. The output file also provides insight in how the mapping was performed and if any adjustments were made, e.g. if only a strength within 10% of the value was available. The input files contain separate columns for the ingredient, dose form, strength, unit, and the Chemical Abstracts Service (CAS) code of the ingredients if available.
2. If needed normalize the strength and units, for example converting to mg
3. Map the dose forms
4. Map to the most granular drug level possible (ideally to Clinical Drug Level and fall back to Clinical Comp, Clinical Form or Ingredient level otherwise).

Results

As shown in Table 1 we were able to successfully map a very large number (97.68%) of the 536854 drugs present in Art 57 to RxNorm or RxNorm Extension. In total 86.96% can be mapped to the Clinical Drug Level.

Table 1. Submission type, abstract length, and page length maximum for OHDSI submissions.

Drug Level	Number	Percentage
Clinical Drug	466852	86.96 %
Clinical Drug Comp	24709	4.60 %
Clinical Drug Form	19974	3.72 %
Clinical Drug Comp & Ingredient	7004	1.31 %
Ingredient	5854	1.09 %
Unmapped	12461	2.32 %

For 1.31% of the drugs a mapping could be made to Clinical Drug Comp Level only for one or more of its components while other components could only be mapped to ingredient level because these were not yet available in the vocabularies.

We found that some ingredients were not available in RxNorm or RxNorm Extensions, e.g. epitizide, that are used in multiple drugs and therefore all these drugs were unmapped. Some of these are already added to the vocabulary to solve this problem.

Conclusion

Preliminary results showed that there is a good coverage of the authorized drugs on the European market that are available in the Article 57 database. Our next steps are to validate and further improve the automated mapping in collaboration with the EMA. This work will also include an assessment of the coverage of Article 57 by comparing it with national coding systems we encounter in the EHDEN project. Furthermore, together with the Vocabulary Team we will further extend the OMOP Standardized Vocabularies to obtain full coverage of all medicines on the European Market.