1. Background

At The Hyve, we have developed conversions of several European datasets to the OMOP CDM. This includes source data analysis, conversion workshops, development and execution of ETL scripts and most importantly data quality assessment. Each source dataset is unique, both in structure (syntactic) and in ontologies (semantic) used.

The semantic mapping is a challenging and imprecise task, which is especially pronounced for non-US datasets. The OMOP standard vocabulary was primarily built around US claims data and therefore in many cases does not support local European ontologies. The OHDSI community is working on support for European ontologies and already incorporated many in the OMOP vocabularies. A great effort is for instance the RxNorm Extension, which includes RxNorm-like concepts for international drug prescriptions.

Still, high quality vocabulary mapping poses a challenge. Here, we present an extension to the existing quality assessment tool, Achilles, to monitor vocabulary mapping quality with additional descriptive statistics. We applied this new extension for the quality analysis of the OMOP conversion of two European datasets:

- **Swedish Electronic Health Records (EHR)**, consisting of four national registries; patient care (inpatient and outpatient), death registry, a subset of the drug registry (only cardiovascular drugs) and a subset of the socio-economic registry.

- **Dutch Primary Care (PC)**, collected from a selection of primary care providers in the Netherlands. The primary care consists of episodes, medication, test results and referrals to secondary care.

3. Results

The Achilles extension outputs a general coverage (Table 1) and a detailed metric per source vocabulary and target concept class. The detailed metric can reveal additional quality information, for example the RxNorm class to which drug concepts were mapped (Figure 1). In addition, the user can choose to export a list of the most occurring mapped and unmapped source concepts. The first serves as a sanity check of the mappings and the second shows how much there is to gain by mapping the top unmapped concepts.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Procedure</th>
<th>Drug</th>
<th>Measurement</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swedish EHR</td>
<td>99%</td>
<td>14%</td>
<td>83%</td>
<td>N/A</td>
</tr>
<tr>
<td>Dutch PC</td>
<td>96%</td>
<td>N/A</td>
<td>65%</td>
<td>100%</td>
</tr>
</tbody>
</table>

The Achilles extension outputs a general coverage (Table 1) and a detailed metric per source vocabulary and target concept class. The detailed metric can reveal additional quality information, for example the RxNorm class to which drug concepts were mapped (Figure 1). In addition, the user can choose to export a list of the most occurring mapped and unmapped source concepts. The first serves as a sanity check of the mappings and the second shows how much there is to gain by mapping the top unmapped concepts.

The Achilles extension outputs a general coverage (Table 1) and a detailed metric per source vocabulary and target concept class. The detailed metric can reveal additional quality information, for example the RxNorm class to which drug concepts were mapped (Figure 1). In addition, the user can choose to export a list of the most occurring mapped and unmapped source concepts. The first serves as a sanity check of the mappings and the second shows how much there is to gain by mapping the top unmapped concepts.

4. Discussion

In order to improve integration into the OHDSI quality analysis workflow, we have started to integrate the outputs into the Data Sources component of Atlas. This is work in progress and a first mockup is shown in Figure 2. Another useful feature would be to enrich the Achilles Heel warnings with information from these new vocabulary mapping metrics. For example by creating a drilldown with which concepts are not mapped and creating new warnings based on the vocabulary mapping metrics presented here. This quality analysis is complementary to the existing Achilles Heel quality analysis.

Lastly, the proposed mapping metrics are all quantitative; they report whether a mapping exists or not. The metrics cannot be used to determine whether an existing mapping is correct. For this purpose an expert review of the individual mappings should be provided. Usagi is a possible tool to extend with this capability, as it already allows an expert to search for and give a new suggestion for a mapping.

References: