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

Transforming Electronic Health Records from Swedish registers to the OMOP CDM

Maxim Moinat, MSc, Marinel Cavelaars, MSc PhD, Kees van Bochove, MSc Bed
The Hyve, Utrecht, Netherlands; Bayer Pharmaceuticals

Abstract ***(100 - 200 words)***

A major challenge for conducting observational studies using Electronic Health Records (EHR) is to apply the same study protocol to several databases. We developed the Extract Transform and Load (ETL) procedure to load Swedish EHR data into the OMOP[1] Common Data Model. Multiple Nordic healthcare coding systems, in particular the NOMESCO coding system, have been translated to the OMOP standard vocabulary. This allows for the use of OHDSI tools on the dataset. Specifically, deployment of OHDSI R packages is being planned for guidance in developing code for statistical analysis.

Introduction

The Nordic countries have one of the best centralized Electronic Health Records (EHR) in Europe. At The Hyve we transformed and loaded the Swedish inpatient, outpatient, drug, death and population (LISA) registers into the OMOP CDM. All the presented work is still under development. It has been successful so far and the work is expected to finalize soon.

Mapping of concepts

The first step was mapping the Swedish vocabularies to the OMOP standard vocabulary. Most work was performed on the NOMESCO[2] vocabulary for chirurgical procedures. The 6012 NOMESCO codes had to be mapped to SNOMED[3] procedures. Some NOMESCO codes were mapped to a higher 'chapter' level and some did not have a direct counterpart in the SNOMED vocabulary.

The Swedish version of ICD10 classification was also mapped to SNOMED concepts in the OMOP standard vocabulary. Other mapping tables that were created were: Drugs (ingredient level), provider specialties, care sites (hospitals) and locations (Swedish counties).

All mappings were stored in separate 'mapping files' that are loaded during the ETL procedure. This allows for manually curation of the mappings. A downside is that new source concepts are not automatically mapped to the OMOP standard vocabulary.

Challenges

The Swedish register data files are supplied as csv files, which are directly loaded in a PostgreSQL database. A challenge was coping with a changing format of the csv files, because the order of variables in the data files and the names of the variables are subject to change. To overcome this problem, we prepared the source data loading scripts according to the heading of each csv file. In addition, the expected variable names can be manually adjusted.

Other difficulties during data loading were the lack of support of the OMOP CDM for multiple causes of death. The Swedish death register can contain up to 48 causes of death and additional qualifiers for e.g. death related to alcohol or work. To solve this, we made a death addendum table to save multiple causes of death and the additional qualifiers. Similarly, the Swedish dataset contained up to 21 secondary diagnoses and 30 procedures. OMOP only supports 'Primary Condition', 'Secondary Condition' and 'First Position Condition'. We chose to apply 'First Position Condition' to the first secondary diagnosis and 'Secondary Condition' to the others.

The observation period was determined from multiple variables for each person. The observation start date was determined from the first date of the dataset (01-01-1997), the birth year and the immigration date. The

observation end date was similarly derived from the end of the dataset (01-08-2015), the year of death and emigration date.

Conclusion

In the future the developed ETL procedure can be used to transform other Swedish healthcare data. Especially the NOMESCO mapping will be a useful tool for all Nordic health data. The mappings could be extended to a more detailed level.

References

1. Observational Medical Outcomes Partnership [Internet]. OMOP. 2016 [cited 22 June 2016]. Available from: http://omop.org/
2. Nordic Medico-Statistical Committee (NOMESCO) 2011; NOMESCO Classification of Surgical Procedure; http://www.socialstyrelsen.se/klassificeringochkoder/atgardskoderkva
3. SNOMED CT [Internet]. IHTSDO. 2016 [cited 18 June 2016]. Available from: http://ihtsdo.org/snomed-ct