Standardizing FAERS Adverse Event Cases For Research Using The OHDSI Athena Standard Vocabularies

Lee D. Evans, LTS Computing LLC, West Chester PA

Abstract

The US FDA Adverse Events Data base (FAERS) is an invaluable resource for generating potential pharmacovigilence drug safety signals for further investigation. However it is challenging to take full advantage of this resource due to the use of 'free text' data entry of drug names. We successfully used the OHDSI Athena Vocabularies to map FAERS drug names into RxNorm standard code ingredients & clinical drug forms (for multi-ingredient drugs). We also used the Athena vocabularies to map FAERS drug indications and reactions from MedDRA preferred terms to SNOMED-CT standard codes. This enables FAERS adverse event case information to be integrated with the OHDSI LAERTES knowledge base and potentially any Electronic Health Record database that is in OMOP CDM Version 5 format.

Declaration of Interests

Introduction

The US FDA Adverse Events Data base FAERS database data sets are publicly available on the FDA web site. There are several quarterly data sets available for download, from January 2004 on-wards. Note. There was a change in the format of the database (in particular the assignment of unique primary keys) in Q4 2012 and so there is a need to take into account the older legacy format (LAERS) and the current FAERS format. We standardized all the legacy and current data using a 4-CPU 15GB Memory Postgresql server hosted on the Google cloud. (Google Compute Engine). We created an ETL process that could potentially be re-run quarterly or twice a year in order to maintain a standardized FAERS database for research by the OHDSI community. The adverse event drug/reaction pair counts we generated from this standardized FAERS database will also be incorporated as one of the data sources for the OHDSI LAERTES knowledge base.

Overall approach used

- Download the data sets from the FAERS web site and load the data into staging tables in the Postgresql database. All subsequent steps utilize SQL code for ETL.

- Remove duplicate cases across the legacy and current data using demographic key fields. Keep the latest available case version based on case id and case version id.

- Map drug names to RxNorm OMOP concepts using a combination of techniques:

  - Active ingredient supplied on some current FAERS records

  - Ingredient look-up using NDA number and the FDA Orange Book reference data set

  - Regular expression matching of the drug name text to OMOP Athena concept name within the RxNorm vocabulary

  - Manual drug name mapping using the OHDSI Usagi mapping tool

- Standardize from RxNorm drug brand names and deleted/updated older RxNorm drug concepts to RxNorm standard ingredient or clinical drug form using Athena Vocabulary concept relationships

- Map indications and reactions from MedDRA preferred terms to SNOMED-CT via Athena Vocabulary concept relationships

- Generate counts of drug/reaction pairs

- Generate drug/reaction pair contingency tables.

- Generate ROR and PRR potential safety signal statistics for further investigation

Conclusion

The OHDSI Athena vocabularies were an essential part of our ETL process for the standardization of the FAERS database. The ETL code we created will be made publicly available as open source on GitHub. The FAERS drug/reaction counts that we produced will be incorporated into the OHDSI LAERTES knowledge base and the standardized FAERS database may at some point in the future be made directly available to OHDSI community researchers, within the OHDSI cloud.

Acknowledgements