Standardization of the French national healthcare database (SNDS) in OMOP-CDM

Introduction

➢ The SNDS main database (Système National des Données de Santé) is one of the world’s largest healthcare database, covering outpatients claims, hospital discharge summaries, and national death registry for the whole French population.

➢ SNDS main database relies on a complex structure (180 tables, 4,500 variables) and numerous French-specific vocabularies: e.g., CCAM and CSARR (procedures), NABM (laboratory tests), LPP (medical devices), CIP and UCD (drugs).

➢ Data standardization is needed to improve reuse of the SNDS main database for real-world evidence generation and sharing of scripts and programs.

➢ Initially focused on a cohort of a few hundred thousand patients, standardization is gradually being applied to larger cohorts, introducing new challenges associated with processing large volumes of data.

Methodology

SNDS main database

 Extraction
3,000,000 persons
7 years (2015-2021)
93 tables
2 To

OMOP-CDM syntactic and semantic standardization

OMOP-CDM Database
16 base tables
310 Gb

Data Quality Checks
Achilles
CDM Inspection
DataQualityDashboard

Results

➢ French ontologies map to 4 major domains in the OMOP-CDM standard domains (Figure 1).

➢ A method of mapping is identified for each terminology.

➢ CCAM, CSARR, and NABM are mapped and reviewed by medical residents.

➢ LPP and UCD are mapped by an automatic tool, and the mapping is reviewed by medical residents.

➢ About 75% of visit records are outpatient visits (Figure 2).

➢ Almost 98% of visit records come from the SNIIRAM (Figure 2).

➢ More than ⅓ of records belong to the Visit domain (Figure 4).

➢ Half of records are part of Measurement, Procedure and Drug domains (Figure 4).

Conclusion

➢ Semantic harmonization was made complex by the level of detail captured by the French ontologies and is currently being improved.

➢ Standardized database has been validated by an OMOP certified SME.

➢ In European studies involving data from millions of patients, distributed computing with Spark allows large volumes of data to be effectively managed. The existing ETL facilitates the conduct of federated real-world studies in the SNDS main database via OHDSI tools, providing significant capabilities for health outcomes research.

Gaëlle Collumeau, Cécile Charles, Lorien Benda, Elena Mylonas, Hakim Randrianarivo, Axelle Menu, Gil Lampe, Stéphanie Combes
opensource@health-data-hub.fr