Success factors to an OMOP CDM ETL: having the right people, planning that vocabulary mapping can have complexity, and proactively preparing to deal with data governance issues.

European Health Data & Evidence Network (EHDEN) – Learnings from Building out a Standardized International Health Data Network

Background: An ETL converts health data to the OMOP CDM. During EHDEN COVID-19 data call, 25 data partners (DPs) transformed health data to the OMOP CDM. This data call provided an opportunity to evaluate ETL development success factors based on timely development and network research involvement. This study investigates OMOP CDM conversion success.

Results

Figure 1 – OHDSI Extract, Transform, & Load (ETL) Development Process with 1A ETL process map, 1B box plots showing the length in days for each step

- 25 DPs participated in this EHDEN data call, representing 11 different countries, collectively covering more than 67 million patient records
- 21 of these DPs, the median time it took to complete the ETL process was 358 days, with the shortest time being 172 days and the longest being 622 days
- Of the 21 DPs, 52% had built their CDM in under 365 days, 43% participated in the COVID-19 AESI study [1], and 33% had both a timely study and participated in the COVID AESI study

Methods

- Each DP was expected to follow the current OMOP CDM ETL development process (Figure 1A).
- The process was tracked in 3 ways (1.) through surveys of DP, (2.) tracking data associated to ETL process, and (3.) DataQualityDashboard results.
- Success was measured as total days to transform source data into the OMOP CDM (success: < 365 days) and if a DP participated successfully in network research (COVID-19 AESI study [1]).