A network of research ready data sets with a Dashboard tracking data quality issues and a Library of Metadata.

DARWIN EU®: Assessing the data quality on onboarding stage

DARWIN EU® (Data Analysis and Real-World Interrogation Network) is setting up a network of health care databases across Europe to support large scale regulatory studies. This requires standardization of a large amount of data sources. Data quality is strongly context dependent therefore data quality assessments should be performed at different stages in the evidence generation workflow. During onboarding of a data source in the DARWIN EU® data network the DARWIN EU® Network Operations pillar collects metadata from the data partners in a proactive manner. This process will be repeated with each update of the CDM. In addition, study type specific tests will be added at a later phase by the development pillar.

Methods

1. Data Quality Dashboard

The Data Quality Dashboard was developed by the European Health Data and Evidence Network (EHREN) project and is maintained in collaboration with the EHRDN community. This dashboard is a web-based system to evaluate over 1,000 data quality checks against a given CDM instance [1] and is used to track data quality issues in a real-time manner. The results helped identifying multiple issues that were then subsequently addressed and if provided opportunities for training on data standardization [2].

2. ACHILLES

ACHILLES [3] is a tool development by OHDSI providing a visualization of the entire database. It lists ACHILLES scores of 200 analyses, covering all elements of the OMOP CDM. Exploration of these characteristics allows identification of potential quality issues. In this knowledge of the database and local biobanking system is needed to evaluate the ACHILLES reports. Therefore, the quality checks must be performed by a source data expert together with an EHRDN expert.

The methods above are applied as part of the onboarding process for each DARWIN EU® data partner. The results of the quality assurance process were part of the recommendation for inclusion of the data partner in the DARWIN EU® network. Relevant recommendations coming from the data quality framework delivered by EMA are incorporated in upcoming versions. The Quality Assurance Package will also be constantly updated, e.g., to cover new quality checks that may become of value when the Catalogue of Analytics grows. The framework we developed can easily be extended to meet our future goals.

References

5. https://github.com/EHDEN/cdminspection

DQ Tracker

The data quality process was applied for the first 10 data partners in the DARWIN EU® network. The QA documents were evaluated, and data quality issues were identified for each data partner, which was given as feedback together with suggested remediations. The issues were recorded in a tracker for follow-up.

Darwin Portal

For the onboarding data partners an entry is created in the Darwin portal, containing information about the processes at the DP site as well as extracted metadata. The metadata is visualized in graphs and is used both for data quality checks and study feasibility analysis.