Data Network Feasibility Tool – Software Demonstration

Frank DeFalco, Clair Blacketer

Background

In the initial stages of study design, it is often challenging to know which data sources from a data network would be useful to include based on a study's requirements. Phenotype development and specific inclusion and exclusion criteria may not be finalized at study inception which prevents a full evaluation from being conducted. A tool for evaluating a network of data sources along a wide range of potential, high level requirements without requiring access to person level data would provide a valuable capability for initial study feasibility assessment.

Methods

The Achilles and Data Quality Dashboard packages provide an existing set of tools for characterizing the content and quality of observational data sources. Using the statistics generated by these packages, a new report was built within the ARES framework to evaluate a set of criteria to provide a feasibility assessment for each data source available in the network. This approach allows for an evaluation to be conducted using summary statistics without requiring access to the full person level data in the data source. The domain requirements criterion (Figure 1) allows the user to understand the proportion of people in a data source that have content across any of the specified domains in the data source. The desired domain criterion (Figure 2) allows the user to specify domains that must have content in the data source. The range requirements criterion (Figure 3) allows the user to specify the age range of people of interest for the study, the range of years required, and minimum cumulative observation time. The visit type requirements criterion (Figure 4) allows the user to select the necessary types of visits that will be required for the study.
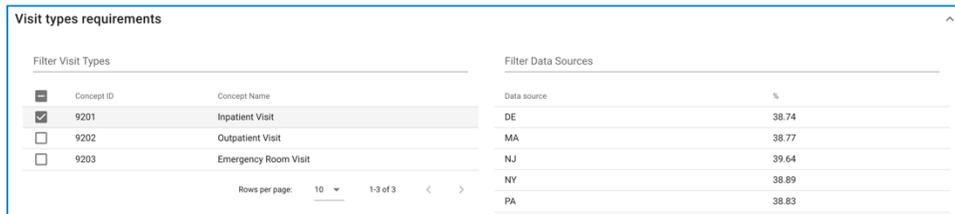

Figure 1

Figure 2

Figure 3

Figure 4

Results

A data source feasibility overview (Figure 5) is generated by combining the results from each specified criterion. An estimated population available for study is calculated from the criteria evaluation for each data source across the available network.

Data Source Feasibility Overview

Data Source	Source population	% Required Domains	% Desired Domains	% Cumulative Observation	% Population by age	% Observed	% Visit types	↓ Estimated population
MA	227	97.00	Present	35.24	100.00	16.93	38.77	5.09
DE	191	96.00	Present	31.41	100.00	15.88	38.74	3.54
PA	206	95.00	Present	28.16	100.00	16.53	38.83	3.54
NY	198	96.00	Present	28.79	100.00	16.48	38.89	3.51
NJ	169	96.00	Present	26.04	100.00	15.33	39.64	2.57

Figure 5

Conclusion

The data source feasibility tool provides a novel capability to evaluate initial study requirements across a network of standardized observational data sources through a simple web interface, without access to person level data. The resulting overview provides a view of potential populations for study as well as criterion specific results for additional insights.

References

1. Achilles (https://github.com/ohdsi/achilles)
2. Data Quality Dashboard (https://github.com/ohdsi/dataqualitydashboard)