Characterizing the OHDSI Evidence Network – A Global Snapshot of Real-World Data Partners

Clair Blacketer[1,2,4], Evanette Burrows[1,4], Ben Gerber[1,5], Vojtech Huser[1,6], Paul Nagy[1,7]

1. OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY, USA
2. Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, NL
3. Department of Biomedical Informatics, Columbia University, New York, NY, USA
4. Johnson & Johnson, Raritan, NJ, USA
5. Department of Population and Quantitative Health Sciences, UMass Chan Medical School, Boston, MA, USA
6. EPAM Odysseus, Newtown, PA, USA
7. Johns Hopkins University, Baltimore, MD, USA

Background

Generating reliable and generalizable real-world evidence (RWE) requires access to diverse, high-quality data from multiple settings. Federated research networks enable such collaboration by allowing data holders to maintain control of their data while contributing to large-scale studies. Examples in the US include PCORnet and the Sentinel Initiative, which have both demonstrated the potential of coordinated infrastructure for national-scale evidence generation [1,2]. In Europe, initiatives like EHDEN and DARWIN EU have further expanded the reach of federated networks built on the OMOP Common Data Model (CDM) [3,4].

In 2024, the OHDSI community launched the OHDSI Evidence Network to organize and coordinate federated research across its global community of OMOP CDM data holders. The initiative leverages OHDSI's open-source tools and opt-in participation model to build a global network of data partners that can support high-quality, efficient evidence generation. Understanding the scope, diversity, and structure of the participating data sources is essential for evaluating the network’s capacity to support a broad research agenda.

Objective

To characterize the composition of the OHDSI Evidence Network after one year of operation, including data source types, care settings, population coverage, and domain completeness.

Methods

Upon joining the network, data partners produced a DbProfile for each database by running the Database Diagnostics R package, an open-source tool that generates standardized summary statistics across key data quality domains. Metrics include counts of persons by age, sex, and race, coverage of essential data domains (e.g., drug exposures, measurements, procedures, and conditions), observation time, and record density. The full list of statistics is available at https://ohdsi.github.io/DbDiagnostics/articles/SummaryStatistics.html. Partners submitted these results to the OHDSI Coordinating Center, for aggregation and use in study feasibility assessments.

Databases were characterized by:

- **Source type**: Databases were categorized as electronic health records (EHRs), administrative claims, or other, as identified by the data partner organization.
- **Visit setting**: Visit concepts represented in each database were mapped to their highest-level ancestor in the OMOP standard vocabulary. Then, the number of persons per ancestor concept was compared across databases.
- **Population size:** Summary statistics describing the population size were produced.
- **Domain completeness**: proportion of patients with available records in key OMOP domains.

Results

As of April 2025, the OHDSI Evidence Network included 44 databases from 26 data partners across North America, South America, Europe, and Asia. 75% of the databases are EHRs, 20% administrative claims, and 5% other, like clinical registries.

Figure 1 shows the representation of visit ancestor concept ids across the databases in the network, with the colors indicating the proportion of persons with at least one record of a visit that rolls up to the ancestor. Almost all databases (97.7%) have outpatient visits and 84% have inpatient visits. Of the 30 databases that report both separate emergency room visits and inpatient visits, only about half (16) utilize the specific “Emergency Room and Inpatient Visit” concept id.

Databases range in size from 2.6k persons to 829M persons. The average person count is 53M and the median is 5.3M. The aggregated population of the network is 52% female.

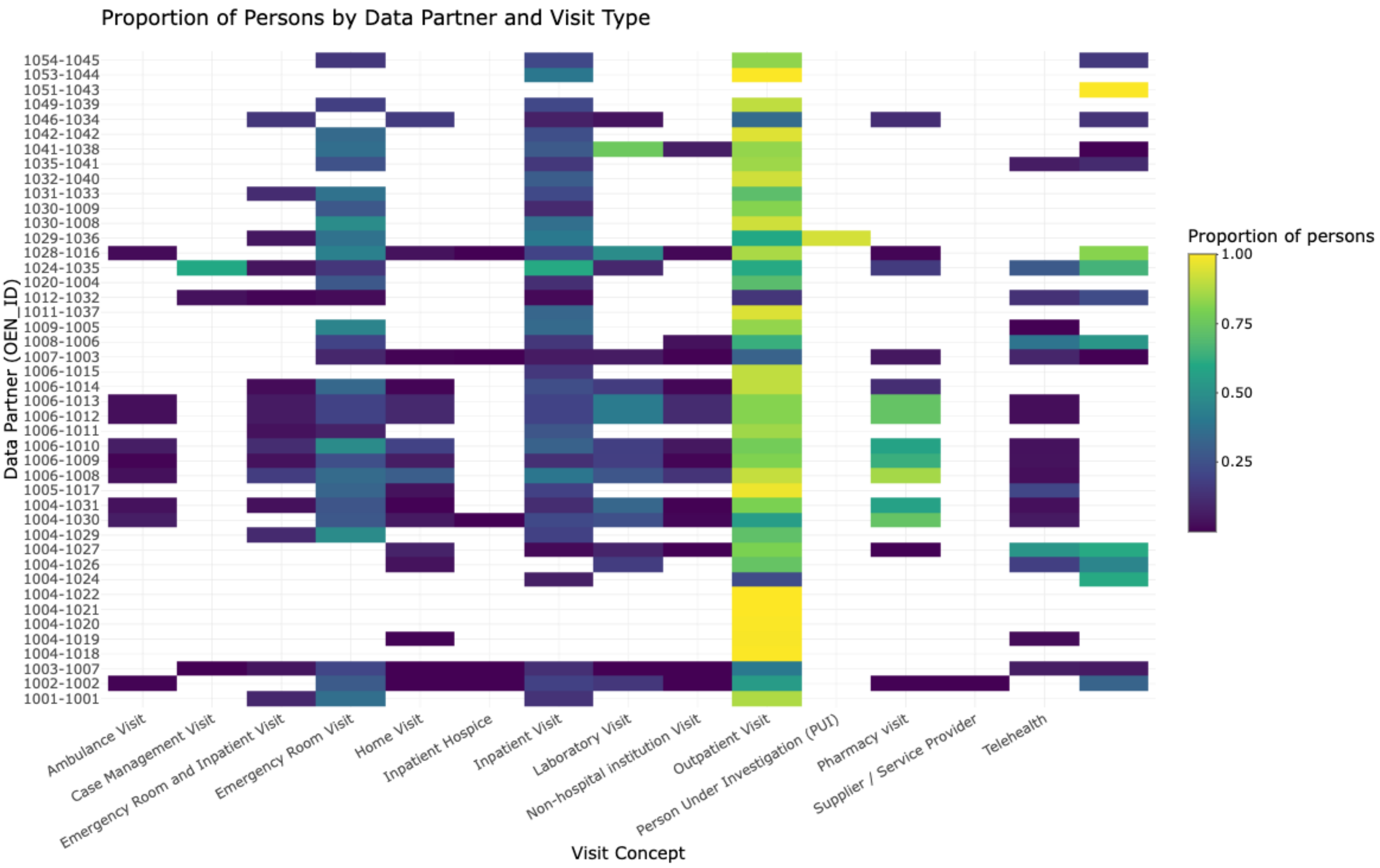

Figure 1: The proportion of persons with at least one visit, by data partner and visit ancestor concept id across the OHDSI Evidence Network.

Figure 2 shows a heat map with the proportion of persons by database with a record in each OMOP CDM domain. Both the condition and observation domain are the most populated, with all but one database showing records in each. The device domain is the least populated by database while the death domain has the lowest proportion of persons. There is one database that shows only records in the observation table. Seven datasets have no death data.

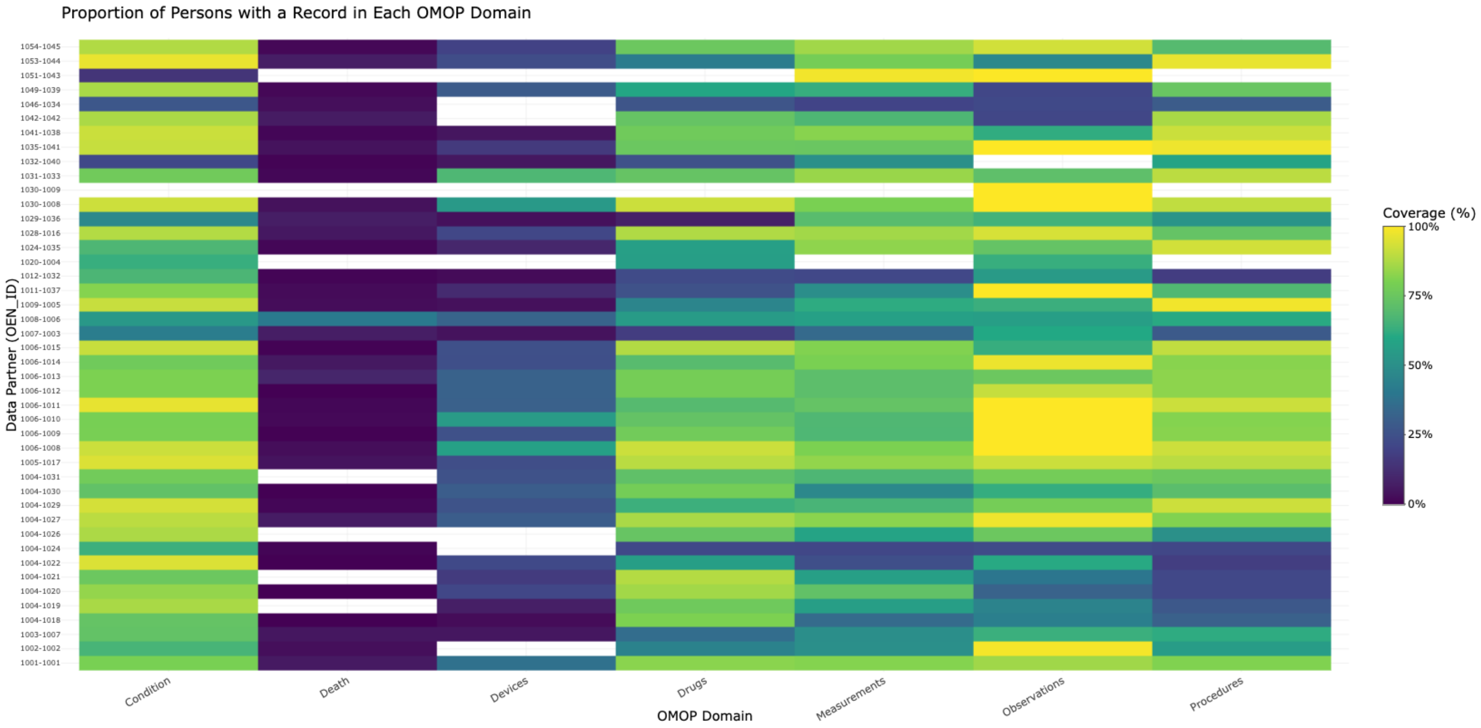

Figure 2: The proportion of records in each OMOP domain by database in the OHDSI Evidence Network.

Discussion

The OHDSI Evidence Network has established a diverse, global set of real-world data sources capable of supporting a wide range of observational studies. With hundreds of millions of patient records, varied visit settings, and comprehensive domain coverage, the network is well-positioned to address international research questions across therapeutic areas. The heterogeneity of the databases in the network underscores the need to utilize more than one source to generate evidence per study question as only one source will only tell part of the story.

Unlike proprietary networks or those bound by formal contracts, the OHDSI Evidence Network relies on open-source tools and opt-in participation, allowing broader inclusion while preserving local autonomy. Continued expansion of the network will further enhance feasibility assessment, accelerate study design, and support scalable real-world evidence generation.

References

1. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc. 2014;21(4):578-582.
2. Brown JS, Maro JC, Nguyen M, et al. The SENTINEL initiative: a national strategy for monitoring medical product safety. Pharmacoepidemiol Drug Saf. 2009;18(1):17-23.
3. European Health Data & Evidence Network (EHDEN). EHDEN Consortium Reports and Deliverables. https://www.ehden.eu (accessed 2025).
4. European Medicines Agency. DARWIN EU™: The Data Analysis and Real World Interrogation Network. https://www.ema.europa.eu (accessed 2025).