

The Wizards of OMOP: Shortening time-to-first insight on OMOP data through guided queries and dashboards in Data2Evidence

Andrew Deonarine¹, Tim Walz^{2,3}, Jerome Ng³, Alicia Koh³, Karthik Seetharaman³, Peter Hoffman³, Pablo Guerrero³, Mario Lawes², Alexia Sampri², Patricia Kovatch¹, Ben Illigens^{2,3,4,5}, Girish Nadkarni¹

¹Mount Sinai Health System, New York City, USA, ²D4L data4life gGmbH, Potsdam, Germany;

³Data4Life Asia Ltd., Singapore; ⁴Hasso Plattner Institut, Potsdam, Germany; ⁵Dresden International University, Dresden, Germany

Background

The OMOP Common Data Model and OHDSI tool stack have standardized observational health analytics across a large federated network, enabling reproducible studies at scale.^{1,2} Yet, even routine descriptive work like incidence, prevalence, mortality, or cohort characterization demands a working knowledge of OHDSI terminology, cohort definitions, concept sets, and an analytic interface.³ This can represent a significant barrier to engaging with OMOP data for clinical researchers, trainees, and operational stakeholders who often hold the most pressing questions but may not yet have the technical background to access it independently. The practical cost is a prolonged time-to-first insight: a simple descriptive question can take a trained analyst a working session, and an untrained clinician far longer, if it is possible at all. Platforms such as the Artificial Intelligence-Ready Mount Sinai (AIR·MS) environment have demonstrated that OMOP data covering millions of patients can be served interactively⁴, but speed of computation does not by itself lower the conceptual barrier to entry. We describe Wizards, a feature of the open-source Data2Evidence⁵ platform that packages common epidemiological measures as guided, fully customizable queries feeding interactive dashboards, with two explicit objectives: (a) to shorten time-to-first insight, and (b) to open OMOP datasets to non-expert users and to operational questions beyond traditional research use. We recognize that opening OMOP data to users without formal epidemiological training carries a risk of misinterpretation; Wizards are intended to augment, not replace, methodological guidance, and surface the underlying cohort logic and data assumptions to counter this.

Methods

A Wizard is a pre-configured “analytic recipe” defined declaratively using JSON in the cohort builder configuration of Data2Evidence, which the backend treats as a flexible object rather than a fixed schema. The user interface (UI) interprets fields such as wizard identifiers, labels, input and dashboard definitions, so new wizards or measures are added through configuration alone, without requiring code changes, making the catalogue fully customizable across datasets and

research sites. Wizards complement ATLAS, HADES, and Strategus: where those tools standardize the execution and sharing of a fully specified cohort and analysis, Wizards standardize how a non-expert first specifies one, emitting standard OHDSI objects that feed directly into that stack.

In the user flow, a simplified Wizard interface collects a few inputs (e.g., a condition and a year range) and constructs a deep link carrying the dataset ID, a compressed cohort definition, and the wizard-only context. Wizards are selected from an initial screen presenting summaries of each wizard and the approach used for calculating a particular measure or outcome, so a non-expert can choose the appropriate wizard before entering any parameters. Data2Evidence's data exploration tool decompresses the query into a virtual linked cohort, maps wizard fields to inclusion criteria, and, when required, materializes the cohort into the OMOP results schema via a streaming, batch-insert pipeline. Calculations for measures such as length of stay and readmission rate are performed using common assumptions and conventions adopted by healthcare organizations such as the AHRQ (Agency for Healthcare Quality and Research).

Dashboards are delivered as static Shinylive applications, where the Python Shiny code is exported as assets that run entirely client-side through Pyodide (CPython compiled to WebAssembly), served by the portal gateway. These static assets contain only the UI, runtime, and visualization logic, not the analytic data. At launch, the parent cohort builder passes runtime context (auth token, dataset ID, cohort ID, wizard config, query) into the dashboard iframe via `postMessage` after origin validation; the Pyodide app then issues authenticated API calls to Data2Evidence services to retrieve cohort-specific data. This deliberately separates dashboard delivery from dashboard data, reducing per-view server hosting burden while keeping every view live and access-controlled.

Results

In a worked example on AIR-MS, a user without SQL or cohort-definition expertise can select the wizard of choice, enter the conditions and time frame of interest, and obtain a live dashboard in a handful of clicks (Figure 1). The depicted dashboard shows cohort size by condition-occurrence year (rising to >105,000 annual patients by 2024), an age-gender population pyramid, and race and ethnicity distribution. In this deployment, time-to-first insight was observed at around 2 minutes as opposed to creating a de-novo analysis notebook that would typically take ~ 30 minutes – 1 hour with the help of AI tools such as LLMs. The current catalogue covers common epidemiological measures (incidence, prevalence, mortality) alongside cross-sectional characterization, with readmission rate, length-of-stay, and loss-to-follow-up on the roadmap. The addition of a new wizard requires no code change, demonstrating full customizability without needing redeployment of a new software version.

Known limitations include edge cases in specific cohort/measure combinations, which are logged and iteratively corrected. We are in the process of gathering qualitative feedback from users on the usability of the wizards as well as their strengths and weaknesses.

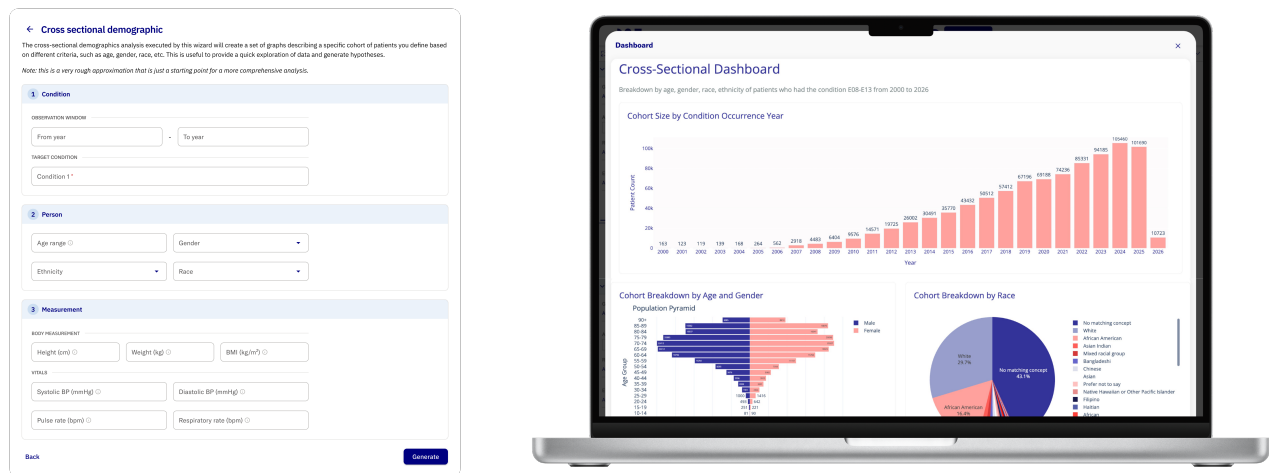


Figure 1. Cross section demographics wizard (left) and respective dashboard (right)

Conclusion

Wizards in Data2Evidence offer the OHDSI community a reusable, extensible, fully customizable pattern for lowering the entry barrier to OMOP analytics without abandoning methodological rigor. Guided queries collapse expert workflows into a few inputs, and static Shiny/Pydroid dashboards deliver live, cohort-specific results without a hosted analytics server per view. Deployed on AIR-MS, a production OMOP platform with more than 12 million patients, the approach demonstrates that standardized measures can be made accessible to clinical and operational users at scale. This extends the reach of OMOP data beyond the methods-expert audience while preserving reproducible cohort definitions and surfacing the assumptions needed to interpret them responsibly. We invite the community to contribute and share Wizard proposals.

References

1. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574–578.
2. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc*. 2012;19(1):54–60.

3. OHDSI. The book of OHDSI: observational health data sciences and informatics. 2021.
Available from: <https://ohdsi.github.io/TheBookOfOhdsi/>
4. Guerrero P, Ernebjerg M, Holst T, et al. The AIR·MS data platform for artificial intelligence in healthcare. JAMIA Open. 2025;8(6):ooaf145.
5. D4L data4life gGmbH. (2026). Data2Evidence [Computer software].
<https://github.com/OHDSI/Data2Evidence>