

Multiverse analysis for large-scale real-world evidence: A framework for transparent and robust observational research using OHDSI tools

Alexia Sampri^{1,2}, Mario Lawes^{1,2}, Ben Illigens^{1,2,3}

¹ Data4Life, Berlin, Germany; ² Hasso Plattner Institut, Potsdam, Germany; ³ Dresden International University, Dresden, Germany

Background

Real-world evidence (RWE) from observational health data is vulnerable to the "garden of forking paths", where defensible choices regarding, for example, cohort definition, covariate selection, and model specification can yield divergent results. This analytical flexibility challenges the reproducibility, interpretability, and credibility of observational findings. Multiverse analysis addresses this issue by systematically evaluating a range of plausible specifications and visualizing their impact on results (1,2). A central requirement is that all specifications must be methodologically sound and target the same underlying causal estimand, so that variation in results reflects uncertainty regarding analytical choices rather than heterogeneity in the research question. Despite growing use in the social sciences, multiverse analysis remains underused in RWE (for exceptions see 3,4). The OHDSI ecosystem, with its OMOP Common Data Model (CDM) and shared set of tools for standardized large-scale analytics, enables systematic exploration of specification spaces that would otherwise require bespoke re-engineering for each study.

Methods

We developed a structured framework for applying multiverse analysis to large-scale RWE studies building upon OHDSI tools and the OMOP CDM. Specifically, we use the Strategus R package to define and execute all analyses across the defined specification space. Results of the specifications are then extracted and rendered into specification-curve visualizations. The proposed pipeline will then be implemented within Data2Evidence (D2E), an open-source platform for scalable, transparent RWE generation (5,6). A key advantage of the framework is that it can be applied across any data partner in the OHDSI network without study-specific re-engineering.

Results

We demonstrate how to conduct multiverse analysis and interpret the resulting specification curve visualizations (see Figure 1 for an example). We further compare multiverse analysis to related approaches, including vibration of effects visualizations and the standard diagnostic and results plots used in large-scale OHDSI studies such as LEGEND-HTN. Finally, we discuss key challenges, including the selection of appropriate specifications, the interpretation of large specification spaces, the utility of multiverse analyses in federated settings, and the effective communication of results to clinical audiences.

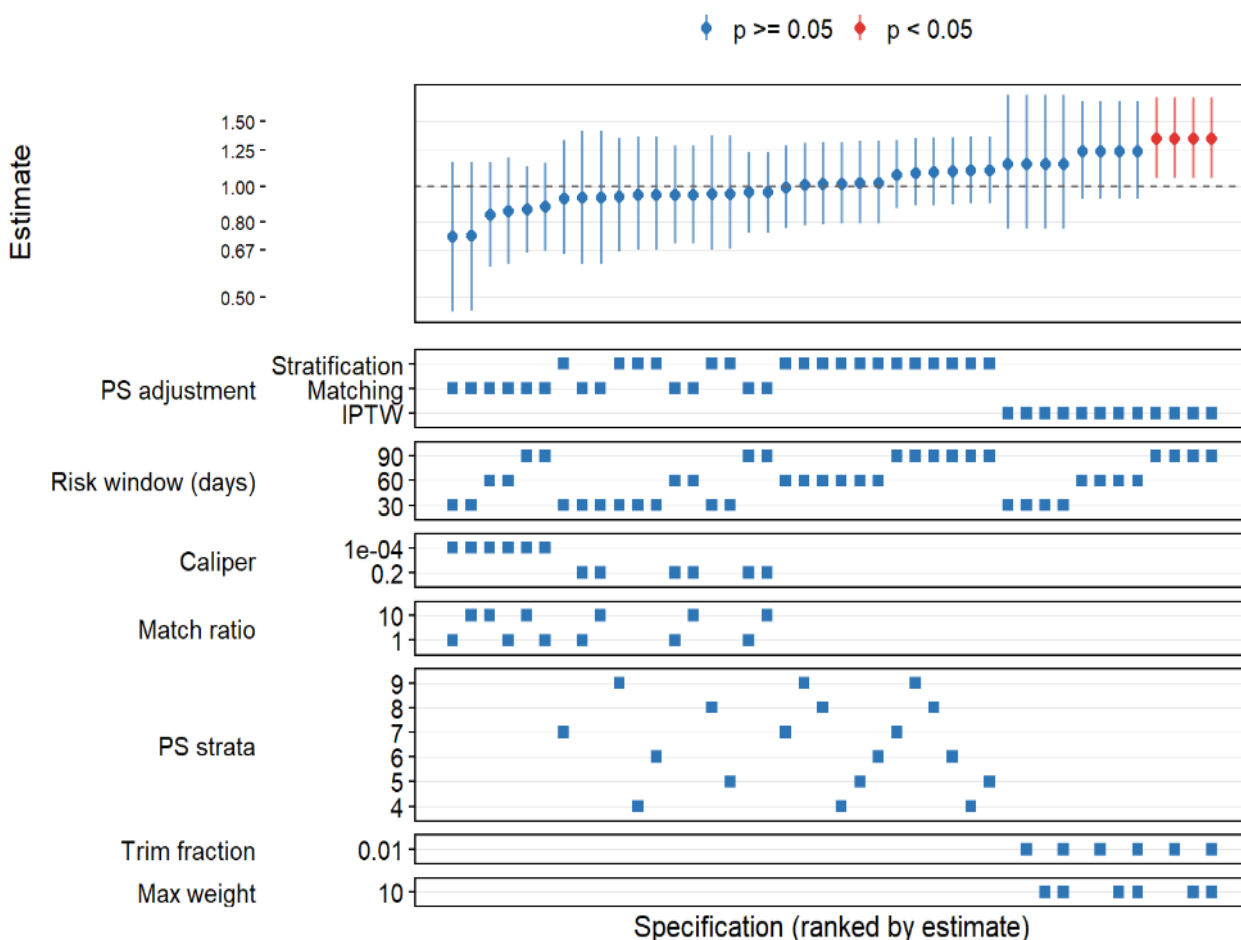


Figure 1. An exemplary specification curve summarizing the results of a multiverse analysis across 42 analytical specifications, with varying risk windows (30, 60, and 90 days) and propensity score methods (i.e., matching with varying caliper widths and matching ratios; weighting with varying trimming fractions and maximum weights; and stratification with a varying number of strata). Hazard ratios ranged from 0.73 to 1.34. In 38 of 42 specifications, the hazard ratio was not statistically significant. The four specifications yielding a significantly increased hazard ratio ($p < 0.05$) shared a common configuration: inverse probability of treatment weighting with a 90-day risk window. Trimming and truncation of weights had no impact on effect estimates across specifications.

The multiverse framework is openly available on our GitHub page:

www.github.com/d4l-data4life/multiverse-ohdsi

Conclusion

Multiverse analysis can increase transparency and robustness in RWE studies, but only when implemented with care. Its value lies in quantifying and visualizing the impact of different analytical specifications and prompting reflection on analytic choices -- not in replacing principled study design. The OMOP CDM and the OHDSI toolchain make such analyses particularly tractable at scale. For the OHDSI community, this

offers a practical path to stress-testing observational findings. Rather than reporting a single point estimate, researchers can present the full distribution of effects across defensible analytical choices, making the robustness of conclusions visible to reviewers, regulators, and clinical audiences. Ultimately, multiverse analysis transforms a key vulnerability of observational research, namely the dependence on analytical choices, into something visible, quantifiable, and open to scrutiny. That is precisely the kind of transparency the OHDSI mission depends on.

References

1. Simonsohn U, Simmons JP, Nelson LD. Specification curve analysis. *Nat Hum Behav.* 2020 Jul 27;4(11):1208–14. doi:10.1038/s41562-020-0912-z
2. Del Giudice M, Gangestad SW. A traveler’s guide to the multiverse: promises, pitfalls, and a framework for the evaluation of analytic decisions. *Adv Methods Pract Psychol Sci.* 2021 Jan;4(1):2515245920954925. doi:10.1177/2515245920954925
3. Levitt M, Zonta F, Ioannidis JPA. Excess death estimates from multiverse analysis in 2009–2021. *Eur J Epidemiol.* 2023 Nov;38(11):1129–39. doi:10.1007/s10654-023-00998-2
4. Patel CJ, Burford B, Ioannidis JPA. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *J Clin Epidemiol.* 2015 Sep;68(9):1046–58. doi:10.1016/j.jclinepi.2015.05.029
5. Data4Life. Data2Evidence [Internet]. GitHub; 2026. Available from: <https://github.com/OHDSI/Data2Evidence>
6. Guerrero P, Ernebjerg M, Holst T, Weese D, DiBello H, Ibing S, et al. The AIR·MS data platform for artificial intelligence in healthcare. *JAMIA Open.* 2025 Nov 3;8(6):ooaf145. doi:10.1093/jamiaopen/ooaf145