Toward test-driven development of OHDSI cohort definitions

William A Baumgartner Jr and Lisa M Schilling
Division of General Internal Medicine, Department of Medicine, University of Colorado Anschutz Medical Campus

Background

In retrospective observational research, patient cohort definitions are foundational. They determine which individuals are included in analyses, directly influencing study outcomes and the evidence base that informs clinical and policy decisions. (1) Despite their critical importance, cohort definitions are frequently developed through iterative refinement, with existing validation approaches—such as clinical adjudication and PheValuator(2–4)—focusing on evaluating completed definitions rather than providing real-time feedback during the development process itself.

This report advocates for enhancing cohort development workflows by treating cohort definitions as executable artifacts and applying principles from test-driven development (TDD), a well-established methodology in software engineering. In TDD, developers write tests before implementing code, ensuring that each component meets predefined expectations from the outset and can be confidently modified throughout development and beyond. (5) Different types of tests serve distinct roles throughout the development lifecycle. During early development, unit tests validate expected behavior and ensure that edge cases are handled appropriately. Once development stabilizes, these same tests function as regression checks, enabling confident updates and refactoring. In mature TDD workflows, tests are automatically triggered by changes in version control systems, reinforcing continuous quality assurance. By adopting this mindset in cohort development—writing validation tests that assert correct inclusion and exclusion logic as definitions are built—we can complement existing cohort validation strategies and enhance the robustness, maintainability, and reproducibility of the cohort development process. This approach aligns seamlessly with OHDSI's commitment to open science, transparency, and reusability. (6)

Methods

This report demonstrates TDD of OHDSI cohort definitions by building on existing tooling to show how TDD practices (Figure 1) can be readily implemented. Central to this methodology is the Darwin-EU TestGenerator R package (7), which already provides most of the necessary infrastructure: a fully-specified OMOP Common Data Model (CDM) instance with complete vocabulary support, along with tools for defining synthetic patient profiles and testing them against cohort definitions. Our contribution is a straightforward extension that enables TestGenerator to operate on individual synthetic patients, allowing us to extract detailed, case-level explanations for why a patient was included or excluded from a cohort. This enhancement turns each test case into a precise diagnostic tool, helping developers detect and resolve logic errors early in the cohort development process.

For this demonstration, cohorts are constructed using ATLAS (8) and exported into the ATLAS JSON format. Cohort construction, however, is not limited to ATLAS—programmatic approaches like Capr (9) or other methodologies could seamlessly be integrated into this workflow. The JSON cohort definition is then loaded into the CDM instance where the cohort is executed, and the CDMConnector (10) attrition machinery is used to evaluate whether synthetic patients are correctly included or excluded. This approach effectively treats each test patient as a unit test for the cohort logic.

This methodology enables iterative, test-first development of cohort definitions, ensuring that inclusion and exclusion criteria behave as intended from the outset. It also promotes reproducibility, transparency, and collaboration by embedding validation directly into the development process.

Results

We introduce the **cohortTdd** R package[1], a lightweight framework designed to bring TDD principles to OHDSI cohort development workflows. While this package introduces new functionality, its implementation is intentionally minimal—built on top of the robust capabilities already provided by the TestGenerator package. Our package simply extends this foundation to support patient-level testing and reporting. Using synthetic OMOP CDM patient records accompanied by expected outcomes for each case, cohortTdd automatically evaluates cohorts in the ATLAS JSON format against these scenarios and produces clear, actionable reports summarizing test results (Figure 2). The primary contribution of this work is not the code itself, but the demonstration of how TDD can be effectively applied to cohort development to improve accuracy, transparency, and reproducibility

To illustrate this approach, we present the **cohortTdd.demo** project[2], which applies TDD to a representative use case. This demonstration includes initial test patients that were used to develop the representative cohort and a continuous integration (CI) workflow that uses these test cases as regression tests, automatically executing them with each committed change to the cohort definition using GitHub Actions. This approach ensures that future modifications do not unintentionally alter cohort behavior.

By introducing a feedback loop during cohort development, test-driven cohort design helps catch logic errors early—errors that might otherwise go unnoticed until post-hoc validation with clinical adjudication through manual chart review or tools like CohortDiagnostics (11) and PheValuator (3,4). Unlike these approaches, which are either resource-intensive or assess cohort performance after construction, the cohortTdd approach validates inclusion and exclusion logic prospectively, using known test cases at

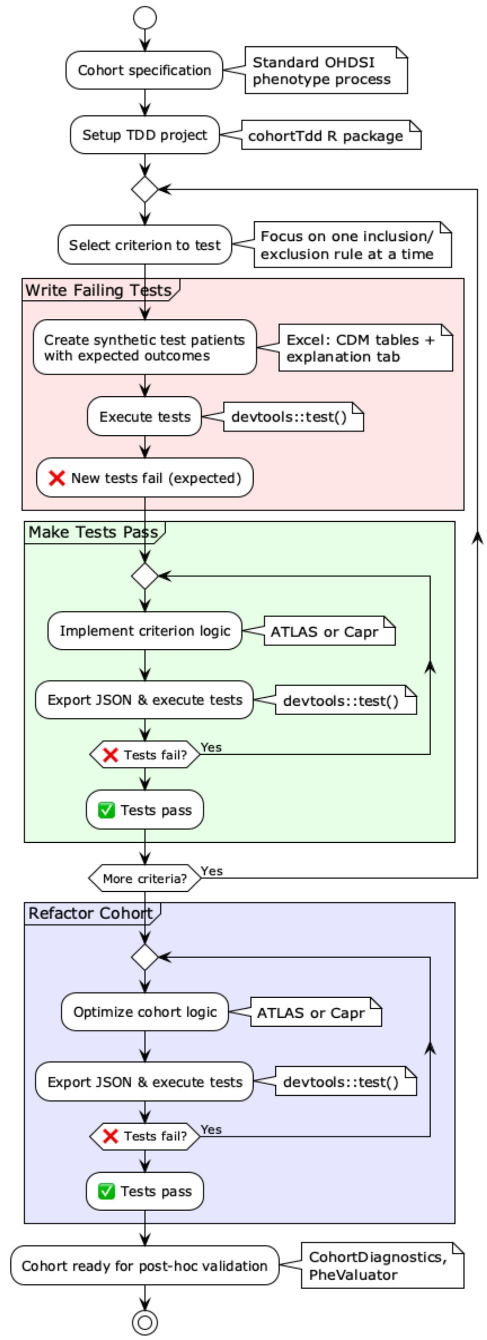

Figure 1. Test-Driven development process for OHDSI cohort definitions. The iterative workflow begins with creating synthetic test patients with known expected outcomes (red), followed by implementing or refining cohort logic until all tests pass (green). This cycle repeats for each inclusion and exclusion criterion, ensuring that cohort definitions are validated throughout development. Once stable, cohort definitions can be refactored to improve readability and maintainability while preserving functionality (blue).

[1] https://github.com/bill-baumgartner/cohortTdd.git
[2] https://github.com/bill-baumgartner/cohortTdd.demo.git

minimal cost. This approach has the potential to improve the reliability of cohort definitions while enhancing transparency: test cases serve as living documentation, clarifying the intent behind cohort logic and enabling others to more confidently reuse and adapt third-party definitions. In this way, TDD of cohort definitions complements and strengthens the existing OHDSI validation ecosystem.

```
======================================================================
                     COHORT VALIDATION SUMMARY
======================================================================
Total patients tested: 7
Tests passed: 6
Tests failed: 1
Success rate: 85.7%

** FAILED TESTS:
--------------------------------------------------
Patient 5: EXCLUDE - has <4 HCT after index date
  Error: Expected patient 5 to be included, but was excluded with reason 3


** DETAILED RESULTS:
--------------------------------------------------
[PASS] Patient 1: INCLUDE - Exact match
[PASS] Patient 2: EXCLUDE - on testosterone, but the 2 testosterone exposures
[PASS] Patient 3: EXCLUDE - due to estradiol exposure
[PASS] Patient 4: EXCLUDE - no HCT before index date
[FAIL] Patient 5: EXCLUDE - has <4 HCT after index date
[PASS] Patient 6: INCLUDE - Exact match
[PASS] Patient 7: INCLUDE - Exact match
```

Figure 2: Cohort Validation Summary Report. Automated test results from the cohortTdd package showing pass/fail status for each synthetic test patient, with detailed explanations for validation failures. The summary provides immediate feedback on cohort logic correctness, enabling rapid identification and correction of inclusion/exclusion criteria errors.

Conclusion

As observational health research increasingly informs real-world decision-making (12,13), it is essential to adopt development practices that ensure reproducibility, transparency, and maintainability. While OHDSI tools like CohortDiagnostics and PheValuator provide robust post-hoc validation, an opportunity exists to complement these approaches by integrating systematic validation during cohort development.

This report introduces a test-driven development approach that addresses this opportunity by embedding logic testing directly into the cohort creation process. TDD ensures that definitions are accurate from the outset and remain reliable over time. The cohortTdd framework demonstrates that these practices can be implemented with minimal changes to existing OHDSI workflows, offering substantial benefits in quality assurance. As studies grow in complexity and scale, integrating TDD represents a practical step toward more rigorous and trustworthy evidence generation.

#AugmentedByAI – The ideas and contributions presented in this report are the authors' own. However, portions of the text were revised and refined with the assistance of AI-based language tools.

References

1. Miroshnychenko A, Zeraatkar D, Phillips MR, Bakri SJ, Thabane L, Bhandari M, et al. Cohort studies investigating the effects of exposures: key principles that impact the credibility of the results. Eye. 2022;36(5):905–6.

2. OHDSI. The Book of OHDSI [Internet]. 2021. Available from: https://ohdsi.github.io/TheBookOfOhdsi/

3. Swerdel JN, Hripcsak G, Patrick B. Ryan, Patrick B Ryan, Ryan PB. PheValuator: Development and evaluation of a phenotype algorithm evaluator. J Biomed Inform. 2019 Sep 1;97:103258–103258.

4. Swerdel JN, Schuemie M, Murray G, Ryan PB. PheValuator 2.0: Methodological improvements for the PheValuator approach to semi-automated phenotype algorithm evaluation. J Biomed Inform. 2022;135:104177.

5. Beck K. Test driven development: By example. Addison-Wesley Professional; 2022.

6. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. In: MEDINFO 2015: eHealth-enabled Health. IOS Press; 2015. p. 574–8.

7. TestGenerator [Internet]. Available from: https://github.com/darwin-eu/TestGenerator

8. ATLAS [Internet]. Available from: https://github.com/OHDSI/Atlas

9. Capr [Internet]. Available from: https://github.com/OHDSI/Capr

10. CDMConnector [Internet]. Available from: https://github.com/cran/CDMConnector

11. Rao GA, Shoaibi A, Makadia R, Hardin J, Swerdel J, Weaver J, et al. CohortDiagnostics: phenotype evaluation across a network of observational data sources using population-level characterization. PloS One. 2025;20(1):e0310634.

12. Alipour-Haris G, Liu X, Acha V, Winterstein AG, Burcu M. Real-world evidence to support regulatory submissions: a landscape review and assessment of use cases. Clin Transl Sci. 2024;17(8):e13903.

13. Burns L, Le Roux N, Kalesnik-Orszulak R, Christian J, Hukkelhoven M, Rockhold F, et al. Real-world evidence for regulatory decision-making: guidance from around the world. Clin Ther. 2022;44(3):420–37.