

Using Artificial Intelligence to Generate a Quarto Front-End for Presenting ATLAS Treatment Pathways: Lessons from an Inflammatory Bowel Disease Epidemiology Study

William J. O'Brien^{1,2}, Robert Suruki¹, Jill Hardin^{1,2}

¹Global Epidemiology, Johnson & Johnson Research and Development, Raritan, NJ, USA; ²Observational Health Data Sciences and Informatics, New York, NY, USA;

Background

ATLAS, the OHDSI open-source analytics platform[1], enables fast computation of treatment pathway analyses across OMOP Common Data Model (CDM) databases [2,3]. A practical challenge in real world epidemiology studies is managing the combinatorial scale that arises when a protocol calls for multiple dimensions such as indications, phenotype variations, demographic subsets, sensitivity analyses, drug granularity levels, and source databases. Each dimension multiplies the number of distinct pathway outputs that must be retrieved, formatted, annotated, and assembled into a stakeholder-ready report. Populating an epidemiology report from this volume of ATLAS output manually represents a substantial and error-prone burden.

Large language model (LLM)-based AI coding assistants offer a means of accelerating this assembly step. However, the practical workflow, prompt engineering requirements, human oversight burden, and reproducibility implications of AI-assisted pathway report generation in a real OHDSI study have not been systematically described. This abstract reports lessons from a real-world application of an AI coding assistant to generate the Quarto front-end for an inflammatory bowel disease (IBD) treatment pathway study, with particular attention to managing large-scale pathway output presentation.

Methods

We conducted a descriptive study using 4 US claims databases mapped to the OMOP CDM. Target cohorts were defined in ATLAS for 2 indications (Crohn's disease [4,5] and ulcerative colitis [6,7]), 2 phenotype definitions per indication (a sensitive condition occurrence phenotype and a specific phenotype based on the Dawwas algorithm [8]), 3 study periods, 4 age strata, and 4 databases. Drug exposure events were defined at 2 levels: individual biologic drug ingredients and drug class/group level.

Pathways were defined and executed in ATLAS and retrieved programmatically using ROhdsiWebApi. Additional analysis for proportion of patients with any biologic utilization was computed with AI-assisted R code by querying the Atlas results database. The Quarto report integrating all outputs was generated with substantial assistance from a GitHub Copilot AI coding assistant (GPT-5 class model) in the Positron IDE. The AI was prompted iteratively to: (a) scaffold the Quarto document based on a prior project template, (b) write parameterized R helper functions for pathway table generation using ROhdsiWebApi and datatable, (c) author custom SQL for biologic drug exposure stratified by age and cohort, and (d) produce narrative summary text. All AI-generated code and prose were reviewed, tested, and committed to version control by the study team. A structured project AI instructions file in the repository encoded HADES-style connectivity patterns, preferred packages, cohort naming conventions, and required clinical sources to maintain consistent AI behavior across sessions.

Results

The combinatorial structure produced n=384 distinct pathway outputs requiring display in a single Quarto report rendered to html. Each output consisted of a pathway frequency table, and an age-stratified biologic utilization summary.

The copilot-assisted workflow successfully scaffolded the Quarto report and generated all core R helper functions for parameterized pathway retrieval and display. More than 70% of report lines were first drafted by AI prior to human review. Iterations requiring the most human correction involved: (1) schema and source-key filtering when joining results across multiple CDM schemas, (2) cohort name extraction from WebAPI response objects whose structure differed from the AI's training assumptions, and (3) SQL requiring dialect translation via SqlRender. Use of a persistent project instructions file substantially reduced the need for correction across AI sessions by encoding OHDSI-specific conventions the AI would otherwise not infer.

Conclusion

AI coding assistants can meaningfully accelerate generation of OHDSI treatment pathway reports in Quarto when studies face combinatorial complexity across multiple levels of indications, phenotypes, age strata, drug granularity levels, and databases. In this study, a copilot-assisted workflow produced the majority of a reproducible Quarto front-end integrating hundreds of distinct ATLAS pathway outputs, substantially reducing analyst effort for code scaffolding, helper function authoring, SQL writing, and narrative generation. Without AI assistance, generating and populating these sections manually would represent multiple days of effort for code writing alone. Key lessons were that: (1) a persistent, structured project instructions file is essential for consistent AI behavior across sessions; (2) human review remains critical for database connectivity details, platform-specific SQL dialect handling, and WebAPI response parsing; and (3) AI assistance did not compromise reproducibility -- all generated code was version-controlled and testable. These methods are broadly applicable to OHDSI studies where the volume of results would otherwise represent a practical reporting bottleneck. Future work could generalize our reporting framework to become a feature of ATLAS.

References

1. <https://github.com/OHDSI/Atlas>
2. 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. PMID: 26262116.
3. 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 Jan-Feb;19(1):54-60. doi: 10.1136/amiajnl-2011-000376. Epub 2011 Oct 28. PMID: 22037893; PMCID: PMC3240764.
4. Tian C, Ranasinghe IR, Hsu R. Crohn Disease. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026. PMID: 28613792. <https://www.ncbi.nlm.nih.gov/books/NBK436021/>

5. Lichtenstein GR, Loftus EV, Isaacs KL, Regueiro MD, Gerson LB, Sands BE. ACG Clinical Guideline: Management of Crohns Disease in Adults. Am J Gastroenterol. 2018;113(4):481-517. PMID: 29610508. DOI: 10.1038/ajg.2018.27

6. Lynch WD, Hsu R. Ulcerative Colitis. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026. <https://www.ncbi.nlm.nih.gov/books/NBK459282/>

7. Rubin DT, Ananthakrishnan AN, Siegel CA, Sauer BG, Long MD. ACG Clinical Guideline: Ulcerative Colitis in Adults. Am J Gastroenterol. 2019;114(3):384-413. PMID: 30840605. DOI: 10.14309/ajg.000000000000152

8. Dawwas GK, Weiss A, Constant BD, et al. Development and Validation of Claims-Based Definitions to Identify Incident and Prevalent Inflammatory Bowel Disease in Administrative Health Care Databases. Inflamm Bowel Dis. 2023;29(12):1993-1996. doi:10.1093/ibd/izad053.

Figure 1. Example of Treatment Pathways Table Reporting

On this page

- 1 Objective
- 2 Executive Summary
- 3 Methods
- 4 Results
 - 4.1 Treatment Pathways
 - 4.1.1 Listing of Computed Pathways
 - 4.1.2 Pathway 351
 - 4.1.2.1 Distinct Pathways
 - 4.1.2.2 Proportion of Subjects in a Pathway
 - 4.1.3 Pathway 352
 - 4.1.4 Pathway 363
 - 4.1.5 Pathway 364
 - 4.1.6 Pathway 355
 - 4.1.7 Pathway 365
 - 4.1.8 Pathway 362
 - 4.1.9 Pathway 358
- 4.2 Percent of Patients Who Used a Biologic at Any Step in a Pathway
- 5 Conclusions
- 6 Limitations
- 7 Appendix

4.1.2 Pathway 351 [↗](#)

4.1.2.1 Distinct Pathways

Show entries

Search:

Database: HealthVerity v3800

Target: [epi1351] [PL] Ulcerative colitis, age 2 to 11, during 2014 to 2017

Cumulative % column is applicable only when table is sorted by descending person count

	Step 1	Step 2	Step 3	Person Count	Row %	Cumulative %
1	[epi1351] Steroids rectal and oral, w 0d prior obsv, 1d gap			491	33.1	33.1
2	[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update			73	4.9	38.0
3	[epi1351] Steroids rectal and oral, w 0d prior obsv, 1d gap	[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update	[epi1351] Steroids rectal and oral, w 0d prior obsv, 1d gap,[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update	46	3.1	41.1
4	[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update	[epi1351] Steroids rectal and oral, w 0d prior obsv, 1d gap,[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update	[epi1351] Steroids rectal and oral, w 0d prior obsv, 1d gap	39	2.6	43.7
5	[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update	[epi1351] Steroids rectal and oral, w 0d prior obsv, 1d gap,[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update		34	2.3	46.0
6	[epi1351] Steroids rectal and oral, w 0d prior obsv, 1d gap	[epi1351] Steroids rectal and oral, w 0d prior obsv, 1d gap,[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update	[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update	31	2.1	48.1
7	[epi1351] Steroids rectal and oral, w 0d prior obsv, 1d gap	[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update		29	2.0	50.1
8	[epi1351] Anti-inflammatory w 0d prior obsv, 30d gap - update	[epi1351] Steroids rectal and oral, w 0d prior obsv, 1d gap		29	2.0	52.0