

Conference: OHDSI 2026 Global Symposium

Deadline: Friday, June 5 by 8pm ET

Length & Format: 200 words – 1000 words with sections: Background, Methods, Results, Conclusions.

Title: “Agentic CohortBuilder: Integrating LLMs with OHDSI/OMOP Standards for Cohort Construction”

Authors: Ram Varma, Ziwei Yin, Eleanor Davies, Juan Aristy, Konstantin Iaroshovets (EPAM), Andrej Snegiriov (EPAM), Gregory Klebanov (EPAM)

Background

Artificial intelligence (AI) has the potential to accelerate real-world evidence (RWE) generation; however, defining transparent, reproducible observational study cohorts remains a persistent bottleneck as well as an effort intense process. Translating narrative clinical protocols or analysis plans into structured cohort logic, standardized concept sets, and executable code is often manual, time-intensive, and difficult to trace. Although the OHDSI ecosystem provides mature tools such as ATLAS, CapR, and the OMOP Common Data Model (CDM), cohort construction still requires substantial expert effort for concept identification, logic translation, and quality control.

To address this challenge, Gilead Sciences, in collaboration with EPAM, is developing an agentic, OHDSI-compatible, human-in-the-loop workflow for cohort construction. The workflow is designed to use large language models (LLMs) to assist with generating structured, human-readable cohort specifications rather than directly producing SQL. By constraining LLM outputs to higher-level OHDSI-aligned formats, such as CapR specifications or validated cohort-definition ATLAS JSON, the approach aims to reduce hallucination risk, improve reviewability, and enable deterministic translation into executable cohort logic after human quality control.

Methods

We designed a three-phase, end-to-end agentic workflow with human-in-the-loop oversight to support reproducible cohort construction from a structured study analysis plan.

Phase 1 – Concept Set Generation and Standardization

Starting from a structured study analysis plan describing the target population, eligibility criteria, exposures, outcomes, covariates, and analysis requirements, an LLM-enabled Concept Set Generation AI Agent is designed to extract clinical intent and identify initial seed concepts. The AI Agent can then query biomedical literature and reference sources, such as PubMed and ClinicalTrials.gov, to recommend additional relevant concepts across OMOP domains including conditions, drugs, procedures, measurements, and observations, while retaining evidence supporting each recommendation.

An OMOP Standardization and Profiling AI Agent is then designed to map candidate concepts to standard OMOP vocabularies, including SNOMED CT, RxNorm, LOINC, and other domain-appropriate standard vocabularies. Where data access is available, the agent may also compute preliminary dataset-level prevalence and coverage metrics to support feasibility assessment and concept-set refinement. The planned output is a version-controlled, machine-readable, evidence-linked concept set that can be reviewed and finalized by clinical and methods experts before downstream cohort-definition development.

Phase 2 – Agentic Cohort Definition

The CohortBuilder AI Agent is designed to translate narrative eligibility criteria and finalized concept sets into a formal OHDSI-compatible cohort specification. The agent encodes inclusion and exclusion criteria, temporal constraints, observation windows, index-event logic, and cohort entry/exit rules into structured outputs such as cohort-definition JSON or validated OHDSI package-based R code, including CapR where appropriate.

To improve transparency and facilitate validation, the AI Agent is also designed to generate intermediate, criterion-level sub-cohorts. These sub-cohorts support stepwise attrition review, allowing analysts to assess how each eligibility criterion affects cohort size and whether the implemented logic aligns with the intended clinical and epidemiologic design.

Phase 3 – Quality Control and Human Review

A Cohort QC AI agent is designed to assess whether the generated cohort definition faithfully reflects the clinical intent documented in the analysis plan. The QC process includes deterministic checks of inclusion and exclusion criteria, temporal logic,

concept-set usage, index-event definitions, and cohort entry/exit rules. Discrepancies are flagged for review, and an embedded LLM can generate natural-language explanations and remediation suggestions to support efficient human quality assurance.

Final approval remains under human oversight, with domain experts and analysts reviewing concept sets, cohort logic, attrition outputs, and QC findings before implementing ~~on~~ in an OMOP CDM environment.

Results

We illustrate the intended workflow using a simulated use-case scenario. When fully implemented, Phase 1 is expected to produce traceable, evidence-linked OMOP concept sets enriched with supporting biomedical references and, where available, preliminary dataset-level prevalence metrics. Phase 2 is designed to generate an OHDSI-compatible cohort definition, along with labeled criterion-level sub-cohorts to support stepwise attrition assessment. Phase 3 is intended to produce a structured QC report evaluating whether the executable cohort logic accurately represents the narrative eligibility criteria from the analysis plan.

Collectively, these planned outputs demonstrate how an agentic workflow could support transparent, version-controlled, and auditable documentation for reproducible cohort design. The simulated scenario also highlights how constraining LLM-generated outputs to OHDSI-aligned specifications may make cohort logic easier to inspect, validate, and translate into executable code compared with workflows that rely on direct generation of SQL.

Conclusions

Gilead Sciences and EPAM designed an agent-driven workflow that integrates LLM-based assistance with OHDSI/OMOP standards to address key bottlenecks in observational study design. By supporting evidence-linked concept generation, structured cohort specification, criterion-level attrition review, and quality control within a governed human-in-the-loop architecture, this framework has the potential to improve the transparency, reproducibility, and efficiency of RWE analytics.

This work establishes a foundation for ongoing implementation and empirical benchmarking, including evaluation of time savings, accuracy, review burden, and concordance with expert-developed cohort definitions. Future work will explore integration with established OHDSI validation tools such as CohortDiagnostics and PHOEBE to support scalable, reliable, and auditable observational research.