Echo: Visual Steering for AI-Assisted Cohort Definition and Comparison in Electronic Health Records

Ruey-Ling Weng*, Vincent J. Zhang*, Lingfei Qian, Vipina K. Keloth, Yujia Zhou, Na Hong, Hua Xu, and Huan He

Department of Biomedical Informatics and Data Science, Yale School of Medicine

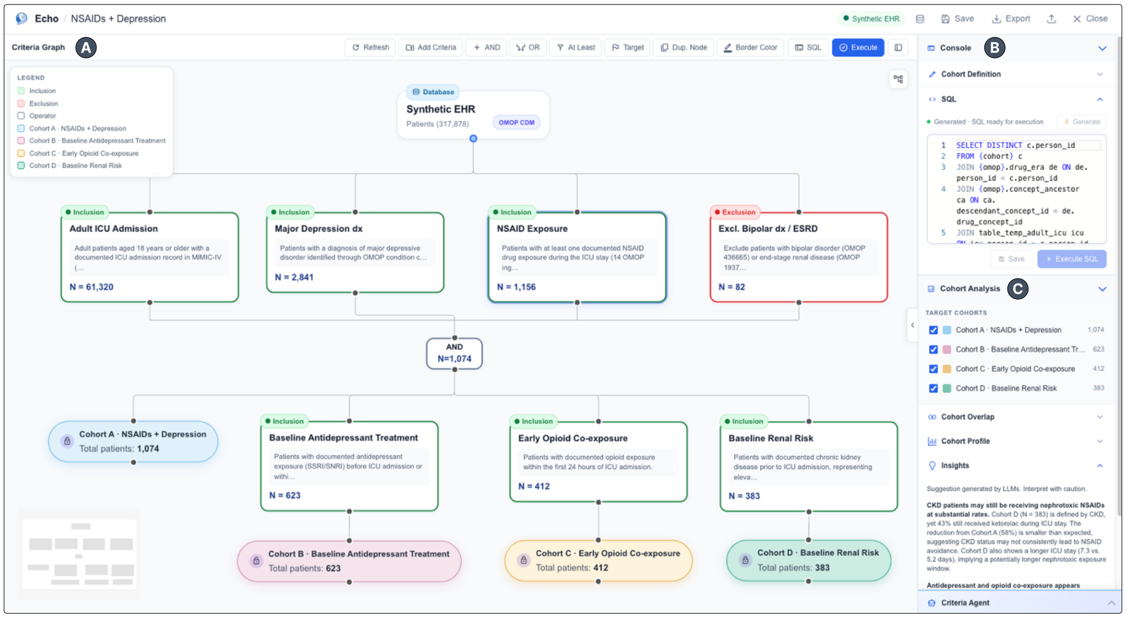

Figure 1: Echo combines three coordinated regions: (A) an editable Criteria Graph for inspecting shared criteria and branching alternatives, (B) a criterion-level SQL and Agent Console for reviewing AI-generated predicates and execution feedback, and (C) Cohort Analysis views for comparing patient-set overlap, cohort profiles, and statistics-grounded insights.

Abstract

Constructing clinical cohorts in Electronic Health Records (EHRs) requires translating clinical eligibility intent into executable database logic while preserving assumptions and downstream effects. Large Language Model (LLM)-based query generation can reduce SQL authoring barriers, but offers limited support for inspecting how criteria are grounded, refined, and compared across alternative cohort definitions. We present Echo, an early-stage visual analytics prototype for steering AI-assisted cohort construction in EHRs. Echo combines an editable Criteria Graph, criterion-level SQL review, and coordinated cohort comparison views. A case study with two domain experts suggests that Echo has the potential to support eligibility-logic inspection and sub-cohort derivation, while surfacing needs for codelist validation, value-based variables, temporal logic, and larger-scale evaluation.

Index terms: Visual analytics, electronic health records, cohort construction, text-to-SQL, human-AI interaction.

e-mail: {ruey-ling.weng, vincent.j.zhang, lingfei.qian, vipina.kuttichikeloth, yujia.zhou, na.hong, hua.xu, huan.he}@yale.edu

* These authors contributed equally to this work.

1 Introduction

Electronic Health Record (EHR)-based cohort construction is a central step in observational studies, retrospective analyses, and real-world evidence generation [1]. Clinical research teams must translate cohort inclusion and exclusion criteria into executable query logic, such as SQL queries, to retrieve patients from EHRs, which requires both clinical expertise and database knowledge.

LLM-based text-to-SQL systems can generate database queries from natural language, but cohort construction is rarely a one-shot task. Researchers need to revise thresholds, test alternatives, reuse shared eligibility paths, and compare how each change affects cohort size and composition. Whole-query generation can obscure criterion relationships, while independent criterion generation can lose the global context needed to preserve shared logic.

This work asks: *How can a visual analytics workspace help researchers intuitively inspect, refine, and compare cohort criteria?* Here, we present Echo, a visual steering workspace for criteria exploration, SQL review, and cohort comparison.

2 Related Work

Prior work supports cohort construction through form-based builders, natural-language eligibility parsing, text-to-SQL, and visual analytics for iterative cohort analysis [1–4].

Unlike prompt-only text-to-SQL systems, Echo does not treat AI output as the final answer. Its contribution is a visual steering loop that links clinical cohort criteria, inspectable execution logic, and downstream cohort comparison within one workspace.

3 ECHO DESIGN

Echo is designed for multidisciplinary clinical research teams: domain experts define study intent, informaticians inspect SQL predicates and concept mappings, and analysts compare alternative definitions. Echo exposes SQL and execution traces for collaborative validation rather than requiring every clinician to edit SQL directly.

Criteria Graph (Fig. 1A). The Criteria Graph represents eligibility logic as an editable directed graph. Criterion nodes encode inclusion/exclusion predicates and intermediate patient counts; operator nodes encode AND, OR, and At-Least-N; target nodes mark final cohorts. The graph makes shared criteria, branch points, and patient drop-off visible. Inclusion/exclusion states use labels and icons to avoid relying on color alone.

SQL and Agent Console (Fig. 1B). Selecting a criterion opens the SQL and Agent Console, which shows the natural-language criterion, AI-generated SQL predicate, execution status, and preview output. Edits update the affected predicate and propagate to downstream counts and comparison views, making the translation from criterion text to executable query traceable at the criterion level.

Cohort Analysis Views (Fig. 1C). Users select target cohorts in the right-side panel, and comparison views update in lockstep with graph edits. The Cohort Overlap (Fig. 2A) view uses an UpSet-style plot [5] to show shared and exclusive patient subsets. The Cohort Profile (Fig. 2B) view summarizes final N, eligibility paths, demographics, and comorbidity statistics. The Insight view generates short observations grounded in computed cohort statistics.

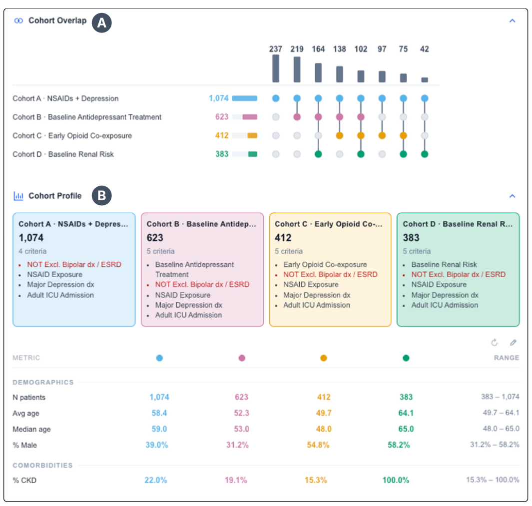

Figure 2: Echo's Cohort Analysis Views: (A) an UpSet-based patient-set overlap view and (B) cohort profile cards with comparative demographic and comorbidity metrics.

4 CASE STUDY

We conducted a case study with two domain experts: one faculty member and one data scientist with clinical-trial and cohort-study experience. Echo was configured using a preprocessed MIMIC-IV Intensive Care Unit (ICU) subset represented in Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) format [1], [6]. OMOP CDM standardizes EHR data into common tables and vocabularies.

Participants constructed and compared cohorts for a clinical question about adult ICU patients with major depression and Nonsteroidal Anti-inflammatory Drug (NSAID) exposure. As shown in Fig. 1A, the shared base cohort included adult ICU admission, major depression diagnosis, NSAID exposure during the ICU stay, and exclusions for bipolar disorder and end-stage renal disease. From this base, participants compared three sub-cohorts: baseline antidepressant treatment, early opioid co-exposure, and baseline renal risk / chronic kidney disease.

Three preliminary findings emerged. First, the Criteria Graph supported logic inspection by helping participants follow filter order, shared criteria, and divergence points. Second, visual forking supported low-friction exploration: participants derived sub-cohorts from a base definition, and one created additional exploratory sub-cohorts. Third, the assessment surfaced validation needs around concept mappings, codelists, temporal criteria, and value-based variables such as lab thresholds and units.

5 DISCUSSION AND LIMITATIONS

Echo aims to frame cohort construction in EHRs as an inspectable, iterative workflow rather than a fully automated query-generation step. Its main value lies in helping users externalize cohort logic, preserve shared context across alternatives, and compare how definitional changes reshape downstream patient populations.

Echo remains an early-stage prototype. Expert validation of codelists and value sets is still necessary when clinical conditions map to many codes or code groups. Value-based variables, including lab thresholds, units, and normalized ranges, require stronger interface support. Temporal logic, including index dates, follow-up windows, event ordering, and registration periods, is not yet fully modeled. Future work will add guidance for decomposing compound criteria, strengthen concept-grounding support, and evaluate usability, correctness, and decision impact in larger task-based studies.

6 CONCLUSION AND FUTURE WORK

By combining a graph-based workspace, text-to-SQL agents, and coordinated cohort comparison views, Echo aims to help teams inspect, refine, and compare alternative cohort criteria. As future work, we plan to conduct more case studies based on real-world clinical trial eligibility criteria, generating a broad spectrum of cohort definition graphs with varying complexity, which allow us to systematically evaluate the effectiveness of our design and its ability to support real-world cohort construction.

REFERENCES

[1] G. Hripcsak et al., "Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers," Studies in Health Technology and Informatics, pp. 574–578, 2015. doi: 10.3233/978-1-61499-564-7-574.

[2] C. Yuan et al., "Criteria2Query: a natural language interface to clinical databases for cohort definition," Journal of the American Medical Informatics Association, vol. 26, no. 4, pp. 294–305, 2019. doi: 10.1093/jamia/ocy178.

[3] J. Park et al., "Criteria2Query 3.0: Leveraging generative large language models for clinical trial eligibility query generation," Journal of Biomedical Informatics, vol. 154, p. 104649, 2024. doi: 10.1016/j.jbi.2024.104649.

[4] Z. Zhang, D. Gotz, and A. Perer, "Iterative cohort analysis and exploration," Information Visualization, vol. 14, no. 4, pp. 289–307, 2015. doi: 10.1177/1473871614526077.

[5] A. Lex et al., "UpSet: Visualization of Intersecting Sets," IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 1983–1992, 2014. doi: 10.1109/TVCG.2014.2346248.

[6] A. E. W. Johnson et al., "MIMIC-IV, a freely accessible electronic health record dataset," Scientific Data, vol. 10, no. 1, p. 1, 2023. doi: 10.1038/s41597-022-01899-x.