

OHDSI Study Agent for Human-AI Team-Based Real-World Evidence Generation in the Veterans Health Administration

Richard D. Boyce, PhD^{1,2}

¹University of Pittsburgh, Pittsburgh, PA, USA, ²VA Informatics and Computing Infrastructure (VINCI), Veterans Health Administration, Salt Lake City, UT, USA

Background

Clinical researchers using real-world data from the Veterans Health Administration (VHA) need tools that support high-quality, reproducible, and rigorous real-world evidence (RWE) generation. VHA is the largest integrated health care network in the United States, serving more than 9 million Veterans across over 1,300 care facilities¹, and its Corporate Data Warehouse (CDW) includes longitudinal patient data extending back to the 1990s. The scale and complexity of CDW-derived datasets make observational study design and execution time-intensive and expertise-dependent, creating opportunities for software that reduces manual burden while preserving scientific rigor. VA VINCI supports OHDSI tools with CDW data transformed to the OMOP common data model (CDW-OMOP).

RWE generation is a high-stakes workflow requiring linked judgments about data fitness, target trial specification, causal assumptions, confounding control, and interpretation under uncertainty.² Agentic AI systems that decompose tasks, retrieve supporting evidence, generate analysis artifacts, and synthesize results may reduce the burden of this work. However, agentic systems can also amplify invalid assumptions and produce plausible but flawed designs when outputs are accepted without careful interrogation. Evidence from multiple fields suggests risks of AI over-reliance, including improved short-term task performance but degraded vigilance and independent critical judgment.³⁻⁶ Recent evidence that scientific and professional accountability layers may fail⁷ raises concerns that uncritical adoption could contaminate the evidentiary substrate used for clinical, operational, and regulatory decisions.

Agentic AI is increasingly being applied to study design and execution, including a trend toward greater autonomy.^{8,9} In high-stakes evidence generation, autonomy should be paired with deliberate safeguards that keep human researchers in the lead and strengthen, rather than erode, their capacity to justify design choices and detect errors. OHDSI Study Agent has been created by an open collaboration between VINCI and the OHDSI collaborative to support human-AI teams by providing structured, checkable assistance across observational study design tasks while maintaining researcher oversight and accountability.

Flows/endpoints	What it does	Key inputs	Key outputs
phenotype_recommendation	Retrieves and ranks candidate phenotypes for a study intent	study_intent, top_k, max_results, candidate_limit	recommendations, candidate_count, candidate_limit
phenotype_make_computable	Converts a narrative or non-computable phenotype into a validated computable one	narrative_statement, confirmed_scope, concept_sets	
phenotype_improvements	Suggests targeted improvements to a selected phenotype definition	protocol_text, cohort, characterization_previous	phenotype_improvements, code_suggestion
concept_sets_review	Reviews a concept set and proposes diffs (adds/removes)	concept_set, study_intent	findings, patches, actions
cohort_critique_general_design	Flags design risks and proposes cohort patches	cohort	findings, patches, actions
phenotype_validation_review	Reviews a single case summary and returns label + rationale	keeper_row, disease_name	label, rationale
workflow_context_dialogue	Workflow-contextual question answering for /ohdsi	user_prompt + workflow context	answer, current_step_guidance, cautions, suggested_next_actions

Table 1. A sample of Study Agent capabilities as flows with stable endpoints and deterministic validation rules (excerpted).

Methods

Study Agent is an agent harness: it manages the context, tools, state, and execution required for agentic workflows and coordinates calls to external generative models (the system does not bundle a model). Design goals align with human–AI teaming functions (reasoning, memory, attention, and meta-coordination/governance) and emphasize calibrated reliance, role clarity, shared mental models, and task-appropriate support rather than autonomy-first workflows.¹⁰ In RWE generation, we target workflow stages where researchers incur high manual effort and design risk (problem framing, phenotype selection and refinement, concept-set review, cohort critique, diagnostics, and artifact generation).

Architecture: *Study Agent* is implemented as a service-based Python system with three separable layers: (1) an Agent Client Protocol (ACP) server that governs interaction policy (dialog state, confirmations, routing, and safety checks), (2) one or more Model Context Protocol (MCP) servers that expose tool contracts and prompt bundles via structured schemas, and (3) a deterministic core library that contains pure validation and transformation utilities. The separation is intentional: MCP supplies bounded retrieval and prompt assets, ACP orchestrates tool/model calls and enforces stage gates, and the core validates structured outputs and applies deterministic constraints. A key governance constraint is that row-level clinical data are not sent to the generative model; model inputs are restricted to sanitized summaries, metadata, and non-sensitive artifacts, with “fail-closed” behavior for PHI/PII sanitation when applicable.

Flow-centric coordination: Rather than using a monolithic “chat agent,” *Study Agent* organizes capabilities as named flows with stable HTTP endpoints (e.g., `/flows/phenotype_recommendation`) and a registry that specifies required inputs, expected outputs, MCP tool dependencies, and validation rules (see **Table 1**). This flow-centric approach enables (i) efficient and repeatable use of generative models only where needed, (ii) deterministic validation and constraints at each stage (e.g., restricting outputs to allowed identifiers; requiring user confirmation before writes), and (iii) auditable composition of services into end-to-end study workflows.

Workflow-contextual assistance via `/ohdsi`: *Study Agent* provides highly contextualized OHDSI study-design assistance through an interactive command (`/ohdsi`) embedded in workflow shells. The shell maintains step-aware context (workflow type, current step, active role, selected cohorts/artifacts, and constraints), flattens this into a stable request, and calls `/flows/workflow_context_dialogue`. ACP fetches the corresponding dialogue prompt bundle from MCP, queries the LLM with the user question plus structured workflow context, and returns a normalized response with step-specific guidance, cautions, and suggested next actions.

What the user might ask (examples)	Why generic chat is risky / unhelpful	What <code>/ohdsi</code> adds via workflow context
“Why are these phenotype recommendations weak here?”	Without knowing the current stage and available candidates, advice drifts into generalities	Answer is tied to workflow type, current step, and the already-selected cohorts/artifacts
“How should I think about time at risk in this incidence design?”	Time-at-risk guidance is stage-dependent; generic advice may conflict with current study setup.	Uses step-aware context (e.g., incidence TAR configuration step) and returns step-specific guidance and next actions
“What is the main risk in this CohortMethod analytic setting?”	Risk depends on what’s already specified and what constraints apply.	Context includes workflow type and current step; response includes cautions + suggested next actions
“How will this Keeper step affect downstream artifacts?”	Generic answers often ignore how specific pipeline steps change outputs	Answer is conditioned on the current step and active role; focuses on downstream artifacts available at that step
“Should I use a more restrictive adjustment set than defaults?”	Adjustment advice must reflect current study intent and what has already been chosen	Uses study intent + current workflow stage; returns structured cautions and next actions rather than a single “yes/no”

Table 2. `/ohdsi` is designed to answer workflow-aware OHDSI study-design questions in context, binding answers to workflow type, step, role, and selected artifacts.

Results

`Study Agent` is at an early functional prototype stage of approx. TRL 4–5 (lab validation with scripted demo workflows). It provides a functional ACP endpoint supporting an end-to-end observational clinical study workflow using R-based `OHDSI Strategus` via R packages that provide `Study Agent` client shells and integration with Atlas3/WebAPI.^{11–14} In the `Strategus` shell, users iteratively select target/outcome cohorts, request improvements, run `Keeper` review steps, and generate HADES-ready scripts and specifications. Atlas3/WebAPI integration is early-stage and currently provides `/ohdsi` support for concept set and cohort creation. Videos can be found on the project's github site that demonstrate generative AI assisted phenotype retrieval, creation, and improvement; case sample concept set creation and review; cohort definition critique; and study design code creation for HADES incidence rate analysis and estimation using the HADES `CohortMethod`.

Conclusions

No existing tool provides agentic AI support tailored to VHA's OMOP-CDW. `Study Agent` provides human-centered agentic support for the design and execution of RWE studies using `OHDSI Strategus` and CDW-OMOP. Key contributions are that it provides users with workflow-aware support for OHDSI study-design questions, and accomplishes safety-bounded generative AI interaction with deterministic validation. The tool is currently being evaluated to confirm that it helps VHA researchers more easily and reliably achieve strong and impactful clinical studies that will advance the health of veterans.

References

1. Internal. Veteran's Health Administration. <https://www.va.gov/health/>. Last Accessed 06/30/202
2. Hernandez RK, Critchlow CW, Dreyer N, Lash TL, Reynolds RF, Sørensen HT, et al. Advancing Principled Pharmacoepidemiologic Research to Support Regulatory and Healthcare Decision Making: The Era of Real-World Evidence. *Clin Pharmacol Ther.* 2025 Apr;117(4):927–37. doi:10.1002/cpt.3563 PubMed PMID: 39807817; PubMed Central PMCID: PMC11924150.
3. Bastani H, Bastani O, Sungu A, Ge H, Kabakcı Ö, Mariman R. Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics [SSRN Scholarly Paper] [Internet]. Rochester, NY: Social Science Research Network; 2024 [cited 2026 May 20]. Available from: <https://papers.ssrn.com/abstract=4895486> doi:10.1073/pnas.2422633122
4. Shen JH, Tamkibn A. How AI assistance impacts the formation of coding skills [Internet]. arXiv; [cited 2026 May 20]. Available from: <https://www.anthropic.com/research/AI-assistance-coding-skills> doi:10.48550/arXiv.2601.20245
5. Dell'Acqua F, McFowland III E, Mollick E, Lifshitz H, Kellogg KC, Rajendran S, et al. Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organ Sci.* 2026;37(2):403–23.
6. Budzyń K, Romańczyk M, Kitala D, Kołodziej P, Bugajski M, Adami HO, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol Hepatol.* 2025 Oct 1;10(10):896–903. doi:10.1016/S2468-1253(25)00133-5 PubMed PMID: 40816301.
7. Topaz M, Roguin N, Gupta P, Zhang Z, Peltonen LM. Fabricated citations: an audit across 2·5 million biomedical papers. *The Lancet.* 2026 May 9;407(10541):1779–81. doi:10.1016/S0140-6736(26)00603-3
8. Moeller-Grell N, Shenzhang S, Jiang ZJ, Dobson RJ, Chandrabalan VV. FastOMOP: A Foundational Architecture for Reliable Agentic Real-World Evidence Generation on OMOP CDM data [Internet]. 2026. Available from: <https://arxiv.org/abs/2604.24572>
9. Sapkota R, Roumeliotis KI, Karkee M. AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges. *Inf Fusion.* 2026 Feb 1;126:103599. doi:10.1016/j.inffus.2025.103599
10. Gonzalez C, Donahue K, Goldstein DG, Heidari H, Jalali MS, Schelble B, et al. Toward a science of human–AI teaming for decision making: A complementarity framework. *PNAS Nexus.* 2026;5(3).
11. OHDSI Generative AI Workgroup and VA VINCI. `Study Agent` GitHub repository. Accessed Oct 9, 2026. <https://github.com/OHDSI/StudyAgent>
12. OHDSI Generative AI Workgroup and VA VINCI. `SlashOhdsiStrategusAssistant` R package GitHub repository. Accessed Oct 9, 2026. <https://github.com/OHDSI/SlashOhdsiStrategusAssistant>
13. OHDSI Generative AI Workgroup and VA VINCI. `OHDSI/SlashOhdsiAcpClient` R package GitHub repository. Accessed Oct 9, 2026. <https://github.com/OHDSI/SlashOhdsiAcpClient>
14. OHDSI Generative AI Workgroup and VA VINCI. Demo of `Study Agent` and Atlas3/WebAPI GitHub repository. Accessed Oct 9, 2026. <https://github.com/OHDSI/AtlasWebAPISandbox/tree/main/demos>