ontada

Integrating Generative AI into Biostatistical Workflows for Trial Emulation: A Framework to Accelerate Real-World Evidence Generation

Saamir Pasha[1], Jessica K. Paulus[1], Zhaohui Su[1]
[1] Ontada, Boston, MA, USA

The US Oncology Network

MSR66

Background

- Trial emulation using real-world data (RWD) is increasingly used to inform healthcare decision-making but remains resource intensive due to high subject matter expertise requirements.
- Acceleration of generative AI (GenAI) development may support application to methodologically complex activities like trial emulation.
- **Objective: Apply GenAI to the execution of a trial emulation leveraging RWD to assess fidelity to the trial protocol, reproducibility, transparency and efficiency gains.**

Methods

Study Design

- **AI-assisted research environment** supporting data curation, QC, and analysis (**Figure 1**).
- **Trial protocols into structured prompts** guiding eligibility mapping, variable derivation, and outcome definitions.
- Applied framework to a metastatic breast cancer trial emulation using electronic health record (EHR) data.
- **AI robustness assessment:** Semantically equivalent prompt variants were issued to a single LLM to generate SQL for an attrition analysis, with consistency across variants used as the indicator of robustness (**Table 1**).
- **Validation:** Outputs were evaluated for logical correctness and cohort fidelity and underwent independent biostatistical subject-matter-expert (SME) review prior to analytic use.

Data Source

- Data sources from structured fields of the US Oncology Network's iKnowMed (iKM) EHR
- The US Oncology Network includes over 2,700 providers in over 600 sites of care across the US, and ~50 non-Network clinics have adopted the iKM EHR and participate in real-world research activities with Ontada

Statistical Analysis

- **Propensity-score matching:** estimated via logistic regression on 15 baseline. Patients were matched 1:1 using greedy nearest-neighbor matching without replacement (caliper = 0.2 × SD of the propensity score)
- **Covariate balance:** mean standardized difference (MSD)
- **Endpoints:** Overall Survival (OS), Time-to-next treatment (TTNT), Time-to-treatment discontinuation (TTD)
- **AI-Assisted Controls (Figure 1)**
  - Eligibility and outcome mapping aligned to target trial definitions
  - Request **chain of reasoning** to understand AI thought process
  - Output **progressive summaries** to review key results at each milestone during analysis
  - All AI-assisted outputs were reviewed with **SME–validated QC checklist** along with biostatistical experts to ensure methodological integrity.

Presented at ISPOR 2026, May 17-20, 2026, Philadelphia, PA

Results

Figure 1. AI-assisted trial emulation framework

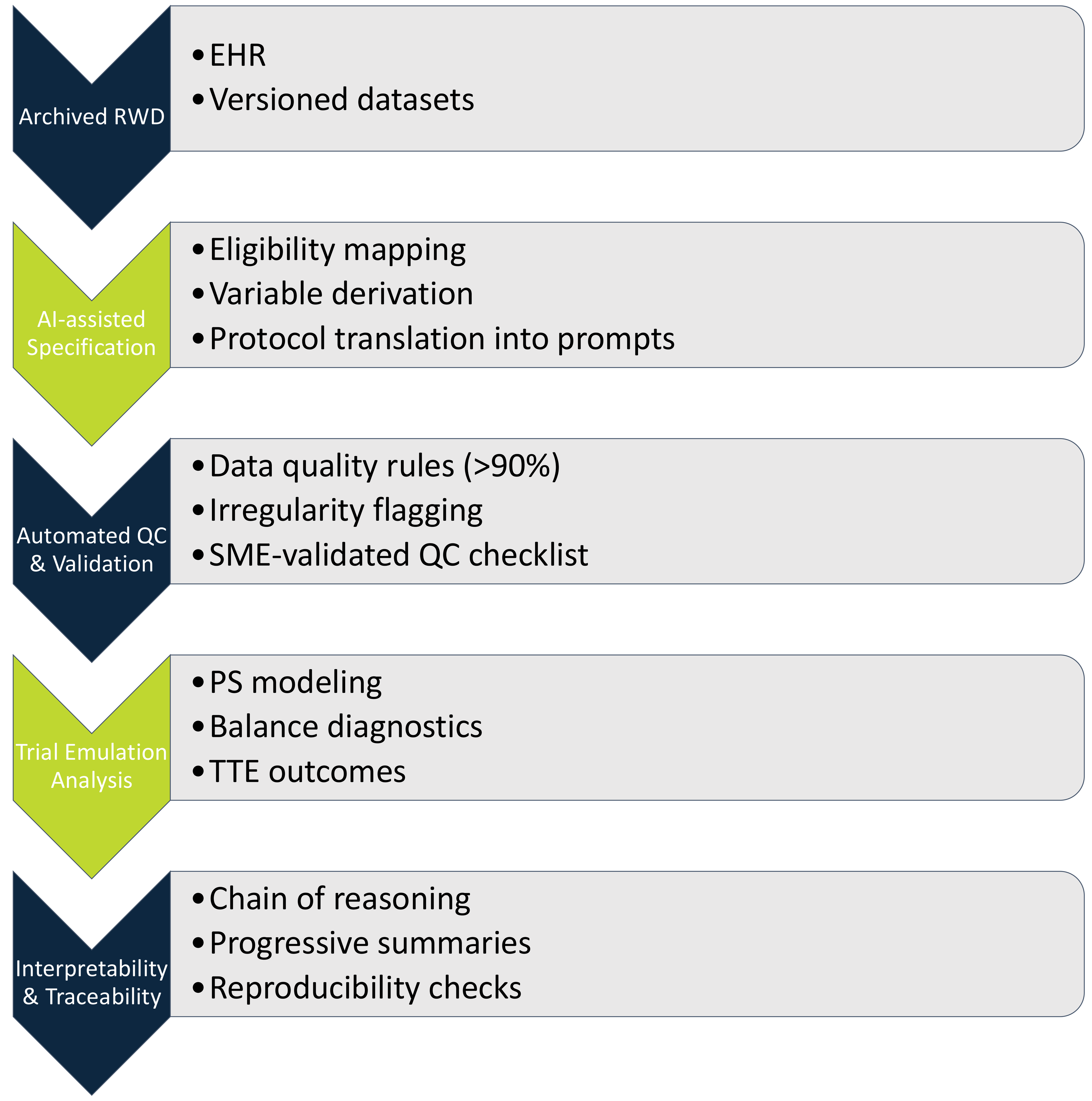
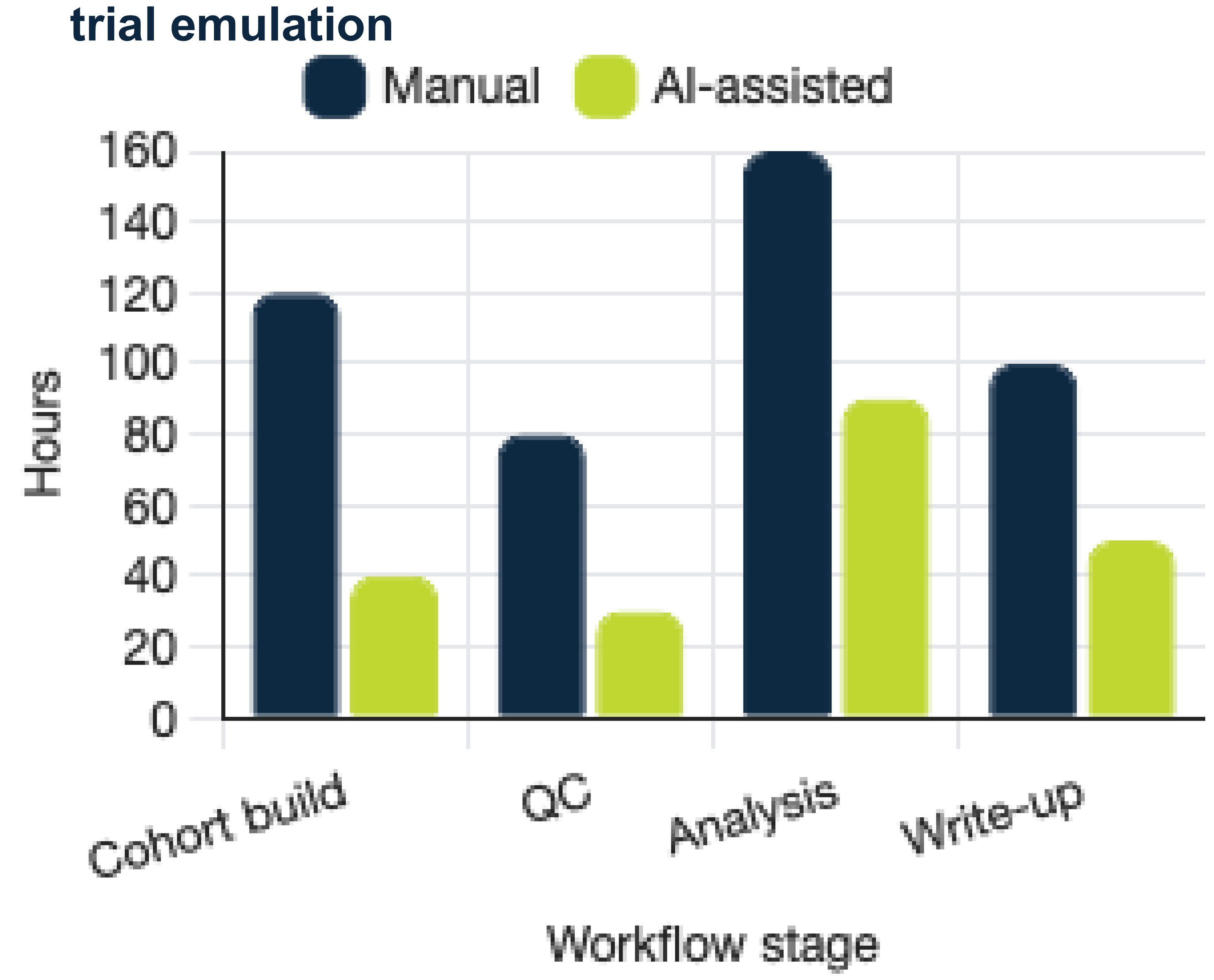


Figure 2. Workflow efficiency gains with AI-assisted trial emulation



Effort estimates shown for given use case; AI-driven efficiency may vary by study context

Table 1. LLM-generated SQL quality assessment for attrition analysis

Prompt variant	Run ID	Key SQL traits	SME Judgement	Quality Score
Original	run_1	DISTINCT patient count; date range 2014–2024; HER2+ filter	Correct	5
Modified 1	run_1	Uses FLOAT cast; diagnosis_date range logic	Correct	4
Modified 2	run_1	YEAR(diagnosis _date) filter; COUNT(*)	Correct	4
Modified 3	run_1	Explicit loss-to-follow-up framing; DISTINCT ID	Correct	5
Original	run_2	Replication of baseline logic; consistent filters	Correct	5

Discussion

AI assistance was most effective for well-specified rule-based analytic tasks (e.g., cohort attrition queries), highlighting the importance of clear protocol formalization when integrating LLMs into RWE workflows

Prompt-based robustness assessment surfaced minor implementation variability (e.g., date handling, aggregating logic) that would be difficult to detect without structured evaluation beyond single-prompt

Human SME review remained critical for contextual clinical and methodological judgement to ensure cohort fidelity and appropriate interpretation of EHR-derived variables.

Limitation

- This evaluation was conducted within a single platform and trial emulation use case; generalizability to other data sources, disease areas, and study designs requires further validation

Conclusions

- **Integrating generative AI into biostatistical trial emulation workflows** can substantially accelerate real-world evidence generation when implemented within a structured, auditable framework.
- **Methodological rigor and reproducibility are preserved** through explicit protocol translation, robustness checks across prompt variants, and mandatory human-in-the-loop validation.
- **The proposed framework is scalable and indication-agnostic**, supporting broader adoption of AI-assisted approaches across RWE studies while maintaining transparency and regulatory credibility.