DAILY ENGLISH READING · 17 SEPTEMBER 2026

Can an LLM audit whether a clinical trial quietly changed its primary outcome?

C1 · 30 minutes · LLM monitoring × endpoint lineage × SP change control

Why selected

A JAMA Network Open study published 16 September 2026 uses an LLM to monitor primary-outcome changes across 15,199 prospectively registered clinical trials. It is a strong SP reading because a technically correct program can still analyze the wrong scientific target if an endpoint change is not propagated from protocol and registry into SAP, ADaM and TFLs.

Reading order

0-3 min · Preview: addition, removal and time-frame modification.

3-14 min · Main article: validation sample, classifier uncertainty, estimates and limitations.

14-20 min · FDA M11: machine-readable structured protocol fields.

20-25 min · CDISC 360i: connected metadata from study design through analysis.

25-30 min · Output: design an endpoint change-control agent.

Brief background

Primary outcomes are prespecified before enrollment to create a transparent reference point. Amendments can be legitimate, but material changes should be visible, dated, approved and explained. The study compared the last prospective registry version before enrollment with the latest version and classified additions, removals and time-frame changes.

The authors validated candidate LLMs on 100 human-annotated trials and propagated measured classifier error into prevalence and regression estimates. This is a useful pattern for AI-assisted QC: model output should carry validation evidence and uncertainty rather than silently becoming ground truth.

For SP work, endpoint lineage should connect: approved protocol objective → endpoint definition → registry version → SAP analysis → ADaM PARAM/PARAMCD and visit/window → TFL shell → final result.

Key vocabulary

Term	中文	Meaning
prospective registration	前瞻性注册	Registering key trial information before participant enrollment begins.
prespecification	预先设定	Defining outcomes, analyses, or rules before seeing trial results.
primary outcome	主要结局	The principal outcome used to answer the trial's main research question.
outcome modification	结局修改	A post-registration addition, removal, or time-frame change to an outcome.
selective reporting	选择性报告	Reporting results based on favorability rather than the original prespecified plan.
version history	版本历史	The archived sequence of changes to a trial registration or controlled document.
manual curation	人工整理	Human review and labeling of records, often accurate but expensive at scale.
classifier	分类器	A model that assigns records to predefined categories.
annotated sample	标注样本	A set of examples labeled by humans and used to evaluate a model.
performance-adjusted estimate	性能校正估计	An estimate that incorporates uncertainty caused by imperfect classifier performance.
probabilistic sampling	概率抽样	Sampling that uses explicit probabilities, here to propagate classification uncertainty.
adjusted odds ratio	校正优势比	An association measure estimated while controlling for other variables.
protocol amendment	方案修订	A formally documented change to an approved clinical trial protocol.
machine-readable protocol	机器可读方案	A protocol represented in structured fields that software can exchange and compare.
change-control rule	变更控制规则	A rule defining which changes require documentation, review, approval, or escalation.

Useful phrases

1. prospective registration creates a reference point for later comparison.

2. an outcome change is not automatically misconduct, but it requires transparent justification.

3. version history can be treated as analyzable data rather than passive documentation.

4. the classifier should surface candidate changes for review.

5. model uncertainty should propagate into downstream estimates.

6. a structured protocol makes automated comparison easier.

7. the audit question is what changed, when it changed, and why.

8. the most important control is the link between the approved plan and the final analysis.

9. automation can scale monitoring without eliminating adjudication.

10. traceability turns a detected difference into an explainable change record.

Comprehension questions

• Why is the last prospective registration before enrollment a useful comparison point?

• What three types of primary-outcome change did the study classify?

• Why did the researchers propagate classifier uncertainty into their estimates?

• Why can technically correct ADaM/TFL programming still implement the wrong target?

• How can structured protocol metadata improve endpoint change control?

Retelling prompts

30 sec · LLM + registry version history = scalable trial audit.

45 sec · Legitimate amendment versus unexplained endpoint drift.

60 sec · Protocol objective → SAP → ADaM → TFL and where semantic drift can occur.

5-minute speaking/output task

Minute 1: define authoritative artifacts. Minutes 2-3: compare endpoint name, measure, time frame, analysis population, estimand, visit/window and statistical method. Minute 4: classify wording-only, approved material and unexplained material changes. Minute 5: explain what must be updated before final TFL delivery.

One sentence to keep

AI can make trial-version monitoring scalable, but the real control is end-to-end endpoint lineage: every material change should be dated, justified, approved, propagated, and visible in the final analysis.

Sources

https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2854124

https://www.fda.gov/regulatory-information/search-fda-guidance-documents/m11-clinical-electronic-structured-harmonised-protocol

https://www.cdisc.org/events/webinar/360i-phase-2-vision-implementation-lessons-learned-and-road-ahead