DAILY ENGLISH READING

Why a perfect-looking SAP can still produce the wrong analysis

15 September 2026 · C1 · LLM agents × statistical QC · 30 minutes

Why selected

A fresh open-access JMIR evaluation follows an LLM agent through a real clinical-data-analysis pipeline. The key result for SP work is that strong SAPs did not reliably predict correct code. Silent errors in cohort boundaries, formulas, censoring and interval structure could survive runtime execution and still yield plausible results.

Reading order

0–3 min: abstract and strengths/failures.
3–14 min: SAP-to-code fidelity, cohort boundaries, formulas, censoring, diagnostics.
14–20 min: ClinAgent architecture.
20–25 min: CDISC standards, traceability and human oversight.
25–30 min: design one mandatory QC gate.

Core lesson

A green log answers only whether the program executed. It does not prove the right population, time origin, censoring rule, formula, data structure or diagnostics were used. Trustworthy automation must validate each transition: SAP → code → structure → model → result → interpretation.

Open-access sources

JMIR main article: https://www.jmir.org/2026/1/e99597
ClinAgent: https://academic.oup.com/biomethods/article/11/1/bpag032/8706433
CDISC AI Innovation Challenge: https://www.cdisc.org/events/ai-innovation-challenge

Key vocabulary

Term	中文	Meaning
SAP-to-code fidelity	SAP 到代码的一致性	Code faithfully implements plan
silent failure	静默错误	Logical error without crash
cohort boundary	队列边界	Eligibility/filter logic
formula composition	模型公式构成	Terms actually in model
censoring definition	删失定义	Follow-up ending rule
counting-process structure	计数过程结构	Start-stop survival intervals
proportional hazards assumption	比例风险假设	Cox model assumption
multicollinearity	多重共线性	Correlated predictors
concordance	一致性指数	Survival discrimination
structural code review	结构性代码审阅	Review filters/formulas/rows
execution success	执行成功	Code runs without error
analytical correctness	分析正确性	Intended logic is implemented
diagnostic metacognition	诊断性元认知	Recognizing needed checks
verification time-tax	验证时间成本	Audit time offsetting speed

Term	中文	Meaning
reference benchmark	参考基准	Validated comparator