

DAILY ENGLISH READING

25 August 2026 - 30-minute active reading pack for Davin

TODAY'S QUESTION

Where should the LLM stop and the statistical method begin?

Difficulty: C1 | LLM-assisted statistics | causal mediation | clinical trials

WHY SELECTED

SurroPilot uses an LLM to reduce interaction cost around a difficult statistical workflow, but formal inference is performed by an established heterogeneous causal-mediation method. A programmer-inspector loop corrects generated R code, and variable selections are checked before execution - a useful boundary pattern for regulated statistical programming.

READING ORDER

0-3 min Preview: where does the LLM stop and the statistical method start?

3-14 min Main: workflow, programmer-inspector loop, ACTG175 and limitations.

14-20 min ISPOR: evidence levels, uncertainty and surrogate validity.

20-25 min FDA/EMA: context of use, risk, governance and lifecycle.

25-30 min Output: design an explicit AI/inference boundary for one SP analysis.

SOURCES

Main: SurroPilot - arXiv:2608.07635

<https://arxiv.org/abs/2608.07635>

Support: ISPOR surrogate-endpoint Good Practices - Value in Health 2026

<https://www.valueinhealthjournal.com/article/S1098-3015%2826%2900050-1/fulltext>

Support: FDA / EMA Guiding Principles of Good AI Practice in Drug Development

BRIEF BACKGROUND

Surrogate endpoints can shorten trials, but their validity may differ across subgroups.

SurroPilot uses an LLM for dataset understanding, preprocessing, variable selection and interpretation.

The inferential engine remains an established heterogeneous causal-mediation framework.

The ACTG175 demonstration shows a reproducible end-to-end workflow, not autonomous methodology.

ISPOR emphasizes evidence, uncertainty and generalizability; FDA/EMA add lifecycle governance.

KEY VOCABULARY

surrogate endpoint	替代终点	A measure used in place of a later or more clinically meaningful outcome.
causal mediation	因果中介分析	A framework for decomposing a treatment effect into pathways through a mediator and other mechanisms.
heterogeneous effect	异质性效应	An effect that differs across patient subgroups or covariate patterns.
mediator	中介变量	A variable lying on a possible causal pathway from treatment to outcome.
effect modifier	效应修饰因子	A variable associated with differences in the size or direction of an effect.
direct effect	直接效应	The treatment effect operating outside the pathway through the mediator.
indirect effect	间接效应	The portion of effect operating through the mediator pathway.
programmer-inspector loop	编程者 - 检查者循环	An iterative pattern in which generated code is reviewed, tested and corrected.
variable selection	变量选择	The process of choosing mediators, covariates or predictors for an analysis.
analytical workflow	分析 workflow	The ordered set of data, method, execution, checking and reporting steps.
reproducibility	可复现性	The ability to regenerate results from controlled data, code and settings.
statistical inference	统计推断	Drawing conclusions about effects or uncertainty from data using formal methods.
validation layer	验证层	An independent step that checks inputs, code, assumptions or outputs.
generalizability	泛化性	How well a method or conclusion applies beyond the data used to develop it.
context of use	使用情境	A precise definition of what an AI system is intended to do and under which conditions.

USEFUL PHRASES

1. assist with analytical reasoning without replacing statistical inference
2. keep the inferential core inside a validated method
3. separate method selection from numerical estimation
4. validate generated variable choices before execution
5. correct code iteratively through an inspector loop
6. interpret subgroup-specific surrogate performance
7. prespecify scientifically defensible analytical choices
8. report uncertainty rather than only point estimates
9. define a bounded context of use for the AI layer
10. retain human responsibility for judgment-dependent decisions

COMPREHENSION QUESTIONS

1. What does the LLM do, and what remains inside the formal statistical framework?
2. Why does subgroup heterogeneity complicate surrogate evaluation?
3. What problems do the programmer-inspector loop and variable checks target?
4. How does ISPOR guidance change interpretation of an AI-assisted workflow?
5. Which decisions still require statistician or programmer approval?

RETELLING PROMPTS

30 sec: Problem -> split architecture -> reliability mechanism.

45 sec: Data -> AI specification -> validated method -> execution -> diagnostics -> interpretation.

60 sec: Explain why orchestration should be automated before methodology.

5-MINUTE SPEAKING / OUTPUT TASK

Minute 1: Choose Cox, logistic, MMRM, subgroup, sensitivity or mediation analysis.

Minutes 2-3: Separate AI assistance, deterministic statistical code and human decisions.

Minute 4: Add variable, population, diagnostic, reproducibility and independent-check gates.

Minute 5: State the evidence required before expanding automation.

ONE SENTENCE TO KEEP

The strongest AI-assisted statistics pattern is to reduce friction around a validated inferential core whose inputs, assumptions, execution and review remain explicit.