Journal of Clinical and Translational Science

www.cambridge.org/cts

A critical juncture: Integrating large language models in biostatistical workflows

Vihaan Sahu

Medicine, Georgian National University SEU, Georgia

Letter

Cite this article: Sahu V. A critical juncture: Integrating large language models in biostatistical workflows. *Journal of Clinical and Translational Science* **10**: e47, 1–2. doi: 10.1017/cts.2026.10713

Received: 25 December 2025
Revised: 25 December 2025
Accepted: 13 February 2026

Corresponding author:
V. Sahu; Email: vsahu@seu.edu.ge

© The Author(s), 2026. Published by Cambridge University Press on behalf of Association for Clinical and Translational Science. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

To the Editor

Grambow et al [1] provide essential, empirical insights into the integration of large language models (LLMs) in biostatistical workflows, revealing both rapid adoption and significant, unmitigated risks. Their survey of biostatisticians demonstrates substantial integration, with 63.8% (44/69) of respondents already using LLMs, nearly half (46.5%) daily [1]. The primary uses are for quantitative coding (77.5%) and writing tasks (76.3% for editing), indicating a profound shift in daily practice [1]. However, this adoption starkly outpaces the development of adequate safeguards. The most critical finding is that 70.7% (29/41) of users reported encountering incorrect LLM outputs with potentially serious consequences, spanning incorrect code generation, statistical misinterpretation, content fabrication, and inappropriate tone [1].

This scenario represents a genuine paradigm shift. LLMs are not merely augmenting workflows but are fundamentally transforming how biostatisticians generate code, communicate findings, and access knowledge, creating a new, hybrid human-AI workflow [2,3]. This transformation necessitates a parallel shift in our approach to verification and quality control. The current reliance on individualized, ad hoc verification strategies such as personal expertise, external checks, and debugging is insufficient [1]. This is compounded by a glaring institutional support gap: while 49.3% of respondents reported organizational encouragement to use LLMs, only 18.8% had access to formal guidance or training [1].

This disconnect between rapid adoption and inadequate safeguards opens a larger series of essential experiments. First, we need rigorous, longitudinal studies to examine how LLM integration affects the reproducibility and quality of clinical research analyses, as initial comparative assessments show alarming variability in endpoint calculations like objective response rate [4]. Second, domain-specific verification frameworks must be developed and validated to systematically detect the high-frequency error types identified [1,5]. Third, comparative evaluations of different LLM architectures and prompting strategies for core biostatistical tasks are urgently required, moving beyond simple accuracy metrics to assess reasoning, hallucination rates, and robustness as recommended by scoping reviews [6,7].

The impact on clinical and translational science is twofold. Properly managed, LLMs could accelerate trial design, analysis, and the communication of complex results. However, the current high error rate and lack of standardized verification pose a direct threat to research integrity and subsequent clinical decision-making [1,8,9]. The strong demand for structured support (75.9% of respondents wanted case studies and 69.0% interactive tutorials) must be met with field-specific resources that emphasize critical evaluation and the eight core principles for responsible use proposed by the authors, including collaborative verification and transparency [1,10].

Grambow et al [1] have laid a crucial evidentiary foundation. Their work should catalyze a coordinated effort to develop the specialized training, robust evaluation frameworks, and institutional policies required to ensure this paradigm shift enhances, rather than undermines, the methodological rigor that is the cornerstone of valid clinical and translational science.

Author contributions. **Vihaan Sahu:** Conceptualization, Writing - original draft, Writing - review and editing.

Competing interests. The author declares no conflicts of interest.

References

1. **Grambow SC, Desai M, Weinfurt KP, et al.** Integrating large language models in biostatistical workflows for clinical and translational research. *J Clin Transl Sci.* 2025;**9**:1–8. doi: 10.1017/cts.2025.10064.
2. **Dell'Acqua F, McFowland E** III, Mollick E, **et al.** Navigating the jagged technological frontier: field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Working Paper, 24-013. Boston, MA: Harvard Business School; 2023. doi: 10.2139/ssrn.4573321
3. **Peng S, Kalliamvakou E, Cihon P, Demirer M.** The impact of AI on developer productivity: evidence from GitHub copilot. 2023. doi: 10.48550/arXiv.2302.06590.

4. **Denecke K, May R, Rivera Romero O.** Potential of large language models in health care: Delphi study. *J Med Internet Res*. 2024;**26**:e52399. doi: 10.2196/52399.
5. **Perlis RH, Fihn SD.** Evaluating the application of large language models in clinical research contexts. *JAMA Netw Open*. 2023;**6**:e2335924. doi: 10.1001/jamanetworkopen.2023.35924.
6. **Komandur R, McDunn J, Nair N, et al.** Artificial intelligence in biomedical data analysis: a comparative assessment of large language models for automated clinical trial interpretation and statistical evaluation. *MedRxiv*. 2025. doi: 10.1101/2025.02.05.25321607.
7. **Lee J, Park S, Shin J, Cho B.** Analyzing evaluation methods for large language models in the medical field: a scoping review. *BMC Med Inform Decis Mak*. 2024;**24**:366. doi: 10.1186/s12911-024-02709-7.
8. **Zhou H, Liu F, Gu B, et al.** A survey of large language models in medicine: progress, application, and challenge. 2023. doi: 10.48550/arXiv.2311.05112.
9. **Thapa S, Adhikari S.** Bard, and large language models for biomedical research: opportunities and pitfalls. *Ann Biomed Eng*. 2023;**51**:2647–2651. doi: 10.1007/s10439-023-03284-0.
10. **Low YS, Jackson ML, Hyde RJ, et al.** Answering real-world clinical questions using large language model-based systems. 2024. doi: 10.48550/arXiv.2407.00541.

https://doi.org/10.1017/cts.2026.10713 Published online by Cambridge University Press