



# Will Artificial Intelligence Replace Biostatisticians? Evolving Tools and Enduring Responsibilities

## Yapay Zekâ Biyoistatistikçilerin Yerini Alacak mı? Gelişen Araçlar ve Kalıcı Sorumluluklar

Özge Pasin

University of Health Sciences Türkiye, Hamidiye Faculty of Medicine, Department of Biostatistics, İstanbul, Türkiye

Cite this article as: Pasin Ö. Will artificial intelligence replace biostatisticians? Evolving tools and enduring responsibilities. Hamidiye Med J. 2026;7(2):63-66

In recent years, as biostatisticians, we have increasingly encountered similar questions from researchers and students: “With the rapid advancement of artificial intelligence (AI) applications, biostatistical analyses can now be performed using systems such as ChatGPT and similar tools. In this context, will the need for biostatisticians decrease in the future? Can statistical analyses be conducted reliably and accurately solely through AI tools without consulting an expert?”

By generating tables and outputs through prompts, it may appear that analyses are being conducted correctly. However, the reality is not as straightforward as it seems. Statistical analysis is not merely a process of entering data into a model and obtaining result tables. What truly matters is asking the correct research question, selecting the appropriate methodology, verifying underlying assumptions, and accurately interpreting the clinical significance of the findings. There is no doubt that AI accelerates many processes. Nevertheless, it does not assume the scientific responsibility arising from the analysis itself. Incorporating AI-generated outputs directly into scientific manuscripts without critical evaluation can be highly misleading and, in some cases, may contribute to the dissemination of scientifically inaccurate conclusions. Consequently, researchers may unknowingly produce manuscripts based on erroneous statistical analyses. At precisely this point, it becomes essential to explain why academics and experts specializing in biostatistics are still indispensable in scientific research. AI systems may perform certain statistical calculations, and the resulting outputs may appear highly convincing. However, determining whether these evaluations are accurate, valid, reliable,

and scientifically defensible still requires the expertise of a qualified biostatistician. Furthermore, it is important to recognize that “statistics is the invisible backbone of science.” Statistics constitutes the fundamental framework that ensures the reliability of scientific knowledge, yet because it often remains in the background, its critical role may go unnoticed.

From past to present, the field of biostatistics has continuously evolved. In earlier periods, analyses were performed manually using calculators; this was later replaced by statistical package software, and more recently by programming-based platforms such as R, which allow researchers to conduct advanced analyses through coding. Just as the transition from calculators to statistical software required rapid adaptation to technological progress, AI has now emerged as a new collaborative tool in the analytical process. At times, AI systems may appear almost like highly talkative, fast-responding, self-confident, rigid, yet creatively decisive individuals. While these characteristics may provide advantages in certain situations, they may also be misleading in others. Producing analyses and presenting results rapidly without critically filtering and evaluating every stage of the process can create a false sense of accuracy and reliability. Consequently, the speed and fluency of AI-generated outputs should not be mistaken for methodological correctness or scientific validity.

When we honestly ask ourselves this question, whether AI will completely eliminate our profession emerges as an important topic of discussion. From my personal perspective, however, the answer to this question is “no.” AI is capable of taking over certain tasks, accelerating various processes, and creating substantial transformations in some areas of



work. Nevertheless, it should not be regarded as a force capable of entirely replacing the fundamental nature and essence of the profession.

Biostatistics is a discipline that involves distinguishing meaningful signals from noise within data, ensuring that appropriate research questions are addressed using suitable methodologies, and evaluating the difference between statistically possible findings and scientifically defensible conclusions. AI applications may accelerate analytical workflows and simplify technical procedures; however, the ultimate judgment regarding which research questions should be asked, how reliable the obtained findings truly are, and whether these findings carry genuine clinical significance still fundamentally depends on human expertise.

In a study conducted by Bin Zhu, it was reported that although ChatGPT-4o was capable of solving certain biostatistical problems, this was achievable only after careful prompting and multiple iterative attempts. The study emphasized that biostatisticians play a crucial role in recognizing when the model produces inaccurate outputs, guiding the system in the correct direction, and managing complex analytical tasks through precise and structured instructions. In this context, the role of the biostatistician in the era of AI is expected to center on guidance, supervision, and critical evaluation. Furthermore, the study suggested that AI-assisted workflows may reduce the burden of repetitive and time-consuming tasks, thereby allowing experts to devote greater attention to problem-solving, formulating appropriate research questions, and developing scientifically robust analytical strategies (1). Similarly, Kim SN argued that AI systems should not be regarded as independent decision-making mechanisms capable of replacing biostatisticians. Rather, these systems should be considered supportive tools that assist research activities, reduce workload, and accelerate analytical processes (2).

In the study conducted by Aleksandar Ignjatović and colleagues, the reliability of ChatGPT as a supportive tool for medical students solving biostatistical problems was investigated. For this purpose, ten biostatistical problems were randomly selected from the Oxford Handbook of Medical Statistics, and different versions of ChatGPT were compared. GPT-3.5 correctly answered only five out of ten questions on the first attempt, whereas GPT-4 correctly solved six out of ten questions during the initial trial. The authors observed that although the responses were presented in a highly organized, logical, and convincing manner, the models were still capable of producing mathematically incorrect results. This issue was considered particularly critical because students may

easily be persuaded by answers that are fluent, coherent, and expressed with high confidence. Moreover, error rates increased substantially when the models were confronted with more complex problems. In multi-step analytical tasks, the systems were found to select inappropriate formulas and misinterpret tabular data. The authors ultimately emphasized that AI may serve as a supportive educational tool that facilitates learning; however, it cannot replace the learning process itself (3).

The study published by Dobler and colleagues provides a highly instructive framework and a comprehensive perspective on this issue. Rather than discussing ChatGPT solely from a theoretical standpoint, the article evaluates the system directly through the types of tasks encountered in the routine practice of biostatisticians. Across a wide range of application areas—including meta-analysis, latent class analysis, sample size calculations, causal inference, data analysis, and code generation—ChatGPT demonstrated remarkably successful performance in certain tasks, while in other situations it produced incorrect or misleading outputs with a high degree of confidence. The diverse use cases presented in the study illustrate that generative large language models (LLMs) may serve as practical and time-saving tools in biostatistical practice. Nevertheless, several fundamental principles must be considered in order to use these tools safely and effectively. First, although LLMs are particularly useful for accelerating routine tasks, it is essential to provide the model with sufficient contextual information to ensure that the task is performed appropriately. Furthermore, these systems should not be assumed to consistently reflect human expertise accurately; therefore, all generated outputs must be critically re-evaluated. Not only textual explanations, but also the consistency between analytical outputs produced in the data analysis environment and the accompanying narrative interpretations should be carefully verified. The study also emphasized that results may often be improved by explicitly pointing out previous errors to the model or by providing clearer and more structured prompts. In addition, due to the inherent stochastic nature of LLMs, the same task may yield different responses across separate sessions; consequently, output stability should, when necessary, be assessed through multiple independent attempts. Finally, the authors noted that the obtained results may vary depending on the programming language used and the specific model version, highlighting that the use of LLMs should not be viewed as a static process, but rather as a dynamic practice that requires continuous adaptation to evolving technological conditions (4).

AI applications provide substantial time advantages for researchers, particularly in routine statistical



processes. They can assist with data cleaning procedures, offer guidance regarding appropriate analytical methods, and facilitate exploratory data analyses in a manner that establishes a shared communicative framework with the researcher. Certain preparatory stages that traditionally required hours to complete may now be carried out within minutes through AI-assisted workflows. Nevertheless, it should be recognized that interpretations presented in a fluent and convincing manner do not necessarily correspond to scientifically accurate or reliable conclusions. In some instances, AI systems may present erroneous or scientifically questionable findings with a high degree of consistency and confidence. Consequently, the apparent coherence and persuasiveness of AI-generated outputs should not be automatically interpreted as indicators of methodological validity or scientific reliability. AI systems often do not independently question whether the fundamental assumptions required for the application of statistical methods have been satisfied. Core assumptions—such as homogeneity of variances, the assumption of normality, and the suitability of the data structure for the selected analytical method—represent critical components of the statistical analysis process. Although AI systems may recommend potentially applicable statistical tests, determining whether these methods are truly appropriate and valid for a given dataset constitutes a separate process that requires specialized expertise. Consequently, the investigation and verification of these methodological considerations should be supervised by qualified experts. In this context, it may be argued that with the advancement of AI technologies, the role of the biostatistician is becoming not less important, but increasingly essential. In the future, less time may be devoted to writing code itself; however, greater intellectual and analytical effort will likely be required for planning analytical workflows, evaluating the validity of statistical methods, and interpreting findings within their appropriate scientific and clinical contexts. Although fewer mechanical and repetitive tasks may remain, the need for methodological supervision, validation, and the development of innovative analytical approaches is expected to increase. This is because the primary aim of biostatistics is not merely to generate numerical outputs, but also to correctly interpret what the data truly represent, evaluate the reliability of the findings, and identify the most appropriate methodological approach for the underlying data structure. AI can generate code for biostatistical analyses and assist with data cleaning procedures, thereby providing substantial advantages in routine and repetitive tasks. However, the fluent and

persuasive language of AI does not necessarily indicate that the statistical analyses it produces are reliable. In statistics, the most dangerous errors are often not the obvious ones, but rather those that appear reasonable and convincing at first glance.

As technology continues to advance, it is essential not to resist these developments, but instead to adapt to them. In this context, the role of the biostatistician will likely evolve: becoming less focused on mechanical coding and more centered on managing critical decision points; less occupied with repetitive technical operations and more engaged in constructing analytical frameworks; less concerned with determining “which command works,” and more focused on answering “which inference is scientifically defensible.” Consequently, biostatisticians will increasingly devote their efforts to interpreting, questioning, and contextualizing data. AI models may generate text, write code, and produce tables. Nevertheless, the ultimate responsibility for determining whether an analysis is truly correct still belongs to humans. Therefore, rather than eliminating the field of biostatistics, the integration of AI is more likely to replace simpler and routine tasks within biostatistical practice.

In conclusion, researchers may substantially benefit from AI in biostatistical analyses. When researchers formulate appropriate questions and construct effective prompts, AI systems can provide access to basic statistical guidance and analytical code generation. Nevertheless, these systems are not consistently reliable and should therefore be considered preliminary supportive tools rather than definitive authorities. Accordingly, all stages of the research process should ultimately be reviewed and critically evaluated in collaboration with a domain expert. Furthermore, teaching students and researchers solely how to use statistical software packages no longer appears sufficient in the current era. Contemporary educational approaches should also encompass AI literacy, including the ability to communicate effectively with AI systems, construct accurate and meaningful prompts, and critically evaluate the correctness and appropriateness of AI-generated outputs. Equally important is equipping researchers with the ability to question whether the responses generated by AI systems are genuinely appropriate for the underlying research question. In this context, the integration of AI applications within a collaborative framework involving both researchers and biostatisticians may represent one of the most effective strategies for achieving faster, more accurate, and more reliable scientific outcomes.

## REFERENCES

---

1. Zhu B. Biostatisticians meet AI: navigating shifts while preserving principles. *Statistics in Medicine*. 2025;44:e70271. [\[Crossref\]](#)
2. Kim SN. Statistical analysis using ChatGPT in medical research. *Obstetrics & Gynecology Science*. 2025;68:467-472. [\[Crossref\]](#)
3. Ignjatović A, Stevanović L. Efficacy and limitations of ChatGPT as a biostatistical problem-solving tool in medical education in Serbia: a descriptive study. *J Educ Eval Health Prof*. 2023;20:28. [\[Crossref\]](#)
4. Dobler D, Binder H, Boulesteix AL, Igelmann JB, Köhler D, Mansmann U, et al. ChatGPT as a tool for biostatisticians: a tutorial on applications, opportunities, and limitations. *Statistics in Medicine*. 2025;44:e70263. [\[Crossref\]](#)