Pharmaceutical Statistics

WILEY

VIEWPOINT

Will the Pharmaceutical Industry Need Statisticians in an AI World?

Chris Harbron

Roche Products, Welwyn Garden City, UK

Correspondence: Chris Harbron (chris_harbron@hotmail.com)

Received: 6 November 2025 | **Revised:** 11 March 2026 | **Accepted:** 8 May 2026

ABSTRACT

Artificial Intelligence (AI) is rapidly becoming more visible in all aspects of our lives. This Viewpoint discusses the impact AI will have on statisticians working within the pharmaceutical industry. While aspects of the statistician's role will change, I propose that AI will make the core components of a statistician's skillset even more critical in the future.

1 | Introduction

Recent years have seen an explosion of interest and speculation around Artificial Intelligence (AI) and the multiple impacts, both positive and negative, it may have upon society. A high-profile concern is the potential impact on jobs, both whether certain jobs will still exist and the nature of those jobs that do. The website willrobotstakemyjob.com gives biostatisticians a moderate risk score of 40% of having their job taken over by AI. I would argue this is an overestimate of the risk, but it is certainly a topic that statisticians within the pharmaceutical industry must consider along with how to evolve their role to ensure they remain relevant.

This article is structured as follows. The following section briefly discusses the use of AI within the wider environment of the pharmaceutical industry and healthcare in general, Section 3 discusses the potential for statisticians to use AI to perform their current activities more efficiently and Section 4 considers the value statisticians bring within the pharmaceutical industry. Section 5 outlines some opportunities for statisticians to increase their contribution to drug development as a result of AI, and the article concludes with some forward-looking remarks.

2 | The Wider Context

While this article focuses on statisticians' contributions to clinical development, AI will also have a major impact on many other aspects of the pharmaceutical industry and healthcare in general. For example, Alphafold [1] demonstrated the potential for AI in drug discovery to model previously intractable molecular problems such as predicting molecular interactions and configurations. Manufacturing [2, 3], pharmacovigilance [4] and supply chain optimization are also areas where the pharmaceutical industry is investing heavily in AI.

Regulatory agencies have recognised the growing application of AI and have developed guidance for the use of AI [5, 6] and are also actively considering how AI could be used to support their own internal work. As both regulators and industry become more comfortable with AI's benefits, this may lead to opportunities for collaboration, particularly in sharing pre-competitive experiences.

More generally across healthcare, there is increasing recognition of the potential for AI to help address some of the challenges being faced by healthcare systems and improve patient outcomes

with applications ranging from earlier and more accurate disease diagnosis to reducing administrative burden to clinical note taking [7, 8]. However, considerable challenges still remain in the widespread adoption of AI, including issues of trust, bias and liability, and ensuring the infrastructure is available to allow the integration of AI into clinical workflows.

3 | Greater Efficiency

Throughout its history, the statistics profession has seen a series of inventions allowing statisticians to perform their tasks more efficiently. We no longer rely on mechanical computational devices or punched cards but instead can access multiple CPU cores though user-friendly Integrated Development Environments from anywhere in the world. We no longer consult statistical tables, but can rely on statistical software functions to give us results at greater levels of accuracy. The latest suite of AI tools enabling activities such as rapid drafting of reports or analysis plans [9] and programming or coding support [10] should be seen as a continuation of the increasing efficiency of our roles, removing many of the more mundane tasks and allowing statisticians to focus more time on those activities where they uniquely add the greatest value. Other opportunities for increased efficiency may come from producing summaries of complex situations from many disparate data sources, for example understanding the competitive trial landscape across multiple endpoints and populations for a disease area or for proposing a draft Clinical Development Plan (CDP) based upon competitor information and target commercial data.

The greatest benefit will come from a partnership where AI is able to support a statistician by rapidly performing tasks which would be mundane or time-consuming for a human, whilst the statistician is always conscious of the tendency of Large Language Models to hallucinate, that is to generating plausible but inaccurate outputs. In fact, recent research suggests that hallucinations are an inevitable consequence of how LLMs are trained [11]. Keeping up to date and learning to responsibly use these tools should naturally be part of the skills we develop to help us both professionally and in our personal lives, just as we take the relatively recent inventions of search engines and email for granted today.

4 | What Do Statisticians Bring to the Table?

So where can statisticians uniquely add the most value? The job of a statistician is often poorly understood by those outside of the profession, often being viewed as a sample-size and p value calculator or eliciting comments such as "lies, damned lies and statistics." Indeed, many statisticians find it hard to articulate their own value.

Figure 1 is an attempt to map some of the unique contributions of statisticians through the lens of working within the pharmaceutical industry. The top cells of Figure 1 highlight four fundamental capabilities of statisticians:

- The understanding of data: features such as its source, its preprocessing, and its intrinsic properties.
- The understanding of uncertainty: a fundamental of statistics through the understanding of distributions and their implications.
- Rigour: allowing us to make precise statements about this uncertainty with appropriate levels of confidence.
- Understanding the practicalities, opportunities and boundaries within drug development. Whilst not unique to statisticians, this is a critical component of what we can contribute, especially when combined with the previous capabilities.

The lower cells of Figure 1 highlight some of the capabilities which derive from the combination of these fundamental capabilities. Key among these is the rigorous understanding of the bias and uncertainty inherent in any AI outputs, and how these impact the operating characteristics of any subsequent decision making.

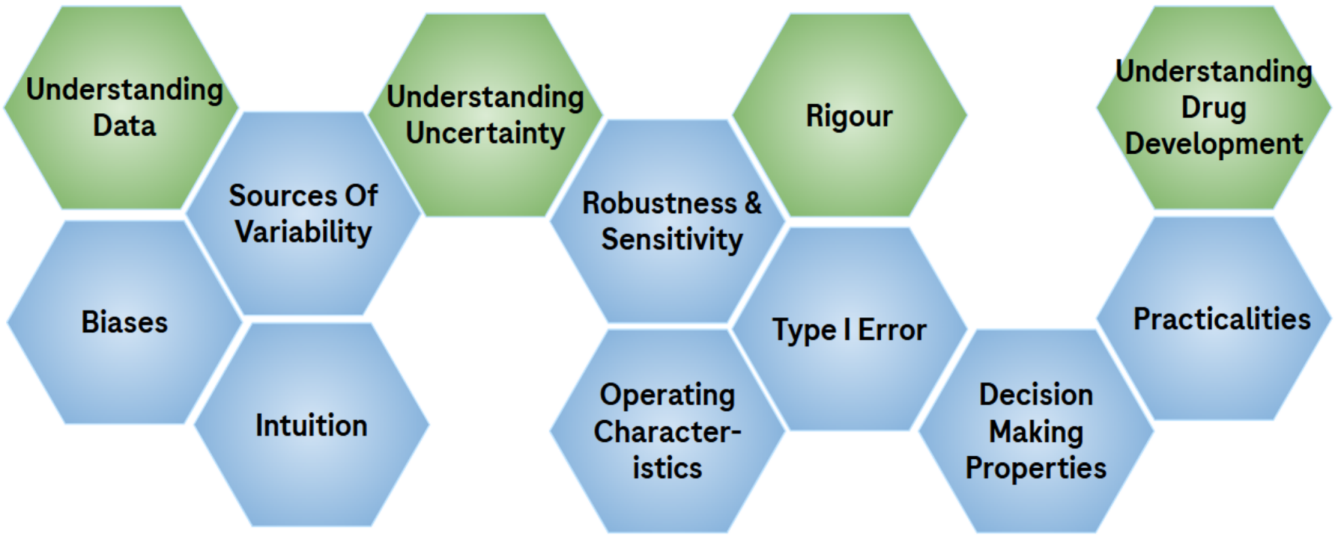

FIGURE 1 | Key capabilities of statisticians.

In addition to their more technical contributions, the training in logical and rigorous thinking that statisticians have undergone can be invaluable to project teams in numerous ways such as challenging assumptions, ensuring the team remains focused on a realistic path to regulatory and commercial success and providing a structured basis for decisions involving trade-offs.

More widely than just the pharmaceutical industry, there is a substantial literature providing guidance and advice on the skills needed and best practices for statistical consulting, supported by numerous real examples demonstrating the value that statistical thinking can add, for example, [12–14]. Some of the most critical themes discussed are having or gaining an understanding of the wider context in which the problem is being posed, the capability to ask beyond the presenting question to address the real underlying challenge, displaying the integrity to stand by the correct answer even if it is unpopular and not what the team wanted to hear and the pragmatism of finding a fit for purpose solution or design.

5 | Statisticians Supporting AI

As AI tools become more prevalent within drug development, applying these unique capabilities to either the generation of AI tools or their outputs is a critical role for statisticians. A non-exhaustive list of examples where statisticians' input is critical is given below:

- AI is already facilitating the development of new endpoints that are more clinically relevant, faster to assess, and less costly. In multiple disease areas, imaging can now generate more complete measures of disease, more rapidly and reproducibly than human experts [15]. Wearable devices allow measurement of a patient's activity throughout their normal lifestyle, as opposed to artificial measurements during a clinic visit on a single day [16]. However, effectively using these new endpoints requires a deep understanding of their background and derivation, properties such as their variability and any biases, as well as consideration of issues such as how to deal with partially or fully missing data.
- Including pre-specified covariates into analyses to increase the precision of estimates of treatment effect has long been a common practice. It is natural to consider extending this approach to covariates previously derived using AI from historical datasets potentially utilising multiple different data modalities [17, 18]. These covariates are still pre-specified in the analysis of treatment effect, but have much greater complexity than many commonly used covariates. While attractive, this raises many questions needing to be addressed, both in the derivation and interpretation of the covariate, but also in how it will be used in the analysis and the interpretation and robustness of any results.
- In April 2025, the FDA published its Roadmap to Reducing Animal Testing in Preclinical Safety Studies [19]. AI may be expected to play a key role in some of the new approach methodologies. Adoption of these methodologies will alter the type and amount of information that will be available prior to commencing first-in-human studies, which will have to be taken into account when designing these studies to optimise their value without compromising patient safety.
- Digital Twins are virtual representations of patients that can simulate real-time personalised responses to medical interventions, which go beyond traditional predictive models in their longitudinal and multivariate capabilities. This approach has been used to deal with complexity in situations where there are accurate and well defined models such as in supply chain optimization and planning, but early attempts within healthcare frequently struggled due to the high model accuracy required. Recent approaches using Large Language Models [20] have shown potential for utilising irregularly collected real-world data at large scale. These new classes of models, with increasingly limited explainability and unclear operating regions, will require a detailed and robust understanding of their properties if they are to be used for decision making.
- Pharmaceutical companies are frequently approached by external companies with their latest developments increasingly either being driven by AI or containing substantial elements of or claims around AI. Just as with in-licensing a medicine, robust due diligence is required before making any commitment or investment. With their combination of technical understanding and practical understanding of the business and drug development, statisticians are ideally positioned to contribute to these assessments.

6 | The Future

My answer to the question posed in the title of this article is a resounding yes. However, there is a caveat: while the expertise, skills and way of thinking of a statistician will most definitely be needed, how statisticians spend their time will evolve. The time spent on many more routine tasks will decrease, allowing more time to be invested into more strategic and high added-value activities. This is a continuation of a long-term trend now being accelerated by AI. The underlying skills and knowledge required for being a statistician, understanding data, uncertainty and drug development with rigour, will remain critical and will continue to need to be supplemented by an understanding of the context in which they are being applied. In the future this will also require a level of understanding of new and emerging AI technologies and a curiosity to experiment with them.

There is a risk of marginalisation of statisticians in that scientists, clinicians, or managers with a query may turn to their favourite Large Language Model rather than consult with a statistician, thereby missing some of the subtle and nuanced discussions around data collection and interpretation which may have profound consequences on business decision making and the likelihood of successful drug development. Mitigating this risk will require ongoing proactive communication of the importance and value of including statisticians in these discussions. The frustration of statisticians not always being included in the right discussions isn't a new problem, but without careful management, it could be heightened by the ubiquitous availability of AI.

How do we work together with AI? What is the continuing role for humans? Consider this scenario: using current technology it would be perfectly feasible for a sponsor to instruct a set of AI agents to process and analyse raw CRF data through a data processing and analysis pipeline into a series of reports and submit these to a regulatory agency. It would also be feasible for the regulatory agency to instruct its own set of AI agents to review these documents, submit questions back to the sponsor's agents and come to a conclusion to approve the drug or not. This scenario would currently not be remotely acceptable to either the sponsor or the regulator for multiple reasons, not least the near certainty of mistakes and hallucinations and the high likelihood of missing key information or signals. However, as a thought experiment this is useful in highlighting where human involvement remains critical and in which parts of the process AI could take on a greater role.

There are good reasons to be optimistic. In 2012 the Harvard Business Review [21] famously described Data Scientists as being "The Sexiest Job of the 21st Century." I see the evolving situation as being captured by this previous quote from 2009 by Hal Varian, the Chief Economist of Google: "The sexy job in the next ten years will be statisticians. Because now we really do have essentially free and ubiquitous data. So the complimentary factor is the ability to understand that data and extract value from it." This quote captures where statisticians need to be positioning themselves as we increasingly move into an AI world.

Acknowledgments

This Viewpoint was inspired by an invitation to give a short presentation to introduce a panel discussion at the EFPSI Regulatory Workshop in Basel in September 2025. During the preparation of this manuscript, the author used Google Gemini for minor language edits, aiming to enhance readability. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the article.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

1. J. Jumper, R. Evans, A. Pritzel, et al., "Highly Accurate Protein Structure Prediction With AlphaFold," *Nature* 596 (2021): 583–589.

2. J. J. Peterson, T. T. Kramer, J. D. Hofer, and G. Atkins, "Opportunities and Challenges for Statisticians in Advanced Pharmaceutical Manufacturing," *Statistics in Biopharmaceutical Research* 11 (2019): 152–161.

3. K. Huanbutta, K. Burapapadh, P. Kraisit, et al., "Artificial Intelligence-Driven Pharmaceutical Industry: A Paradigm Shift in Drug Discovery, Formulation Development, Manufacturing, Quality Control, and Post-Market Surveillance," *European Journal of Pharmaceutical Sciences* 203 (2024): 106938.

4. A. Nagar, J. Gobburu, and A. Chakravarty, "Artificial Intelligence in Pharmacovigilance: Advancing Drug Safety Monitoring and Regulatory Integration," *Therapeutic Advances in Drug Safety* 16 (2025): 20420986251361435.

5. "Reflection Paper on the Use of Artificial Intelligence (AI) in the Medicinal Product Lifecycle," European Medicines Agency (2024).

6. "Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products," US FDA (2025).

7. Philips, "Future Health Index 2025: Building Trust in Healthcare AI, tech. rep., 10th Annual Report," Philips (2025).

8. N. Maslej, L. Fattorini, R. Perrault, et al., "The AI Index 2025 Annual Report, tech. Rep. (Institute for Human-Centered AI, Stanford University, Stanford, CA" (2025).

9. R. Landman, B. Emir, R. Zhang, et al., "Leveraging Generative AI to Transform Statistical Analysis Plan Authoring in Clinical Trials," *Clinical Trials* https://doi.org/10.1177/17407745261422365.

10. D. Dobler, H. Binder, A.-L. Boulesteix, et al., "ChatGPT as a Tool for Biostatisticians: A Tutorial on Applications, Opportunities, and Limitations," *Statistics in Medicine* 44 (2025): e70263.

11. A. T. Kalai, O. Nachum, S. S. Vempala, and E. Zhang, "Why Language Models Hallucinate" (2025).

12. D. J. Finney, "The Questioning Statistician," *Statistics in Medicine* 1 (1982): 5–13.

13. C. M. Anderson-Cook, "Statistical Consulting: A Guide to Effective Communication," *Technometrics* 44 (2002): 89.

14. D. J. Hand and B. Everitt, *The Statistical Consultant in Action* (Cambridge University Press, 2008).

15. M. Sebastian, H. Batra, M. L. Saini, et al., "Applications and Challenges of Utilizing Digital Pathology and AI-Enabled Workflows in Clinical Trials," *Journal of Pathology Informatics* 20 (2026): 100542.

16. L. Servais, D. Eggenspieler, M. Poleur, et al., "First Regulatory Qualification of a Digital Primary Endpoint to Measure Treatment Efficacy in Dmd," *Nature Medicine* 29 (2023): 2391–2392.

17. N. Anegondi, S. S. Gao, V. Steffen, et al., "Deep Learning to Predict Geographic Atrophy Area and Growth Rate From Multimodal Imaging," *Ophthalmology Retina* 7 (2023): 243–252.

18. B. Holzhauer and E. T. Adewuyi, ""Super-Covariates": Using Predicted Control Group Outcome as a Covariate in Randomized Clinical Trials," *Pharmaceutical Statistics* 22 (2023): 1062–1075.

19. "Roadmap to Reducing Animal Testing in Preclinical Safety Studies," US FDA (2025).

20. N. Makarov, M. Bordukova, P. Quengdaeng, et al., "Large Language Models Forecast Patient Health Trajectories Enabling Digital Twins," *npj Digital Medicine* 8 (2025): 588.

21. "Data Scientist: The Sexiest Job of the 21st Century," Harvard Business Review (2012).