**Description**

# Executive Summary

AI is reshaping peer review, particularly in statistical evaluation. Beyond text generation, AI can check consistency, rerun analyses, and flag questionable practices, easing reviewer fatigue. Used responsibly with bespoke, secure systems, AI can streamline first-pass checks, allowing reviewers to focus on interpretation, originality, and clinical significance.

# Rethinking Statistical Rigor in the Age of AI-Powered Peer Review

Despite the widespread use of AI in various aspects of academic publishing by both authors and publishers, its full potential is probably not being realized. Frankly, the potential for generative AI packages based on large language models beyond the generation of text, checking grammar, spelling, the use of English and searching for information is probably unknown by many. For example, some who criticize generative AI for providing false information, hallucinating references and its occasional errors, even for checking spelling and grammar, have no idea that even common packages can be used to analyze and produce spreadsheets, run complex calculations and even perform statistical analysis. These facilities of AI lead directly to the possibility of academic publishers using AI as an aid to review the statistical aspects of manuscripts reporting quantitative studies. AI is here, it is being used and it will not go away, we must learn to use it responsibly.

While packages such as ChatGPT should not be used to upload and analyze manuscripts submitted to academic journals for reasons of copyright infringement, I regularly use ChatGPT to analyze published articles by uploading the article and interrogating specific aspects of the reported study. For example, I am especially interested in the effectiveness of COVID-19 vaccines and ChatGPT can be used to calculate absolute risk reduction and numbers needed to treat for the vaccines where such studies commonly only report relative risk reduction and rarely report numbers needed to treat. The package will run the calculation, showing the steps in the calculation, meaning that the results can be checked.

Statistical analysis has evolved rapidly in recent decades and sophisticated commercially available packages, such as SPSS® and SAS®, and public domain statistical packages such as *R*, have put the

ability to conduct sophisticated and complex statistical analyses in the hands of non-statisticians. This has led to a large volume of studies, translated into manuscripts for publication, being submitted to academic journals. Often the same methods are used repeatedly without much consideration of how appropriate they are and without expert statistical oversight.

As a result, there has been a concomitant increase in the scrutiny of quantitative studies by journal editors, an increase in the standards of statistical analysis required and increased rigor in peer reviewing processes. For example, where the outcomes of randomized clinical trials were traditionally reported using only the statistical significance to justify the effectiveness of interventions, there has been a move in favor of reporting effect sizes of differences between control and intervention groups along with 95% confidence intervals to provide a better estimation of where the true difference between the groups lies.

In addition to the frequentist statistical methods traditionally applied to clinical trials, there has been some advocacy for the use of Bayesian methods. Researchers using Bayesian methods are required to take prior knowledge into account in estimating how effective an intervention may be and then examining the outcome in terms of the expected and actual outcomes, the posterior outcome. They can also estimate the credible interval – a range of values – within which the effect of the intervention lies. However, while the absolute number of such studies using Bayesian methods is increasing, the percentage remains constant and very small.

With the volume of manuscripts submitted to academic journals doubling approximately every 15 years pre-COVID – and with the massive increase during COVID – academic publishers and editors have long been turning to AI for solutions to help reviewers cope with the burden of various aspects of the review process, including statistical review. Publishers and editors are also concerned with maintaining rapid copy flow and, since peer review is the rate-limiting step in the peer review process, they are always looking for ways to expedite reviews as rapidly as possible. Always, of course, with an eye on the quality of the reviews that are submitted. Academics are busier than ever and, as the number of manuscripts increases, they have less time to review the manuscripts they receive leading to reviewer fatigue. This threatens the quality of peer reviewing, including statistical reviewing.

AI has considerable potential to help streamline and accelerate the process of reviewing the statistical aspects of manuscripts. Some examples of where AI could be used include the following automated steps:

- Pre-screening: standards check (CONSORT, STROBE, PRISMA)
- Consistency Check: cross-verify reported numbers with submitted databases
- Method Validation: ensure tests match study design
- Sample Size and Power: recalculate and confirm
- Checking veracity of and re-running submitted statistical coding such as Python® and *R*
- Effect Size and Interpretation: statistical versus clinical meaning
- Bias and questionable research practices: look for p-hacking, outcome switching
- Bayesian statistics: verify whether Bayesian results are interpreted correctly and priors are clearly stated
- Draft reviewer notes: structured comments and feedback
- Decision support: confidence score and need for statistical reviewer

Naturally, there are limitations and caveats to be addressed in the use of AI in the process of statistical

reviewing. Commercial packages such as ChatGPT should not be used as the coding and algorithms are not in the public domain and anything which is uploaded to these packages may be shared with and used by others. Therefore, bespoke in-house AI packages are being developed by publishers. These will require training, involving qualified statisticians to ensure that the results they produce are both reliable and valid. In that light, AI should be seen as augmenting the statistical review process, not replacing expert statistical judgment.

Taken together, these questions point to a future where AI is not a replacement for human reviewers but a catalyst for re-thinking how we balance speed, rigor, and transparency. If AI can reliably handle the first pass of statistical checking, flagging problems and re-running code, then editors and reviewers can focus more sharply on interpretation, originality, and clinical relevance. Whether this hybrid model becomes the new normal will depend on how well bespoke, trustworthy AI tools are developed, and how much confidence journals place in them to support—not supplant—the human judgement at the heart of peer review.

**Category**

1. Thought Leadership

**Date Created**
2025/09/17
**Author**
rogerw