

Bayes and the Evidence Base: Reanalyzing Trials Using Many Priors Does Not Contribute to Consensus

Harm-Jan de Grooth¹ and Olaf L. Cremer²

¹Department of Intensive Care Medicine, Amsterdam UMC – Location VUmc, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands; and ²Department of Intensive Care Medicine, University Medical Center Utrecht, Utrecht, the Netherlands

ORCID ID: 0000-0002-7499-076X (H.-J.d.G.).

Recent years have seen an increase in Bayesian reanalyses of clinical trials. Proponents argue that this helps the interpretation of trial results (1–3). Bayesian inference has also attracted skepticism, most of which is poorly justified (4). The stale criticism that subjectivism is at odds with the scientific method is misguided: conclusions based on the frequentist paradigm are equally dependent on the prior likelihood of hypotheses, albeit less transparently so (5). Nonetheless, we believe that some apprehension against Bayesian reanalyses is justified. Our concerns are not of a statistical nature, nor do they relate to studies that are designed around Bayesian methods from the outset. Rather, we are concerned that the many-priors mode of Bayesian reanalyses disregards the communal aspects of evidence-based medicine, in which consensus among clinicians is gradually developed through the accumulation of generalizable study data.

The Many-Priors Paradigm Frustrates Consensus Building

A key reason for conducting randomized controlled trials is to resolve uncertainty and disagreement in the clinical community (6). Equipose derives not from the uncertainty of any individual but from the uncertainty in the community of expert clinicians: when some experts believe that a treatment is beneficial whereas others believe that it is harmful or wasteful, successful clinical trials

forge consensus among reasonable parties who previously disagreed (7, 8).

Unfortunately, the results of many randomized controlled trials are not convincing. So investigators increasingly turn to Bayesian methods to reanalyze trials using a range of priors as representations of the *a priori* opinions in the expert community (9–11). Even though this can be insightful, there is no widely accepted framework to translate the spectrum posterior results into a consensus viewpoint about treatment efficacy. For example, a Bayesian reanalysis may indicate 98% probability of a clinically important benefit under strongly enthusiastic prior assumptions but also 37% probability of a negligible effect or harm under strongly skeptical prior assumptions (10). Parties with opposing suspicions about the treatment can each see their position strengthened by the results, especially if the intervention is costly or associated with severe adverse effects.

Bayesian methods are also used to derive probabilities for different treatment effect sizes. The product of evaluating multiple effect sizes using multiple priors is an often very large set of posterior probabilities. It is common for reports to contain as many as 30–80 distinct probabilities for a single primary endpoint (9, 10). This is highly transparent yet utterly perplexing to impartial readers. When there is no guiding principle to choose which of these many estimates is most relevant, it appears as if each can choose his own. The complexity is further increased by the heavy use of technical graphs and jargon, requiring

readers to decipher probability density curves and priors expressed on the log-odds scale (12). Such representations of trial results are transparent only to those well versed in statistics.

Finally, the hyperquantitative way of presenting results distracts from the *qualitative* assessment that should be central to every trial interpretation. A skeptical prior (i.e., an expectation of small effects) does not provide a safeguard against bias or poor generalizability. Because bias distorts the very way in which the data are generated, a large and biased trial will easily overwhelm a strong skeptical prior.

In all, it is not surprising that the “pick your own prior” paradigm of Bayesian reanalyses appears vacuous to many. With strong data, no reasonable prior will make a material difference to the interpretation of a trial. With weak data, using different priors will only entrench disagreeing parties in their respective beliefs. In both cases, no progress toward consensus is made, while the potential points of disagreement grow exponentially in the cross-evaluation of multiple priors, multiple effect sizes, and multiple cutoff probabilities. As one critic has argued, it would be much simpler just to urge clinicians to enlist their own beliefs when faced with unconvincing data (13).

Yet the use of subjective priors is not necessarily at odds with the communal nature of evidence-based medicine. The idea of using priors that represent different beliefs in the community of experts (“a community of priors”) was born from the view that

(Received in original form August 20, 2023; accepted in final form November 3, 2023)

Ⓐ This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0. For commercial usage and reprints, please e-mail Diane Gern (dgern@thoracic.org).

Correspondence and requests for reprints should be addressed to Harm-Jan de Grooth, M.D., Ph.D., Department of Intensive Care Medicine, Amsterdam UMC – Location VUmc, De Boelelaan 1117, 1081 HV Amsterdam, the Netherlands. E-mail: h.j.s.degrooth@umcutrecht.nl.

This article has a related viewpoint.

Am J Respir Crit Care Med Vol 209, Iss 5, pp 483–484, Mar 1, 2024

Copyright © 2024 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.202308-1455VP on November 3, 2023

Internet address: www.atsjournals.org

consensus arises when the data are strong enough to convince even a reasonable adversary (14–16).

The Adversarial Perspective

It was the perspective of the hypothetical *other*, not the quantification of one's own beliefs, that motivated early proponents of Bayesian methods for clinical trials.

By using an adversarial perspective, a Bayesian reanalysis can be used to specifically assess whether the available data are strong enough to forge consensus among reasonable and well-intentioned experts (6). If the point estimate of a trial indicates benefit, regardless of the *P* value, consensus can arise only if the data are strong enough to convince those with reasonable pessimistic prior expectations. Conversely, if the trial results indicate harm or a negligible effect, it is the position of the reasonable optimists that requires careful evaluation (14–16).

Focusing Bayesian reanalyses squarely on the adversarial prior reduces the overwhelming amount of posterior probabilities in study reports and limits the potential points of disagreement to two important questions: is the prior representative of a “reasonable adversary” of the observed effect, and what amount of posterior uncertainty about the treatment effect is acceptable? The answers to both questions are rooted in factors such as the importance of the outcome, previous evidence, the costs and harms of the intervention, and a qualitative assessment of risk of bias and generalizability.

The adversarial principle may also help overcome an important asymmetry in the published literature. Bayesian methods are now used to reanalyze trials with borderline nonsignificant results, but reanalyses of trials with borderline significant results are suspiciously less prevalent. This one-sidedness is unnecessary and wasteful, as the interpretation of small trials with spectacular

results may change most from an adversarial perspective (1).

Conclusions

Although personalized decision making is at the core of evidence-based medicine, the evidence base itself cannot be made contingent on personal beliefs. No critical care clinician can practice in isolation from his colleagues or without regard for cost-benefit considerations at the societal level. Yet the many-priors mode of reporting *post hoc* Bayesian analyses makes it seem as if each can choose his own treatment effect. This is confusing, overly complicated, and unnecessary. The discourse about the effectiveness of a treatment is better served by focusing on a single question: can the matter be considered settled, or is the evidence too weak to reach consensus? ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1. Yamell CJ, Abrams D, Baldwin MR, Brodie D, Fan E, Ferguson ND, *et al*. Clinical trials in critical care: can a Bayesian approach enhance clinical and scientific decision making? *Lancet Respir Med* 2021;9:207–216.
2. Zampieri FG, Casey JD, Shankar-Hari M, Harrell FE Jr, Harhay MO. Using Bayesian methods to augment the interpretation of critical care trials: an overview of theory and example reanalysis of the Alveolar Recruitment for Acute Respiratory Distress Syndrome trial. *Am J Respir Crit Care Med* 2021;203:543–552.
3. Yamell CJ, Granton JT, Tomlinson G. Bayesian analysis in critical care medicine. *Am J Respir Crit Care Med* 2020;201:396–398.
4. Gelman A. Objections to Bayesian statistics. *Bayesian Anal* 2008;3:445–450.
5. Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
6. London AJ. Self-defeating codes of medical ethics and how to fix them: failures in COVID-19 response and beyond. *Am J Bioeth* 2021;21:4–13.
7. London AJ. Equipose in research: integrating ethics and science in human research. *JAMA* 2017;317:525–526.
8. van der Graaf R, van Delden JJM. Equipose should be amended, not abandoned. *Clin Trials* 2011;8:408–416.
9. Goligher EC, Tomlinson G, Hajage D, Wijeyesundera DN, Fan E, Jüni P, *et al*. Extracorporeal membrane oxygenation for severe acute respiratory distress syndrome and posterior probability of mortality benefit in a *post hoc* Bayesian analysis of a randomized clinical trial. *JAMA* 2018;320:2251–2259.
10. Levy B, Girerd N, Amour J, Besnier E, Nessler N, Helms J, *et al*; HYPO-ECMO Trial Group and the International ECMO Network (ECMONet). Effect of moderate hypothermia vs normothermia on 30-day mortality in patients with cardiogenic shock receiving venoarterial extracorporeal membrane oxygenation: a randomized clinical trial. *JAMA* 2022;327:442–453.
11. Granholm A, Munch MW, Myatra SN, Vijayaraghavan BKT, Cronhjort M, Wahlin RR, *et al*. Dexamethasone 12 mg versus 6 mg for patients with COVID-19 and severe hypoxaemia: a pre-planned, secondary Bayesian analysis of the COVID STEROID 2 trial. *Intensive Care Med* 2022;48:45–55.
12. de Grooth H-J, Elbers P. Pick your prior: scepticism about sceptical prior beliefs. *Intensive Care Med* 2022;48:374–375.
13. Aberegg SK. *Post hoc* Bayesian analyses. *JAMA* 2019;321:1631–1632.
14. Kass RE, Greenhouse JB. [Investigating therapies of potentially great benefit: ECMO]: comment: a Bayesian perspective. *Stat Sci* 1989;4:310–317.
15. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. *J R Stat Soc Ser A Stat Soc* 1994;157:357.
16. Spiegelhalter DJ. Incorporating Bayesian ideas into health-care evaluation. *Stat Sci* 2004;19:156–174.