VIEWPOINT: TURNING THE AIR BLUE

What Is the Point of Bayesian Analysis?

Ewan C. Goligher[1,2] and Michael O. Harhay[3,4]

[1]Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, Ontario, Canada; [2]Division of Respirology, Department of Medicine, University Health Network, Toronto, Ontario, Canada; [3]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania; and [4]MRC Clinical Trials Unit, University College London, London, United Kingdom

ORCID IDs: 0000-0002-0990-6701 (E.C.G.); 0000-0002-0553-674X (M.O.H.).

In her compelling historical account of the development of Bayesian statistical theory, Sharon Bertsch McGrayne recounts how Bayesian thinking has long engendered controversy and negative reactions (1), particularly within a statistical (and broader scientific) community that for much of the 20th century was dominated by a few key thought leaders fiercely opposed to the approach (i.e., "anti-Bayesians"). Now, in the 21st century, modern computing has made Bayesian analysis more feasible and accessible, and the ways in which Bayesian thinking can advance scientific inquiry, including in medical research, have begun to be realized. During the coronavirus disease (COVID-19) pandemic, the Bayesian approach was used in adaptive clinical trial designs, enabling more timely and efficient evaluation of potential therapies. Some of these trials changed clinical practice, suggesting that the clinical community is ready to embrace Bayesian analysis in clinical trials. Yet, some still view the Bayesian approach with a measure of skepticism because it has led to marked (and sometimes favorable) reinterpretations of data, particularly for randomized trials. That different statistical frameworks, one of which explicitly admits prior information, yield competing interpretations seems to introduce a kind of epistemic fog into our work. Can we establish the *truth* about an intervention's effect? And are the seemingly positive results of Bayesian analysis too good to be true?

Ironically, such skepticism is itself a kind of Bayesian analysis of Bayesian analysis—a skeptical prior belief is being brought to bear on results that seem too good to be true. Bayesian analysis is just how we all think, but with a concerted effort to represent our beliefs in quantitative and probabilistic terms. In any case, the results are not really too good to be true: in the vast majority of cases, Bayesian analysis does not conclude superiority of a therapy under study any more than traditional frequentist analysis, as shown in a systematic review of 10 years of clinical trials in critical care by Yarnell and colleagues (2). For trials that do not conclude superiority, Bayesian analyses often yield a somewhat clearer message as to whether potential benefit has been entirely ruled out. And this, we suggest, is the main point of Bayesian analysis: to clarify the meaning of the data in hand by quantifying how much information the evidence provides (i.e., the posterior distribution) and the resulting level of confidence or uncertainty about a hypothesis (i.e., the posterior probability).

In this issue of the *Journal*, de Grooth and Cremer (pp. 483–484) bring a thoughtful critique to bear on the common approach to Bayesian analysis used over the last several years in critical care medicine, an approach they refer to as the "many priors mode" (3). de Grooth and Cremer are not anti-Bayesians; they support the use of Bayesian analysis and recognize the appropriateness of explicitly incorporating prior beliefs in quantitative

analysis. They are concerned that reporting multiple posterior probabilities under multiple competing priors and for multiple effect sizes leads to a wide range of possible conclusions about an intervention's effect that seems to obfuscate, rather than clarify, the meaning of the data. In other words, they argue that this approach directly contradicts the whole point of Bayesian analysis: to clarify the meaning of the data. They worry that this undermines, rather than advances, consensus about the data and leads individual readers to think they can just decide for themselves whether an intervention works or does not work.

We understand their concern and appreciate their point. Having both been involved in Bayesian analyses that used the many priors mode, we are guilty as charged. And we recognize the concerns about presenting many priors and many effect sizes; it is certainly more challenging for the reader to interpret than simply reporting "$P < 0.05$" or "$P > 0.05$". The unfortunate problem is that when conducting a Bayesian RE-analysis, one cannot prespecify one's priors or the clinically relevant effect size before having the data. The use of widely varying priors is aimed at compensating for this lack of prespecification intrinsic to the reanalysis of data.

We could point out some slight inconsistencies in their critique. They recognize that varying priors are used to "represent different beliefs in the community of experts" (exactly!) but suggest that using multiple competing priors leads to excessive

Table 1. Recommendations for Designing a Bayesian Analysis

Advice	Explanation
Avoid performing an unplanned secondary Bayesian analysis of your data if possible.	Bayesian reanalysis of clinical research is suboptimal, as the strength of inference is weakened if the analysis is designed after the results are available. When the interpretation of a Bayesian reanalysis differs from the primary frequentist interpretation, there is a risk of (perceived) bias from Bayesian reanalysis, particularly if studies selected for reanalysis are those in which Bayesian reanalysis is most likely to change the interpretation (a type of selection and publication bias). Nevertheless, prespecification is not possible for trials that are already completed. Provided the priors for analysis are specified carefully and transparently with a clear justification, Bayesian reanalysis provides relevant results to clarify the interpretation of data.
Make Bayesian analysis a standard part of the prespecified analysis plan, even if it is not the primary analysis.	Prespecifying a Bayesian analysis as the primary or secondary analysis before seeing the data reduces the risk of perceived bias in the results. Full prespecification of priors and analysis parameters (such as the minimum clinically important effect size) before awareness of data strengthens scientific inference. If positioned as a secondary analysis, the Bayesian analysis will provide complementary information to assist in clarifying the meaning of the data.
Priors need to be clearly justified.	Provide a detailed scientific rationale for the specification of each prior, appealing to previous data as much as possible and explaining exactly how the prior was derived and numerically specified. When describing each prior, consider reporting the "equivalent sample size" (i.e., the number of patients that would be required to generate a probability distribution with the same variance as the prior) to provide a concrete way of assessing the amount of information represented by the prior (4).
In general, use a neutral prior for the primary analysis.	Unless there is strong justification from previously and rigorously collected data, the primary prior for a Bayesian analysis should generally be a neutral (slightly skeptical) prior with most of the probability density located in reasonable proximity to the null, consistent with clinical equipoise. This makes extreme values for benefit or harm less likely (because they are, in fact, less likely) and constitutes a conservative analysis of data.
Avoid the use of "standard" priors.	Although there are guidelines for selecting enthusiastic or skeptical priors, there is no simple "standard cookbook recipe" for prior specification. Available guidelines should be treated as guidelines, not hard and fast rules. Where possible, prior specification should rely on data and reasonable assumptions, rather than simple standard parameters. The goal is to ensure that the specified priors represent the actual range of plausible beliefs in the expert community, especially those of potential "adversaries" (i.e., skeptics inclined to doubt a positive result, and enthusiasts inclined to doubt a negative interpretation).
Treat priors as a sensitivity analysis on the primary analysis.	Prespecify one prior as the primary and use the others as sensitivity analyses to assess how much the posterior is influenced by the other prior(s). This allows you to evaluate whether you have reached a definitive conclusion, provided the priors represent the reasonable range of prior beliefs about treatment effects.
Include an "adversarial" prior.	A conclusion of benefit is bolstered if the posterior probability of benefit remains high even under a skeptical adversarial prior. The evidence for benefit with this prior should be sufficient to convince a reasonable skeptic. Similarly, a conclusion of futility is bolstered if the posterior probability of futility remains high even under an enthusiastic adversarial prior. The evidence of futility should be sufficient to convince a reasonable enthusiast.
Evaluate trial methodology using standard published methods.	Bayesian analysis does not affect nor negate in any way the importance of valid study design and conduct, including issues such as allocation concealment, randomization, blinding, loss to follow-up, etc. Methodological evaluation remains essential to reaching an appropriate posterior conclusion. Concerns about methodological quality can serve as the basis for a skeptical prior.

individualism in interpretation. They also recognize that "with strong data, no reasonable prior will make a material difference to the interpretation of a trial." It seems then that their concern about the many priors mode only applies to situations where the available information is insufficient to reach a conclusion. Yet, in this situation, the dependence of the posterior result on prior information just serves to confirm that consensus is likely not appropriate. And here we find the basis for a Bayesian definition of a definitive trial: if the posterior conclusion is essentially the same regardless of the prior, then the question seems to be settled, and we have a strong basis for consensus. Versions of this notion were argued in the Bayesian analysis of the ECMO to Rescue Lung Injury in Severe ARDS (EOLIA) trial (4) and in the systematic review of Yarnell and colleagues (2), where the vast majority of trials were found to yield the same conclusion independent of the prior; most often, that conclusion was "we need more information." Indeed, many thoughtful scholars and frequentist-minded readers, especially in critical care, often suggest that a negative (i.e., P value > 0.05 for the null hypothesis) superiority trial is better termed indeterminate (5).

In any case, we are forced to attend to the most difficult part of executing a Bayesian analysis (indeed, the only difficult part, really): selecting priors (6–8). Part of the challenge with prior selection is that generating mathematical or probabilistic representations of our beliefs may not be intuitive (and we receive little or no training for this in conventional statistics education). But being forced to face this challenge is actually one of the virtues of Bayesian analysis, for it forces us to think hard about the strength of available evidence and its relevance to the question at hand. Furthermore, we must decide what we mean by clinically relevant benefit and how much certainty we need to decide whether to declare that a treatment should be used—it is conceivable that many patients or caregivers do not need a probability of benefit >97.5% (equivalent to a P value < 0.05 in conventional statistics) to decide to accept a treatment, depending on the intervention. Prior selection is so important because it is the only place that a Bayesian analysis can go wrong. Unlike conventional frequentist statistics, Bayesian analysis does not aim to establish the single true value for treatment effect (the point estimate) and it does not treat study data as if it was equivalent to a coin toss, possibly resulting by mere chance. Hence, it is less susceptible to concerns about false-positive (type I error) or false-negative (type II error) conclusions. Provided the data are collected in an unbiased and rigorous fashion (and all the issues around evaluating the rigor of trial methodology still apply), the prior is the only source of error in the posterior.

In Table 1, we offer some advice for designing a Bayesian analysis and selecting priors. We agree with the suggestion by de Grooth and Cremer that an "adversarial" prior should be included. The most important piece of advice we might offer is to not do a Bayesian reanalysis: design your study up front under a Bayesian framework so that you can prespecify all the parameters of the analysis and avoid the need for the many priors mode altogether (although multiple priors may still be appropriate to represent the range of beliefs among the community of experts, at least as a sensitivity analysis). The second most important advice, we believe, is to provide a well-justified rationale for each specified prior, appealing as much as possible to previously published observations. Priors should represent well-informed, thoughtful judgments about the evidence available before a study is undertaken. In general, the primary prior for analysis should be neutral, with its probability mass clustered about the null, unless there is strong justification for doing otherwise. This makes a Bayesian analysis more conservative than frequentist analysis (with its implicit, unstated, and obviously false prior assumption that all values for treatment effect, no matter how large or small, are equally likely, i.e., the "flat" or "noninformative" prior) and coheres with the clinical equipoise that motivated the study.

Bayesian thinking proved especially useful for adaptive trial design during the pandemic, with the successful completion and publication of multiple Bayesian trials of therapies for COVID-19 (9, 10). In the future, Bayesian analysis could (we hope) become a standard element of grant applications by enabling quantitative arguments demonstrating residual uncertainty and the need for more information. To ensure that trials are designed to provide definitive conclusions, sample size computations could be based on the number of observed patients and events required to yield a posterior probability distribution that is independent of the range of priors, including the adversarial prior (closely akin to current approaches to Bayesian adaptive trial design). The use of Bayesian statistics for study design and analysis will (we hope) grow in the coming years, as more and more investigators come to understand the point of Bayesian analysis: to clarify the meaning of the data and to decide when we can finally reach consensus, even if absolute certainty remains elusive. ■

References

1. McGrayne SB. The theory that would not die. New Haven: Yale University Press; 2011.
2. Yarnell 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.
3. de Grooth HJ, Cremer OL. Bayes and the evidence base: re-analyzing trials using many priors does not contribute to consensus. *Am J Respir Crit Care Med* 2024;209:483–484.
4. Goligher EC, Tomlinson G, Hajage D, Wijeysundera 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.
5. Sackett DL. Superiority trials, noninferiority trials, and prisoners of the 2-sided null hypothesis. *ACP J Club* 2004;140:A11.
6. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. *J R Stat Soc Ser A Stat Soc* 1994;157:357–387.
7. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Methods in health service research: an introduction to Bayesian methods in health technology assessment. *BMJ* 1999;319:508–512.
8. 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.
9. Gordon AC, Mouncey PR, Al-Beidh F, Rowan KM, Nichol AD, Arabi YM, *et al.*; REMAP-CAP Investigators. Interleukin-6 receptor antagonists in critically ill patients with Covid-19. *N Engl J Med* 2021;384:1491–1502.
10. Goligher EC, Bradbury CA, McVerry BJ, Lawler PR, Berger JS, Gong MN, *et al.*; REMAP-CAP Investigators; ACTIV-4a Investigators; ATTACC Investigators. Therapeutic anticoagulation with heparin in critically ill patients with Covid-19. *N Engl J Med* 2021;385:777–789.