Received: 9 September 2020 | Revised: 25 September 2020 | Accepted: 1 October 2020

DOI: 10.1111/aas.13725

REVIEW

acta Anaesthesiologica Scandinavica

Interpreting the results of clinical trials, embracing uncertainty: A Bayesian approach

Steven A. Frost[1,2,3,4] | Evan Alexandrou[1,2,3,4,5] | Luis Schulz[1] | Anders Aneman[1,4,6]

[1]Intensive Care Unit, Liverpool Hospital Sydney Australia, Liverpool, NSW, Australia
[2]Western Sydney University, Penrith South, NSW, Australia
[3]South Western Sydney Centre for Applied Nursing Research, Ingham Institute of Applied Medical Research, Liverpool, NSW, Australia
[4]South Western Clinical School, University of New South Wales, Sydney, NSW, Australia
[5]Griffith University, Brisbane, QLD, Australia
[6]Faculty of Health Sciences, Macquarie University, Sydney, NSW, Australia

Correspondence
Steve A. Frost, South Western Sydney Centre for Applied Nursing Research, 1-3 Campbell Street, Liverpool BC 2170, Liverpool, NSW, Australia.
Email: steven.frost@health.nsw.gov.au

Abstract

Most clinical trials use null hypothesis significance testing with frequentist statistical inference to report *P* values and confidence intervals for effect estimates. This method leads to a dichotomisation of results as 'significant' or 'non-significant'. A more nuanced interpretation may often be considered and in particular when the majority of the confidence interval for the effect estimate suggests benefit or harm. In contrast to the frequentist dichotomised approach based on a *P* value, the application of Bayesian statistics allocates credibility to a continuous spectrum of possibilities and for this reason a Bayesian approach to inference is often warranted as it will incorporate uncertainty when updating our current belief with information from a new trial. The use of Bayesian statistics is introduced in this paper for a hypothetical sepsis trial with worked examples in the R language for Statistical Computing environment and the open-source statistical software JASP. It is hoped that this general introduction to Bayesian inference stimulates some interest and confidence among clinicians to consider applying these methods to the interpretation of new evidence for interventions relevant to anaesthesia and intensive care medicine.

1 | INTRODUCTION

Clinicians must interpret existing knowledge and new evidence as it arises, from well designed, conducted and reported clinical trials to guarantee the best quality of patient care. Clinical evidence is typically collected in an incremental and iterative process where new information is added to existing knowledge. However, the reporting of results from many trials often leads to uncertainty among clinicians on how to interpret a trial's outcomes with the translation of research into practice at times also challenged by prior established practices and beliefs. Traditionally, clinical trials are reported using *P* values and confidence intervals (CI) relevant to the study hypothesis (most commonly the null hypothesis of zero difference) and effect estimate (often the odds or risk ratio). This approach to inference of trial results is referred to as frequentist and uses data from a single trial in isolation and assigns the probability that the observed outcome has arisen by chance from a hypothetical number of repetitions of the trial (but not that the findings are erroneous or that the hypothesis is false). By convention, a threshold probability of $P < .05$ given the power of the trial is used as a compromise between type-1 (false positive) and type-2 (false negative) errors to reject the assumption of a chance finding. This approach furthermore leads to a dichotomisation of results that are reported as 'significant' or 'non-significant' purely based on frequentist statistical inference with the 'non-significant' result often receiving a connotation of a 'negative' trial or showing an 'absence' of effect.[1,2] It comes as no surprise that the *P* value has been criticised with calls made for an alternative approach to inference.[3-7] Using a Bayesian approach of inference, new trial results are considered in the context of existing information (please refer to the Appendix S1 for an explanation of Bayes theorem in an

A/B test as used in this text). It aims to update current knowledge or the prior probability of trial results gained from previous studies into a posterior probability revised by the new trial result.[8-11] The belief in the prior probability will, and rightly so, vary among clinicians but the Bayesian approach appears intuitive to the well-informed clinician eager to consider new trial data.[12] Whether doubtful or optimistic about the study results, any analysis to assess difference, superiority, inferiority or futility should be able to convince a sceptic, or an optimist alike. Fundamentally, Bayesian inference is reallocation of credibility across possibilities and therefore a key step is to define the dataset of possibilities for which credibility is allocated. In contrast to the frequentist dichotomised approach based on a *P* value, the Bayesian approach allocates credibility to a continuous spectrum of possibilities with 95% of the most credible range of the posterior distribution of possibilities contained within the highest density interval (HDI). For this reason, a Bayesian approach to inference is superior to traditional frequentist approaches as it will incorporate uncertainty when updating our current belief with information from a new trial.[13] Some recent examples of trials with the primary outcome reported as non-significant using frequentist inference that still demonstrated evidence for benefit using Bayesian inference include EOLIA,[14] OPTIMISE[15] and ANDROMEDA-SHOCK.[16]

In this commentary, we outline a Bayesian approach to inference with an illustrative hypothetical trial and provide guidance to some available statistical resources. Importantly, we aim to show that the uncertainty of the results of clinical trials can be better presented by a Bayesian approach to inference. We will explore the hypothetical trial data as an introduction to Bayesian inference using the R language for Statistical Computing environment[17] and the open-source statistical software JASP[18] (the reader is encouraged to download either software for use with the example and clinical trial data provided in the Supplement). The examples are conducive to the 'A/B test' that is common to many clinical trials evaluating the proportion of successes or failures in an intervention group compared to a control group. We hope that this general introduction will stimulate more clinicians to evaluate and incorporate new knowledge from recent clinical trials into current practice beyond the limitations of the frequentist *P* value.

2 | A NEW THERAPY FOR SEPTIC SHOCK

An open-source code that can be copied and pasted into the R console is provided in Supplement (file name 'SepsisTrial'). In JASP, the Bayesian A/B test is found under the test menu 'Frequencies' and has been set as default in the Supplemental file (SepsisTrial.jasp). Files relevant to the recent clinical Bayesian studies are also available in the Electronic Supplement (file names including the study acronyms, cf. Table 1). Consider a fictional clinical trial evaluating a new therapy in the intensive care setting to reduce 90-day mortality in patients with septic shock. Currently, the 90-day mortality in septic shock is approximately 40%.[19] The investigators propose the new therapy can reduce this rate to 30% (10% absolute risk reduction[20]) with statistical significance set at 0.05 (type-1 error rate) using two-sided tests and statistical power set at 0.80 (type-2 error rate of 20%). It is estimated that approximately 360 study participants are required in each arm of the trial (being randomly allocated 1:1).

Editorial comment

Traditional significance testing based on *P* value limits and a statistical result derived from one set of observations can be misleading in some situations, and there is a lively general discussion about the value of significance testing in this way in general. This narrative review demonstrates how using prior probabilities and Bayesian analysis can be used to help assess new findings where traditional significance testing may not be easily interpreted.

3 | RESULTS FROM THE NEW STUDY

The trial is completed and the rates of 90-day mortality reported for the intervention and control groups are 32.2% (116/360) and 38.8% (140/360) respectively (Table 1, top shaded row).

The estimate of effect of the new intervention versus usual care is an odds ratio of 0.75 for decreasing 90-day mortality with the 95% CI estimated to range from 0.55 to 1.01 ($P = .062$ for a chi-squared test, $P = .073$ for the Fisher's exact test) (Please note that JASP reports the odds ratio as the natural logarithm depicted as 'log', whereas this manuscript text uses 'ln').

4 | INTERPRETATIONS OF THE STUDY RESULTS

A frequentist view of the *P* value would lead to an interpretation that the new treatment had no effect since the null hypothesis of zero difference cannot be rejected based on a non-significant $P > .05$ and exclude the intervention from clinical practice. Many would agree that other factors deserve consideration as well, in particular when the majority of the confidence interval suggests a survival benefit.[21] In a Bayesian analysis, the trial results are treated as a probability distribution. The mean of the distribution would be equal to the natural logarithm (ln) of the odds ratio, ln (0.75) = −0.288, with an approximate standard deviation on either side of the mean of [ln(1.01) − ln(0.55)]/(1.96[1] × 2) = 0.159 (Figure 1, left).

The proportion of the lower tail of this distribution below the null-hypothesis, that is there is no treatment benefit (an odds ratio of ≤1 and hence ln 1 ≤ 0) for a one-sided test is 0.073/2 = 0.0365. This *P* value is a probability statement to describe the chance of the

[1] The number 1.96 is the standard score for 97.5% of the normal distribution being within that limit based on a 95% confidence interval with 5% split between both end-tails.

TABLE 1 Observed events of 90-day mortality among hypothetical trial participants (shaded row). The data from recent clinical trials that have been re-analysed using Bayesian inference are listed below. All data analyses files are provided in the Supplement

Study	No. of events/N	Event rate	Estimate (95% CI)	*P* value
Hypothetical sepsis trial				
Control	140/360	38.9%	OR = 0.75 [0.55-1.01]	χ^2 test = .062
Intervention	116/360	32.2%		Fisher's exact test = .073
EOLIA[14]				
Control	57/125	45.6%	RR = 0.78 [0.57-1.06]	χ^2 test = .104
Intervention	44/124	35.5%		Fisher's exact test = .121
OPTIMISE[15]				
Control	158/364	43.4%	RR = 0.75 [0.56-1.01]	χ^2 test = .061
Intervention	134/366	36.6%		Fisher's exact test = .070
ANDROMEDA- SHOCK[16]				
Control	92/212	43.4%	OR = 0.70 [0.63-1.02]	χ^2 test = .073
Intervention	74/212	34.9%		Fisher's exact test = .091

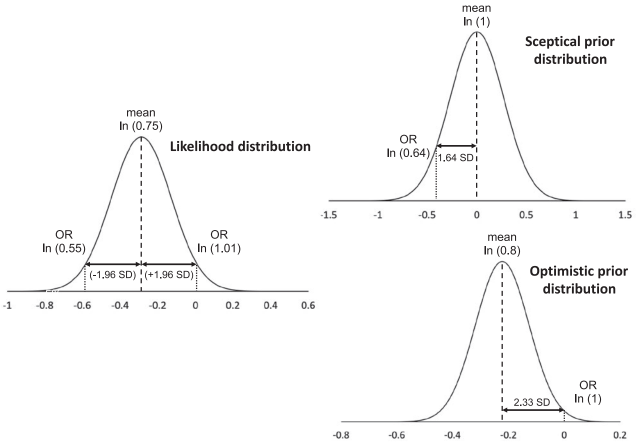

FIGURE 1 Normal probability distributions for a hypothetical sepsis trial. The trial results are shown on the left side with the probability distribution for the odds ratio (OR) of 0.75 and its 95% confidence interval of 0.55-1.01. A sceptical prior is shown top right with an assumed OR of 1 and a 1/20 chance of the treatment effect exceeding an OR of 0.64. An optimistic prior is shown bottom right with an assumed OR of 0.8 and a 1/100 chance of the treatment effect leading to harm OR > 1

trial results (*y*) given no beneficial treatment effect (*noTE*). This conditional probability can be represented as follows:

$$p(y|noTE) = 0.0365.$$

This new trial result (*y*) or information is referred to as the likelihood in Bayesian terminology.[13] A frequentist interpretation of this *P* value is often given as: in the long-run frequency of repeated trials of a similar size, conditional on no-effect of the intervention and sampling randomly from a null population, it is expected that in approximately 4 of every 100 instances would the results demonstrate a lack of improvement or worsened odds for survival. However, a purist would say it is the probability of the test-statistic, not the data itself.[13]

Now let us consider how we could incorporate some prior belief of potential treatment effect. The rationale is outlined in the Appendix S1 ‘Bayes theorem for general quantities’. Let us first take a very sceptical view, in that we consider there is no treatment effect (odds ratio = 1.0), and that there is less than a 1/20 (5%) chance of the treatment effect being greater, that is lesser odds ratio, than the one used to estimate the required sample size for the fictional trial above (30% vs 40%, an odds ratio = (0.30/0.70)/(0.40/0.60) = 0.64). The distribution of this prior belief would have a mean of no-effect, ln(1.0) = 0, with a standard deviation of ln (1.0) − ln (0.64))/(1.64[2]) = 0.272 (Figure 1, top right).

A logarithmic scale is used so the distribution can be considered Gaussian and is equally applicable to rate ratios and hazard ratios from reported trials. This assumption of normality of the data makes it an easy step to apply standard probability theory, in that ±1.96 standard deviations from the mean value, would represent the 95% confidence interval of a given distribution.

A Bayesian approach to inference will now combine the sceptical view as the prior and the study result as the likelihood to estimate

[2] The number 1.64 is the standard score for 95% of the normal distribution being within that limit based on a 1/20 chance.

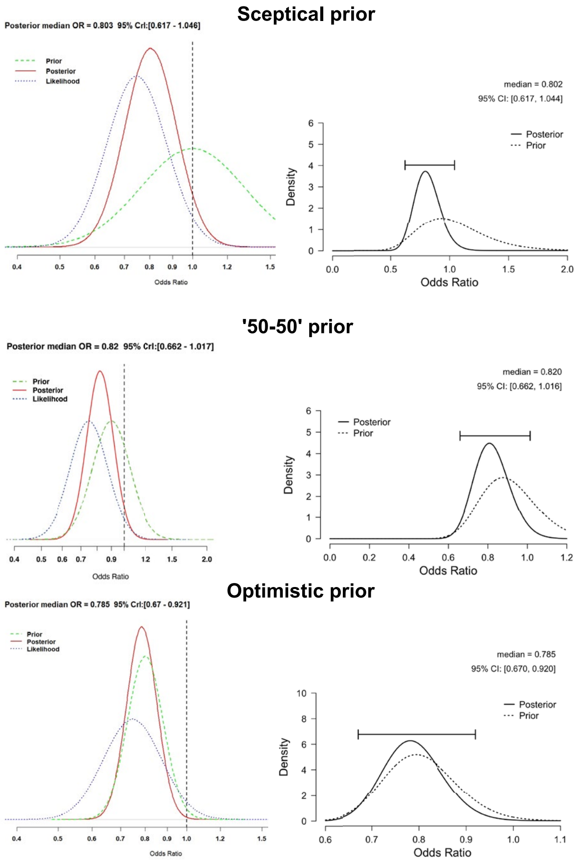

FIGURE 2 Prior, likelihood (for the output generated in R only) and posterior distribution of a hypothetical sepsis trial. R output on the left and JASP output on the right (please refer to files in the Supplement). Sceptical prior with a mean OR = 1.0 and a 1/20 chance of OR < 0.64 (top). The posterior probability for an OR < 1.0, 0.9, 0.8, 0.7, 0.64 being 0.95, 0.80, 0.49, 0.16 and 0.05 respectively. Optimistic prior with a mean OR = 0.80 and 1/100 change of OR > 1.0 (bottom). The posterior probability for an OR < 1.0, 0.9, 0.8, 0.7, 0.64 being 0.99, 0.95, 0.59, 0.08 and 0.006 respectively. Prior representing clinical equipoise ('50-50') as an average of the sceptical and optimistic prior odds ratios with a mean OR = 0.90 (middle). The posterior probability for an OR < 1.0, 0.9, 0.8, 0.7, 0.64 being 0.97, 0.80, 0.41, 0.08 and 0.012 respectively [Colour figure can be viewed at wileyonlinelibrary.com]

what is now the posterior distribution. The prior and the likelihood curves are multiplied and the total area under the resulting curve is made equal to 1.[8] This posterior distribution is then used as a probability distribution to estimate a posterior probability of a given treatment effect.

Figure 2 gives the prior, likelihood (in the R output) and posterior distribution using a sceptical prior (top). The posterior probability of *any* treatment effect (odds ratio < 1.0) was estimated to be 95%, and 80% for a treatment effect of at least 10% (odds ratio < 0.90). However, the prior to posterior analysis, has estimated that it is unlikely the treatment effect is beyond 20% (odds ratio < 0.80), as the posterior probability was estimated to be only 49%.

Let us then take a look at the posterior distribution using an optimistic prior of a treatment effect of 20% (odds ratio = 0.8) and a 1/100 (1%) chance of harm, that is the upper limit exceeding the null, >1.0 or ln (1) − ln (0.8)/(2.33[3]) = 0.096 (Figure 1 bottom right).

The posterior probability distribution from this optimistic prior also supports a treatment effect, but importantly, it communicates that even with an optimistic prior belief of a treatment effect, there is little evidence that the effect is as extreme as that proposed from the results of the trial alone (posterior probability of odds ratio < 0.80, was estimated to be 59%, 8% for an odds ratio < 0.7 and 0.6% for a OR < 0.64) (Figure 2, bottom).

Another potential prior could be a mixture between the above sceptical and optimistic priors, an example of true clinical equipoise. This prior would have a mean odds ratio of 0.9, with a ln(standard deviation) of 0.156, based on a trial among 720 study participants (see Supplement for method of calculation). The posterior probability from this mixed prior also suggests a treatment effect, with little evidence of an effect of more than 20% (posterior probability of odds ratio < 1.0, was estimated to be 97%, 80% for an odds ratio < 0.9, and 0.41% for a OR < 0.8) (Figure 2, middle).

Finally, let us consider the potential argument that our fictional study, despite its sample size calculation, did not reach sufficient power to demonstrate an effect still perceived to be real, a case not seldom made by clinicians practically adopting a Bayesian perspective of an optimistic prior. If we were to increase the sample size of the trial, from 360 to 500 participants, with the same ratio of events between the new treatment and usual care groups, the odds ratio is similar for both trials (0.75), but as expected the 95% CI narrows to [0.57-0.96] and the associated *P* value (0.025) is now statistically significant at the <0.05 level. Table 2 shows the effect of the increased sample size and essentially an increased weighting of the trial data when combined with the sceptical prior to generate the distribution of posterior probability. Importantly, this shows how a traditional frequentist approach to inference would consider the treatment effective with these results in a larger cohort while the Bayesian estimates deliver a consistent message of a high probability of a treatment effect, that is unlikely to exceed a 30% reduction in risk.

The calculation of the posterior distribution and the so-called Bayesian approach to inference is given in the detailed text by David

TABLE 2 Potential treatment effects and posterior probability from a sceptical prior belief for the hypothetical trial with its initial (360 participants) and expanded (500 participants) sample size

Treatment effect Odds ratio	Posterior probability 360 participants	Posterior probability 500 participants
<1.0	95%	98%
<0.9	80%	87%
<0.8	49%	56%
<0.7	16%	16%
<0.64	5%	4%

[3] The number 2.33 is the standard score for 99% of the normal distribution being within that limit based on a 1/100 chance.

Spiegelhalter et al,[13] with many examples and references to many other applications beyond clinical trials.

CONFLICTS OF INTEREST

None of the authors has any conflicts of interest to declare relevant to this manuscript. No funding was received for the work presented in this manuscript.

ORCID

Steven A. Frost https://orcid.org/0000-0002-8879-0486
Anders Aneman https://orcid.org/0000-0003-2096-5304

REFERENCES

1. Charlesworth M, Choi SW. Non-inferiority studies: is 'better' the enemy of 'good enough'? *Anaesthesia*. 2018;73:1162-1164.
2. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995;311:485.
3. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019;567:305-307.
4. Laber EB, Shedden K. Statistical significance and the dichotomization of evidence: the relevance of the ASA statement on statistical significance and p-values for statisticians. *J Am Stat Assoc*. 2017;112:902-904.
5. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2:e124.
6. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. *R Soc Open Sci*. 2014;1:140216.
7. Ioannidis JPA. What have we (not) learnt from millions of scientific papers with p values? *Am Stat*. 2019;73:20-25.
8. 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.
9. Ferreira D, Barthoulot M, Pottecher J, Torp KD, Diemunsch P, Meyer N. Theory and practical use of Bayesian methods in interpreting clinical trial data: a narrative review. *Br J Anaesth*. 2020;125:201-207.
10. Sidebotham D. Are most randomised trials in anaesthesia and critical care wrong? An analysis using Bayes' theorem. *Anaesthesia*. 2020;75(10):1386-1393.
11. Charlesworth M, Pandit JJ. Negative outcomes in critical care trials: applying the wrong statistics - or asking the wrong questions? *Anaesthesia*. 2020;75(10):1284-1288.
12. Gill CJ, Sabin L, Schmid CH. Why clinicians are natural bayesians. *BMJ*. 2005;330:1080-1083.
13. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. John Wiley & Sons; 2004.
14. Goligher EC, Tomlinson G, Hajage D, 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.
15. Ryan EG, Harrison EM, Pearse RM, Gates S. Perioperative haemodynamic therapy for major gastrointestinal surgery: the effect of a Bayesian approach to interpreting the findings of a randomised controlled trial. *BMJ Open*. 2019;9:e024256.
16. Zampieri FG, Damiani LP, Bakker J, et al. Effects of a resuscitation strategy targeting peripheral perfusion status versus serum lactate levels among patients with septic shock. A Bayesian reanalysis of the ANDROMEDA-SHOCK trial. *Am J Respir Crit Care Med*. 2020;201:423-429.
17. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2017.
18. Team J. JASP. Version 0.13.1 edn, 2020.
19. Vincent JL, Jones G, David S, Olariu E, Cadwell KK. Frequency and mortality of septic shock in Europe and North America: a systematic review and meta-analysis. *Crit Care*. 2019;23:196.
20. Huang DT, Angus DC, Barnato A, et al. Harmonizing international trials of early goal-directed resuscitation for severe sepsis and septic shock: methodology of ProCESS, ARISE, and ProMISe. *Intensive Care Med*. 2013;39:1760-1775.
21. Young PJ, Nickson CP, Perner A. When should clinicians act on non-statistically significant results from clinical trials? *JAMA*. 2020;323(22):2256.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Frost SA, Alexandrou E, Schulz L, Aneman A. Interpreting the results of clinical trials, embracing uncertainty: A Bayesian approach. *Acta Anaesthesiol Scand*. 2021;65:146–150. https://doi.org/10.1111/aas.13725