

Bayesian meta-analysis in the 21st century: Fad or future of evidence synthesis?

Shruthi C. Iyer¹ MBBS, Qai Ven Yap² BSc, John Soong^{3,4} MBChB, Matthew Cove⁴ MBChB

ABSTRACT

Meta-analyses are pivotal in evidence-based medicine. They synthesise findings from multiple studies to inform clinical practice. Traditionally, frequentist statistical methods have dominated this field. However, growing recognition of their limitations in interpreting uncertainty and integrating prior knowledge has led to increased interest in Bayesian methods. We compared the Bayesian and frequentist approaches to meta-analysis, assessed their respective strengths and limitations, and evaluated how each technique influences the interpretation of clinical trial data, particularly when evidence is ambiguous. A narrative review was conducted to understand and compare key aspects of frequentist and Bayesian statistics. Real-world clinical examples were used to illustrate the differences between how each statistical paradigm interprets the same dataset. Frequentist methods were assessed based on null hypothesis testing and *P* values, while Bayesian methods were evaluated for their use of prior distributions and posterior probabilities. We demonstrated how statistical significance of the obtained result may differ based on whether a frequentist or Bayesian statistical analysis was employed. In most cases, Bayesian methods enabled increased intuitiveness and flexibility of interpretations by incorporating uncertainty and pre-existing evidence. While frequentist methods remain standard for simpler univariate analyses, the Bayesian approach offers a more holistic framework for complex meta-analyses, as it aligns more closely with practical clinical decision-making. This places Bayesian inference a valuable tool in the future of evidence synthesis.

Keywords: Bayesian, biostatistics, frequentist, meta-analysis, statistics

INTRODUCTION

Evidence-based medicine is the cornerstone of modern clinical decision-making, but not all evidence is practice-changing. To aid clinicians, there is an accepted hierarchy to classify published evidence, often described as the evidence pyramid. Weaker study designs are found at the bottom of the pyramid,

with more robust study methods towards the top (Fig. 1).^{1,2} Systematic reviews and meta-analyses synthesise existing evidence and sit at the top of the pyramid, significantly influencing education, future research and clinical care delivery.

A meta-analysis is a statistical tool for compiling results from multiple scientific studies seeking to answer the same research question. This is essential because the results of one study may not apply under different circumstances.^{3,4} Combining studies through a meta-analysis helps to draw nuanced conclusions as it allows researchers to estimate the mean effect size across different populations while performing subgroup analyses. Meta-analyses also provide an assessment of the reliability of results.^{5,6}

There are 2 primary approaches to statistics in meta-analyses: the frequentist and Bayesian methods. They are differentiated by their interpretation of uncertainty, which refers to the degree of inherent variability in a measured value.⁶ Until recently, the reporting of meta-analyses was largely dominated by frequentist statistics. However, with the emergence of platform trials and a growing recognition of the limitations of arbitrary significance thresholds, Bayesian methods have gained traction.³ Moreover, the Bayesian approach is viewed as more intuitive and aligns better with cumulative science.⁷ How do the 2 approaches differ, and what are the implications for the future of meta-analyses?

An overview of frequentist statistics

Frequentist statistical inference has long been the preferred method for clinical trials and meta-analyses. This approach assumes that a fixed null hypothesis is true, following which long-term probabilities are calculated using only the data obtained (i.e. assessing how probable the dataset is given the null hypothesis). Therefore, any uncertainty in a frequentist analysis arises solely from sampling error. This can be understood by considering the outcome of flipping a coin indefinitely. The null hypothesis posits that the

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ Yong Loo Lin School of Medicine, National University of Singapore, Singapore

² Department of Biostatistics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

³ Department of Medicine, National University Hospital, Singapore

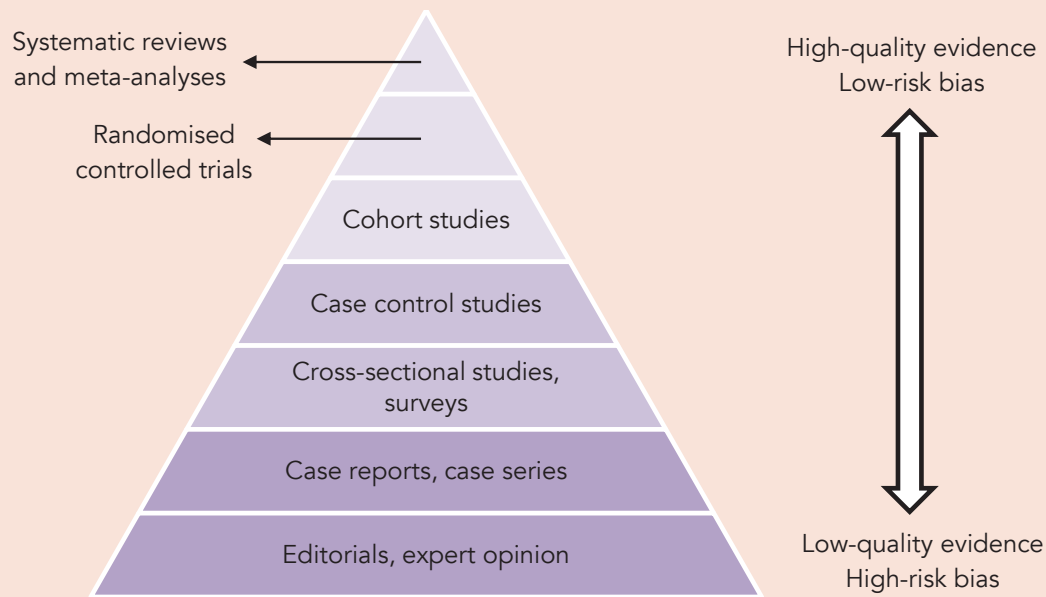
⁴ Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Correspondence: Ms Shruthi C Iyer, Department of Medicine, National University Hospital, NUHS Tower Block, Level 10, 1E Kent Ridge Road, Singapore 119228. Email: shruthi.ciyer@u.nus.edu

Accepted: 16 June 2025

Published Online First: 9 July 2025

Fig. 1. The evidence pyramid.



outcome would be heads half the time (all else being equal). Flipping the coin a few times may yield a different observation by chance (e.g. all heads). If this pattern persists in the long run, it represents a departure from the expected data, leading to the rejection of the null hypothesis.⁸ The researcher concludes that other factors must be responsible for the disparity between the observed and expected results.⁹

In a frequentist analysis, the probability of obtaining another dataset as extreme as the one collected is known as the *P* value. *P* values inform researchers of the likelihood of observing a specific dataset and help assess the compatibility of the obtained data with the null hypothesis. They do not quantify the probability of the null hypothesis itself or measure the magnitude of the treatment effect.¹⁰ A *P* value of less than 0.05 is typically regarded as an acceptable threshold for significance, allowing for the rejection of the null hypothesis. This arises from the fact that there is less than a 5% probability that the observed result is due to chance alone. However, this standard is essentially arbitrary, yet it is considered reasonable by most researchers. Challenges arise when a *P* value is very close to 0.05, as considerable doubt arises regarding the significance of the result. This may lead to a type 1 error, which occurs when the researcher erroneously rejects the null hypothesis.¹¹ The *P* value of 0.05 merely indicates the threshold at which this is deemed acceptable; it does not eliminate the risk of drawing an incorrect conclusion. In such instances, prior

knowledge becomes crucial, although frequentist inference does not provide pathways to leverage previous studies.^{12,13} The Bayesian approach addresses these shortcomings.

An overview of Bayesian statistics

The Bayesian method of statistical inference treats parameters and hypotheses as probability distributions, while the data remain fixed. This sharply contrasts with the frequentist approach, where probabilities are assigned to the observed data, and the null hypothesis is fixed and assumed to be true. Bayesian statistics is founded on Bayes' theorem, which details a method for updating probabilities using current data to build upon previous knowledge. Prior knowledge informs the prior distribution, which summarises what is already known about the subject. It can be derived from expert opinion, historical data or a combination of both. The probability of an outcome after completing a Bayesian analysis is referred to as the posterior probability (or post-test distribution). It encapsulates all the knowledge researchers have about a topic based on previous observations and the most recent experimental data.⁹

Knowledge of prior information can influence the outcome, even when all other conditions remain unchanged.⁹ A real-life example of this would be in the treatment of acute respiratory distress syndrome (ARDS).¹⁴ In 2018, the results of the extracorporeal membrane oxygenation (ECMO) to rescue lung injury in severe ARDS (EOLIA) trial were interpreted using

frequentist statistics. Subsequent interpretation of the same results using Bayesian methods yielded different outcomes.¹⁵ This emphasised how using varying pre-test probabilities in Bayesian analyses of clinical trial data can provide important information for study interpretation. This is particularly so when evidence from the trial is close to the significance threshold and it is uncertain whether current clinical practice should be changed.¹⁵

The EOLIA international clinical trial randomly assigned patients with very severe ARDS to receive immediate venovenous ECMO on diagnosis (ECMO group) or continued conventional treatment (control group), with crossover to ECMO made available for patients in the control group who had refractory hypoxaemia. The primary outcome was 60-day mortality. The trial found that 44 of 124 patients in the ECMO group and 57 of 125 patients in the control group had died. Using a frequentist statistical analysis, this is reflected as a relative risk (RR) of 0.76, 95% confidence interval (CI) of 0.55–1.04, and a *P* value of 0.09. Hence, the authors concluded that there was no significant difference in the 60-day mortality of ARDS patients treated with ECMO versus those treated with conventional mechanical ventilation that included ECMO as a rescue therapy.¹⁴

However, the results of the EOLIA trial were subsequently re-interpreted in a Bayesian analysis.¹⁵ This was because the clinically important point estimate of the absolute risk difference (11%), the

near statistical significance of the effect, and the wide divergence of pre-existing views regarding the benefits of ECMO had made interpretation of the trial controversial.¹⁵ This new study aimed to demonstrate the potential utility of Bayesian analyses by estimating the posterior probability, under various assumptions, that early ECMO was associated with reduced mortality in patients with very severe ARDS. Statistical prior distributions were developed to represent diverse levels of support for ECMO in treating lung injury in patients with severe ARDS.¹⁵ For each of the priors, the corresponding RR, credible interval (CrI) and absolute risk reduction (ARR) were estimated using Bayesian analysis. This means that the post-hoc analysis estimated the posterior probabilities, for which ECMO treatment had exceeded a range of potential values for the minimum, clinically important treatment effect (i.e. $RR < 1$, $RR < 0.9$, $RR < 0.8$, $RR < 0.67$ and $ARR \geq 2\%$, $ARR \geq 4\%$, $ARR \geq 6\%$, $ARR \geq 8\%$, $ARR \geq 10\%$ and $ARR \geq 20\%$). The results of the Bayesian inference are summarised in Table 1.

Overall, the RR of 60-day mortality in patients who received early ECMO treatment compared to conventional mechanical ventilation varied based on the prior that was used in the analysis. The more strongly the pre-test probabilities suggest that early ECMO intervention is lifesaving in severe ARDS, the higher the likelihood that the same results will strongly favour ECMO over mechanical ventilation, and vice versa. If insufficient data are available on a

Table 1. Bayesian analysis of the results of the EOLIA trial.

Prior belief	Assumed median RR	Probability of treatment effect equal to or greater than specified threshold			
		RR<1	RR<0.9	RR<0.8	RR<0.7
Non-informative (Equivalent to no prior data)	1.0	50%	50%	49%	49%
Strongly enthusiastic towards ECMO (Equivalent to previous RCT enrolling 100 patients finding a 33% RR reduction)	0.67	95%	89%	77%	58%
Moderately enthusiastic towards ECMO (Equivalent to previous RCT enrolling 264 patients finding a 22% RR reduction)	0.78	95%	83%	57%	24%
Sceptical towards ECMO (Equivalent to previous RCT enrolling 100 patients finding a 0% RR reduction)	1.0	50%	33%	18%	7%
Strongly sceptical towards ECMO (Equivalent to previous RCT enrolling 264 patients finding a 0% RR reduction)	1.0	50%	24%	7%	1%
Prior informed by previous studies	0.71	99%	48%	Not stated	Not stated

ECMO: extracorporeal membrane oxygenation; EOLIA: extracorporeal membrane oxygenation to rescue lung injury in severe acute respiratory distress syndrome; RCT: randomised controlled trial; RR: relative risk

topic, then an unbiased, non-informative prior with an assumed median RR can be used, and the outcome is driven more by the new experimental data than prior beliefs. Most notably, when the prior was informed by previous studies that contained the latest experimental data on using ECMO for ARDS as of 2018, the results were convincing and conclusive that early ECMO was superior to mechanical ventilation in treating severe ARDS (RR 0.71, 95% CrI 0.55–0.94).¹⁵ This contrasts with the results of the frequentist analysis with the same results.

Bayesian analysis of clinical trial data may provide useful information in interpreting uncertain evidence. It can provide information about the posterior probability of mortality benefit under a broad set of assumptions that can be reflective of the diverse clinical scenarios and patient populations to which trial results will be applied. Incorporating prior knowledge enables researchers to evaluate how new information alters their overall understanding of a treatment. Hence, a Bayesian analysis can refine the interpretation of an outcome and guide changes in clinical practices.⁹

Comparing the Bayesian and frequentist methods

Fundamentally, frequentist statistics aim to determine how probable the observed dataset is, given that the null hypothesis is true, and provide information on our confidence in the accuracy of the obtained result. In contrast, the Bayesian method assigns probabilities to parameters and hypotheses, enabling researchers to assess the likelihood that a treatment is superior. A key advantage of the Bayesian approach is that it mirrors the nature of clinical decision-making, which is generally grounded in probability. Consequently, it is more intuitive than frequentist inference, which reports the probability of obtaining a specific dataset under the assumption that the treatments are equal.⁹

Interpreting the frequentist 95% CI is similarly less intuitive than the Bayesian 95% CrI, as it merely reflects the long-run probability of accurately estimating a parameter rather than the actual value of interest. Reporting the frequentist 95% CIs alongside the *P* value provides insights on the size and direction of an effect, with the width of the CI indicating precision.^{16,17} It informs researchers that 95% of these computed CIs would include the true value of interest if the experiment was repeated infinitely, therefore quantifying the uncertainty surrounding effect estimates.¹⁸ In contrast, the Bayesian 95% CrI indicates that there is a 95% probability that the true value of interest lies within the CrI. The interval is fixed, while the parameter of interest is treated as the random variable. It is common for the frequentist 95% CI to be misinterpreted as conveying this meaning. However, the 95% CI and the 95% CrI

are fundamentally different means of quantifying uncertainty and may be used to estimate statistical significance when they exclude null values but should not be confused between each other.¹⁸ Only the Bayesian CrI offers a direct probability statement about the actual value of interest, allowing for instinctive understanding.

Frequentist methods are more beneficial in simple univariate analyses due to their straightforward nature, whereas Bayesian methods offer versatility in more sophisticated analyses involving complex models.¹⁹ Frequentist methods are the standard for conventional univariate meta-analysis, which assesses each pair of interventions for each outcome separately. These analyses primarily concentrate on understanding the characteristics of a single variable, reducing the necessity to incorporate prior information. Frequentist statistics are also more effortlessly integrated into various software packages, making their results simpler to interpret.^{20,21} However, Bayesian methods enable more effective modelling of complicated variance-covariance structures in multivariate meta-analysis when comparing multiple outcomes. Improved computing technology has led to more programmes offering user-friendly functions for conducting Bayesian meta-analyses, thus contributing to the increasing simplicity of calculating otherwise intricate posterior probabilities and facilitating Bayesian inference.²²

Additionally, as frequentist analyses rely on conclusions drawn solely from the collected data, the conclusion remains fixed irrespective of pre-existing knowledge.⁹ In contrast, the Bayesian method integrates prior knowledge. If robust priors are available from earlier evidence, the new study data must be compelling enough to shift the posterior probability in a different direction. This illustrates how clinicians interpret new evidence in practice, and this makes Bayesian analysis more useful to inform guideline panels, guide reimbursement decisions and assist in health technology assessments.

If there is insufficient prior knowledge on the topic, an uninformative prior or multiple priors may be employed. In such instances, statistical inference is driven more by the acquired data than by prior knowledge, making it easier to overturn preconceived beliefs.⁹ In this situation, the outcomes of a Bayesian inference will not vary substantially from frequentist methods. To obtain balanced and representative results, it is pivotal to use a widely accepted prior that is informed by objective and thorough experimental data. Hence, a key limitation of using Bayesian analysis is that the difficulty of determining an accurate prior, especially when there is a paucity of pre-existing data, leaves more opportunity for the outcomes to be influenced by biased pre-test probabilities.

Finally, Bayesian meta-analyses facilitate the propagation of uncertainty in the heterogeneity parameter. In random effects models, heterogeneity is a critical parameter that reflects the variability in treatment effects across studies but can be challenging to estimate when the number of studies is limited. In frequentist methods, heterogeneity is assumed to be fixed at a certain value, which can lead to biased conclusions. In contrast, Bayesian methods treat heterogeneity as a random variable with its own distribution. Furthermore, the risk of bias is incorporated with greater accuracy. This enables the full uncertainty in the heterogeneity parameter to be carried forward throughout the analysis, resulting in more realistic conclusions.¹⁸

CONCLUSION

The choice between frequentist and Bayesian inference depends on numerous factors. Frequentist statistics ascertain whether the observed result is attributable to chance, making it a logical selection for univariate analyses. Conversely, the Bayesian approach more intuitively captures the complex interpretation of new study data by the clinician making decisions at the bedside. Consequently, it may be more suitable for a complex, multivariate meta-analysis, where the goal is to determine whether new evidence alters current beliefs.

Ethics statement

Not applicable.

Declaration

The authors declare they have no affiliations or financial involvement with any commercial organisation with a direct financial interest in the subject or materials discussed in the manuscript. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. No generative artificial intelligence technologies were used.

REFERENCES

1. Yetley EA, MacFarlane AJ, Greene-Finestone LS, et al. Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group. *Am J Clin Nutr* 2017;105:249S-85S.
2. Murad MH, Asi N, Alsawas M, et al. New evidence pyramid. *Evid Based Med* 2016;21:125-7.
3. Dekkers OM. Meta-analysis: Key features, potentials and misunderstandings. *Res Pract Thromb Haemost* 2018;2:658-63.
4. Howick J, Glasziou P, Aronson JK. Can understanding mechanisms solve the problem of extrapolating from study to target populations (the problem of 'external validity')? *J R Soc Med* 2013;106:81-6.
5. Hackenberger BK. Bayesian meta-analysis now - let's do it. *Croat Med J* 2020;61:564-68.
6. Thompson CG, Semma B. An alternative approach to frequentist meta-analysis: A demonstration of Bayesian meta-analysis in adolescent development research. *J Adolesc* 2020;82:86-102.
7. Zagaria A, Lombardi L. Bayesian versus Frequentist approaches in Psychometrics: a bibliometric analysis. *Discov Psychol* 2024;4:61.
8. Lammers D, Richman J, Holcomb JB, et al. Use of Bayesian Statistics to Reanalyze Data From the Pragmatic Randomized Optimal Platelet and Plasma Ratios Trial. *JAMA Netw Open* 2023;6:e230421.
9. Fornacon-Wood I, Mistry H, Johnson-Hart C, et al. Understanding the Differences Between Bayesian and Frequentist Statistics. *Int J Radiat Oncol Biol Phys* 2022;112:1076-82.
10. Tony O'Hagan. Dicing with the Unknown. *Significance* 2004;1:132-3.
11. Shreffler J, Huecker MR. Type I and Type II Errors and Statistical Power. [Updated 2023 Mar 13]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025.
12. du Prel JB, Hommel G, Röhrig B, et al. Confidence interval or p-value?: part 4 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2009;106:335-9.
13. Kwak S. Are Only p-Values Less Than 0.05 Significant? A p-Value Greater Than 0.05 Is Also Significant! *J Lipid Atheroscler* 2023;12:89-95.
14. Combes A, Hajage D, Capellier G, et al. Extracorporeal Membrane Oxygenation for Severe Acute Respiratory Distress Syndrome. *N Engl J Med* 2018;378:1965-75.
15. 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;320:2251-9.
16. Ranstam J. Why the P-value culture is bad and confidence intervals a better alternative. *Osteoarthritis Cartilage* 2012;20:805-8.
17. Hoekstra R, Morey RD, Rouder JN, et al. Robust misinterpretation of confidence intervals. *Psychon Bull Rev* 2014;21:1157-64.
18. Hespanhol L, Vallio CS, Costa LM, et al. Understanding and interpreting confidence and credible interval around effect estimates. *Braz J Phys Ther* 2019;23:290-301.
19. Al Amer FM, Thompson CG, Lin L. Bayesian Methods for Meta-Analyses of Binary Outcomes: Implementations, Examples, and Impact of Priors. *Int J Environ Res Public Health* 2021;18:3492.
20. Grant RL. The uptake of Bayesian methods in biomedical meta-analyses: A scoping review (2005–2016). *J Evid Based Med* 2019;12:69-75.
21. Borenstein M, Hedges LV, Higgins JPT, et al. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 2010;1:97-111.
22. Wei Y, Higgins JPT. Bayesian multivariate meta-analysis with multiple outcomes *Stat Med* 2013;32:2911-34.