Patel and Green *Critical Care* (2025) 29:73
https://doi.org/10.1186/s13054-025-05307-9

Critical Care

PERSPECTIVE

Open Access

Death by *p*-value: the overreliance on *p*-values in critical care research

Sharad Patel[1*] and Adam Green[1]

Abstract

The *p*-value has changed from a versatile tool for scientific reasoning to a strict judge of medical information, with the usual 0.05 cutoff frequently deciding a study's significance and subsequent clinical use. Through an examination of five critical care interventions that demonstrated meaningful treatment effects yet narrowly missed conventional statistical significance, this paper illustrates how rigid adherence to *p*-value thresholds may obscure therapeutically beneficial findings. By providing a clear, step-by-step illustration of a basic Bayesian calculation, we demonstrate that clinical importance can remain undetected when relying solely on *p*-values. These observations challenge current statistical paradigms and advocate for hybrid approaches—including both frequentist and Bayesian methodologies—to provide a more comprehensive understanding of clinical data, ultimately leading to better-informed medical decisions.

Keywords Bayesian analysis, Fisher's exact test, *P*-value, Critical care research, Re-analysis

Background

Medical research stands at a crossroads where statistical significance, often judged by *p*-values, risks overshadowing clinical relevance. Originally popularized by Sir Ronald Fisher in the early 20th century [1], the *p*-value was intended as a measure of how incompatible observed data might be with a null hypothesis [2, 3]. Over the years, the *p*-value shifted from being a flexible exploratory tool to an unofficial guardian of scientific "truth" in clinical research [1]. Many journals and funding bodies continue to use the 0.05 threshold as a defining standard of significance, prompting investigators to plan and interpret their studies primarily around meeting this benchmark [1, 2].

This phenomenon is especially notable in critical care. Intensive care units face highly diverse patient populations, complex diseases, and urgent treatment decisions [4–6]. Obtaining the large, homogeneous samples required for consistently achieving $p < 0.05$ can be challenging [7]. As a result, numerous trials yield findings suggestive of clinically meaningful benefits but do not attain "significance" by strict frequentist criteria, leaving potentially valuable interventions underused [2, 3, 8, 9].

In this paper, we focus on two main issues. First, we highlight the pitfalls of relying solely on a rigid *p*-value threshold, which can understate the clinical potential of interventions that fail to cross the arbitrary 0.05 line. Second, we present a straightforward Bayesian approach for estimating the probability of true benefit, even when the *p*-value is slightly above 0.05. To make these points tangible, we discuss five critical care interventions—each with compelling biological plausibility and near-miss *p*-values—to illustrate how probabilistic reanalysis may reveal high likelihoods of clinically important effects.

Main text

Search strategy and selection

In planning a structured approach to identify and evaluate underrecognized interventions in critical care, we

*Correspondence:
Sharad Patel
patel-sharad@cooperhealth.edu
[1] Department of Critical Care Medicine, Cooper University Health Care and Cooper Medical School of Rowan University, 1 Cooper Plaza, Camden, NJ 08103, USA

balanced transparency with practicality. Although this project did not fulfill all requirements of a formal systematic review (e.g., PRISMA registration or comprehensive risk-of-bias checks), we employed reproducible steps, sufficient detail, and predefined inclusion criteria so that interested readers could adapt or replicate our method [10].

We searched PubMed, Embase, and the Cochrane Library (2010–2023) for studies evaluating critical care interventions in adult populations (e.g., sepsis, acute respiratory distress syndrome [ARDS], or other severe Intensive Care Unit [ICU]-relevant conditions). We focused on trials reporting at least one of the following:

- A 10% or greater relative reduction in mortality or another clinically significant ICU outcome.
- A decrease of 2 or more days in ICU or hospital length of stay.
- A 2-point or greater improvement in a validated organ dysfunction score (e.g., sequential organ failure assessment [SOFA]).

Because our goal was to highlight near misses, we included only studies where the primary outcome's *p*-value landed between 0.05 and 0.10, reasoning that such borderline results exemplify the pitfalls of rigid significance thresholds [1, 11]. We further restricted our selection to prospective, randomized trials published in peer-reviewed journals [2, 3].

After compiling an initial set of potential articles, we purposively chose five studies for illustrative depth. These examples highlight how a Bayesian perspective can uncover clinically meaningful probabilities of benefit, even when *p*-values exceed 0.05.

Bayesian approximation

Because patient-level data were unavailable, we conducted approximate Bayesian analyses using each paper's published point estimates and 95% confidence intervals [8, 11, 12]. Specifically, we:

- Defined a neutral prior centered at zero effect (e.g., 0% mortality benefit) with a standard deviation (SD) reflecting typical effect sizes in critical care (± 10%).
- Constructed a likelihood based on the reported effect size and derived the standard error from the 95% confidence interval, assuming approximate normality [11, 12].
- Applied Bayes' theorem to combine our neutral prior with the likelihood, producing a posterior distribution for the intervention's true effect.
- Calculated the probability that the intervention's actual benefit exceeded a clinically relevant threshold, such as a 10% mortality reduction or a 2-day decrease in length of stay.

While not a full Bayesian reanalysis—ideally requiring patient-level data—this simplified approach can still highlight real clinical value masked by borderline *p*-values [8, 10, 12, 13].

Etomidate vs ketamine for rapid sequence intubation

Context

In critically ill patients requiring emergency intubation, the choice of induction agent can affect both hemodynamics and organ function. Etomidate maintains stable blood pressure but may cause adrenal suppression. Ketamine preserves blood pressure but has potential neuropsychiatric effects. In a multicenter randomized controlled trial, patients received either etomidate (0.3 mg/kg) or ketamine (2 mg/kg) for rapid sequence intubation across multiple French ICUs [14].

Key finding

The primary outcome of this study was max SOFA score during a three-day period. The etomidate group demonstrated a SOFA score max of 10.3 (SD 3.7), while the ketamine group had a max of 9.6 (SD 3.9) with a difference of 0.7 points (95% CI, 0.0–1.4, $p = 0.056$). Notably adrenal insufficiency was significantly more common with etomidate (OR 6.7, 95% CI 3.5–12.7).

Bayesian reanalysis

- Neutral prior: Mean difference 0 SOFA points, SD 2 points.
- Posterior probability: ~ 75–80% chance that ketamine lowers the 3-day SOFA score by ≥ 0.5 points.

Interpretation

Despite a borderline *p*-value (0.056), the modest improvement in organ function with ketamine and the high rate of adrenal insufficiency from etomidate suggest that ketamine may be preferable in certain critically ill patients, especially when adrenal compromise is a concern.

Beta-blockade in septic shock

Context

Septic shock can involve profound hyperdynamic circulation. Morelli et al. explored whether controlling heart rate with a short-acting beta-blocker could improve hemodynamic stability without causing hypoperfusion [15]. Although the idea of giving beta-blockers to hypotension-prone patients was initially controversial, the physiological rationale is to reduce harmful tachycardia.

Key finding

The investigators observed a 12% absolute reduction in 28-day mortality with beta-blockade, but the 95% CI crossed zero, leading to $p = 0.07$. Under frequentist conventions, this near-miss p-value might discourage clinicians from exploring beta-blockers further.

Bayesian reanalysis

- Neutral prior: Mean 0%, SD 10%.
- Posterior probability: ~ 85–90% that beta-blockers confer a ≥ 10% mortality benefit.

Interpretation

The high probability of benefit, coupled with a strong physiological basis, underscores why borderline results should not be dismissed outright. Further prospective trials or subgroup analyses could clarify patient populations most likely to benefit from carefully titrated beta-blockade in septic shock.

Early vs late tracheotomy

Context

Among critically ill patients expected to require prolonged mechanical ventilation, determining the optimal timing for tracheotomy remains a significant clinical question. Early tracheotomy (around days 6–8 of intubation) may reduce sedation needs, facilitate weaning, and potentially lower the incidence of ventilator-associated pneumonia (VAP), but it also involves procedural risks. Terragni et al. conducted a multicenter, randomized trial across 12 Italian ICUs (600 adult patients) to compare early (days 6–8) vs. late (days 13–15) tracheotomy in reducing pneumonia and resource utilization (e.g., ventilator and ICU days) [16].

Key finding

VAP were diagnosed in 14% of patients in the early tracheotomy group vs 21% in the late group ($p = 0.07$). Hazard ratio for developing VAP with early vs late tracheotomy was 0.66 (95% CI, 0.42–1.04). Although the difference narrowly missed conventional significance ($p < 0.05$), the direction suggests a possible reduction in pneumonia with early tracheostomy. The study also showed a reduction in time on mechanical ventilation (HR 0.70, 95% CI 0.56–0.87), time in the ICU (HR 0.73, 95% CI, 0.55–0.97, and death (HR 0.70, 95% CI 0.56–1.15).

Bayesian reanalysis

- Neutral prior: Assume a 0% difference in VAP incidence, with a moderate standard deviation to encompass plausible clinical effects (e.g., ± 10%).
- Posterior probability: Given the observed 7% absolute difference (14% vs. 21%) but a p-value of 0.07, a Bayesian approach might still yield a meaningful probability (> 75%) that early tracheotomy reduces pneumonia by ≥ 5%.

Interpretation

Despite not achieving *frequentist* significance, the trend toward reduced VAP in the early tracheotomy group may be clinically relevant. The hazard ratios also suggest fewer ventilator and ICU days with earlier intervention. Because even modest reductions in nosocomial infections and ICU stay can translate into substantial benefits and cost savings, these borderline findings warrant careful consideration and further study. The decision on timing should balance the procedural risks against the potential for decreased ventilator exposure and pneumonia.

Hydrocortisone in severe traumatic brain injury

Context

Pneumonia is a frequent and serious complication in patients with severe traumatic brain injury (TBI). Asehnoune et al. evaluated whether low-dose hydrocortisone could lower the incidence of hospital-acquired pneumonia (HAP) [17]. By tempering the physiologic stress response, corticosteroids might reduce inflammatory damage and improve respiratory outcomes.

Key finding

HAP by day 28 was 45.8% with hydrocortisone vs. 53.3% with placebo (hazard ratio 0.75; 95% CI, 0.55–1.03; $p = 0.07$). Although the result fell short of $p < 0.05$, a 10% absolute reduction could meaningfully impact morbidity, ICU length of stay, and healthcare costs.

Bayesian reanalysis

- Neutral prior: Mean hazard ratio = 1, SD = 0.25.
- Posterior probability: ~ 87% chance that hydrocortisone reduces pneumonia risk by at least 10%.

Interpretation

Considering the high burden of pneumonia in TBI and the potential for improved neurological recovery when secondary infections are minimized, a p-value of 0.07 should not halt further exploration. The likelihood of benefit is far from negligible.

Continuous vs interrupted chest compressions

Context

High-quality cardiopulmonary resuscitation (CPR) is vital for improving outcomes in out-of-hospital cardiac arrest. Nichol et al. compared continuous chest compressions with interrupted compressions (brief pauses for ventilation) in a large cluster-randomized trial [18]. While continuous compressions maintain coronary perfusion pressure, the trade-off is potential inadequate ventilation.

Key finding

Survival to hospital discharge was 9.0% with continuous compressions vs. 9.7% with interrupted compressions (adjusted difference −0.7 percentage points; 95% CI, −1.5 to 0.1; $p=0.07$). Although small in absolute terms, any difference in cardiac arrest outcomes can be important when applied to large populations.

Bayesian reanalysis

- Neutral prior: Mean 0%, SD 2%.
- Posterior probability: ~75% that interrupted compressions provide a ≥0.5 percentage-point improvement in survival.

Interpretation

Even a half-percent increase in survival can be meaningful in a condition with high mortality rates. A strict reliance on $p<0.05$ could discourage further investigation into the nuances of CPR technique that might save additional lives.

Synthesis of findings

Taken together, these five examples suggest that an overreliance on the $p<0.05$ cutoff may obscure high-probability signals of clinical benefit. By incorporating an approximate Bayesian view, we found substantial chances—often in the 75–90% range—that each intervention surpassed clinically relevant thresholds for mortality, ventilator-free days, or infection reduction.

All five interventions had plausible physiological underpinnings and rigorous methodologies yet have seen limited uptake in practice, seemingly because their p-values did not cross 0.05. The magnitude of these potential benefits, supported by mechanistic rationales, raises concerns that strict significance criteria may unintentionally stifle the development of promising therapies [1–3, 8].

Conclusions

Two themes emerge from our findings. First is the concern that a singular fixation on $p<0.05$ can be misleading, especially in complex critical care settings where large sample sizes are difficult to achieve [2, 4, 5]. Traditional hypothesis testing assesses how frequently results as extreme as those observed could appear by chance, rather than determining the probability that a therapy truly works [8, 9, 11, 19]. As a result, clinically important signals may be dismissed as "non-significant" despite robust biological plausibility and near-miss confidence intervals.

Second, incorporating Bayesian methods can clarify whether an intervention is likely to provide tangible benefits [8, 11, 12, 20]. By uniting prior knowledge with new data, Bayesian analysis offers a posterior probability that a treatment surpasses a clinically meaningful threshold—precisely the information needed at the bedside. Although sophisticated Bayesian re-analyses would ideally use patient-level data and possibly hierarchical or adaptive models [12, 13, 16, 21], our simplified approach shows that even basic posterior probability calculations can alter how borderline findings are perceived.

Importantly, this is not a call to abandon frequentist methods entirely. Both Bayesian and frequentist perspectives have value and combining them can yield a fuller understanding of trial results [1–3]. Nonetheless, the examples here illustrate that p-values just above 0.05 might mask high probabilities of meaningful clinical gains. In critical care, where every percentage point of mortality reduction or day saved in the ICU can matter, adopting a more flexible and probabilistic viewpoint may enhance patient outcomes.

Our analysis is constrained in multiple ways. First, we relied on published summary statistics and thus could not perform more precise Bayesian models that incorporate patient-level variability. Second, while we followed a reproducible search strategy, we did not register our methods or perform a comprehensive systematic review. Other "near-miss" trials might meet our inclusion criteria but remain unidentified. Finally, our neutral priors could be disputed by experts who might prefer different assumptions or more domain-specific prior distributions [8, 11].

Nonetheless, these limitations do not negate our core argument. Even a straightforward Bayesian assessment can shed light on potential treatment benefits hidden by borderline p-values, suggesting that more detailed Bayesian analyses—and prospective Bayesian trial designs—could bring added clarity and accelerate therapeutic progress in critical care [12, 15].

Moving forward, research groups might develop larger-scale re-analyses of borderline trials, employing both

simple and advanced Bayesian frameworks. Prospective trials can also adopt Bayesian endpoints from the outset, reducing the risk of discarding meaningful results due to *p*-values narrowly exceeding 0.05. Finally, evaluating how these near-miss interventions fare in real-world practice—through feasibility projects, pilot studies, or implementation science initiatives—can determine whether their probabilistic promise holds up under actual clinical conditions [12, 22–25].

Relying solely on $p < 0.05$ can lead to the underestimation of therapies that hold substantial promise in critical care. The five trials detailed here showed biologically credible rationales, clinically meaningful effect sizes, and high posterior probabilities of real benefit—yet did not meet the 0.05 threshold for statistical significance. By embracing both frequentist and Bayesian perspectives, focusing on the probability of exceeding clinically relevant benchmarks, and remaining open-minded about borderline findings, critical care researchers and clinicians can better align statistical practice with patient-centered outcomes. Shifting beyond a single cutoff stands to improve how we interpret evidence, prioritize further study, and ultimately serve patients in the ICU.

Abbreviations

ARDS	Acute respiratory distress syndrome
ICU	Intensive care unit
SOFA	Sequential organ failure assessment
SD	Standard deviation
VAP	Ventilator associated pneumonia
TBI	Traumatic brain injury
HAP	Hospital-acquired pneumonia
CPR	Cardiopulmonary resuscitation

Author contributions

AG and SP conceptulized the paper. SP is responsible for constructing the methods and obtaining the results. SP wrote the initial draft. AG critically edited the paper and prepared it for publication.

Data availability

No datasets were generated or analysed during the current study.

Competing interests

The authors declare no competing interests.

Received: 20 January 2025 Accepted: 1 February 2025
Published online: 11 February 2025

References

1. Wasserstein RL, Lazar NA. The ASA statement on *p*-values: context, process, and purpose. Am Stat. 2016;70(2):129–33.
2. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2(8): e124.
3. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, *p* values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31(4):337–50.
4. Vincent JL. Critical care—where have we been and where are we going? Crit Care. 2013;17(Suppl 1):S2.
5. Young PJ, Bellomo R. Statistical analysis of clinical trials in critical care medicine. Intensive Care Med. 2018;44(11):1901–3.
6. Iwashyna TJ, Burke JF, Sussman JB, et al. Implications of heterogeneity of treatment effect for reporting and analysis of randomized trials in critical care. Am J Respir Crit Care Med. 2015;192(9):1045–51.
7. Ridgeon EE, Young PJ, Bellomo R, et al. The fragility index in multicenter randomized controlled critical care trials. Crit Care Med. 2016;44(7):1278–84.
8. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov. 2006;5(1):27–36.
9. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019;567(7748):305–7.
10. Kalil AC, Sun J. Low-dose steroids for septic shock and severe sepsis: the use of Bayesian statistics to resolve clinical trial controversies. Intensive Care Med. 2011;37(3):420–9.
11. Gelman A, Carlin JB, Stern HS, et al. Bayesian Data Analysis. 3rd ed. New York: Chapman and Hall/CRC; 2013.
12. Harhay MO, Wagner J, Ratcliffe SJ, et al. Outcomes and statistical power in adult critical care randomized trials. Am J Respir Crit Care Med. 2014;189(12):1469–78.
13. Rubin DB. For objective causal inference, design trumps analysis. Ann Appl Stat. 2008;2(3):808–40.
14. Jabre P, Combes X, Lapostolle F, et al. Etomidate versus ketamine for rapid sequence intubation in acutely ill patients: a multicentre randomised controlled trial. Lancet. 2009;374(9686):293–300.
15. Morelli A, Ertmer C, Westphal M, et al. Effect of heart rate control with esmolol on hemodynamic and clinical outcomes in patients with septic shock. JAMA. 2013;310(16):1683–91.
16. Terragni PP, Antonelli M, Rumagalli R, et al. Early vs late tracheotomy for prevention of pneumonia in mechanically ventilated adult ICU patients: a randomized controlled trial. JAMA. 2010;303(15):1483–9.
17. Asehnoune K, Seguin P, Allary J, et al. Hydrocortisone and fludrocortisone for prevention of hospital-acquired pneumonia in patients with severe traumatic brain injury (Corti-TC): a double-blind, multicentre phase 3, randomised placebo-controlled trial. Lancet Respir Med. 2014;2(9):706–16.
18. Nichol G, Leroux B, Wang H, Callaway CW, Sopko G, Weisfeldt M, et al. Trial of continuous or interrupted chest compressions during cardiopulmonary resuscitation. N Engl J Med. 2015;373(23):2203–14.
19. Benjamin DJ, Berger JO, Johannesson M, et al. Redefine statistical significance. Nat Hum Behav. 2018;2(1):6–10.
20. Berry SM, Carlin BP, Lee JJ, Muller P. Bayesian adaptive methods for clinical trials. Boca Raton: Chapman and Hall/CRC; 2010.
21. Bhatt DL, Mehta C. Adaptive designs for clinical trials. N Engl J Med. 2016;375(1):65–74.
22. McShane BB, Gal D, Gelman A, et al. Abandon statistical significance. Am Stat. 2019;73(sup1):235–45.
23. Cook D, Lauzier F, Rocha MG, et al. Serious adverse events in academic critical care research. CMAJ. 2008;178(9):1181–4.
24. Ospina-Tascón GA, Büchele GL, Vincent JL. Multicenter, randomized, controlled trials evaluating mortality in intensive care: doomed to fail? Crit Care Med. 2008;36(4):1311–22.
25. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester: Wiley; 2004.

Publisher's Note