How Much Should A Null Trial Change Our Minds? A Bayesian Re-Analysis of a Large Randomised Trial of Conservative Oxygen Therapy in the Intensive Care Unit

Hoger Anwar Sadeeq

Duhok Technical Institute, Duhok Polytechnic University, Kurdistan, Iraq

Corresponding author: hoger.anwer@dpu.edu.krd

Received: 01 January 2026; Revised: 11 March 2026; Accepted: 17 March 2026; Published: 03 May 2026

Abstract

Background: A large pragmatic randomised trial of conservative oxygen therapy in mechanically ventilated critically ill adults reported an adjusted absolute risk difference in 90-day mortality of 0.7 percentage points (95 % confidence interval −0.7 to 2.0; P = 0.28) and was reported as showing no significant reduction in mortality. A P value above 0.05 states only that the data are compatible with no effect; it does not quantify how probable any particular effect now is, and it does not indicate how a clinician holding a prior belief should revise it.

Objective: To re-express the trial result as posterior probabilities of clinically framed statements under a range of explicitly stated prior beliefs, and to determine how strong a prior conviction would be required to sustain a belief in benefit after the trial.

Methods: A Bayesian re-analysis was performed using only published summary data. The adjusted risk difference and its confidence interval were used to construct a normal likelihood with a standard error of 0.689 percentage points. Five prior distributions were specified in advance: non-informative; sceptical and strongly sceptical priors centred on no effect; an optimistic prior centred on the 2.5-percentage-point benefit the trial was designed to detect; and an evidence-based prior derived from a published meta-analysis of earlier trials. Conjugate normal updating produced posterior distributions, from which probabilities of benefit, of harm and of practical equivalence were computed.

Results: Under a non-informative prior the posterior probability of any benefit from conservative oxygen therapy was 15.5 %, the probability of a benefit of at least 1 percentage point was 0.7 %, and the probability of a benefit as large as the design target was below 0.1 %. The probability that the true effect lies within one percentage point of zero in either direction was 66.2 %, rising to 92.4 % under a strongly sceptical prior. Under the optimistic prior the posterior probability of benefit was 51.5 % — an even split — with only a 5.3 % probability of a benefit of at least 1 percentage point. To retain a 95 % posterior probability of benefit, a clinician with a prior standard deviation of 1.28 percentage points would have needed a prior belief in a 6.8-percentage-point mortality reduction.

Conclusions: The trial does more than fail to demonstrate benefit: it makes a clinically important benefit improbable under any defensible prior, while leaving a modest harm not excluded. Re-expressing null trials in this form is inexpensive, uses only published summary data, and conveys information that a P value cannot.

Keywords: Bayesian Analysis, Oxygen Therapy, Critical Care, Clinical Trial Interpretation, Posterior Probability, Statistical Inference.

1. INTRODUCTION

A large randomised trial reports a difference that does not reach statistical significance, and the reader is told that the intervention did not reduce mortality. What the reader usually wants to know is different: given this result, how likely is it that the intervention helps at all, how likely that it helps enough to be worth

doing, and how likely that it does harm? The frequentist apparatus in which trials are reported answers none of these directly.

The confusion is well documented and is structural rather than careless. A P value is the probability of data at least as extreme as those observed, computed under the assumption that the null hypothesis is true; it is not the probability that the null hypothesis is true. A 95 % confidence interval is a statement about the long-run coverage of a procedure, not a range within which the true value lies with 95 % probability. Both statements are routinely read as though they were their Bayesian counterparts, which is the interpretation clinicians actually need and which the reported analysis does not supply.

This matters most for large null trials, because they are the trials whose interpretation is most contested. A small trial that finds nothing is uninformative and everyone agrees. A trial of 16,500 patients that finds nothing is making a substantive claim, but the claim is buried: it lies in the width of the interval and in what that interval excludes, and the abstract usually reports the P value instead.

The subject of this re-analysis is a large pragmatic trial of conservative oxygen therapy in mechanically ventilated critically ill adults, which randomised 16,500 patients across 97 intensive care units and reported an adjusted absolute risk difference in 90-day mortality of 0.7 percentage points, with a 95 % confidence interval from −0.7 to 2.0 and P = 0.28. The conclusion was that minimising oxygen exposure did not significantly reduce all-cause mortality at 90 days. This paper asks what that result implies for a clinician's belief, and how much the answer depends on what the clinician believed beforehand.

The purpose is methodological rather than polemical. Nothing here disputes the trial's design, conduct or analysis, all of which were exemplary, and the re-analysis uses the trial's own adjusted estimate. The intention is to demonstrate a form of reporting that any trial team could add to any primary report at negligible cost, and to show what it reveals that the conventional presentation does not.

2. METHODS

2.1 Data

Only published summary data were used; no individual patient data were obtained and no approval was required. The trial reported an adjusted absolute risk difference of 0.7 percentage points, conservative minus usual care, with a 95 % confidence interval from −0.7 to 2.0 percentage points. Throughout this paper a negative risk difference favours conservative oxygen therapy.

The standard error of the adjusted estimate was recovered from the interval as

$$SE = (\text{upper} - \text{lower}) / (2 \times 1.96) = (2.0 - (-0.7)) / 3.92 = 0.689 \text{ percentage points} \quad (1)$$

For reference, the unadjusted event data — 2,908 deaths among 8,211 patients receiving conservative therapy and 2,858 among 8,183 receiving usual care — give an unadjusted risk difference of 0.49 percentage points with a standard error of 0.75, closely consistent with the adjusted figures. The adjusted estimate is used throughout, as it is the trial's primary result.

2.2 Priors

Five priors were specified before any posterior was computed, and are listed in Table 1. They were chosen to span the range of positions a reasonable reader might have held before the trial reported, rather than to produce any particular conclusion.

Table 1. Prior distributions, expressed on the absolute risk difference scale (percentage points)

Prior	Mean	SD	Interpretation
Non-informative	0	—	No prior opinion; posterior determined entirely by the data
Sceptical	0	1.28	No effect expected; 95 % prior probability that the true effect lies within ±2.5 pp
Strongly sceptical	0	0.64	No effect expected; 95 % prior probability within ±1.25 pp
Optimistic	−2.5	1.28	Centred on the benefit the trial was designed to detect

Evidence-based	+0.7	0.71	Derived from a published meta-analysis of five earlier trials (pooled RR 1.02, 95 % CI 0.98–1.06) converted to the risk-difference scale at 35 % baseline mortality

Note. pp = percentage points. The sceptical priors are constructed by the conventional device of placing a stated proportion of prior probability within a range of plausible effects, and encode doubt rather than a belief in harm.

2.3 Analysis

With a normal likelihood and a normal prior, the posterior is normal with conjugate updating,

$\mu_{post} = (\mu_0/\sigma_0^2 + y/\sigma^2) / (1/\sigma_0^2 + 1/\sigma^2)$, $\sigma_{post}^2 = 1 / (1/\sigma_0^2 + 1/\sigma^2)$ (2)

where y and σ are the observed estimate and its standard error and μ_0 and σ_0 the prior mean and standard deviation. Posterior probabilities were computed for six clinically framed statements: any benefit, benefit of at least 1 percentage point, benefit as large as the 2.5-point design target, any harm, harm of at least 1 percentage point, and practical equivalence defined as a true effect within one percentage point of zero in either direction. A final analysis determined, for several prior standard deviations, the prior mean that would be required to retain a 95 % posterior probability of benefit after the trial. All computation was performed in Python and reproduces from the published summary figures alone.

3. RESULTS

3.1 Prior, likelihood and posterior

Figure 1 shows how each of three representative priors combines with the trial likelihood.

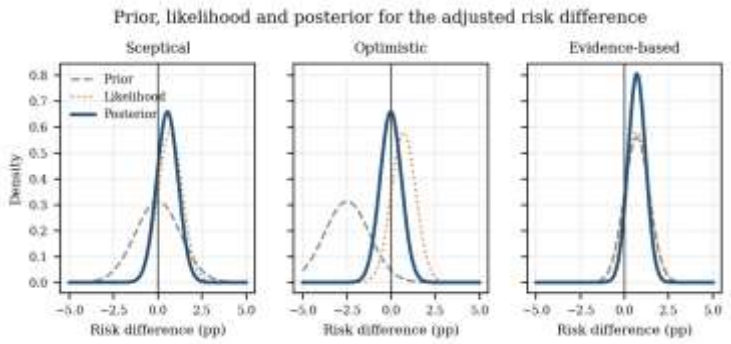

Figure 1. Prior, likelihood and posterior distributions for the adjusted risk difference under three prior specifications.

The likelihood is narrow, which is the consequence of a trial of this size, and it dominates every prior shown. Under the sceptical prior the posterior shifts only slightly toward zero. Under the optimistic prior, centred on a 2.5-point benefit, the posterior is pulled almost entirely onto the data and comes to rest close to zero: a prior expectation of substantial benefit does not survive contact with 16,500 patients. Under the evidence-based prior, which was already centred near the observed value, prior and likelihood reinforce one another and the posterior narrows.

3.2 Posterior estimates

Table 2. Posterior distributions of the adjusted risk difference (percentage points)

Prior	Posterior mean	Posterior SD	95 % credible interval
Non-informative	+0.70	0.689	−0.65 to 2.05
Sceptical	+0.54	0.606	−0.65 to 1.73
Strongly sceptical	+0.32	0.468	−0.59 to 1.24
Optimistic	−0.02	0.606	−1.21 to 1.17
Evidence-based	+0.70	0.496	−0.27 to 1.67

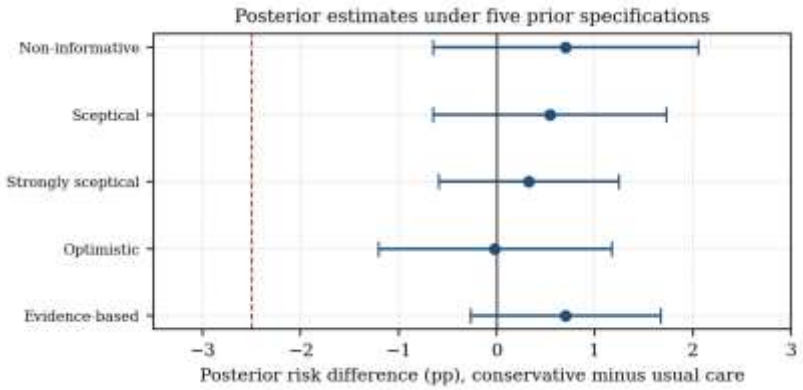

Figure 2. Posterior means and 95 % credible intervals under the five priors, with the trial's design target marked

Two features of Table 2 are worth drawing out. Every posterior credible interval excludes the 2.5-point benefit the trial was designed to detect, by a wide margin and under every prior including the optimistic one. And every posterior interval includes zero, so none of them supports a confident claim of harm either. The range of posterior means across the five priors is 0.72 percentage points, which is smaller than the width of any individual credible interval: the choice of prior matters less than the residual uncertainty, which is the expected behaviour when the data are this plentiful and is itself a reassuring finding about the robustness of the conclusion.

3.3 Probabilities of clinically framed statements

Table 3 and Figure 3 translate the posteriors into the statements a clinician actually wants.

Table 3. Posterior probabilities (%) of clinically framed statements

Statement	Non-informative	Sceptical	Strongly sceptical	Optimistic	Evidence-based
Any benefit (RD < 0)	15.5	18.6	24.5	51.5	7.9
Benefit ≥ 1 pp	0.7	0.5	0.2	5.3	< 0.1
Benefit ≥ 2.5 pp (design target)	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Any harm (RD > 0)	84.5	81.4	75.5	48.5	92.1
Harm ≥ 1 pp	33.2	22.5	7.4	4.6	27.3
Practical equivalence (|RD| < 1 pp)	66.2	77.0	92.4	90.1	72.7

Note. RD = absolute risk difference in 90-day mortality, conservative minus usual care; a negative value favours conservative oxygen therapy. pp = percentage points.

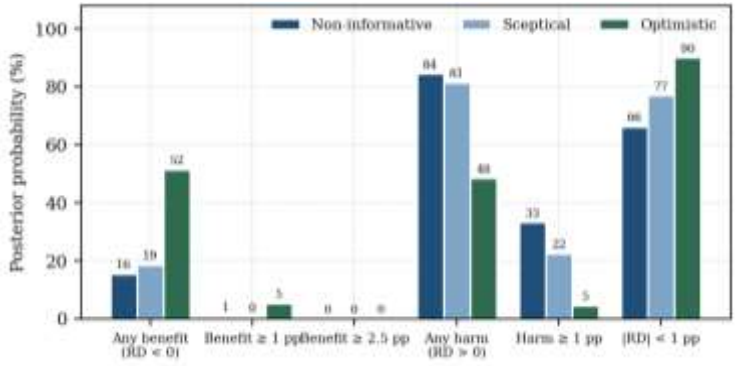

Figure 3. Posterior probabilities of six clinically framed statements under three priors

The third row is the most consequential and is invariant across priors: the probability that conservative oxygen therapy delivers the benefit the trial was designed to detect is below one in a thousand under every prior specification, including the one centred on that very benefit. This is a stronger statement than the reported P value of 0.28 conveys, and it is the statement a guideline committee needs.

The second row is nearly as strong. The probability of a benefit of even one percentage point is 0.7 % under a non-informative prior and 5.3 % under the optimistic one. A clinician who continues to believe that conservative oxygen therapy saves a meaningful number of lives is holding a position that the data assign a probability of around one in twenty at best, and that requires the prior belief to have been strong to begin with.

The first and fourth rows are where the conventional reading of this trial is most often overstated. The trial is frequently summarised as showing no difference, but under a non-informative prior the posterior probability that conservative therapy is harmful to some degree is 84.5 %, and the probability of harm of at least one percentage point is 33.2 %. Those are not negligible figures. The correct summary is not that the two strategies are equivalent but that a benefit of clinical size is improbable while a modest harm remains quite possible — a materially different message, and one that points in the opposite direction to the hypothesis the trial set out to test.

The final row supplies the constructive finding. The probability that the true effect lies within one percentage point of zero in either direction is 66.2 % under a non-informative prior and 92.4 % under a strongly sceptical one. If a difference smaller than one percentage point would not change practice — a judgement for clinicians rather than statisticians — then this trial has come close to establishing practical equivalence, which is a positive result rather than a null one and is invisible in the original presentation.

3.4 How the conclusion varies with the prior

Figure 4 shows the posterior probability of any benefit across the whole plane of prior means and standard deviations, and Figure 5 the full posterior cumulative distributions.

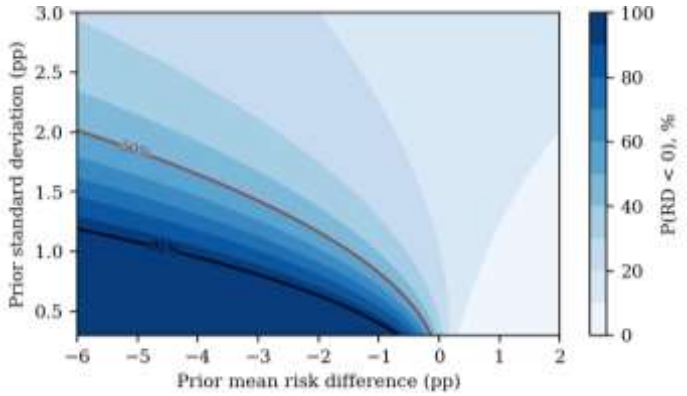

Figure 4. Posterior probability of any benefit as a function of the prior mean and prior standard deviation

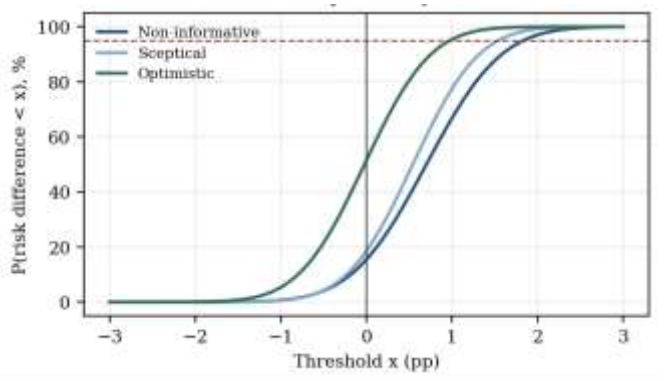

Figure 5. Posterior cumulative probability that the true risk difference falls below a given threshold

The 95 % contour in Figure 4 lies far to the left, and Table 4 locates it precisely. To retain a 95 % posterior probability of benefit after this trial, a clinician with a prior standard deviation of 1.28 percentage points would have needed to hold a prior mean of −6.8 percentage points — that is, to have believed before the trial that conservative oxygen therapy reduces 90-day mortality by nearly seven percentage points in absolute terms. No such belief is defensible: it exceeds the effect of any intervention in the history of critical care and is roughly three times the effect the trial itself was designed to detect.

Table 4. Prior mean required to retain a 95 % posterior probability of benefit

Prior standard deviation (pp)	Posterior SD (pp)	Prior mean required (pp)
0.64 (strongly sceptical width)	0.468	−2.03
1.28 (sceptical width)	0.606	−6.82
2.00	0.651	−16.01
3.00	0.671	−35.33

Note. A narrower prior requires a less extreme prior mean because it resists updating, but a narrow prior centred on a large benefit is itself a very strong claim. A prior standard deviation of 0.64 pp centred at −2.03 pp asserts 95 % prior confidence that the benefit lies between 0.8 and 3.3 pp, which no evidence before the trial supported.

This is the sense in which a large null trial is informative. It does not merely fail to find an effect; it constrains the set of prior beliefs that can survive it, and in this case the surviving set is empty of defensible positions.

4. DISCUSSION

4.1 Principal findings

Re-expressed as posterior probabilities, this trial states three things that its conventional report does not. A clinically important benefit from conservative oxygen therapy is improbable: below 1 % for a benefit of one percentage point, and below 0.1 % for the trial's own design target, under every prior examined. A modest harm is not excluded: the probability of harm of at least one percentage point is 33.2 % under a non-informative prior. And the result is close to establishing practical equivalence within a one-point margin, with probability between 66 % and 92 % depending on the prior.

The robustness across priors is itself a finding. Posterior means ranged over 0.72 percentage points across five very different prior positions, less than the width of any single credible interval. When a trial is this large, the choice of prior is not the decisive input, and the common objection that Bayesian analysis merely encodes the analyst's opinion has little force here — which is precisely why stating the priors explicitly is a strength rather than a vulnerability.

4.2 Implications for reporting

Table 5. What each form of reporting conveys

Quantity reported	What it states	What it cannot state
P = 0.28	The data are compatible with no effect	How probable any given effect now is
RD 0.7 pp (95 % CI −0.7 to 2.0)	A procedure covering the truth 95 % of the time gave this range	That the truth lies in this range with 95 % probability
"No significant reduction in mortality"	The null was not rejected	Whether benefit, harm or equivalence is most probable
P(benefit ≥ 1 pp) = 0.7 %	A clinically important benefit is improbable	Nothing further; but this is the statement clinicians need
P(|RD| < 1 pp) = 66–92 %	The two strategies are probably practically equivalent	Whether a 1 pp margin is the right one — a clinical judgement

Prior mean required for 95 % belief in benefit	How strong a prior conviction the trial defeats	Whether anyone actually held that prior

The practical recommendation is modest and cheap. Any trial reporting a null primary result could add a short Bayesian appendix presenting posterior probabilities under a small set of pre-specified priors, using nothing beyond the summary estimate and its standard error. It requires no additional data, no additional consent and no additional computation of any consequence, and it converts a statement about the compatibility of data with a hypothesis into statements about the plausibility of clinically meaningful effects. The pre-specification matters: priors chosen after seeing the result are not credible, and should be registered alongside the frequentist analysis plan.

A second implication concerns the interpretation of large null trials generally. The reflex reading — that the intervention does not work and the question is closed — conflates two quite different situations: a trial too small to detect anything, and a trial large enough to make an important effect improbable. Only the second warrants changing practice, and distinguishing them requires attention to what the interval excludes rather than to whether it includes zero. The analysis presented here makes that distinction explicit.

4.3 Limitations

Five limitations bound these results. The analysis uses summary data and a normal approximation to the likelihood, which is well justified at this sample size and for a risk difference near 0.5 percentage points but would be less appropriate for a rare outcome or a small trial. The standard error was recovered from the reported confidence interval rather than taken from the primary analysis, which introduces rounding error of the order of a hundredth of a percentage point and does not affect any conclusion.

The priors, although pre-specified within this analysis, are constructions of the present author and were not elicited from clinicians. A formal prior elicitation exercise among intensivists would be more defensible and would probably produce a range of positions wider than the five examined here; the contour in Figure 4 is provided so that readers can locate their own prior and read off the implied posterior directly.

The analysis addresses the average effect on 90-day mortality in the trial population and says nothing about subgroups. Subgroup-specific effects, in particular in patients with hypoxic-ischaemic encephalopathy or sepsis, are being examined in dedicated trials and are not constrained by this re-analysis; a null average effect is fully compatible with offsetting subgroup effects, and the present findings should not be read as closing those questions.

The one-percentage-point margin used to define practical equivalence is a stipulation rather than a consensus. Some readers will regard half a point as important at population scale, given the number of patients ventilated annually, and for them the equivalence probabilities reported here are too generous. The cumulative distribution in Figure 5 allows any margin to be substituted.

Finally, this is a re-analysis of one trial and does not incorporate the other completed trials in this field except through the evidence-based prior. A formal Bayesian meta-analysis across all trials, and in due course the results of the ongoing 40,000-patient programme, would supersede it.

5. CONCLUSION

A trial reported as showing no significant mortality reduction with conservative oxygen therapy makes, when re-expressed, a considerably stronger claim than its abstract suggests. The probability of a benefit as large as the trial was designed to detect is below one in a thousand under every prior examined, including one centred on that benefit. The probability of a benefit of even a single percentage point is under one per cent unless the reader began with substantial optimism, and to retain a 95 % belief in benefit after the trial would have required a prior conviction of a seven-point mortality reduction — a position no evidence supported.

At the same time, the trial does not establish that the two strategies are identical. A modest harm from conservative therapy remains possible, with a one-third posterior probability of harm exceeding a

percentage point under a non-informative prior, and that is not the message conveyed by the phrase no significant difference.

The wider point is that this re-analysis required no new data, no new patients and a few lines of arithmetic on two published numbers. Trials reporting null results should be asked to supply it as a matter of routine, because a P value tells the reader what the data would look like if nothing were true, and the reader wants to know what is probably true.

Declarations

Ethical approval: Not applicable.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of interes: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author contributions:

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Hoger Anwar Sadeeq	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

6. REFERENCES

1. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. J R Stat Soc Ser A. 1994;157(3):357-416. doi.org/10.2307/2983527
2. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester: Wiley; 2004. doi.org/10.1002/0470092602
3. 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(21):2251-2259. doi.org/10.1001/jama.2018.14276
4. Zampieri FG, Casey JD, Shankar-Hari M, Harrell FE, Harhay MO. Using Bayesian methods to augment the interpretation of critical care trials. Am J Respir Crit Care Med. 2021;203(5):543-552. doi.org/10.1164/rccm.202006-2381CP
5. Yarnell CJ, Abrams D, Baldwin MR, et al. Clinical trials in critical care: can a Bayesian approach enhance clinical and scientific decision making? Lancet Respir Med. 2021;9(2):207-216. doi.org/10.1016/S2213-2600(20)30471-9
6. Wijeysundera DN, Austin PC, Hux JE, Beattie WS, Laupacis A. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. J Clin Epidemiol. 2009;62(1):13-21. doi.org/10.1016/j.jclinepi.2008.07.006
7. 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):337-350. doi.org/10.1007/s10654-016-0149-3
8. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. Am Stat. 2016;70(2):129-133. doi.org/10.1080/00031305.2016.1154108
9. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature. 2019;567(7748):305-307. doi.org/10.1038/d41586-019-00857-9

10. Kruschke JK. Rejecting or accepting parameter values in Bayesian estimation. Adv Methods Pract Psychol Sci. 2018;1(2):270-280. doi.org/10.1177/2515245918771304
11. Schjørring OL, Klitgaard TL, Perner A, et al.; HOT-ICU Investigators. Lower or higher oxygenation targets for acute hypoxemic respiratory failure. N Engl J Med. 2021;384(14):1301-1311. doi.org/10.1056/NEJMoa2032510
12. ICU-ROX Investigators and the Australian and New Zealand Intensive Care Society Clinical Trials Group. Conservative oxygen therapy during mechanical ventilation in the ICU. N Engl J Med. 2020;382(11):989-998. doi.org/10.1056/NEJMoa1903297
13. Young PJ, Arabi YM, Bagshaw SM, et al. Protocol and statistical analysis plan for the mega randomised registry trial research program comparing conservative versus liberal oxygenation targets in adults receiving unplanned invasive mechanical ventilation in the ICU (Mega-ROX). Crit Care Resusc. 2022;24(2):137-149. doi.org/10.51893/2022.2.OA4