Enhancing Meta-Analysis interpretation in Anesthesia Literature: A trial sequential analysis perspective

Kerremans O.[1], Vanoverschelde H.[1], De Pauw E.[1], Wouters P.[1,2], Wyffels P.[1]

[1]*Department of Anaesthesiology and Perioperative Medicine, Ghent University Hospital, Ghent, Belgium;* [2]*Department of Basic and Applied Sciences, Ghent University, Ghent, Belgium.*

Corresponding author: Oliver Kerremans, MD, Department of Anaesthesiology and Perioperative Medicine, Ghent University Hospital, Corneel Heymanslaan 10, 9000 Ghent, Belgium. E-mail: oliver.kerremans@uzgent.be

Abstract

Background: **Many meta-analyses in anesthesiology are small, heterogeneous, and prone to random error. Methods: We reappraised the conclusiveness and statistical robustness of 1,300 Cochrane anesthesiology meta-analyses using trial sequential analysis (TSA), which adjusts for repeated significance testing and calculates a required information size (RIS) for 80 % power. For each meta-analysis we assessed whether the cumulative evidence crossed monitoring boundaries for benefit or futility, determined whether the required information size (RIS) was achieved, and quantified underpowered syntheses together with their type I and II error risks.**
Results: **Only 17.9 % of meta-analyses showed a confirmed treatment effect, 27.4 % crossed futility boundaries, and 54.7 % remained inconclusive; overall, 85.9 % failed to reach the required information size (RIS). Among underpowered analyses, 62.8 % of conventionally significant findings were vulnerable to type I error and 63.9 % of non-significant findings were vulnerable to type II error.**
Conclusions: **Most anesthesiology meta-analyses lack definitive evidence, and TSA frequently downgrades apparently positive or null conclusions. Incorporating sequential monitoring and information-size considerations can enhance the reliability of evidence synthesis in anesthesiology.**
Trial Registration: **None.**

Keywords: **Meta-Analysis as Topic, Anesthesiology, Bias (Epidemiology), Randomized Controlled Trials as Topic, Sequential Analysis.**

Introduction

In modern evidence-based medicine, meta-analysis plays a pivotal role by aggregating results from clinical trials to generate a more precise estimate of an intervention's effect. This statistical synthesis not only increases the power to detect clinically relevant differences but also reduces uncertainty surrounding the outcome. Within the evidence hierarchy, rigorously and transparently conducted meta-analyses and systematic reviews occupy the highest level of evidence[1,2]. They may serve as a lens through which current evidence is viewed[3].
Milestones over the past decades have steadily tightened review quality.

DerSimonian-Laird random-effects (1986) enabled pooling of heterogeneous trials[4]. Egger funnel plots (1997) and Duval-Tweedie trim-and-fill (2000) tackled publication bias[5,6]. Higgins-Thompson's τ^2 (2002) and I^2 (2003) then quantified heterogeneity[7,8]. GRADE (2008) formalized certainty-of-evidence appraisal[9], CONSORT (2010) standardized randomized-trial reporting[10], and AMSTAR 2 (2017) refined critical appraisal of systematic reviews[11]. Finally, the PRISMA guidelines evolved from improving review reporting in 2009[12], to promoting protocol registration in 2015[13], and modernizing standards for complex evidence synthesis in 2020[14].

However, it seems that conventional meta-analysis suffers from some significant limitations which may have been severely underestimated[15-17]. Underpowered meta-analyses, containing insufficient trials or participants, increase both type

Previous presentation: Parts of this work were presented as a poster at Euroanaesthesia 2024 (ICM, Munich, Germany), 25–27 May 2024 — "Enhancing meta-analysis interpretation in anesthesia literature: a trial sequential analysis perspective" (Kerremans O.).

I (false positive) and type II (false negative) error risks18-20. Such analyses may produce exaggerated findings due to random error or publication bias or alternatively fail to detect genuine effects. Both scenarios compromise evidence integrity with direct consequences for clinical practice[15-17].

Cumulative meta-analysis presents additional challenges. As meta-analyses incorporate new trials over time, each update effectively performs a new hypothesis test[19,20]. Without proper statistical correction, repeated significance testing inflates type I error risk[20].

Conventional p-values, designed for single hypothesis testing, inadequately account for the complexities of this error accumulation across updates, potentially leading to premature intervention endorsement or inappropriate dismissal[21].

Trial Sequential Analysis (TSA) was developed to address these methodological concerns21,22. TSA adapts interim analysis methodologies from clinical trials to meta-analysis contexts by introducing sequential monitoring boundaries for benefit, harm, and futility. These statistical thresholds must be crossed before drawing definitive conclusions. TSA calculates a Required Information Size (RIS)—analogous to sample size calculation in randomized controlled trials—representing participant numbers needed to reliably detect predefined clinically relevant effects with controlled error rates[23]. When between-study heterogeneity is present, TSA inflates RIS by the estimated diversity (D^2) to yield the diversity-adjusted required information size (DARIS), which captures reduced precision and serves as the benchmark we report throughout[22].

TSA fundamentally reframes meta-analytic interpretation by evaluating evidence adequacy for specific conclusions[24]. If the cumulative Z-statistic crosses the monitoring boundary for benefit, this signals sufficient evidence even before reaching the RIS. Conversely, crossing a futility boundary suggests additional data collection is unlikely to alter outcomes[22,24].

In anesthesiology, where trials typically have modest sample sizes and substantial clinical heterogeneity, premature conclusion risks are particularly concerning[25,26]. Despite increasing meta-analysis use in this field, limited systematic assessment exists regarding how frequently anesthesia-related meta-analyses are underpowered or vulnerable to unreliable conclusions from random error[15,24,25].

Applying Trial Sequential Analysis to Cochrane reviews in the Pain & Anesthesia subdomain, this thesis quantifies how often meta-analyses reach the required information size versus remain within uncertainty zones, revealing the prevalence of underpowered studies and demonstrating TSA's value in flagging premature conclusions and research inefficiencies.

Materials and Methods

Study Selection

To evaluate the statistical robustness and conclusiveness of meta-analytic evidence in anesthesiology, a systematic search and screening of Cochrane reviews was conducted by three reviewers. (O.K./ H.V./ E.D.P)
This process targeted reviews from the Cochrane Database of Systematic Reviews (CDSR) regarded as the gold standard for evidence synthesis27,28—within the subdomain of Pain and Anesthesia, covering a ten-year period from January 1, 2012, to December 31, 2022. The reviews were selected from the following pre-existing subdomains:

1. Peri-anesthetic / peri-operative care
2. Post-anesthetic unit / intensive care unit
3. Drugs in anesthesia & intensive care

The aim was to include systematic reviews that matched predefined eligibility criteria:

- Included binary (dichotomous) outcome data, and
- Contained at least three independent randomized controlled trials (RCTs) per meta-analysis, to ensure a minimum level of statistical reliability and heterogeneity assessment.

Trial Sequential Analysis (TSA) Approach

Figure 1 illustrates a hypothetical conventional two-sided trial sequential analysis (TSA) of a cumulative meta-analysis comprising four randomized trials (Studies A–D). The horizontal axis depicts the cumulative sample size (number of patients), while the vertical axis shows the successive Z-score for the pooled treatment effect (positive values favor benefit, negative values favor harm). Curved red dashed lines represent the TSA monitoring boundaries that maintain the overall two-sided type-I error at 5 %: crossing the upper boundary establishes benefit, whereas crossing the lower boundary establishes harm.

The two horizontal yellow dashed lines correspond to the conventional $p < 0.05$ thresholds ($Z \approx \pm 1.96$) used in ordinary meta-analysis. The vertical red dashed line marks the required information size (RIS, ~1 000 patients) calculated a priori to achieve adequate power. The blue wedge denotes the futility zone, within which further evidence of benefit is considered unlikely.

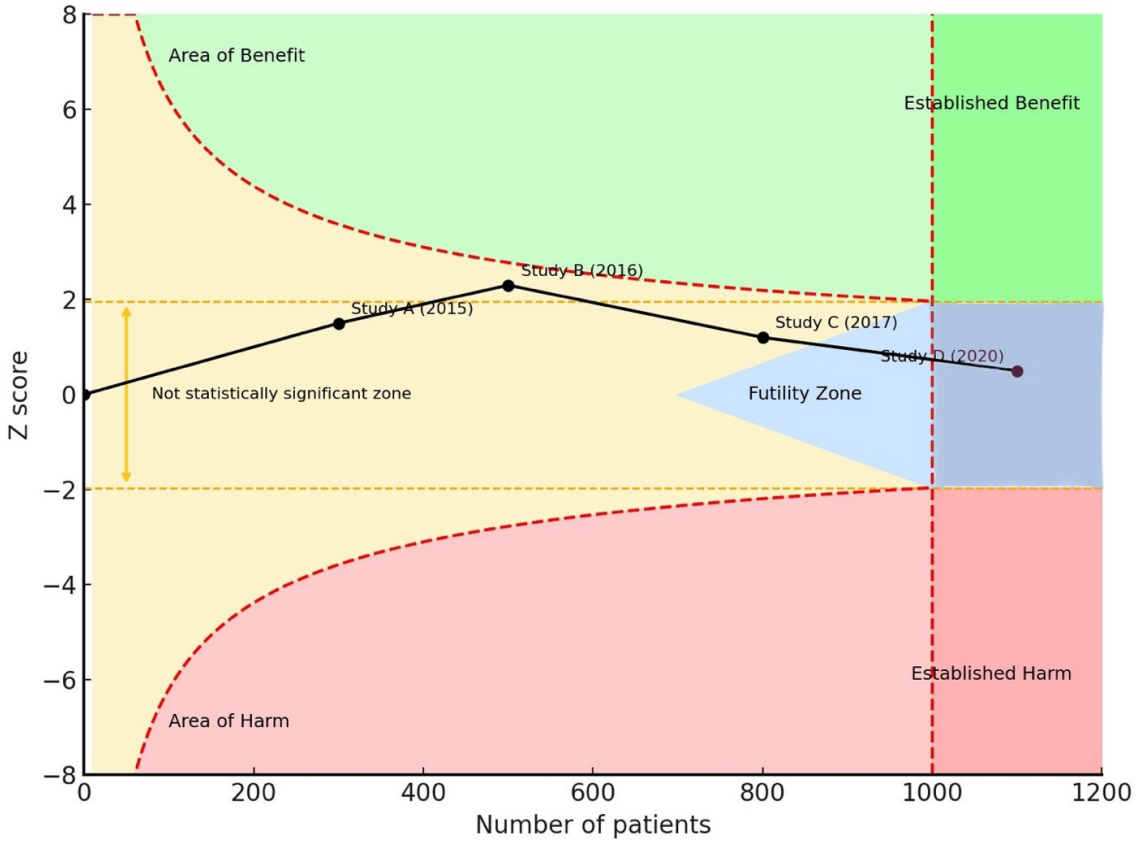


Fig. 1 — Trial Sequential Analysis zones.

After the first trial (Study A, ~300 patients) the cumulative Z-score is approximately +1.4, suggestive of benefit but still below the conventional significance threshold. Addition of Study B (~500 patients in total) raises the Z-score to about +2.2, just exceeding the conventional p-value boundary; nevertheless, the curve remains beneath the stricter TSA benefit boundary, indicating that early significance may be spurious. Incorporating Study C (~800 patients) reduces the Z-score to ~ +1.7, returning the result to non-significance. When Study D brings the cumulative sample to the RIS (~1 100 patients), the Z-score declines further to ~ +1.0, entering the futility zone.

A One-sided trial sequential analysis, using a one-sided, absolute cumulative Z-score framework can be seen in Figure 2. In this configuration the graph shows how far the accumulating evidence strays from the null hypothesis, without distinguishing whether the intervention ultimately helps or harms. This decision was taken to enhance the interpretability of the results.

Each TSA result was classified into one of following six zones (Figure 2):

1. Inconclusive zone Risk of type I error
2. Effect confirmed before RIS (early boundary crossing)
3. Effect confirmed after RIS (post-hoc confirmation with full sample) established effect
4. Inconclusive zone Risk of type II error (non-significant and underpowered)
5. Futility declared before RIS
6. Futility declared after RIS – established futility

This classification allowed nuanced interpretation of the TSA outcomes beyond simple p-value threshold.

Data Extraction

Each included systematic review underwent detailed data extraction using a standardized protocol developed for this project. For every eligible meta-analysis, the following data elements were extracted:

- Identifiers: Title, Cochrane Review ID, publication year, outcome name/type
- Study count: Number of included trials
- Per-study data:
  - Events and total participants in treatment vs. control arms
  - Risk of bias (Cochrane ROB tool)
- Original meta-analysis effect: Odds ratio or relative risk ± 95% CI

TSA Settings

- Software: TSA v0.9.5.10 Beta (Copenhagen Trial Unit)[29]
- Model: Random-effects (DerSimonian-Laird)
- Error rates: $\alpha = 0.05$; $\beta = 0.20$ (80% power)

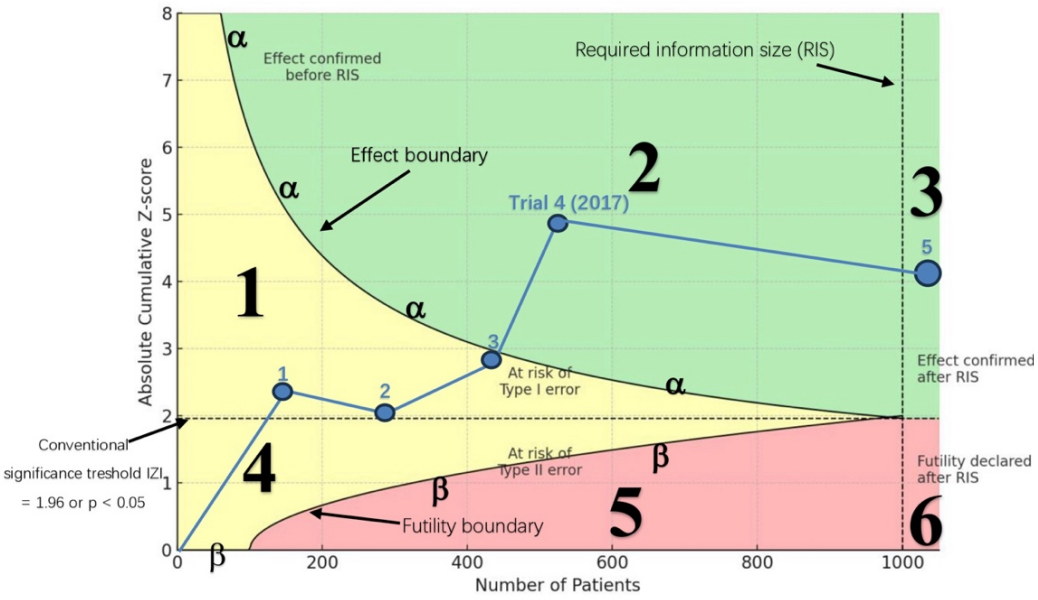
- Anticipated effect: Observed RRR from the meta-analysis
- Control event rate: From low-risk-of-bias trials
- Heterogeneity adjustment: D^2 estimated from the accumulated data
- Monitoring boundaries: Lan-DeMets α-spending with O'Brien–Fleming rules

We analyzed the Cochrane meta-analyses with the TSA batch tool (v0.9.5.10, Copenhagen Trial Unit)[29] to improve reproducibility by automating data import, parameter settings, RIS computation, and Z-curve plotting. For our primary approach, we based the analysis on the evidence at hand. Using a random-effects (DerSimonian-Laird) model, we set the anticipated intervention effect to the observed relative risk reduction (RRR) from each meta-analysis and computed the Required Information Size (RIS) and—accounting for heterogeneity—the diversity-adjusted RIS (DARIS) directly from it (RIS assumes no between-trial heterogeneity; DARIS adjusts this target upward when heterogeneity is present). By construction, this DARIS —using TSA's diversity (D2) estimated from the accumulated data and the control event rate (CER) from low-risk-of-bias trials —is mechanically dependent on the pooled effect. This approach standardizes across topics and asks whether the current evidence is large enough to support its own estimate. The trade-off is potential circularity: if the pooled effect is upwardly biased (e.g., from small-study effects or publication bias), the DARIS will be underestimated and the apparent strength of evidence will be inflated[30], as a larger effect requires a smaller sample size to confirm. Alternative calibration choices for the anticipated effect —not applied here —include a Minimal Clinically Important Difference (MCID) (patient-centered but context-specific and often unavailable)[31], a pre-specified key-RCT effect (concrete yet sensitive to trial selection/external validity)[32], and a uniform a priori effect (transparent yet arbitrary across topics)[30].

Results

Overview of Included Reviews and Meta-Analyses

Across 222 Cochrane reviews initially screened in the "Pain and Anesthesia" domain, 126 met eligibility criteria (binary outcomes, publication 2012-2022, ≥3 trials per meta-analysis). These comprised peri-anesthetic/peri-operative care (63 reviews; 750 meta-analyses), post-anesthesia/ICU (39; 504), and pharmacologic interventions (24; 348). In total, 1,602 meta-analyses were considered; 1,300/1,602 (81.1%) provided sufficient data for Trial Sequential Analysis (TSA;



X-axis: Total number of patients with indication of required information size
Y-axis: Cumulative Z-statistic – Absolute cumulative Z-statistic; the horizontal dashed line represents the conventional significance threshold, Z ± 1.96 (p < 0.05)
Sequential monitoring: Collected data is continuously added; this allows for earlier detection of effect or futility – before reaching the required information size (RIS)
Required information size (RIS): the total number of data needed to detect or reject a predefined intervention effect
Alpha spending (α): Statistical approach used to control overall type I error rates in multiple interim analyses
Benefit boundary: Intervention shows evidence of efficacy if crossed
Beta spending (β): Statistical approach that helps to control the cumulative probability of type II errors across multiple interim analyses
Futility boundary: Intervention unlikely to prove effective if crossed

Fig. 2 — One-sided TSA decision zone graph.

relative risk reduction and diversity D^2 available), while 302/1,602 (18.9%) were too sparse to estimate monitoring boundaries or a required information size (RIS) (Figure 3).

Conventional Meta-Analytic Significance and Power

Among the 1,300 TSA-amenable meta-analyses, conventional testing yielded significance ($p \leq 0.05$) in 368/1,300 (28.3%) and non-significance in 932/1,300 (71.7%). However, only 183/1,300 (14.1%) met an information-size benchmark consistent with 80% power at two-sided $\alpha = 0.05$; the remaining 1,117/1,300 (85.9%) were underpowered (Figure 4). Approximately one in seven meta-analyses met the information-size threshold, whereas six in seven did not, indicating that many conclusions remain vulnerable to random error despite nominal significance.

Trial Sequential Analysis (TSA): Primary Classification

TSA classified 233/1,300 (17.9%) as confirmed benefit, 356/1,300 (27.4%) as futile, and 711/1,300 (54.7%) as inconclusive (Table I; Figure 4). Compared with the 28.3% deemed significant under conventional testing, only 17.9% were confirmed by TSA—a 10.4 percentage-point absolute decrease (~37% relative reduction) in effects confirmed under sequential monitoring. Among conventionally significant findings, 233/368 (63.3%) were confirmed and 135/368 (36.7%) were reclassified as inconclusive; none were reclassified to futility. Among conventionally non-significant findings, 356/932 (38.2%) were classified as futile and 576/932 (61.8%) as inconclusive. When conventional results were non-significant but information-sufficient (30/932), TSA consistently indicated futility, suggesting adequately powered null results typically reflect no meaningful effect rather than false negatives.

Error Risk Among Underpowered Meta-Analyses

Within the underpowered stratum (1,117/1,300), inferential risks were comparable. Of 215 underpowered analyses with $p \leq 0.05$, 135/215 (62.8%) failed to cross the TSA benefit boundary and had not reached the diversity-adjusted required information size (DARIS), indicating susceptibility to false-positive inference if judged on p-values alone. Of 902 underpowered, non-significant analyses, 576/902 (63.9%) remained inconclusive

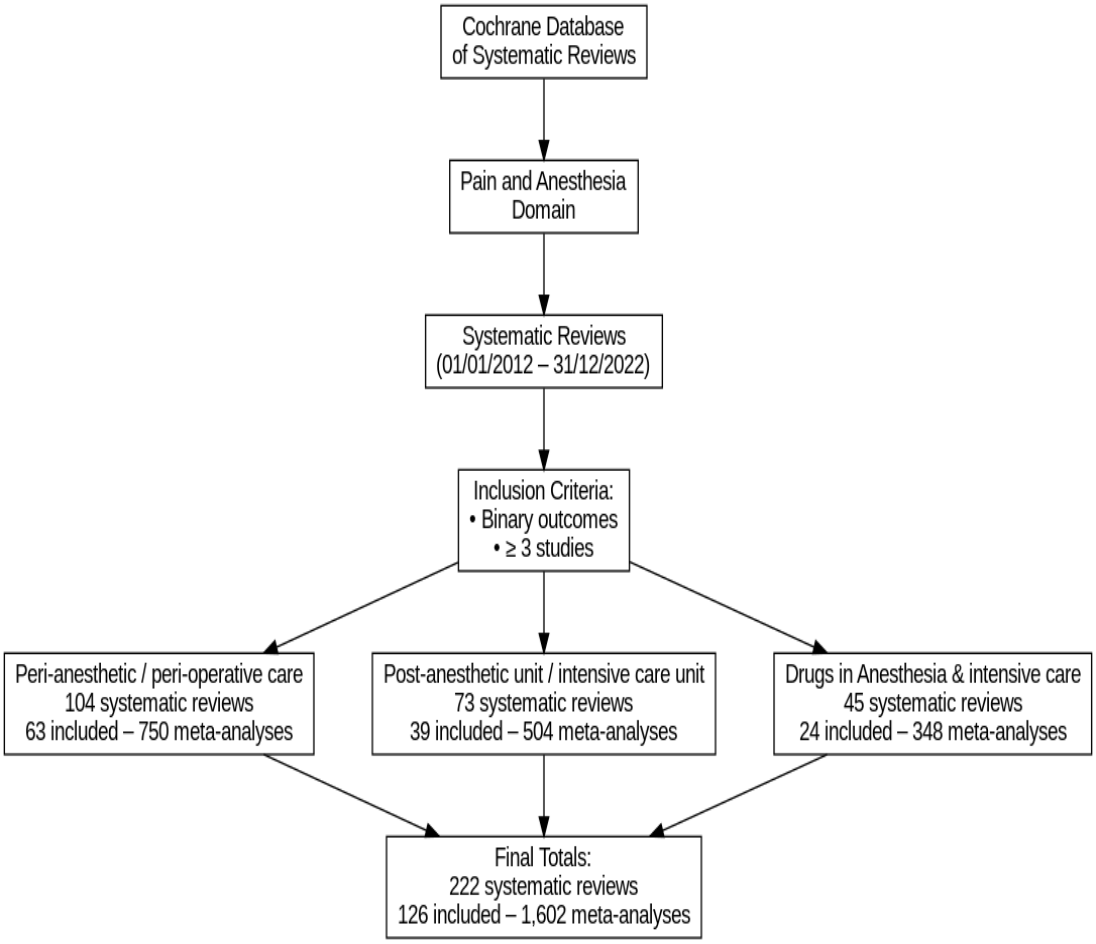


Fig. 3 — A PRISMA-style flow diagram detailing the study-selection process.

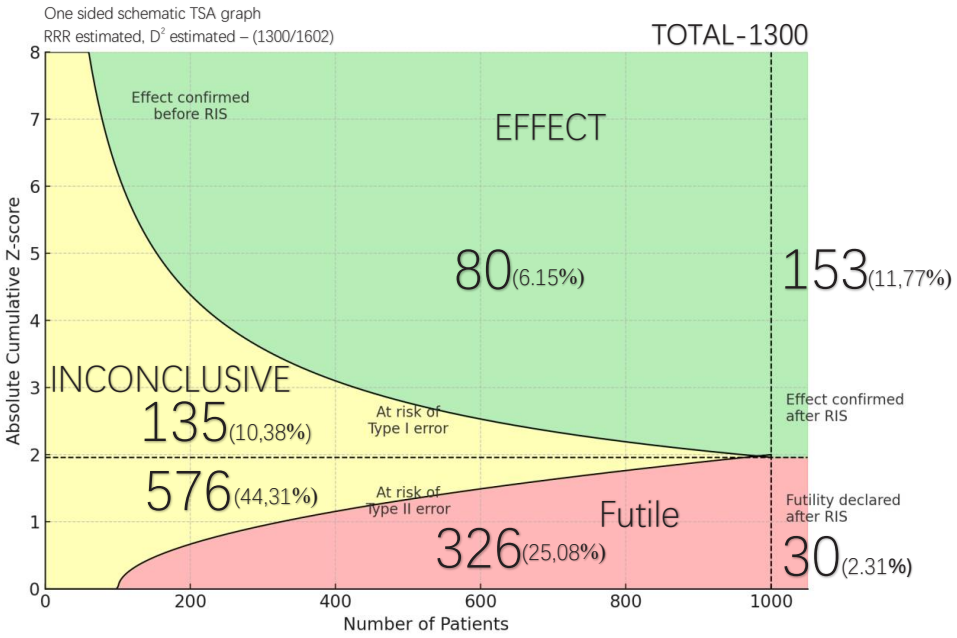


Fig. 4 — A one-sided schematic TSA graph with our results included.

without reaching the DARIS, consistent with heightened false-negative risk due to insufficient information. Insufficient information size thus threatens reliability on both sides of the significance divide.

Discussion

Interpretation of Key Findings

Across the 1,300 Cochrane meta-analyses re-appraised with conventional statistics and TSA, decisive evidence was uncommon. Only 17.9% crossed both the conventional $p \leq 0.05$ threshold and the diversity-adjusted required information size (DARIS), consistent with benefit; 27.4% met futility; and 54.7% remained inconclusive due to imprecision. Nearly two-thirds of findings that would appear "positive" (62.8%) or "negative" (63.9%) under conventional testing lacked adequate information to limit type I or type II error inflation. Collectively, these patterns indicate that meta-analytic evidence in anesthesiology is often underpowered and that reproducible findings are uncommon.

Our results echo—and quantitatively extend—meta-epidemiological work across clinical medicine. Brok et al. showed that many apparently conclusive meta-analyses failed to reach the information size for reliable inference, with TSA downgrading a large share of nominally significant results[18]. Jakobsen et al. embedded TSA within an eight-step appraisal framework for Cochrane reviews, revealing that numerous findings revert to indeterminate once random-error risk is addressed[33]. Wetterslev et al. formalized sequential monitoring boundaries and demonstrated superior control of both false-positive and false-negative conclusions compared with conventional random-effects pooling[21]. In critical-care exemplars, conventionally "conclusive" signals frequently became inconclusive once information-size criteria were applied (e.g., 72%)[34]. Recent audits also report limited reproducibility of information-size calculations in published TSAs (fewer than one-third)[35].

In interpreting our findings, we prioritized an objective, TSA-based benchmark over direct comparisons with narrative conclusions in review abstracts. Such comparisons risk conflating the adequacy of information size with variability in abstract reporting, including the potential for "spin"—reporting strategies that make findings appear more favorable than the results justify[36]. Even in Cochrane Reviews, which generally maintain higher reporting standards, observational

Table I.

TSA Classification	N	% of Total
Confirmed treatment effect	233 (80+153)	17.9%
Inconclusive (Type I &II Risk)	711 (576+135)	54.7%
Futility (no effect likely)	356 (326+30)	27.4%
Total	1300	100.0%

analyses show that abstract reporting still incompletely adheres to PRISMA for Abstracts, with roughly 6 of 12 items reported on average[37]. We therefore focused on whether the (DA)RIS was met. As a practical implication, review abstracts should explicitly state whether the (DA)RIS was achieved and, if not, the distance to (DA)RIS, so that readers can appraise certainty immediately.

Clinical implications

In anesthesiology, where trials are often small and heterogeneous, clinicians depend heavily on meta-analyses to guide care, making error control consequential[38,39]. TSA adds practical guardrails by pairing a diversity-adjusted required information size (DARIS) with benefit and futility boundaries via α-spending, separating no effect (futility crossed) from no evidence yet (no boundary crossed with insufficient information). This framework translates into a tiered response for different stakeholders. For clinicians at the bedside, if a benefit boundary is crossed, adopting an intervention is reasonable; if futility is crossed, de-implementation is appropriate; if neither boundary is crossed and information is insufficient, maintain current standards, use shared decision-making, and state uncertainty explicitly. For guideline developers, matching recommendation strength to evidence maturity curbs over-interpretation of immature signals in living guidelines. For researchers and funders, DARIS provides a concrete enrollment target to reach decisive answers and a principled basis to retire low-yield questions when futility is indicated. Adopting a simple dual-reporting standard, presenting conventional results alongside TSA metrics, makes this evidence maturity explicit with minimal additional burden.

Recommendations for Systematic Review Methodology

To enhance meta-analytic reliability, we recommend a concise dual-reporting standard (conventional results alongside TSA metrics such as the diversity-adjusted required information size (DARIS), adjusted confidence intervals, and monitoring boundaries), pre-registered TSA parameters applied consistently (anticipated effect, D^2 assumptions, and error rates specified in the protocol and, where relevant, in a pre-specified Supplement), and explicit abstract reporting of whether the DARIS was achieved, and if not, the distance to DARIS.

Methodological Limitations

This investigation is subject to several methodological constraints. Because it relies exclusively on Cochrane reviews, its findings may not generalize to non-Cochrane syntheses that follow different standards. Continuous outcomes, such as pain scores and hemodynamic parameters, which comprise a substantial share of anesthesiology endpoints, were excluded because trial sequential analysis (TSA) is optimized for binary data. The analysis employs one-sided absolute Z-scores, a choice that can under-represent bidirectional effects in certain intervention settings, and it depends on published review data, which may carry forward extraction errors from the primary studies. Moreover, TSA's inherent reliance on modeled assumptions can diverge from the empirical effect sizes and heterogeneity observed in subsequent trials. Although human error may have crept into the initial analyses, our large sample size offers partial mitigation. Despite these limitations, the broad scope of 1,300 analyses and the use of a systematic evaluation protocol bolster the credibility of our conclusions. We did not examine how TSA findings align with abstract narratives. Future pre-registered research could map the prevalence of narrative "spin" and evaluate concordance with TSA-based evidence thresholds; narrative "spin" is an important phenomenon that merits dedicated investigation.

Conclusions

This investigation demonstrates quantitatively that most anesthesiology meta-analyses (85.9%) lack requisite statistical power to reliably inform evidence-based clinical decisions. TSA methodology offers a critical quantitative framework to differentiate robust evidence from premature inferential claims, particularly within medical specialties prone to fragmented evidence bases. By implementing rigorous information size thresholds and error control mechanisms, systematic reviews can evolve beyond simple data aggregation to function as statistically valid foundations for clinical practice guidelines. While meta-analyses represent a powerful evidence synthesis methodology, their reliability fundamentally depends on adequate statistical power, appropriate multiple testing adjustment, and transparent analytical protocols.

Acknowledgements and potential conflicts of interest: The authors declare that they have no known financial, professional, or personal conflicts that could have influenced the work reported in this manuscript. No external funding was received, and no commercial or institutional relationships exist that might be perceived to bias the study's design, analysis, interpretation, or presentation.

References

1. Ahn EJ, Kang H. Introduction to systematic review and meta-analysis. Korean J Anesthesiol 2018 71:103–112.
2. Haidich AB. Meta-analysis in medical research. Hippokratia 2010 14:29–37.
3. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med 2016 21:125–127.
4. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 7:177–188.
5. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 315:629–634.
6. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 56:455–463.
7. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 21:1539–1558.
8. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 327:557–560.
9. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 336:924–926.
10. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 340:c332.
11. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 358:j4008.
12. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 6:e1000097.
13. Moher D, Shamseer L, Clarke M, et al.; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 4:1.
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 372:n71.
15. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PloS One 2013 8:e59202.
16. Ioannidis JPA. Why most published research findings are false. PloS Med 2005 2:e124.
17. De Oliveira GS Jr, Chang R, Kendall MC, Fitzgerald PC, McCarthy RJ. Publication bias in the anesthesiology literature. Anesth Analg 2012 114:1042–1048.
18. Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive—trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. Int J Epidemiol 2009 38:287–298.
19. Schmid C, Chandler J. Should Cochrane apply error-adjustment methods when conducting repeated meta-analyses? [Internet]. 2025 [cited 2025 Apr 1]. Available from: https://methods.cochrane.org/sites/default/files/public/uploads/tsa_expert_panel_guidance_and_recommendation_final.pdf
20. Borm GF, Donders AR. Updating meta-analyses leads to larger type I errors than publication bias. J Clin Epidemiol 2009 62:825–830.e10.
21. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol 2008 61:64–75.
22. Thorlund K, Engstrøm J, Wetterslev J, Brok J, Imberger G, Gluud C. User Manual for Trial Sequential Analysis (TSA). 2nd ed. Copenhagen, Denmark: Copenhagen Trial Unit; 2017.
23. Claire R, Gluud C, Berlin I, Coleman T, Leonardi-Bee J. Using trial sequential analysis for estimating the sample sizes of further trials: example using smoking cessation intervention. BMC Med Res Methodol 2020 20:284.
24. Thorlund K, Devereaux PJ, Wetterslev J, et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? Int J Epidemiol 2009 38:276–286.
25. Imberger G, Gluud C, Boylan J, Wetterslev J. Results in systematic reviews of anaesthesia interventions with high type I error risk [poster]. Presented at: Cochrane Colloquium; 2015; Vienna, Austria.
26. Umberham B, Hedin R, Detweiler B, Kollmorgen L, Hicks C, Vassar M. Heterogeneity of studies in anesthesiology systematic reviews: a meta-epidemiological review and proposal for evidence mapping. Br J Anaesth 2017 119:874–884.
27. Seidler AL, Hunter KE, Cheyne S, Berlin JA, Ghersi D, Askie LM. Prospective meta-analyses and Cochrane's role in embracing next-generation methodologies. Cochrane Database Syst Rev 2020 10:ED000145.
28. Shaheen N, Shaheen A, Ramadan A, Hefnawy MT, Ramadan A, Ibrahim IA, et al. Appraising systematic reviews: a comprehensive guide to ensuring validity and reliability. Front Res Metr Anal 2023 8:1268045.
29. Copenhagen Trial Unit. Trial Sequential Analysis (TSA) [computer program]. Version 0.9.5.10 Beta. Copenhagen, Denmark: Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen University Hospital – Rigshospitalet; 2021.
30. Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. Journal of Clinical Epidemiology. 2008;61(8):763-769.
31. Mouelhi Y, Jouve E, Castelli C, Gentile S. How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. Health and Quality of Life Outcomes. 2020; 18:136.
32. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications tatist external validity of trial results. Trials 2015; 16:495.
33. Jakobsen JC, Wetterslev J, Winkel P, Lange T, Gluud C. Thresholds for tatistica land clinical significance in systematic reviews with meta-analytic data. BMC Med Res Methodol 2014 14:120.
34. Koster T, Wetterslev J, Gluud C, Jakobsen JC. Apparently conclusive meta-analyses on interventions in critical care may be inconclusive—a meta-epidemiological study. J Clin Epidemiol 2019 114:1–10.
35. Xing X, Wang Y, Wang Y, Murad MH, Lin L. Lack of reproducibility of trial sequential analyses: a meta-epidemiological study. Trials 2025 26:93.
36. Lazarus C, Haneef R, Ravaud P, Boutron I. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. BMC Med Res Methodol 2015; 15:85.
37. Helbach J, Wandscher K, Pieper D, Hoffmann F. Cochrane review abstracts are getting longer, but this has no large impact on the reporting quality. J Evid Based Med. 2023;16(3):294-302.
38. Umberham B, Hedin R, Detweiler B, Kollmorgen L, Hicks C, Vassar M. Heterogeneity of studies in anesthesiology systematic reviews: a meta-epidemiological review and proposal for evidence mapping. Br J Anaesth 2017 119:874–884.
39. Kim SY, Park JE, Lee YJ, et al. Introduction to systematic review and meta-analysis. Korean J Anesthesiol 2018 71:103–112.

doi.org/10.56126/76.S.07