Acta Anaesth. Bel., 2025, 76 (4): *247-260*

Original study

Exploring discordance in evidence from meta-analyses and subsequent large-scale randomized controlled trials in perioperative medicine

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

[1]*Department of Anesthesiology, University Hospital Ghent, Belgium.*
[*]*The first three authors contributed equally to this work.*

Corresponding author: Vanoverschelde Heleen, Corneel Heymanslaan 10, 9000 Ghent, Belgium.
E-mail: Heleen.Vanoverschelde@UZGent.be

Abstract

Background: **Meta-analyses and randomized controlled trials (RCTs) play pivotal roles in evidence-based medicine. However, meta-analyses are increasingly criticized for overestimating treatment effects and lacking agreement with large RCTs, potentially resulting in misleading or premature conclusions that influence clinical guidelines. Small, early-phase trials and publication bias contribute to type I and type II errors, raising concerns about the strength of meta-analytic findings. Trial Sequential Analysis (TSA) is a statistical tool designed to assess the robustness of cumulative evidence by adjusting for random errors and required information size.**
Objective: **This study evaluates the agreement between meta-analyses and subsequent large RCTs in perioperative medicine published between 2015 and 2022. Additionally, it investigates whether TSA alters the interpretation of meta-analytic findings.**
Methods: **A systematic search identified large RCTs (≥1,000 participants, with at least one major dichotomous clinical outcome) and their corresponding preceding meta-analysis. TSA was applied to each outcome to determine whether the meta-analysis had reached a reliable conclusion and to classify results into distinct evidence zones.**
Results: **Of the 23 outcome comparisons assessed, 78.3% of meta-analyses correctly predicted the results of the corresponding large RCTs. However, TSA reclassified several initially 'accurate' predictions as inconclusive or potentially false positive, particularly under assumptions of higher relative risk reductions.**
Conclusion: **Although meta-analyses often align with subsequent RCTs, they carry a substantial risk of false positives, especially when based on small studies. TSA adds important nuance by identifying when cumulative evidence is insufficient for firm conclusions. These findings support a cautious interpretation of meta-analyses in clinical decision-making and emphasize the need for large, well-powered RCTs before changing clinical practice.**

Key words: **Meta-analysis, Randomized Controlled Trials, Perioperative Care, Sequential Analysis.**

Introduction

Meta-analyses and randomized controlled trials (RCTs) play a fundamental role in evidence-based medicine, shaping clinical decision-making and guidelines in perioperative medicine. Due to logistic challenges and the costs of large clinical trials, the vast majority of published trials in perioperative medicine and anesthesia are small, single-center studies, lacking sufficient statistical strength to reliably assess major morbidity endpoints and mortality[1,2]. The increasing number of published meta-analyses, especially in anesthesia and perioperative medicine, has led to concerns about methodological quality and duplication. Bartels and Sessler recently reported in the British Journal of Anaesthesia that the growth of meta-analyses is surpassing the production of new clinical trial data, metaphorically describing the situation as 'making more lemonade, but from only slightly more lemons'[3]. The majority of published meta-analyses may be unnecessary or misleading because of

Manuscript will be submitted at University of Ghent as a thesis.

selective inclusion of studies, publication bias, and poor reporting practices[4].

Reliable clinical guidelines are important in perioperative medicine, where clinical decisions often must be made quickly and sometimes under uncertain conditions. High-quality guidelines thus contribute to patient safety, promote the use of effective therapies, and encourage cost-effective care. In this complex and resource-intensive environment, inappropriate or insufficiently evidence-based interventions can result in avoidable complications, increased healthcare costs, or suboptimal outcomes. Consequently, it is essential that the evidence base supporting these guidelines, particularly when derived from meta-analyses, is both strong and consistent. Yet recent studies question the reliability of meta-analyses, as they sometimes have results that are later contradicted by large-scale trials[2,5–8]. This issue is of significant concern, as mistakes or premature conclusions derived from underpowered meta-analyses have the potential to shape clinical guidelines and treatment decisions, in ways that may ultimately harm patients.

Type I and type II errors related to repeated significance testing, heterogeneity and publication bias are among the threats to the validity of meta-analyses[2,3,5]. As statistical testing in meta-analyses is based on the null hypothesis, a type I error is made when the null hypothesis is incorrectly rejected, resulting in a claim of a treatment effect that does not actually exist (a false positive). Similarly, a type II error occurs when the meta-analysis concludes that there is no treatment effect, when in reality one does exist (a false negative). Meta-analyses often use small RCTs, which may suffer from methodological limitations, leading to overestimated treatment effects[9]. Meta-analyses dependent on relatively limited data, e.g. a small number of trials or few events, are particularly vulnerable to random variation and imprecision, increasing the risk of drawing incorrect conclusions[10–12]. Because these analyses are often updated with data from new trials, they are subjected to repeated statistical testing. This increases the risk of detecting false-positive results, an issue called 'multiplicity due to repeated significance testing'. This concept is well-known in the context of RCTs, where it has been established that replication of an accumulation of data increases the overall probability of type I errors[11]. Previous studies estimate the risk of a type I error in meta-analyses to range between 10% and 30%, implying that 1 to 3 out of 10 interventions may be falsely reported as beneficial[6,12,13]. Additionally, Møller et al. critically commented on the diminishing value of systematic reviews and meta-analyses in medical research, especially in a time where there are notable increases in poor quality and redundant studies in medical literature[14]. Furthermore, the dichotomous interpretation of p-values, commonly used in both primary studies and meta-analyses, has received increasing criticism. Statisticians have argued for abandoning 'statistical significance' in favor of more nuanced interpretations of evidence[15].

To address these issues, Trial Sequential Analysis (TSA) has emerged as a methodological innovation designed to improve the reliability of cumulative meta-analyses by adjusting for random errors and determining the required information size for reliable conclusions[10,16–18]. TSA operates similarly to interim analyses in RCTs. As new trials are sequentially added to a meta-analysis, a cumulative z-score is calculated after each addition, representing the strength of evidence for a treatment effect at that point in time. The z-score is a standardized statistic that quantifies how far the observed effect deviates from the null hypothesis (typically "no effect"), expressed in units of standard deviation. This z-score is plotted across the x-axis (number of participants/events), forming a z-curve (Fig. 1). The curve is then evaluated against pre-defined monitoring boundaries (e.g., O'Brien-Fleming or Lan-DeMets boundaries), which indicate whether the accumulated evidence is statistically convincing (boundary crossed), inconclusive (no boundary crossed), or suggests futility (futility boundary crossed)[10,17,19,20]. TSA can be performed using two-sided or one-sided testing, depending on the hypothesis. Two-sided testing evaluates whether any difference exists (benefit or harm), whereas one-sided testing is used when a directional effect is hypothesized (e.g., that a treatment improves outcomes)[10,17,21]. In this study, we performed one-sided TSA, as the included meta-analyses predominantly tested interventions hypothesized to reduce perioperative risk. This directional focus justified using one-sided boundaries, aiming to detect beneficial effects rather than harm or neutral outcomes.

There has been an increase in interest in sequential methods to improve the reliability of systematic reviews, meta-analyses and RCTs[22,23]. In this discussion, however, we focus specifically on TSA.

Sivakumar et al. examined the predictive value of meta-analyses in perioperative medicine, published from 2000 to 2014, by comparing them with subsequent large RCTs2. They found that in approximately 40% of cases, the results were inconsistent, often due to the overestimation of treatment effects in the meta-analyses. They

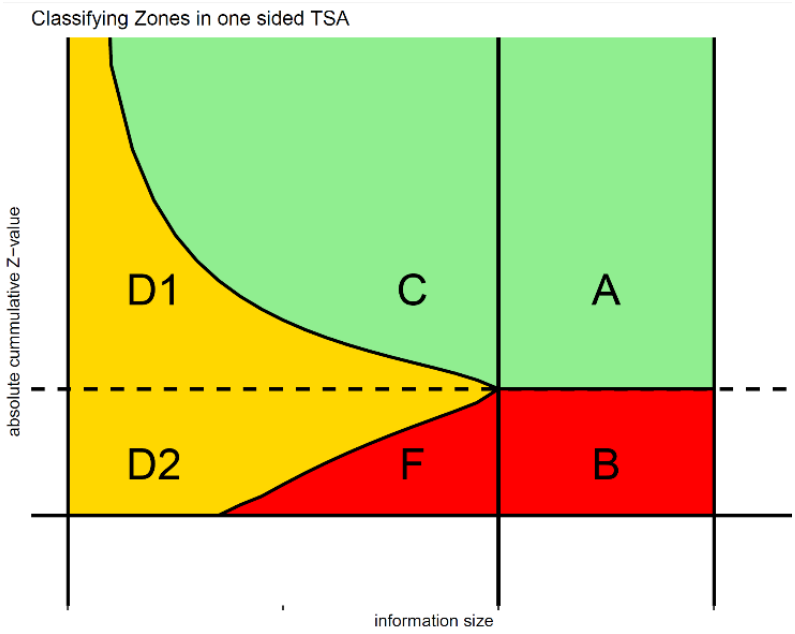

Fig. 1 — Illustration of a Trial Sequential Analysis (TSA) graph. The x-axis represents a linear scale of the cumulative number of patients analyzed across included studies (i.e., the accrued information size). The y-axis shows the absolute cumulative Z-value, but no numerical scale is displayed. The black horizontal line reflects a Z-value of 1.96, which corresponds to statistical significance in a two-sided test ($p < 0.05$). The curved lines in the figure represent the superiority boundary (above) and the futility boundary (below), which are used to classify the strength and conclusiveness of the evidence. Zone A (True Positive): The required information size is reached and the superiority boundary is crossed, indicating a reliable and conclusive positive effect. Zone B (True Negative): The required information size is reached and the futility boundary is crossed, confirming the absence of a meaningful effect. Zone C (Significant but inconclusive): The cumulative Z-curve crosses the superiority boundary before reaching the required information size. Although statistically significant, the evidence remains insufficient for a firm conclusion, and further studies are needed. Zone D1 (False Positive): The Z-curve crosses the conventional significance threshold ($p<0.05$) but not the TSA superiority boundary, suggesting a potentially misleading conclusion based on premature significance. Zone D2 (False Negative): The Z-curve remains within both boundaries without reaching the required information size, leaving uncertainty about the presence or absence of an effect. Zone F (Non-significant but inconclusive): The Z-curve crosses the futility boundary before reaching the required information size, indicating a non-significant effect, but the evidence is not yet strong enough to rule out a potential benefit or harm.

concluded that meta-analyses should rarely be regarded as definitive evidence without confirmation from large, well-conducted RCTs. This study builds upon the research of Sivakumar et al. by extending their analysis to include data from 2015 to 2022. By systematically evaluating the concordance between meta-analyses and subsequent large RCTs over this period, this study aims to determine whether the trends observed in previous research persist. Additionally, by integrating TSA into the meta-analytic framework, we seek to determine whether previous conclusions remain valid when adjusted for sequential monitoring and required information size.

Methodology

The consistency between meta-analyses and subsequent large randomized trials in perioperative medicine was explored using a structured and transparent approach. A comprehensive search was conducted in Medline to identify studies in the fields of anesthesia and perioperative medicine, using the search terms “anesthesia” and “perioperative medicine”. The search was limited to RCTs and multicenter studies, published between 2015 and 2022. All abstracts were screened using a custom text-scoping algorithm developed in R, specifically designed to identify RCTs enrolling at least 1,000 participants. The algorithm used regular expressions and keyword-based filters to scan article titles and abstracts for indicators of study design (e.g.’randomized’, ‘controlled trial’) and study size (e.g. ‘n = 1000’, ‘≥1000 participants’, or numerically reported sample sizes). The output was manually validated during development to ensure accurate identification of large multicenter trials. Any uncertainties or

borderline cases were resolved through manual review. Trials were included in the analysis if they investigated a clinical intervention and reported at least one major dichotomous clinical endpoint, such as all-cause mortality, myocardial infarction, acute renal failure, or other serious perioperative complications. Trials were excluded if they lacked a clearly defined intervention or relevant endpoint, were single-center, or included fewer than 1,000 participants. Screening and eligibility assessment were conducted by a single reviewer, with any uncertainties resolved in consultation with a second independent assessor.

For each eligible large RCT, additional searches were performed in Medline, PubMed, and the bibliography of the trial to identify relevant meta-analyses published prior to the RCT. Meta-analyses were eligible for inclusion if they focused exclusively on RCTs, investigated a comparable clinical intervention in a similar patient population, and reported at least one similar major dichotomous clinical endpoint. Endpoints were included irrespective of whether they were defined as primary or secondary in the meta-analysis or the RCT, provided they were considered clinically important perioperative outcomes. In cases where multiple eligible meta-analyses were identified for the same research question, the most recent or highest quality study was selected.

TSA was performed using the R-package RTSA (v 0.2.2). For each included outcome, two separate TSA models were applied using predefined assumptions. The required information size was estimated with an alpha level of 5%, a beta level of 20%, and an assumed relative risk reduction (RRR) of 20% in one model and 30% in the other. The proportion of events in the control group was derived from trials assessed as low risk of bias, based on the Cochrane Risk of Bias tool, and heterogeneity was adjusted using the D^2 correction method, as recommended by Wetterslev et al.[18]. Significance and futility monitoring boundaries were calculated using the Lan-DeMets O'Brien-Fleming method, and meta-analyses were performed using a random-effects model according to DerSimonian and Laird[19].

The results of each TSA model were visualized to aid interpretation and are presented as TSA graphs. A TSA graph is assessed by examining the trajectory of the cumulative Z-score in relation to the predefined monitoring boundaries (Fig. 1). The x-axis of the graph reflects the total number of participants accumulated across studies (the information size), while the y-axis represents the cumulative Z-score. The Z-score is a statistical measure that quantifies how far the observed effect deviates from the null hypothesis, expressed in standard deviations. In a two-sided test, a Z-score exceeding +1.96 or falling below –1.96 corresponds to a p-value less than 0.05, indicating statistical significance. If the Z-curve crossed the superiority boundary, this was taken as an indication of conclusive evidence of benefit; crossing the futility boundary suggested that further trials were unlikely to alter the conclusion. If the Z-curve remained within both boundaries without reaching the required information size, the findings were considered inconclusive. These interpretations were then compared with the outcomes of the corresponding large RCTs to evaluate the predictive value of TSA and assess whether the conclusions of earlier meta-analyses would have remained valid when adjusted for sequential monitoring and required information size.

Results

The Medline search yielded a total of 16,271 studies, of which 12,437 related to 'anesthesia' and 3,834 to 'perioperative medicine' (Fig. 2). After applying the inclusion criteria, 23 RCTs were identified. These studies each included more than 1,000 participants and involved a clinical intervention targeting at least one major dichotomous morbidity-related endpoint. However, for several of these RCTs, no corresponding meta-analysis meeting the inclusion criteria could be identified. As a result, 13 RCTs were excluded due to the absence of a relevant preceding meta-analysis focusing exclusively on randomized trials and comparable clinical endpoints. A parallel search for meta-analyses resulted in 44 studies, of which 15 were retained as they represented the most recent or highest quality meta-analysis preceding the corresponding RCT. Two meta-analyses were excluded due to the inclusion of non-randomized trials. In total, we identified 10 large RCTs and 13 meta-analyses, each addressing at least one eligible endpoint comparison[24–46] (Table IV,Table V). These studies evaluated the effects of nine clinical interventions on 23 major morbidity-related outcomes (Fig. 3).

For each endpoint, the results of the large RCTs were compared with those of the most recent and comprehensive preceding meta-analysis. Among the 23 assessed endpoint outcomes, 18 (78.3%) were accurately predicted by the meta-analyses (Table I).

Several endpoint predictions initially classified as 'accurate' based on binary significance showed a shift in interpretation when assessed through

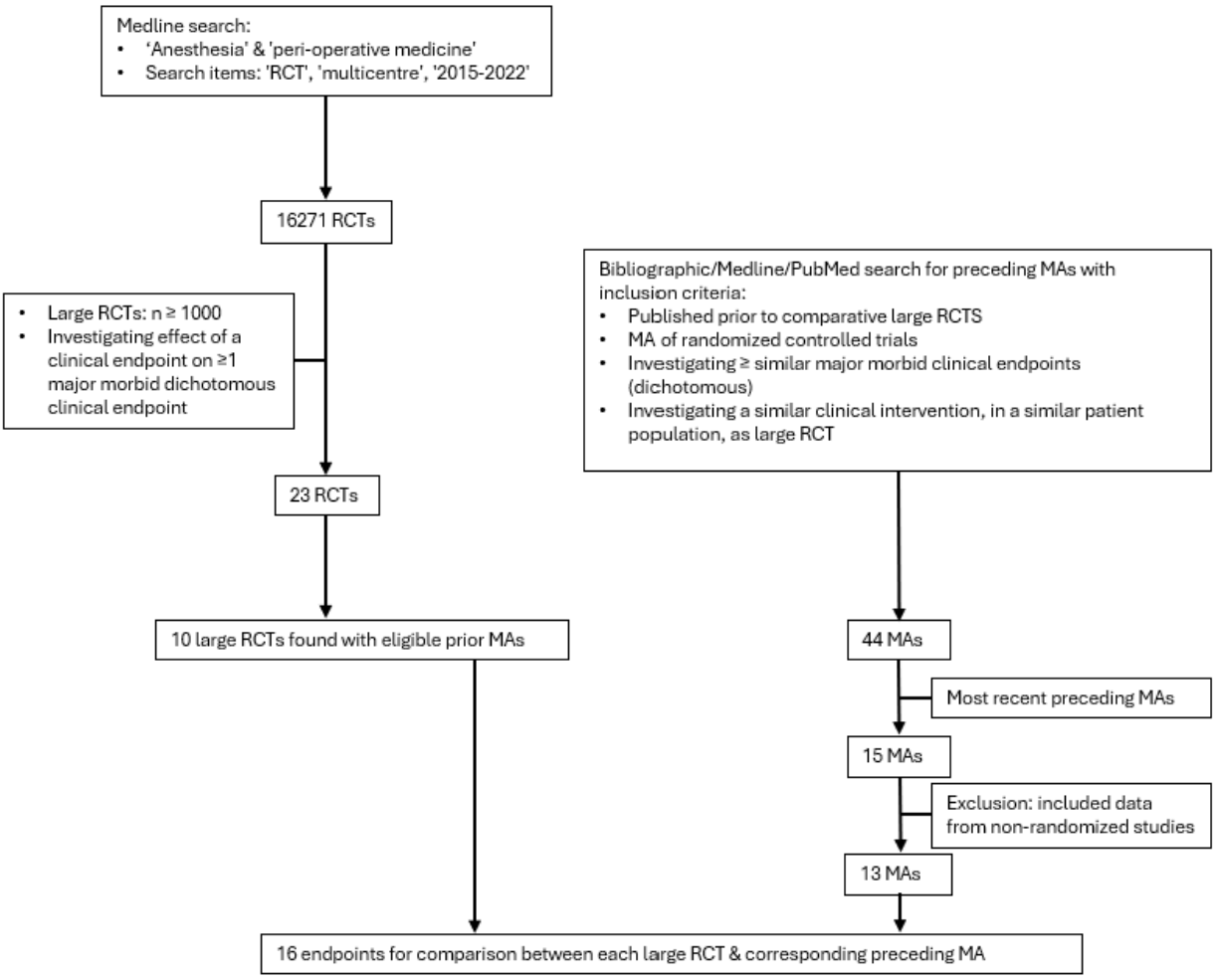

Fig. 2 — Study method flow diagram. RCT: Randomized controlled trial; MAs: Meta-analyses.

Table I. — Analysis of endpoint comparisons between large randomized controlled trials (RCTs) and the most recent preceding meta-analysis of randomized trials examining those endpoints, where study findings were classified as 'significant' or 'non-significant'.

	Large RCTs	
Meta-analyses	Significant	Non-significant
Significant	2	5
Non-Significant	0	16

the lens of TSA. In particular, endpoints initially considered concordant between the meta-analysis and the RCT (e.g., both statistically significant) were reclassified into distinct TSA zones, such as Zone C (significant but inconclusive) or Zone D1 (false positive) (Table II, Fig. 4). These shifts were more pronounced when assuming higher clinical risk differences (Table III, Fig. 5).

Table II. — Classification of endpoint comparisons between meta-analyses and large randomized controlled trials (RCTs) using Trial Sequential Analysis (TSA) zones, based on an assumed clinical difference (CD) of 70%. Study findings were categorized according to concordance between meta-analyses and subsequent RCTs, and further classified as 'significant' or 'non-significant' based on statistical outcomes.

CD 70%	Large RCTs	
Meta-analyses	Significant	Non-significant
A	1	1
C	1	0
D1	0	4
B	0	1
F	0	0
D2	0	15

Table III. — Classification of endpoint comparisons between meta-analyses and large randomized controlled trials (RCTs) using Trial Sequential Analysis (TSA) zones, based on an assumed clinical difference (CD) of 80%. Study findings were categorized according to concordance between meta-analyses and subsequent RCTs, and further classified as 'significant' or 'non-significant' based on statistical outcomes.

CD 80%	Large RCTs	
Meta-analyses	Significant	Non-significant
A	0	1
C	1	0
D1	1	4
B	0	0
F	0	0
D2	0	16

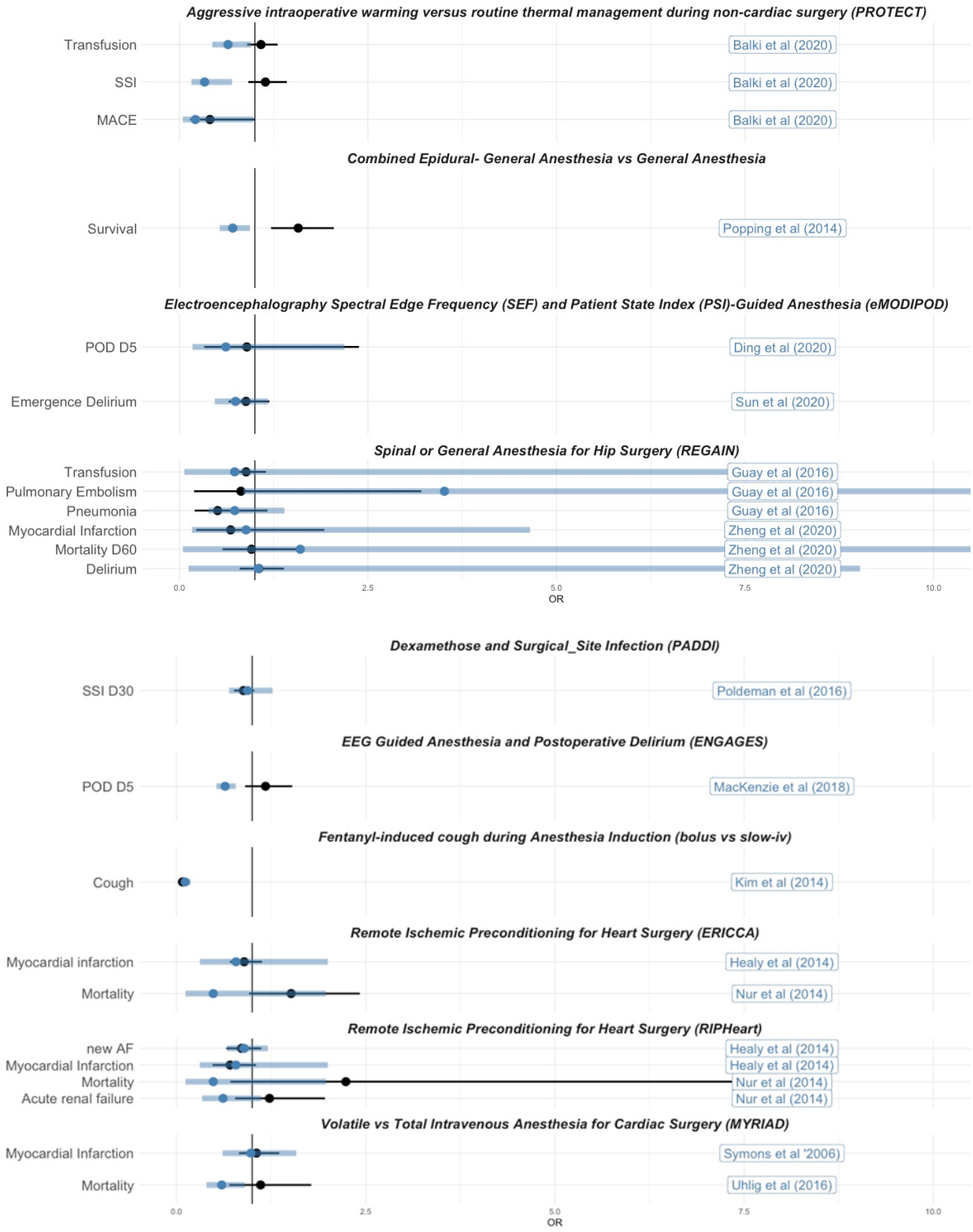

Fig. 3 — Forest plot derived from the present analysis, comparing effect estimates from meta-analyses (blue) and subsequent large randomized controlled trials (black) across multiple perioperative outcomes. Each outcome is plotted with its corresponding odds ratio (OR) and 95% confidence interval. The vertical line at OR = 1 indicates no effect. Overlapping confidence intervals suggest agreement between evidence sources, whereas non-overlapping intervals highlight inconsistency. SSI: Surgical site infection; MACE: Major adverse cardiovascular events; POD D: Postoperative delirium on day 5; AF: Atrial fibrillation.

Discussion

Meta-analysis vs RCTs

This analysis found that meta-analyses were more likely to find a statistical significant treatment effect than subsequent large RCTs evaluating similar interventions in perioperative medicine. When compared to the RCTs, meta-analyses correctly predicted 78.3% of outcomes. However, this means that 21.7% of the outcomes were contradicted by later trial results. The forest plots highlight a fundamental issue: meta-analyses based on small

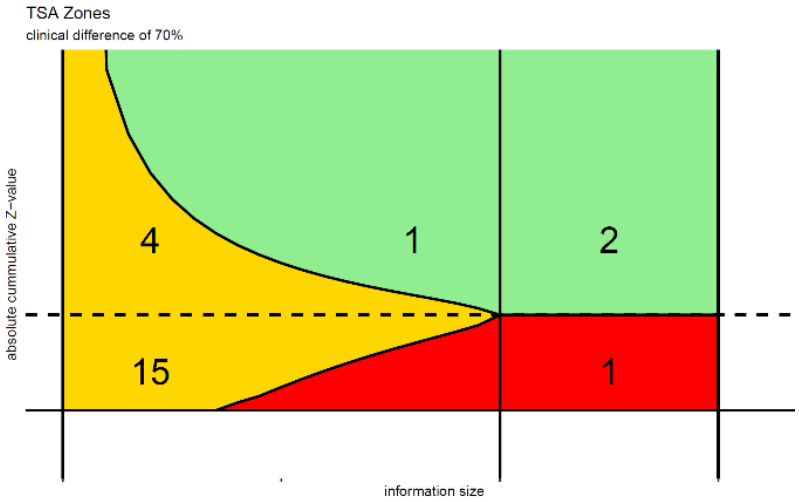

Fig. 4 — Trial Sequential Analysis (TSA) zones applied to the set of meta-analyses, based on a prespecified clinical difference of 70%. The x-axis represents the cumulative information size, while the y-axis shows the absolute cumulative Z-value, indicating the cumulative strength of statistical evidence. Meta-analyses are categorized into three zones according to TSA monitoring boundaries: green indicates a statistically significant result (n = 1+2), yellow reflects inconclusive evidence (n = 19), and red denotes a statistically insignificant result (n = 1). The numbers represent the number of meta-analyses falling within each evidence zone. Notably, within the green zone, one meta-analysis did not reach the required information size, meaning that although the result is statistically significant and likely reliable, additional studies are warranted to confirm the conclusion with sufficient power.

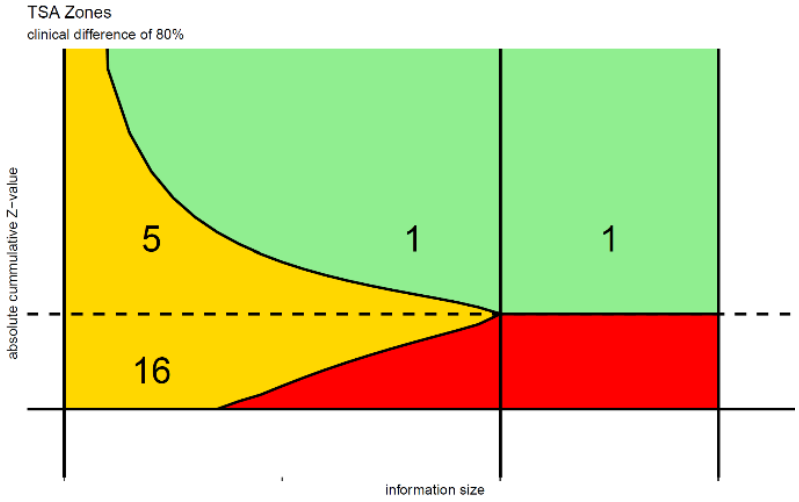

Fig. 5 — Trial Sequential Analysis (TSA) zones applied to the set of meta-analyses, based on a prespecified clinical difference of 80%. The x-axis represents the cumulative information size, while the y-axis shows the absolute cumulative Z-value, indicating the cumulative strength of statistical evidence. Meta-analyses are categorized into three zones according to TSA monitoring boundaries: green indicates a statistically significant result (n = 1+1), yellow reflects inconclusive evidence (n = 21), and red denotes a statistically insignificant result (n = 0). The numbers represent the number of meta-analyses falling within each evidence zone. Notably, within the green zone, one meta-analysis did not reach the required information size, meaning that although the result is statistically significant and likely reliable, additional studies are warranted to confirm the conclusion with sufficient power.

studies may overestimate effect sizes or suggest significance where none exists (Fig. 3). These results are in line with those by Sivakumar et al., and extend their conclusions by incorporating data from 2015 to 2022. This reinforces the concern that most meta-analyses do not have sufficient statistical power to reliably confirm or exclude substantial intervention effects[2,5–7,14,21,47,48].

Several factors may contribute to the higher likelihood of meta-analyses reporting significant

treatment effects. First, an important issue is the inclusion of multiple secondary endpoints from component trials, which increases the risk of type I errors. This is especially important since statistical correction for multiple comparisons is rarely applied in meta-analyses due to data complexity[2,49]. This complexity arises from differences in how outcomes are defined, measured, and reported across studies, making it difficult to apply uniform correction methods. Trials often report secondary endpoints inconsistently or selectively, and researchers may lack access to individual patient data, limiting their ability to harmonize outcome definitions. Furthermore, overlapping or correlated endpoints can complicate the assessment of statistical independence, thereby undermining the assumptions required for standard multiplicity adjustments. The heterogeneity in follow-up duration, measurement tools, and reporting formats across trials further exacerbates the analytical challenges. As a result, decisions regarding which outcomes to include and how to handle multiplicity are often made post hoc, increasing the risk of bias and inflated effect estimates[5,14].

Second, the common misinterpretation of non-significant results as evidence of no effect presents another major challenge in the interpretation of evidence. In conventional meta-analysis, a p-value >0.05 is often (incorrectly) taken as proof of equivalence or futility, even when confidence intervals remain wide and event rates are low. This false reassurance may lead to premature abandonment of potentially effective interventions[15,50–52].

Third, the lack of blinding in many trials included in meta-analyses increases the risk of observer and performance bias, potentially leading to exaggerated treatment effects[2,5,53]. This is especially relevant for perioperative interventions, such as anesthetic techniques, fluid management, or recovery protocols, which are often difficult to blind. Pooling such studies makes meta-analyses more susceptible to inflated effect estimates, contributing to discrepancies when compared with larger, well-controlled RCTs.

Fourth, heterogeneity across studies further undermines the reliability of meta-analytic findings. Differences in patient populations, intervention protocols, outcome definitions, follow-up durations, and study quality can introduce substantial clinical and methodological variability[5,14]. Although statistical methods such as random-effects models aim to account for heterogeneity, high between-study variability can still distort pooled estimates and limit the generalizability of conclusions. In perioperative medicine, where interventions and patient characteristics vary widely, this heterogeneity poses a significant challenge to drawing robust and consistent conclusions from meta-analyses.

Fifth, the strengths and weaknesses of meta-analyses and large RCTs differ fundamentally. Large RCTs are often multicenter, pragmatic trials with broader patient populations, robust randomization, blinding, and research governance that help minimize bias and statistical error. Meta-analyses, on the other hand, typically include smaller, single-center efficacy trials. These smaller trials often lack sufficient statistical power, increasing the risk of type II errors by failing to detect true treatment effects[2,14].

Lastly, positive or exciting results are more likely to be submitted and accepted for publication, a phenomenon consistently demonstrated across various fields, including perioperative medicine[2,5–7]. This selective reporting inflates the proportion of significant findings in the available literature, indirectly increasing the risk of type I errors when such biased data are pooled in meta-analyses. Although publication bias remains a concern, recent evidence suggests some progress in its recognition and management. Khan et al. identified publication bias in leading anesthesiology journals in 2016 and 2025[54,55]. They reported that a higher proportion of meta-analyses in perioperative medicine now explicitly discuss (67.8% vs. 55.1%) and formally assess (58.4% vs. 43%) publication bias, indicating increased author awareness. Funnel plot usage and grey literature searches have also become more common. Despite these improvements, publication bias was still identified in 17.9% of recent studies, underscoring that bias in pooled estimates remains a relevant concern requiring continued attention[55].

Trial Sequential Analysis

To better interpret the predictive value of meta-analyses, this study incorporated TSA, which introduces a more nuanced framework beyond the binary distinction of statistical significance. Whereas traditional RCT outcomes are often interpreted simply as ‘significant’ or ‘non-significant’, TSA classifies evidence into zones (A, B, C, D1, D2, F), offering a gradient of evidential strength. This approach accounts for the risks of random error and insufficient information size, contextualizing findings within the broader accumulation of data over time[10,16,17].

When applying TSA to our data, the outcome of the analysis depends on the assumed clinical risk difference. With an assumed clinical risk difference of 70%, we observe that only three analyses yield a conclusive result, having reached the required

information size Table II, Fig. 4). One result falls into zone C, indicating a statistically significant finding, but with insufficient data to support a reliable conclusion. Notably, one endpoint that was previously classified as statistically significant in the original meta-analysis was reclassified as false positive by TSA, reflecting the influence of sequential monitoring on evidence strength. Out of 23 analyses, 19 fall into either the false-positive or false-negative zone, meaning the data are insufficient to draw firm conclusions. When the assumed clinical risk difference is increased to 80%, representing a stricter and more conservative threshold (i.e., requiring a larger treatment effect to be considered clinically meaningful), the required information size increases. Consequently, even more analyses fail to reach conclusiveness and shift into the inconclusive zones. This illustrates that stricter assumptions about the minimal clinically important difference raise the evidentiary bar and may expose apparent effects as statistically fragile or unsupported (Table III, Fig. 5).

These results align with findings from previous research[6–8,10,16,17,56]. Imberger et al. identified false positives in 7% (95% CI 3% to 14%) of sufficiently powered and ultimately negative Cochrane meta-analyses assessing binary outcomes[13]. Notably, the same study found that TSA prevented these false positives in 93% of cases (95% CI 68% to 98%) when credible, clinically grounded parameters were applied.

Importantly, the classification of an outcome may shift between TSA zones depending on the assumed clinical risk difference (e.g., 70% or 80% of the expected effect). The assumed clinical risk difference refers to the minimum effect size, expressed as an absolute difference in risk between intervention and control groups, that is considered clinically meaningful. In the context of perioperative medicine, this may involve a reduction in the risk of major complications such as myocardial infarction, acute renal failure, postoperative delirium, or mortality. This threshold is specified a priori and has a direct impact on the shape and position of TSA monitoring boundaries. A larger assumed clinical risk difference (e.g., expecting a 30% relative reduction) makes it statistically easier to reach a boundary indicating benefit or harm, while assuming a smaller, more realistic effect (e.g., 20% reduction) requires a larger cumulative sample size to reach the same level of statistical certainty. Therefore, the selection of this factor reflects clinical judgment about what constitutes a relevant treatment effect in a given context and plays a pivotal role in how robust or conclusive the evidence appears in TSA[10,16,56]. The relevance of a smaller or larger prespecified intervention effect may vary depending on the context of each individual meta-analysis[10].

However, despite its theoretical appeal, TSA is not without criticism. Trial sequential analysis is a complex statistical tool that can be misused, and its application is not universally adopted. In fact, a recent Cochrane collaboration has recommended against the routine use of sequential methods in primary analyses[57]. They argued that systematic reviews often assess multiple outcomes and subgroups, each with different thresholds of clinical relevance, a complexity that TSA is not well suited to accommodate. TSA is typically applied only to a single, preselected primary outcome, which limits its generalizability and leaves the risk of invalid findings for secondary outcomes unaddressed. Additionally, although TSA draws conceptual parallels with interim analyses in clinical trials, meta-analyses are fundamentally retrospective and observational in nature. Unlike trialists, meta-analysts cannot influence the timing, design, or quality of the included studies, making it impossible to retrospectively uphold the strict pre-specified assumptions that TSA relies on[11]. Nonetheless, the Cochrane Collaboration states that TSA can be used as a component of secondary analyses[57].

These findings carry important implications for clinical guideline development and evidence-based practice. Meta-analyses that suggest benefit based on early and small-scale data may lead to premature changes in clinical practice, only to be reversed when large RCTs later fail to confirm the effect. By incorporating TSA, clinicians can more confidently distinguish between genuine signals and statistical noise, potentially avoiding the adoption of ineffective or harmful interventions. This also underscores the importance of critical appraising of meta-analytic evidence, especially in fields like perioperative medicine, where many trials remain small and underpowered.

Despite the insights gained, this study has several limitations. First, the analysis was restricted to large RCTs published between 2015 and 2022, potentially excluding recent or ongoing trials with relevant endpoints. Second, although the inclusion of TSA provides greater nuance, the conclusions remain sensitive to assumptions such as expected effect size and event rates, which may vary across settings. TSA is only as reliable as the meta-analyses it is based on, and must be interpreted in the light of their inherent limitations[20]. Lastly, endpoints were matched pragmatically between RCTs and meta-analyses, which may introduce subjective bias, particularly when definitions of outcomes were not perfectly aligned. Nonetheless,

Table IV. — Characteristics and outcomes of included randomized controlled trials (RCTs).

Outcome	Intervention	Control	RCT Name	RCT Author	RCT Year	Effect Intervention	Total Intervention	Effect Control	Total Control
MACE	Active body warming	Routine thermal management	PROTECT	Sessler(24)	2022	6	2493	15	2486
SSI	Active body warming	Routine thermal management	PROTECT	Sessler(24)	2022	178	2487	157	2479
Transfusion	Active body warming	Routine thermal management	PROTECT	Sessler(24)	2022	254	2494	236	2486
POD D5	EEG guidance	No EEG guidance	eMODIPOD	Wang(45)	2020	8	771	9	774
Emergence delirium	EEG guidance	No EEG guidance	eMODIPOD	Wang(45)	2020	91	771	102	774
Mortality D60	Spinal anesthesia	General anesthesia	REGAIN	Neuman(29)	2021	30	768	32	784
Delirium	Spinal anesthesia	General anesthesia	REGAIN	Neuman(29)	2021	130	633	124	629
Myocardial infarction	Spinal anesthesia	General anesthesia	REGAIN	Neuman(29)	2021	6	783	9	793
Pneumonia	Spinal anesthesia	General anesthesia	REGAIN	Neuman(29)	2021	8	783	16	793
Transfusion	Spinal anesthesia	General anesthesia	REGAIN	Neuman(29)	2021	130	782	146	793
Pulmonary embolism	Spinal anesthesia	General anesthesia	REGAIN	Neuman(29)	2021	4	783	5	793
Survival	Combined epidural-general anesthesia	General anesthesia		Du(26)	2021	355	498	326	533
SSI D30	Dexamethasone 8mg	Placebo	PADDI	Corcoran(32)	2021	354	4350	394	4328
POD D5	EEG guidance	No EEG guidance	ENGAGES	Wildes(34)	2019	157	604	140	609
Myocardial infarction	Ischemic preconditioning	Sham intervention	RIPHeart	Meybohm(39)	2015	47	645	63	630
New AF	Ischemic preconditioning	Sham intervention	RIPHeart	Meybohm(39)	2015	147	543	160	530
Acute renal failure	Ischemic preconditioning	Sham intervention	RIPHeart	Meybohm(39)	2015	42	650	35	658
Mortality	Ischemic preconditioning	Sham intervention	RIPHeart	Meybohm(39)	2015	9	683	4	689
Mortality	Volatile anesthesia	Total intravenous anesthesia	MYRIAD	Landoni(42)	2019	38	2709	34	2691
Myocardial infarction	Volatile anesthesia	Total intravenous anesthesia	MYRIAD	Landoni(42)	2019	134	2682	127	2680
Cough	slow intravenous fluid line fentanyl	Direct injection of fentanyl		Liu(36)	2017	52	573	316	565
Mortality	Ischemic preconditioning	Sham intervention	ERICCA	Hausenloy(38)	2015	47	801	32	811
Myocardial infarction	Ischemic preconditioning	Sham intervention	ERICCA	Hausenloy(38)	2015	173	801	191	811

RCT: Randomized Controlled Trial; MACE: Major Adverse Cardiovascular Events; SSI: Surgical Site Infection; POD: Postoperative Delirium; AF: Atrial Fibrillation.

Table V. — Characteristics and outcomes of included meta-analyses (MAs).

Outcome	Intervention	Control	MA Author	MA Year	Effect Intervention	Total Intervention	Effect Control	Total Control	RR	OR	CI
MACE	Active body warming	Routine thermal management	Balki(25)	2020	2	187	10	204		0,21	0,05-0,98
SSI	Active body warming	Routine thermal management	Balki(25)	2020	10	298	27	293		0,34	0,16-0,74
Transfusion	Active body warming	Routine thermal management	Balki(25)	2020	76	495	97	505		0,64	0,44-0,95
POD D5	EEG guidance	No EEG guidance	Ding(28)	2020	411	2214	528	2237	0,79		0,77-0,88
Emergence delirium	EEG guidance	No EEG guidance	Sun(46)	2020	193	1190	280	1209	0,7		0,60-0,83
Mortality D60	Spinal anesthesia	General anesthesia	Zheng(31)	2020	11	363	9	389		1,34	0,56-3,21
Delirium	Spinal anesthesia	General anesthesia	Zheng(31)	2020	26	400	33	409		1,05	0,27-4,00
Myocardial infarction	Spinal anesthesia	General anesthesia	Zheng(31)	2020	2	363	3	389		0,88	0,70-4,65
Pneumonia	Spinal anesthesia	General anesthesia	Guay(30)	2016	19	373	26	388	0,77		0,45-1,31
Transfusion	Spinal anesthesia	General anesthesia	Guay(30)	2016	52	99	61	103	0,9		0,49-1,66
Pulmonary embolism	Spinal anesthesia	General anesthesia	Guay(30)	2016	6	319	0	323		7,51	1,51-37,38
Survival	Combined epidural-general anesthesia	General anesthesia	Pöpping(27)	2014	80	3911	122	3855		0,96	0,51-0,92
SSI D30	Dexamethasone 8mg	Placebo	Polderman(33)	2016	172	2516	167	2415		1,01	0,87-2,95
POD D5	EEG guidance	No EEG guidance	MacKenzie(35)	2018	219	1361	315	1293		0,62	0,51-0,76
Myocardial infarction	Ischemic preconditioning	Sham intervention	Healy(40)	2014	25	883	44	894	0,69		0,34-1,40
New AF	Ischemic preconditioning	Sham intervention	Healy(40)	2014	152	814	167	825	0,92		0,76-1,12
Acute renal failure	Ischemic preconditioning	Sham intervention	Nur(41)	2014	75	318	102	319	0,74		0,53-1,02
Mortality	Ischemic preconditioning	Sham intervention	Nur(41)	2014	2	686	5	687	0,5		0,12-2,05
Mortality	Volatile anesthesia	Total intravenous anesthesia	Uhlig(44)	2016	36	1695	52	1510		0,55	0,35-0,85
Myocardial infarction	Volatile anesthesia	Total intravenous anesthesia	Symons(43)	2016	51	1569	28	840		0,98	0,61-1,58
Cough	Slow intravenous fluid line fentanyl	Direct injection of fentanyl	Kim(37)	2014	82	620	157	462		0,29	0,21-0,39
Mortality	Ischemic preconditioning	Sham intervention	Nur(41)	2014	2	686	5	687	0,5		0,12-2,05
Myocardial infarction	Ischemic preconditioning	Sham intervention	Healy(40)	2014	25	883	44	894	0,69		0,34-1,40

RR: Risk Ratio; OR: Odds Ratio; CI: Confidence Interval; MACE: Major Adverse Cardiovascular Events; SSI: Surgical Site Infection; POD: Postoperative Delirium; AF: Atrial Fibrillation.

the consistency of patterns observed across multiple interventions suggests robustness in the overall conclusions.

While TSA provides a structured way to assess cumulative evidence, it is not the only method available. Alternatives such as the GRADE framework, p-curve analysis, and Bayesian analysis also aim to address uncertainty and bias in meta-analyses[58–61]. However, each of these methods has its own limitations and relies on different assumptions. Combining TSA with qualitative tools like GRADE may offer a more balanced and comprehensive assessment, especially when evidence is used to guide critical clinical decisions.

Future research should continue validating TSA across specialties and explore how its integration into meta-analyses might enhance predictive validity. Studies should also investigate how often clinical guidelines are shaped by early meta-analyses that are later contradicted, and whether TSA could be used prospectively to flag such risks. However, meta-analyses should rarely be considered definitive. Rather than investing resources in small underpowered RCTs, clinical trial networks should prioritize large, collaborative, and independent studies capable of producing high-quality evidence, and reducing reliance on fragile syntheses prone to error[5].

Conclusion

This study demonstrates that meta-analyses in perioperative medicine often report significant treatment effects that are not confirmed by subsequent large RCTs. These discrepancies largely stem from the inclusion of small, early-phase studies prone to type I and type II errors. By comparing recent meta-analyses with subsequent large RCTs and applying TSA, this study offers a more nuanced understanding of their predictive value. Although most endpoints appeared concordant, TSA revealed several to be inconclusive or potentially false positive under stricter clinical assumptions. While TSA helps mitigate random error and improve interpretability, it has limitations and should be applied cautiously. Overall, meta-analyses should be viewed as hypothesis-generating rather than conclusive, and further investment in large, well-powered RCTs remains essential for reliable clinical guidance.

Acknowledgments: This research received no external funding, and no contributions were made by individuals other than the listed authors. The authors declare no conflicts of interest.

References

1. Myles PS. Why we need large randomized studies in anaesthesia. Br J Anaesth [Internet]. 1999 Dec;83(6):833–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10700777
2. Sivakumar H, Peyton PJ. Poor agreement in significant findings between meta-analyses and subsequent large randomized trials in perioperative medicine. Vol. 117, British Journal of Anaesthesia. Oxford University Press; 2016. p. 431–41.
3. Bartels K, Sessler DI. Meta-analyses of clinical trials: are we getting lemonade from lemons? Br J Anaesth [Internet]. 2022 Feb 3;128(2):233–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/34742542
4. Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. Vol. 94, Milbank Quarterly. Blackwell Publishing Inc.; 2016. p. 485–514.
5. Doleman B, Williams JP, Lund J. Why most published meta-analysis findings are false. Tech Coloproctol. 2019 Sep 1;23(9):925–8.
6. LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between Meta-Analyses and Subsequent Large Randomized, Controlled Trials. New England Journal of Medicine [Internet]. 1997 Aug 21;337(8):536–42. Available from: http://www.nejm.org/doi/abs/10.1056/NEJM199708213370806
7. Kastrati L, Raeisi-Dehkordi H, Llanaj E, Quezada-Pinedo HG, Khatami F, Ahanchi NS, et al. Agreement between Mega-Trials and Smaller Trials: A Systematic Review and Meta-Research Analysis. JAMA Netw Open. 2024 Sep 6;7(9):e2432296.
8. Riberholt CG, Olsen MH, Milan JB, Hafliðadóttir SH, Svanholm JH, Pedersen EB, et al. Major mistakes or errors in the use of trial sequential analysis in systematic reviews or meta-analyses – the METSA systematic review. BMC Med Res Methodol. 2024 Dec 1;24(1).
9. Rerkasem K, Rothwell PM. Meta-analysis of small randomized controlled trials in surgery may be unreliable. British Journal of Surgery. 2010 Apr;97(4):466–9.
10. Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. J Clin Epidemiol. 2008 Aug;61(8):763–9.
11. Shah A, Smith AF. Trial sequential analysis: adding a new dimension to meta-analysis. Vol. 75, Anaesthesia. Blackwell Publishing Ltd; 2020. p. 15–20.
12. Borm GF, Donders ART. Updating meta-analyses leads to larger type I errors than publication bias. J Clin Epidemiol. 2009;62(8).
13. Imberger G, Thorlund K, Gluud C, Wetterslev J. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. Available from: http://dx.doi.org/
14. Møller MH, Ioannidis JPA, Darmon M. Are systematic reviews and meta-analyses still useful research? We are not sure. Vol. 44, Intensive Care Medicine. Springer Verlag; 2018. p. 518–20.
15. Wasserstein RL, Schirm AL, Lazar NA. Moving to a World Beyond "p < 0.05." Vol. 73, American Statistician. American Statistical Association; 2019. p. 1–19.
16. 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(1):287–98.
17. 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 Jan;61(1):64–75.

18. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Med Res Methodol. 2009;9.
19. Thorlund K, Engstrøm J, Wetterslev J, Brok J, Imberger G, Gluud C. User Manual for Trial Sequential Analysis (TSA).
20. Payne T, Moran B, Loadsman J, Marschner I, McCulloch T, Sanders RD. Importance of sequential methods in meta-analysis: implications for postoperative mortality, delirium, and stroke management. Vol. 130, British Journal of Anaesthesia. Elsevier Ltd; 2023. p. 395–401.
21. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Med Res Methodol. 2017 Mar 6;17(1).
22. Stallard N, Todd S, Ryan EG, Gates S. Comparison of Bayesian and frequentist group-sequential clinical trial designs. BMC Med Res Methodol. 2020 Jan 7;20(1).
23. Higgins JPT, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. Stat Med. 2011 Apr 30;30(9):903–21.
24. Sessler DI, Pei L, Li K, Cui S, Chan MTV, Huang Y, et al. Aggressive intraoperative warming versus routine thermal management during non-cardiac surgery (PROTECT): a multicentre, parallel group, superiority trial. The Lancet. 2022 May 7;399(10337):1799–808.
25. Balki I, Khan JS, Staibano P, Duceppe E, Bessissow A, Sloan EN, et al. Effect of Perioperative Active Body Surface Warming Systems on Analgesic and Clinical Outcomes: A Systematic Review and Meta-analysis of Randomized Controlled Trials. Anesth Analg. 2020 Nov 1;131(5):1430–43.
26. Du YT, Li YW, Zhao BJ, Guo XY, Feng Y, Zuo MZ, et al. Long-term Survival after Combined Epidural-General Anesthesia or General Anesthesia Alone: Follow-up of a Randomized Trial. Anesthesiology. 2021 Aug 1;135(2):233–45.
27. Pöpping DM, Elia N, Van Aken HK, Marret E, Schug SA, Kranke P, et al. Impact of epidural analgesia on mortality and morbidity after surgery: Systematic review and meta-analysis of randomized controlled trials. Ann Surg. 2014;259(6):1056–67.
28. Ding L, Chen DX, Li Q. Effects of electroencephalography and regional cerebral oxygen saturation monitoring on perioperative neurocognitive disorders: A systematic review and meta-analysis. BMC Anesthesiol. 2020 Sep 30;20(1).
29. Neuman MD, Feng R, Carson JL, Gaskins LJ, Dillane D, Sessler DI, et al. Spinal Anesthesia or General Anesthesia for Hip Surgery in Older Adults. New England Journal of Medicine. 2021 Nov 25;385(22):2025–35.
30. Guay J, Parker MJ, Gajendragadkar PR, Kopp S. Anaesthesia for hip fracture surgery in adults. Vol. 2016, Cochrane Database of Systematic Reviews. John Wiley and Sons Ltd; 2016.
31. Zheng X, Tan Y, Gao Y, Liu Z. Comparative efficacy of Neuraxial and general anesthesia for hip fracture surgery: A meta-analysis of randomized clinical trials. BMC Anesthesiol. 2020 Jun 30;20(1):AR.
32. Corcoran TB, Myles PS, Forbes AB, Cheng AC, Bach LA, O'Loughlin E, et al. Dexamethasone and Surgical-Site Infection. New England Journal of Medicine. 2021 May 6;384(18):1731–41.
33. Polderman JAW, Farhang-Razi V, Van Dieren S, Kranke P, Devries JH, Hollmann MW, et al. Adverse side effects of dexamethasone in surgical patients. Vol. 2018, Cochrane Database of Systematic Reviews. John Wiley and Sons Ltd; 2018.
34. Wildes TS, Mickle AM, Abdallah A Ben, Maybrier HR, Oberhaus J, Budelier TP, et al. Effect of electroencephalography-guided anesthetic administration on postoperative delirium among older adults undergoing major surgery the engages randomized clinical trial. JAMA - Journal of the American Medical Association. 2019 Feb 5;321(5):473–83.
35. MacKenzie KK, Britt-Spells AM, Sands LP, Leung JM. Processed electroencephalogram monitoring and postoperative delirium a systematic review and meta-analysis. Vol. 129, Anesthesiology. Lippincott Williams and Wilkins; 2018. p. 417–27.
36. Liu M qiang, Li F xian, Han Y kun, He J yong, Shi H wen, Liu L, et al. Administration of fentanyl via a slow intravenous fluid line compared with rapid bolus alleviates fentanyl-induced cough during general anesthesia induction. J Zhejiang Univ Sci B. 2017 Nov 1;18(11):955–62.
37. Kim JE, Min SK, Chae YJ, Lee YJ, Moon BK, Kim JY. Pharmacological and nonpharmacological prevention of fentanyl-induced cough: A meta-analysis. Vol. 28, Journal of Anesthesia. Springer-Verlag Tokyo; 2014. p. 257–66.
38. Hausenloy DJ, Candilio L, Evans R, Ariti C, Jenkins DP, Kolvekar S, et al. Remote Ischemic Preconditioning and Outcomes of Cardiac Surgery. New England Journal of Medicine. 2015 Oct 8;373(15):1408–17.
39. Meybohm P, Bein B, Brosteanu O, Cremer J, Gruenewald M, Stoppe C, et al. A Multicenter Trial of Remote Ischemic Preconditioning for Heart Surgery. New England Journal of Medicine. 2015 Oct 8;373(15):1397–407.
40. Healy DA, Khan WA, Wong CS, Moloney MC, Grace PA, Coffey JC, et al. Remote preconditioning and major clinical complications following adult cardiovascular surgery: Systematic review and meta-analysis. Int J Cardiol. 2014;176(1):20–31.
41. Haji Mohd Yasin NAB, Herbison P, Saxena P, Praporski S, Konstantinov IE. The role of remote ischemic preconditioning in organ protection after cardiac surgery: A meta-analysis. Vol. 186, Journal of Surgical Research. 2014. p. 207–16.
42. Landoni G, Lomivorotov V V., Nigro Neto C, Monaco F, Pasyuga V V., Bradic N, et al. Volatile Anesthetics versus Total Intravenous Anesthesia for Cardiac Surgery. New England Journal of Medicine. 2019 Mar 28;380(13):1214–25.
43. Symons JA, Myles PS. Myocardial protection with volatile anaesthetic agents during coronary artery bypass surgery: A meta-analysis. Vol. 97, British Journal of Anaesthesia. Oxford University Press; 2006. p. 127–36.
44. Uhlig C, Bluth T, Schwarz K, Deckert S, Heinrich L, De Hert S, et al. Effects of volatile anesthetics on mortality and postoperative pulmonary and other complications in patients undergoing surgery: A systematic review and meta-analysis. Anesthesiology. 2016 Jun 1;124(6):1230–45.
45. Wang E, Wang L, Ye C, Luo N, Zhang Y, Zhong Y, et al. Effect of Electroencephalography Spectral Edge Frequency (SEF) and Patient State Index (PSI)-Guided Propofol-Remifentanil Anesthesia on Delirium After Laparoscopic Surgery: The eMODIPOD Randomized Controlled Trial. J Neurosurg Anesthesiol. 2022 Apr 1;34(2):183–92.
46. Sun Y, Ye F, Wang J, Ai P, Wei C, Wu A, et al. Electroencephalography-Guided Anesthetic Delivery for Preventing Postoperative Delirium in Adults: An Updated Meta-analysis. Anesth Analg. 2020 Sep 27;131(3):712–9.
47. Sylvester RJ, Canfield SE, Lam TBL, Marconi L, MacLennan S, Yuan Y, et al. Conflict of Evidence: Resolving Discrepancies When Findings from Randomized Controlled Trials and Meta-analyses Disagree. Eur Urol. 2017 May 1;71(5):811–9.
48. Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. 1995.
49. Afshari A, Wetterslev J. When may systematic reviews and meta-analyses be considered reliable? Vol. 32, European Journal of Anaesthesiology. Lippincott Williams and Wilkins; 2015. p. 85–7.
50. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence

intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016 Apr 1;31(4):337–50.
51. Sharma H. Statistical significance or clinical significance? A researcher's dilemma for appropriate interpretation of research results. Vol. 15, Saudi Journal of Anaesthesia. Wolters Kluwer Medknow Publications; 2021. p. 431–4.
52. Smith RJ. P >.05: The incorrect interpretation of "not significant" results is a significant problem. Vol. 172, American Journal of Physical Anthropology. Wiley-Liss Inc.; 2020. p. 521–7.
53. Savovic´, J, Jones HE, Altman DG, Harris RJ, Jü P, Pildal J, et al. Influence of Reported Study Design Characteristics on Intervention Effect Estimates From Randomized, Controlled Trials [Internet]. Vol. 157, Annals of Internal Medicine. 2012. Available from: www.annals.org
54. Hedin RJ, Umberham BA, Detweiler BN, Kollmorgen L, Vassar M. Publication Bias and Nonreporting Found in Majority of Systematic Reviews and Meta-analyses in Anesthesiology Journals. Anesth Analg. 2016 Oct 1;123(4):1018–25.
55. Khan A, Marchbanks J, Hedin R, Vassar M. Publication Bias in Leading Anesthesiology Journals: A Systematic Review. Anesth Analg [Internet]. 2025 Apr 10; Available from: https://journals.lww.com/10.1213/ANE.0000000000007527
56. Thorlund K, Devereaux PJ, Wetterslev J, Guyatt G, Ioannidis JPA, Thabane L, et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? Int J Epidemiol. 2009;38(1):276–86.
57. Thomas J, Askie LM, Berlin JA, Elliott JH, Ghersi D, Simmonds M, Takwoingi Y, Tierney JF, Higgins HPT. Chapter 22: Prospective approaches to accumulating evidence [last updated October 2019]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.5. Cochrane, 2024. Available from cochrane.org/handbook. In.
58. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008 Apr 26;336(7650):924–6.
59. Hackenberger BK. Bayesian meta-analysis now – let's do it. Croat Med J. 2020 Dec 1;61(6):564–8.
60. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. Stat Methods Med Res. 2001 Aug 1;10(4):277–303.
61. Simonsohn U, Nelson LD, Simmons JP. P-curve: A key to the file-drawer. J Exp Psychol Gen. 2014 Apr;143(2):534–47.

doi.org/10.56126/76.4.35