Trial sequential analysis

Shengping Yang PhD, Gilbert Berdine MD

Meta-analysis is commonly regarded as a dependable method for synthesizing research, as it employs statistical techniques to combine and examine data from multiple studies focused on a specific subject. However, in order to enhance the reliability of their findings, and to avoid the potential for errors and false-positive results, researchers are increasingly turning to trial sequential analysis (TSA). Would you introduce TSA?

TSA refers to trial sequential analysis, a statistical method that assesses the reliability of results in meta-analyses and other systematic reviews.

As noted in the question above, meta-analyses are considered the best way to combine evidence from multiple studies because they use all available data in the literature to increase the statistical power for detecting an intervention's likely effect. However, this does not guarantee that the available evidence is either sufficient or strong, and thus the reliability of meta-analytic results can be questionable. For example, if a meta-analysis includes a small number of studies and participants, its findings may be either spurious (a type I error; α) or null (lack of statistical power; a type II error; β). Moreover, when multiple meta-analyses are conducted on the same research subject using largely the same existing evidence in the literature, inflated type I errors can occur.[1-4]

To address these issues, sequential methods have been proposed. Sequential methods refer to a statistical approach used in randomized clinical trials to monitor accumulating evidence and determine when a trial should be stopped early or continued. This concept can be adopted in meta-analyses to evaluate the reliability of the findings by monitoring the cumulative data and determining when enough evidence has been gathered to reach a conclusion. Trial sequential

analysis is one such sequential method that incorporates a sequential monitoring boundary to control the risk of type I and type II errors associated with repeated testing, making it an effective tool for assessing the robustness and validity of meta-analytic findings.

1. SEQUENTIAL ANALYSIS

The concept of sequential analysis was first introduced by Abraham Wald as a tool for efficient industrial quality control.[5] Later, Peter Armitage introduced the use of sequential analysis in medical research, particularly in clinical trials.[6] Since then, sequential methods have become increasingly popular in medicine, with Stuart Pocock's work providing clear recommendations on how to control type 1 error rates in sequential designs.[7,8]

Specifically, sequential analysis allows for statistical estimation or decision making in real time as data are being collected, rather than retrospectively on a fixed sample size. The final number of subjects analyzed is not predetermined, but rather determined by a predetermined stopping rule, such as achieving a particular level of statistical power. This method often requires a smaller sample size than traditional statistical methods because it allows for monitoring of the accumulating data as the trial progresses, enabling early stopping of the trial if a significant effect is detected or if continuing the trial would be unlikely to produce a significant result. As a result, sequential analysis can potentially reduce costs, effort, and resource requirements, while better satisfying ethical considerations.

In general, sequential analysis has the following characteristics, as described by Whitehead: (a) a series of interim analyses (i.e., planned analyses of data that are conducted at predefined times during a study) are performed on the accumulating data at different times during the trial, to test hypotheses and make decisions, rather than waiting until all the data has been collected; (b) each analysis includes an assessment of the effect of the same intervention

of interest; and (c) each analysis has the potential to lead to stopping the trial.

There are many considerations in sequential analysis; this review will focus on the aspects that are more relevant to trial sequential analysis:

1.1 STOPPING RULE

Sequential analysis requires prespecified stopping rules, which define if the study will be terminated at an interim analysis, based on the accumulating data. These rules are typically based on statistical criteria, such as reaching a certain level of statistical significance or futility. In other words, a trial may be halted due to either the detection of intervention efficacy or the unlikelihood of the intervention having an effect.

1.2 ALPHA SPENDING FUNCTION

An alpha spending function is a mathematical function that specifies the distribution of type I error (or α; false positive) across the interim analyses, and it is closely related to the development of stopping rules in a sequential analysis. During each interim analysis, it is possible to declare the efficacy of an intervention and have a type I error. Thus, criteria for declaring statistical superiority must be calibrated to control for the risk of type I error in the overall trial. The thresholds can be specified as a sequence of P values or by using another test statistic for a series of specifically timed analyses. Alternatively, an alpha spending function may be used to distribute the overall risk of a false-positive conclusion across the interim and final analyses of the trial.

The choice of alpha spending function depends on the specific goals and design of the trial. Examples of alpha spending functions include those developed by Lan and DeMets and by Kim and DeMets.[9,10] Other designs, such as the Pocock design, require prespecifying the number and timing of the interim analyses and use the same or more stringent P value stopping criteria.

1.3 SAMPLE SIZE/POWER CALCULATION

Unlike traditional methods, in sequential analysis, the final sample size is not fixed, but rather determined by a prespecified stopping rule. Therefore, sample size determination is an important consideration to ensure that the study has sufficient power to detect a meaningful effect size. The mathematical details involved in these considerations are quite complex and are not the primary focus of this article. Instead, the next sections will now explore how the concept of sequential analysis can be applied in meta-analyses.

2. TRIAL SEQUENTIAL ANALYSIS

Since its introduction in 2005, TSA has become an increasingly popular method to improve the quality of meta-analyses by controlling the risk of type I and II errors.[3,4] Studies have shown that TSA can identify insufficient information size and potentially false discoveries in many meta-analyses.[11] Compared to conventional meta-analyses, which often suffer from 5% type I errors due to significant results by chance and type II errors due to failing to detect an effect, TSA is a cumulative meta-analysis method that considers both α and β errors to estimate when the effect is sufficiently large and unlikely to be affected by further studies.[12] It should be noted that when the number of participants and trials in a meta-analysis is small, it may lead to higher type I (due to publication bias, etc.) and II errors (less statistical power).

Trial sequential analysis is a cumulative meta-analysis method that allows for updates of the analysis as new trial results become available while considering the total required sample size and accrued information. Each cumulative meta-analysis uses a prespecified threshold, similar to the alpha spending function for sequential analysis, to determine statistical significance. Thus, TSA minimizes the risk of type I errors due to multiple testing or sparse data, resulting in more reliable findings. In addition, TSA can estimate the required information size (RIS), which is the sample size required to reach a reliable conclusion about the treatment effect.

Figure 1 presents a visual representation of the TSA. The green (red) solid line represents the O'Brien-Fleming (OBF) boundary for benefit (harm). These boundaries are wider at earlier analyses compared to conventional boundaries to account for the cumulative nature of statistical testing and reduce the risk
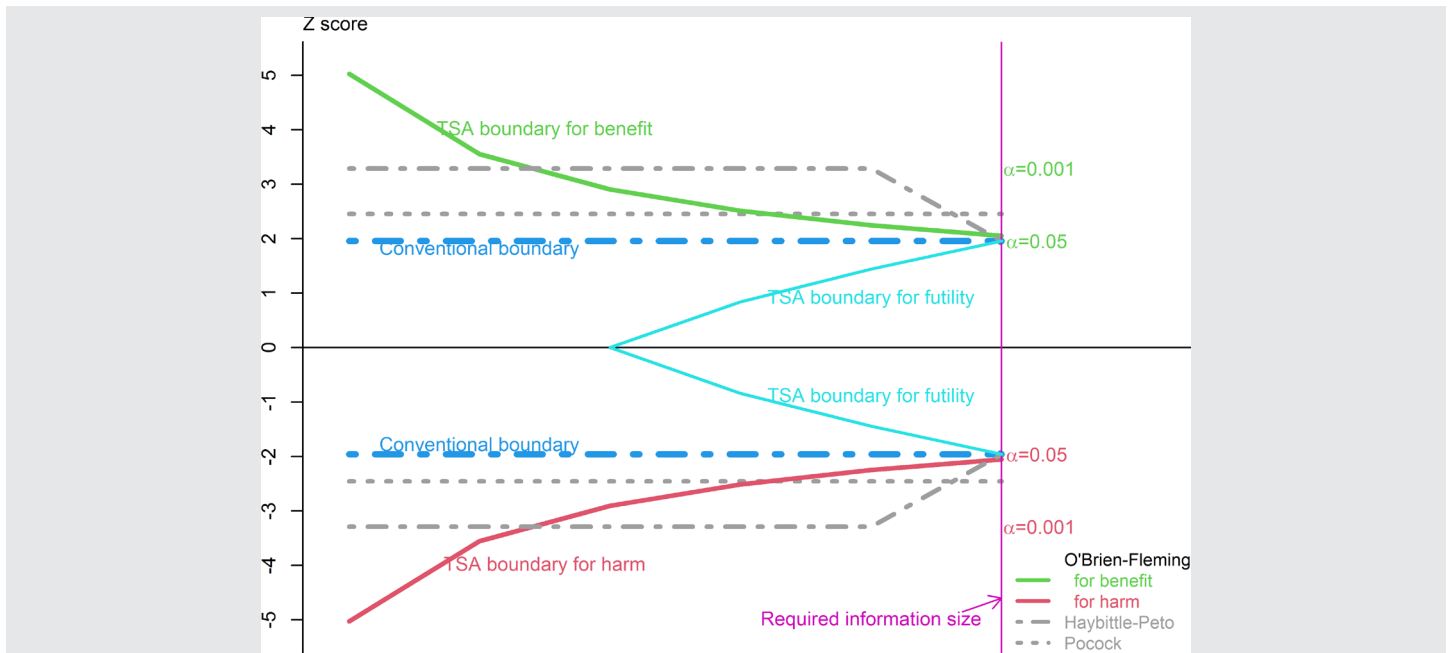
Figure 1. Provides a visual representation of the TSA, which includes various boundaries for efficacy and futility. The O'Brien-Fleming boundary for benefit (green solid line) and harm (red solid line) are shown, along with the conventional boundaries for efficacy (two horizontal blue lines) and the TSA binaries for futility (cyan lines). The required information size is represented by the vertical purple line. Additionally, as different alpha spending functions can be used to define the boundaries, we have included the Haybittle-Peto (dotdash) and Pocock (dotted) boundaries using gray lines for reference.

of stopping the meta-analysis prematurely based on chance findings. The figure also includes the Haybittle-Peto (dotdash) and Pocock (dotted) boundaries using gray lines for reference. The cyan lines are the boundaries for futility.

If the cumulative z-score reaches either the boundary for benefit or harm at any meta-analysis, a conclusion can be made based on the result from the analysis. Alternatively, if the z-score touches any of the futility boundaries, it is recommended to stop the meta-analysis. This is because any future meta-analysis is unlikely to show a significant difference, even though more clinical trials may be conducted until the required information size (vertical purple line) is achieved.

2.1 Required information size

Similar to sample size calculation or power analysis in a single randomized trial, the RIS represents the minimum number of participants and studies needed in a meta-analysis to achieve a prespecified level of statistical power and a prespecified level of alpha, such as 0.05 for a two-sided test. If the total sample size of the included studies and participants in the meta-analysis is smaller than the RIS, the meta-analysis is underpowered and may produce overestimated or underestimated intervention effects due to a lack of precision and power.[13]

Trial sequential analysis software packages, such as TSA (TSA – ctu.dk), can calculate the RIS automatically based on the input parameters specified by the user. The package also provides a visual display of the TSA with the estimated RIS and the cumulative z-curve to assess the cumulative evidence and the risk of random errors.

2.2 Control type I errors

Trial sequential analysis controls the risk of type I error by applying a type of multiple testing correction, known as the OBF method or the Lan-DeMets method.

Similar to adjusting the alpha spending function in interim analyses, TSA determines the boundary for statistical significance at each cumulative meta-analysis. However, unlike in clinical trials where the timing and sample size of interim analyses are pre-specified, meta-analyses are usually updated when data from new clinical trials become available, with arbitrary intervals between trials and unpredictable sample size for each new trial. Therefore, the alpha spending function proposed by O'Brien and Fleming and later developed by Lan and DeMets is more appropriate for TSA.[9,10] By controlling the type I error rate, TSA helps reduce the risk of false-positive results and improve the reliability of conclusions drawn from meta-analyses and clinical trials.

2.3 GUIDE THE DESIGN OF NEW CLINICAL TRIALS

Trial sequential analysis can assist in the design of a new clinical trial by providing information on the optimal sample size and stopping rules for the trial based on the estimated treatment effect from previous studies. It can also determine the stopping rules for the new trial by specifying interim analyses and the alpha spending function. Appropriately specified stopping rules can save time and resources by allowing the trial to be stopped early if the treatment effect is evident. By incorporating TSA in the trial design, researchers can increase the chances of obtaining reliable results and reduce the risk of false-positive or false-negative findings.[1]

In addition, the TSA can be used to evaluate the robustness of the trial design against potential sources of heterogeneity. Sensitivity analyses can be performed to assess how changes in the assumptions or parameters affect the conclusions of a trial.

In summary, TSA is an extension of conventional meta-analysis and allows the calculation of the RIS for a given effect size, adjusts for multiple comparisons, and controls for type I and type II errors. Trial sequential analysis uses alpha spending functions to allocate the overall risk of a false-positive conclusion across multiple meta-analyses. Trial sequential analysis can be a valuable tool for researchers in evaluating the reliability and validity of findings in meta-analyses and can enhance the overall quality of research in various fields. In addition, TSA can guide the design of new clinical trials by providing information on optimal sample size, stopping rules, etc.

REFERENCES

1. Claire R, Gluud C, Berlin I, et al. Using trial sequential analysis for estimating the sample sizes of further trials: example using smoking cessation intervention. BMC Med Res Methodol 2020 Nov 30;20(1):284. doi: 10.1186/s12874-020-01169-7.
2. Simmonds M, Salanti G, McKenzie J, et al. Living systematic reviews: 3. Statistical methods for updating meta-analyses. J Clin Epidemiol. 2017;91:38–46.
3. Thorlund K, Engstrøm J, Wetterslev J, et al. User manual for trial sequential analysis (TSA). Copenhagen Trial Unit, Centre for Clinical Intervention research, Copenhagen, Denmark. 2011:1–115. Available from www.ctu.dk/tsa
4. Thorlund K, Wetterslev J, Brok J, et al. Trial sequential analyses of six meta-analyses considering heterogeneity and trial weight. In: Corroboree. Abstracts of the 13th Cochrane Colloquium; 2005:22–26 Oct; Melbourne, Australia.
5. Abraham W. Sequential tests of statistical hypotheses. Annals Mathematical Statistics 1945;16(2):117–186. doi:10.1214/aoms/1177731118. JSTOR 2235829.
6. Armitage P. Sequential medical trials. Oxford: Blackwell. 1960 ASIN: B0007IX5Q
7. Jennison C, Turnbull BW. Group sequential methods with applications to clinical trials. 2000. Taylor & Francis. ISBN 9780849303166.

8. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika 1977;64(2):191–199.

9. Kim K, Demets DL. Design and analysis of group sequential tests based on the type I error spending rate function Get access Arrow. Biometrika 1987;74(1):149–154.

10. Lan, KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70:659–663.

11. Brok J, Thorlund K, Gluud C, et al. 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. doi: 10.1016/j.jclinepi.2007.10.007.

12. De Cassai A, Pasin L, Boscolo A, et al. Trial sequential analysis: plain and simple. Korean J Anesthesiol 2021 Aug;74(4): 363–365. doi: 10.4097/kja.20637.

13. Kang H. Trial sequential analysis: novel approach for meta-analysis. Anesth Pain Med (Seoul). 2021 Apr;16(2):138–150. doi: 10.17085/apm.21038.