Letter to the Editor

Corresponding author:
Alessandro De Cassai, M.D.
UOC Anesthesia and Intensive Care Unit,
University Hospital of Padua, Gallucci St,
35121 Padua, Italy
Tel: +39-0498213090
Fax: +39-0498213091
Email: alessandro.decassai@aopd.veneto.it
ORCID: https://orcid.org/0000-0002-9773-1832

Trial sequential analysis: plain and simple

Alessandro De Cassai[1], Laura Pasin[1], Annalisa Boscolo[1], Michele Salvagno[2], Paolo Navalesi[1,2]

UOC Anesthesia and Intensive Care Unit, [1]University Hospital of Padua, [2]Department of Medicine-DIMED, University of Padua, Padua, Italy

The use of trial sequential analysis (TSA) in the medical literature is increasing in recent times. However, not all readers may be familiar with this statistical technique.

This correspondence aims to provide readers with the essentials to understand and interpret TSA.

Adequately conducted meta-analyses (MAs) are considered the best evidence in the scientific literature. Nonetheless, MAs are exposed to misleading significant results (type I errors; α) or erroneously insignificant results (type II errors; β) caused by low quality or inadequately powered trials, publication bias, and repeated significance testing [1].

TSA is a cumulative MA method developed [1] to weigh α and β errors while estimating when the effect is large enough to be unlikely to be affected by further studies.

TSA is displayed as a Cartesian graph with cumulative z-score on the y-axis and number of patients on the x-axis, subdivided into four zones by four lines: monitoring boundaries for benefit, and harm, and two futility boundaries (Fig. 1). Two lines parallel to the x-axis are usually displayed, showing the conventional statistically significant line at z, corresponding to 1.96.

The cumulative z statistic line is constructed adding a study sequentially with chronological criteria. The end of the line corresponds to the lastly added study. It will lie in one of the following zones: "benefit", "harm", "inner wedge" or "not statistically significant", representing a statistically significant result for the first two areas ("benefit" and "harm") or a strong evidence that further studies will hardly be able to change the no-effect results ("inner wedge" area). Presence in the "not statistically significant" area means that further studies are needed.

Control of α and β errors may be managed by decreasing the test statistic using a penalizing factor λ (law of the iterated logarithm) or adjusting the significance threshold. The last described strategy is managed in TSAs using α-and β-spending functions.

The α spending function determines both the benefit and harm boundaries, while the beta spending function is displayed on the graph as the futility boundaries.

The spending functions used in TSA are based on the O'BrienFleming's function. Although several examples of such functions have been described, O'BrienFleming's function is the only function implemented in the TSA software.

The spending function is a monotonically increasing function that distributes the α error along the entire analysis for a pre-decided α. The function is defined from 0 to 1, where 0 corresponds to "no patient enrolled" and 1 to the "reached information size" with the information fraction (IF) as the independent variable. The IF is given by the accumulated sample divided by the required sample.

The used α-spending function is

$$\alpha(\text{IF}) = 2 - 2\Phi(Z_{\alpha\text{-}2}/\sqrt{\text{IF}})$$
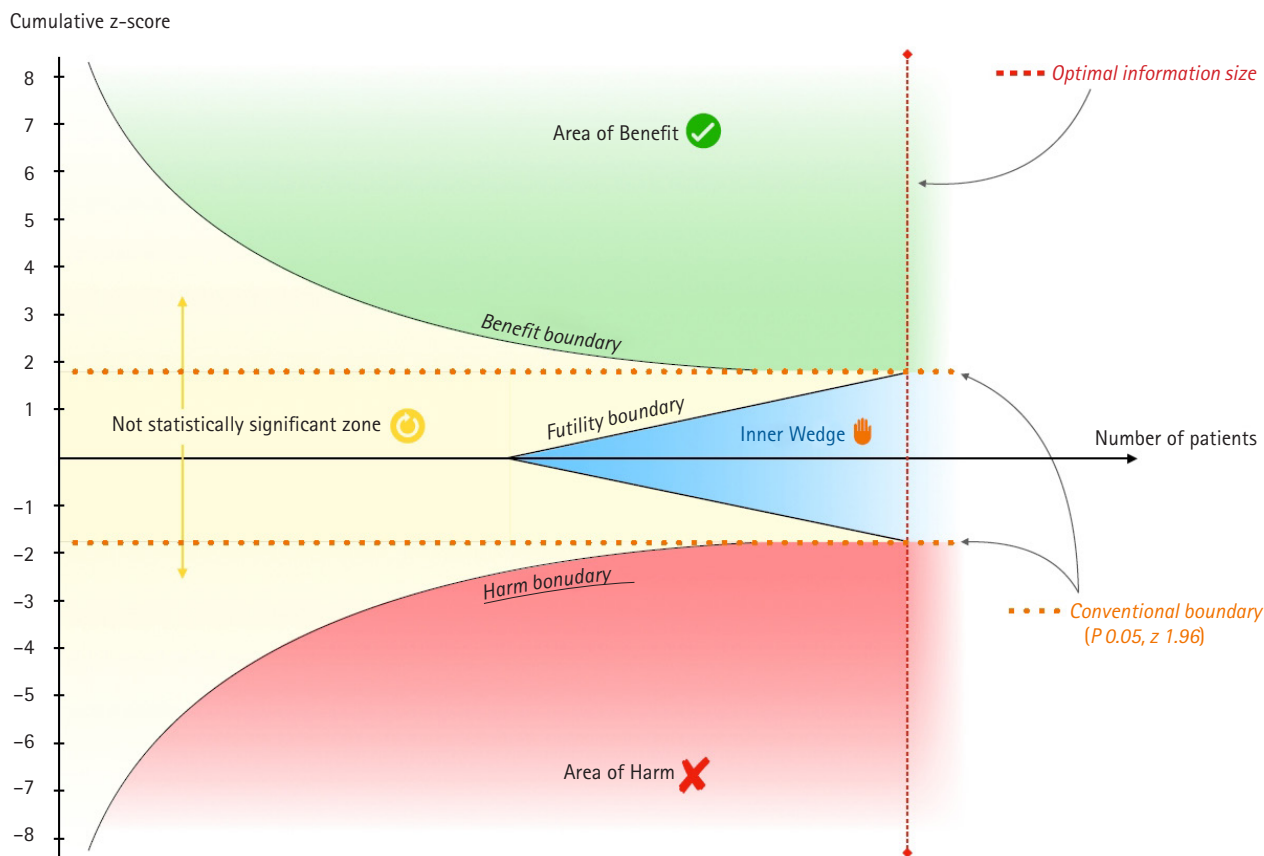
Fig. 1. Trial sequential analysis graph. The graph is subdivided into four zones: area of benefit, area of harm, inner wedge, not statistically significant zone.

where Φ is the standard normal cumulative distribution function [2]. This function represents a generalization of the formula proposed by Lan-De Mets, allowing non-constant and flexible IF increments among trials.

Similarly, the β-spending function is monotonically increasing and defined from 0 to 1, where 1 corresponds to the threshold for the maximum β error chosen for the non-superiority and non-inferiority tests.

Standard MA does not consider if the significance obtained is provided by an adequate cumulative information size (total number of patients among the trials). However, this is a question of paramount importance that is inadequately considered.

Choosing an adequate information size is the corner-stone in TSAs. Nonetheless, there is no standardized way or consensus to establish an adequate information size.

Similar to randomized controlled trials, information size calculation is based on the choice of a priori relative risk reduction (RRR) and of a maximum type I and II error.

RRR is the reduction of the event rate in the treatment group (Pt) compared to the control group (Pc), described as a percentage

$(Pt/Pc)/Pt \times 100\%$.

The choice of RRR is critical and should be based on a realistic and clinically meaningful effect of the intervention. This should be based on previous literature, but when there is insufficient clinical experience (e.g., pilot studies), data from related areas may be used.

It seems reasonable to state that the information size of an MA should be at least as large as the sample size of an adequately powered trial investigating that specific outcome. However, researchers may be more conservative in choosing a higher power (i.e., 90–99%) and a lower α (i.e., 1–5%), (given that MA is at the top of the science hierarchy).

It is preferable to estimate the RRR from the analysis of the low-bias risk trials, by excluding the high-risk of bias studies that could overestimate the intervention effect. [1].

Another more conservative post-hoc approach is to consider the least likely intervention effect (lower confidence limit of the intervention effect) as RRR [3].

MA should compare the effect of identical studies without any

difference in the protocol, population, or outcome assessment. However, this is utopist, and a certain degree of clinical heterogeneity leading to statistical heterogeneity has to be taken into account and accepted. A correction factor for IF derived by heterogeneity magnitude is deemed necessary.

While MA usually uses inconsistency (I^2) as the measure between trial variance, TSA uses diversity (D^2). D^2 is defined as the proportion of the total variance in a random effect model contributed by the between-trial variation despite its estimator [4]. D^2 is always higher than I^2 unless all the weights in the fixed-effect model are equal; particularly, D^2 is 0 only when I^2 is 0 [4].

While the use of D^2 has the advantage of correcting the IF to maintain the anticipated risk of both α and β errors, it does not consider any adjustment in IF for any bias.

Recently, a Cochrane expert panel recommended against the use of TSA and analogous sequential methods in MA [5]. Cochrane highlighted that an interpretation based on estimated intervention effect and its accompanying uncertainty is preferable and recommended instead of the binary interpretation proposed by TSA.

The use of sequential analysis in MA, a retrospective analysis without any control on study design by meta-analysts, makes it impossible to establish the stopping rules that are typical of a pre-planned set of interim analyses.

TSA is usually performed on the primary outcome; however, cumulative evidence from secondary outcomes would be penalized from a premature stopping rule. A striking example is depicted using network meta-analysis, where cumulative evidence will continue to affect some networks when the main effects are already well estimated.

Despite its limitations, and in particular its dichotomous interpretation, TSA is a useful tool in 'researchers' armamentarium.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Author Contributions

Alessandro De Cassai (Conceptualization; Writing – original draft; Writing – review & editing)
Laura Pasin (Writing – original draft; Writing – review & editing)
Annalisa Boscolo (Writing – original draft; Writing – review & editing)
Michele Salvagno (Writing – original draft; Writing – review & editing)
Paolo Navalesi (Supervision; Writing – review & editing)

ORCID

Alessandro De Cassai, https://orcid.org/0000-0002-9773-1832
Laura Pasin, https://orcid.org/0000-0001-5757-4791
Annalisa Boscolo, https://orcid.org/0000-0003-3409-4159
Michele Salvagno, https://orcid.org/0000-0002-8415-4247
Paolo Navalesi, https://orcid.org/0000-0002-3733-3453

References

1. 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.
2. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. Stat Med 1994; 13: 1341-52.
3. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Med Res Methodol 2017; 17: 39.
4. 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: 86.
5. Cochrane Scientific Committee. Repeated Meta-analyses [Internet]. London: Cochrane methods; 2020 Dec [cited 2020 Nov 29]. Available from https://methods.cochrane.org/methods-cochrane/repeated-meta-analyses