This article has been accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination, and proofreading processes, which may lead to differences between this version and the version of record.

Please cite this article as https://doi.org/10.4097/kja.20637
Trial Sequential Analysis: Plain and Simple

Alessandro De Cassai¹; Laura Pasin¹; Annalisa Boscolo ¹; Michele Salvagno ²; Paolo Navalesi ¹,²

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

3. Running title: TSA: Plain and Simple

4. Corresponding author: Alessandro De Cassai, MD. UOC Anesthesia and Intensive Care Unit, University Hospital of Padova, Padova, Italy, via Giustiniani 2, 35128 Padova, Italy Tel +390498213090. Fax +390498213091 mail: alessandro.decassai@aopd.veneto.it

5. Previous presentation in conferences: Not applicable

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

7. Funding: None

8. Acknowledgments: None

9. IRB number: Not applicable

10. Clinical trial registration number: Not applicable

Keywords: Trial sequential analysis; TSA; Meta-analysis; Review; Statistical round

Authors' contributions: ADC designed, wrote and edited the paper; LP wrote and edited the paper; AB wrote and edited the paper, MS wrote and edited the paper; PN wrote and edited the paper.
Dear Editor,

The use of Trial Sequential Analysis (TSA) in medical literature is increasing in the last few years, however, not all readers may be familiar with this statistical technique. The aim of this correspondence is to provide readers the essentials to understand and interpret TSA.

Adequately conducted meta-analyses (MA) are considered the best evidence in scientific literature. Nonetheless, MA are exposed to misleading significant results (type I errors; α) or erroneously not significant results (type II errors; β) caused by low quality or inadequately powered trials, publication bias and repeated testing of significance [1].

TSA is a cumulative MA method developed [1] in order 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 y-axis and number of patients on x-axis, subdivided in four zones by four lines: monitoring boundary for benefit, monitoring boundary for harm, and two futility boundaries (Fig.1). Two parallel lines 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 study by study with a chronological criteria. End of the line corresponds to the last study and will lie in one of the following areas: “benefit”, “harm”, “inner wedge” or “not statistically significant” zone, 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). Lying 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 by a penalizing factor λ (law of the iterated logarithm) or by adjusting significance threshold. The last described strategy is managed in TSA by 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.

Spending functions used in TSA find their bases in the O’BrienFleming’s one. Although several examples of such functions were described, only the O’BrienFleming’s one is implemented in the TSA software. Given a pre-defined α, the spending function is a monotonically increasing function distributing the α error along the whole analysis. 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. IF is given by the accumulated sample divided by the required sample.

The used α-spending function is

\[ \alpha(\text{IF})=2-2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{\text{IF}}}\right) \]  

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

Similarly, the β-spending function is a monotonically increasing function 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 take into account 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 too often is not adequately considered.

Choosing an adequate information size is the cornerstone in TSA. Nonetheless, there is not a standardized way nor a consensus to establish what is an adequate information size.

Similarly to randomized controlled trial, information size calculation is based on the choice of an 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 in comparison to the rate of event in the control group, described usually as a percentage

\[
\frac{(Pt/Pc)}{Pt} \times 100\%.
\]

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 literature is insufficient clinically experience (in example pilot studies) or data from related areas may be used.

It seems reasonable to state that the information size of a MA should be at least as large as the sample size of an adequately powered trial investigating that specific outcome, but researchers may be more conservative choosing a higher power (ie 90%-99%) and a lower \(\alpha\) (ie 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, 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 protocol, population or outcome assessment. However, this is utopian and a certain degree of clinical heterogeneity leading to
statistically heterogeneity has to be taken into account and accepted, and a correction factor for IF derived by heterogeneity magnitude is deemed necessary.

While MA usually uses inconsistency ($I^2$) as the measure of between trial variance, TSA uses diversity ($D^2$). The definition of $D^2$ is 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; in particular $D^2$ is 0 only when $I^2$ is 0 [4].

While the use of $D^2$ has the advantage to correct the IF in order to maintain the anticipated risk of both $\alpha$ and $\beta$ errors, it does not take into account any adjustment in IF for any type of 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 a sequential analysis in MA, that by definition is a retrospective analysis without any control on study design from meta-analyst, makes impossible to establish the stopping rules that are typical of a preplanned set of interim analyses.

TSA is usually performed on primary outcome, however cumulative evidence from secondary outcomes would be penalized from a premature stopping rule. A striking example is depicted by network meta-analysis, where cumulative evidence will continue to add effect to some networks also 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.
References

Figure 1. Trial Sequential Analysis graph. The graph is subdivided in four zones: area of benefit, area of harm, inner wedge, not statistically significant zone.