Should Clinical Trials Be Terminated Early?

David L. Streiner, PhD, CPsych

Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada

ABSTRACT

In many trials, interim analyses are often performed to decide whether the trial should be terminated early, for a variety of reasons. These reasons may include an imbalance in adverse events in one group versus the other(s), a situation in which, given the sample size, it would be impossible to demonstrate superiority of one treatment (futility) or a significant difference between groups. This commentary argues that ending a trial prematurely for the latter reason is fraught with problems and often results in an overestimation of the effect that would have been obtained were the trial allowed to continue. It concludes that stopping a trial early for apparent superiority of a treatment should be avoided. (*Clin Ther*. 2019;41:1889–1891)

Keywords: ending trials, research design, statistics, randomized controlled trials.

INTRODUCTION

Randomized, controlled trials are generally regarded as the "gold standard" for studies designed to evaluate the efficacy or effectiveness of a new intervention. However, there are ethics-related issues that arise from them. If the treatment is effective, then those in the comparison group are deprived of its benefit, at least for the duration of the trial. On the other hand, if the intervention is associated with significant adverse events (AEs), then those in the treatment group are at risk for injury and even death. For this reason, large trials, and especially those that extend for a long duration, often put mechanisms into place to minimize any risk to the participants. These mechanisms include preplanned interim analyses of the benefits and AEs, which are often monitored by an independent data safety and monitoring board (DSMB). Although the composition of the DSMB differs from one trial to the next, it usually consists of 3–7 members with expertise in the clinical aspects of the disorder, in the patient population being studied, in biostatistics, and in conducting randomized controlled trials.[1] There are 3 ethics-related reasons for which a DSMB (or a study's own team) may stop a trial before its scheduled completion: tolerability concerns, futility, and benefit.[2]

With regard to tolerability, the US Food and Drug Administration's criteria for the approval of study protocols by an ethics board states that "Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects and the importance of the knowledge that may reasonably be expected to result."[3] However, if there are unexpected serious AEs or deaths during the course of the trial, the DSMB may recommend to the institutional review board (research ethics board in Canada) that it be terminated early. Second, if an interim analysis shows that the hypotheses cannot be proved with the designed sample size, then the study could be terminated because of futility. On the other hand, if the interim analysis shows clear superiority in one arm of the trial (this injunction pertains even if the study has more than two arms), then it could be stopped early so that participants are not unnecessarily deprived of the benefits of the treatment. (It should be noted that *adaptive trials*—those with procedures modified partway through the study period—use interim analyses, but only for the discontinuation of treatments that may have proved to be futile, for modification of the subject-allocation schemes, or for changing the sample size, not for ending the study early if it appears that one treatment is superior.[4]) This last consideration—terminating a study early because of

Accepted for publication July 3, 2019
https://doi.org/10.1016/j.clinthera.2019.07.004
0149-2918/$ - see front matter

apparent superiority—is the focus of this correspondence.

One potential problem with interim analysis is the inflation of the *P* level, resulting in an increased probability of a Type I error; that is, declaring that there is a significant difference between groups when in fact there is not.[5] As Armitage et al[6] pointed out over 50 years ago, if there are 4 interim analyses performed in addition to the final analysis, each using the 5% α level, then the probability of finding significance, assuming the null hypothesis is true, is actually 18%. To avoid inflating the alpha level, various schemes have been developed to maintain the nominal α level for the entire trial. There are a number of "α-spending" approaches, which vary according to how the α level is divided among the interim analyses. For example, the Pocock method divides the α level so that it is the same for all analyses, while with the O'Brien-Fleming and Haybittle-Peto methods, it is much more difficult to declare significance for earlier analyses.[7]

However, deciding which stopping rule to use begs the question of whether trials should be stopped at all based on the apparent superiority of a treatment. The concern is that trials that have been terminated early because the beneficial effect of the treatment was overestimated, a phenomenon documented over 3 decades ago by Pocock and Hughes.[8] Simulating early stopping rules using data from an actual study, they concluded that "clinical trials that stop early are prone to exaggerate the magnitude of treatment differences" (p. 219S).[8] This exaggeration of effect was documented empirically by Wheatley and Clayton[9] in a study comparing 4 versus 5 courses of consolidation therapy in patients with acute myeloid leukemia. Because the study was relatively large ($n = 1078$) and recruited patients over 8 years, 6 interim analyses were planned in addition to the final analysis. The first analysis, performed after 202 patients had been enrolled, showed a statistically significant odds ratio (OR) of 0.57 in favor of 5 courses of treatment. The next 3 interim analyses were also significant and favored the 5-course arm, albeit with ever-decreasing ORs. The final analysis showed a nonsignificant OR of −0.09 favoring the 4-course condition. Thus, had the trial been terminated based on the first 4 interim analyses, it would have come to the wrong conclusion.

Montori et al[10] performed a systematic review of data from 143 trials that were ended early due to benefit. They reported that the number of such studies has been increasing, and concluded that "Trials stopped early for benefit, particularly those with few events, often report treatment effects that are larger than typical of interventions that have been definitively studied" (p. 2209).[10] Later work by the same team indicated that: large overestimates of benefit were shown when there were fewer than 200 events; with 200−500 events, the overestimates were smaller but "important"; and the overestimates were smallest when the number of events exceeded 500.[11]

CONCLUSIONS

It is tempting, from the perspectives of both ethics and cost, to use interim analyses to look for early signs of superiority in one arm of a study, and for the trial to be terminated if they are found. However, theoretic and empirical work shows that early termination would be a false savings, in that the result would be, at best, an overestimate of the effectiveness of the intervention, and at worst, a declaration in favor of the wrong treatment, as was seen in the Wheatley and Clayton[9] example. Ending a trial early because of futility or an excess of AEs in one group is legitimate, but should be avoided based on apparent superiority in one arm of a study.

CONFLICTS OF INTEREST

The author has indicated that he has no conflicts of interest with regard to the content of this article.

REFERENCES

1. National Institutes of Health, National Institute of Dental and Craniofacial Research. Data and safety monitoring board (DSMB) guidelines. Available at: https://www.nidcr.nih.gov/research/human-subjects-research/interventional-studies/data-and-safety-monitoring-board-guidelines. Accessed June 13, 2019.
2. Deichman RE, Krousel-Wood M, Breault J. Considerations for stopping a clinical trial early. *Ochsner J*. 2016;16:197−198.
3. US Food and Drug Administration. *Institutional Review Boards. Publication no. 21 CFR §56.111*; April 1, 2015. Available at: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr1/456.111. Accessed June 13, 2019.

4. Pallmann P, Bedding AW, Choodari-Oskooei B, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med*. 2018;16:29. https://doi.org/10.1186/s12916-018-1017-7.
5. Streiner DL. Best (but oft-forgotten) practices: the multiple problems of multiplicity—whether and how to correct for many statistical tests. *Am J Clin Nutr*. 2015;102:721–728. https://doi.org/10.3945/ajcn.115.113548, 2015.
6. Armitage P, McPherson CK, Rowe BC, et al. Repeated significance tests on accumulating data. *J Roy Statist Soc Ser A*. 1969;132:235–244. 10.2307/2343787.
7. Pocock SJ. When to stop a clinical trial. *Br Med J*. 1992;305:235–240. https://doi.org/10.1136/bmj.305.6847.235.
8. Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Control Clin Trials*. 1989;10(suppl 4): 209S–221S. https://doi.org/10.1016/0197-2456(89)90059-7.
9. Wheatley K, Clayton D. Be skeptical about unexpected large apparent treatment effects: the case of an MRC AML12 randomization. *Control Clin Trials*. 2003;24:66–70. https://doi.org/10.1016/S0197-2456(02)00273-8.
10. Montori VM, Devereaux PJ, Adhikari NKJ, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294: 2203–2209. https://doi.org/10.1001/jama.294.17.2203.
11. Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA*. 2010;303:1180–1187. https://doi.org/10.1001/jama.2010.310.

Address correspondence to: Dr. David L. Streiner, McMaster University, Department of Psychiatry and Behavioural Neurosciences, 100 West 5th Street, Hamilton, ON L8N 3K7, Canada. E-mail: streiner@mcmaster.ca