Trials

COMMENTARY   **Open Access**

Stopping guidelines for an effectiveness trial: what should the protocol specify?

CrossMark

Jon E. Tyson[1*], Claudia Pedroza[1], Dennis Wallace[2], Carl D'Angio[3], Edward F. Bell[4] and Abhik Das[5]

Abstract

Background: Despite long-standing problems in decisions to stop clinical trials, stopping guidelines are often vague or unspecified in the trial protocol. Clear, well-conceived guidelines are especially important to assist the data monitoring committees for effectiveness trials.

Main text: To specify better stopping guidelines in the protocol for such trials, the clinical investigators and trial statistician should carefully consider the following kinds of questions:

1. How should the relative importance of the treatment benefits and hazards be assessed?
2. For decisions to stop a trial for benefit:

    (a) What would be the minimum clinically important difference for the study population?
    (b) How should the probability that the benefit exceeds that difference be assessed?
    (c) When should the interim analyses include data from other trials?
    (d) Would the evidence meet state-of-the-art standards for treatment recommendations and practice guidelines?

3. Should less evidence be required to stop the trial for harm than for benefit?
4. When should conventional stopping guidelines for futility be used for comparative effectiveness trials?

Conclusion: Both clinical and statistical expertise are required to address such challenging questions for effectiveness trials. Their joint consideration by clinical investigators and statisticians is needed to define better stopping guidelines before starting the trial.

Keywords: Clinical trials, Effectiveness trials, Stopping guidelines, Data monitoring committees, Futility

Background

Despite long-standing problems in decisions to stop clinical trials [1–6], stopping guidelines are often vague or unspecified in the trial protocol. The protocol commonly does indicate the number of interim analyses to be performed and the method used to ensure that the type I error does not exceed 0.05 with multiple interim analyses of the primary outcome [7–9]. However, there is usually little effort to clearly indicate when the findings for both the treatment benefits and hazards [6, 10–12] would justify stopping the trial. Moreover,

this issue has not been explicitly addressed by the International Conference on Harmonization or such agencies as the Food and Drug Administration [13]. Yet clear, well-conceived guidelines are especially important to assist the data monitoring committees for effectiveness trials [14] intended to determine whether therapies should be used in clinical practice.

Main text

Definitive stopping rules that cover all circumstances are not possible. However, based partly on our experience in the NICHD Neonatal Research Network, we believe better stopping guidelines can be developed if the following basic kinds of questions are jointly and carefully considered by the clinical investigators and statisticians and addressed in preparing the study protocol:

* Correspondence: jon.e.tyson@uth.tmc.edu
[1]Center for Clinical Research and Evidence-Based Medicine, Department of Pediatrics, University of Texas Health Science Center at Houston, 6431 Fannin St, MSB 2.106, Houston, TX 77030, USA
Full list of author information is available at the end of the article

1. How should the relative importance of the treatment benefits and hazards be assessed?
   The importance (value) of treatment hazards relative to the benefits [15, 16] must be judged in defining stopping guidelines for either benefit or hazards. Although difficult to make, these judgments are implicit in any stopping guideline (as well as any treatment recommendation or practice guideline) [17–20] and are generally based on the views of investigators or clinicians. Ideally these judgments would be based on prior formal assessments of the views of patients (or their surrogates) who have experienced or are most likely to experience these benefits or hazards. This issue is a fertile area for research in increasing meaningful patient involvement in designing clinical trials [21].
2. What evidence should be required to stop the trial for benefit?
   (a) What would be the clinically important difference (CID)—the minimum magnitude of the treatment benefit large enough to offset the treatment harms [16]—for the patient population in the trial?
   The CID depends on the rate and relative importance of treatment benefits and hazards. Clearly and explicitly defined stopping guidelines can specify how the observed rates of treatment benefits and harms can be used in assessing whether the evidence of net benefit or harm is strong enough to justify stopping the trial.
   (b) How should the probability that the benefit exceeds the CID be assessed?
   While interim analyses are usually performed using only standard frequentist analyses, Bayesian analyses are needed to directly estimate not only the probability of any benefit but also the probability that it exceeds benefit of a specific magnitude such as the CID [22–25]. They may also be used if it is appropriate to incorporate data from prior trials of the same intervention to estimate the updated (posterior) probability of benefit based on the current trial. [See item (c).] In addition, Bayesian analyses are also likely to be more understandable than frequentist analyses to clinicians [23–26]. For these reasons, Bayesian analyses can be recommended for use with conventional frequentist analyses in assessing interim as well as final results [26, 27]. Even so, judgment will still be required in interpreting the analyses and deciding whether the trial should be stopped.
   (c) When should the planned interim analyses include data from other relevant trials?

   The effect size associated with low p values in interim results is likely to be larger than that obtained if the trial is continued to its end [1, 28–31]. At least for trials where early stopping is considered, consideration of data from prior trials and, if appropriate, incorporation into the interim analyses may well be warranted to avoid premature trial termination and erroneous conclusions due to misleading interim data.
   (d) Would the evidence of benefit meet state-of-the-art standards for treatment recommendations and practice guidelines?
   These standards address many factors beyond statistical significance [19, 20, 32]. At least for comparative effectiveness trials that compare commonly used therapies and are conducted to provide definitive results to guide clinical practice, a strong case can be made to continue the trials if such standards have not been met despite meeting conventional stopping guidelines [1, 2].
3. Should less evidence be required to stop the trial for harm than benefit?
   The prudent answer to this question is likely to be "Yes" based on the ethical maxim of *primum non nocere* and the need to protect patient safety [33, 34] and to focus the limited resources for clinical trials on therapies most likely to be beneficial.
   Yet it is uncertain exactly how much evidence of harm should be required, or in Bayesian terms, what probability of net harm is high enough to stop a trial. For treatment harms as for benefits, interim findings extreme enough to be statistically significant or to have high Bayesian probabilities are likely to be more extreme than those obtained if more patients are studied [6, 28, 35, 36]. Misleading findings are especially likely if the analyses are repeated at frequent intervals. Large numbers of future patients, including some patients who otherwise would have been included in the trial, may then be harmed if one erroneously concludes that a truly beneficial therapy is harmful. Quantitative methods to better address this dilemma are now being explored [37].
4. When should conventional stopping guidelines for futility be used for comparative effectiveness trials?

Conditional power analyses of interim results or Bayesian predictive probabilities [38] help identify when a significant difference between treatment groups is unlikely to be identified if the trial continues to the pre-planned sample size. However, such findings would not necessarily indicate that a trial should be stopped. Even in the absence of significant differences, completion of comparative effectiveness trials comparing two widely used therapies may promote greater use of the therapy

with the higher likelihood of benefit, particularly if it is also less invasive, hazardous, expensive, or inconvenient. Moreover, important therapeutic questions are rarely answered in a single trial, and early trial termination will reduce the power of later meta-analyses of all relevant trials to identify important treatment effects for all patients, important subgroups, and patients at differing risk [39]. Such analyses are crucial for applying the results of clinical trials to individual patients [15, 16, 39]—a cutting-edge issue important to augmenting the value of clinical trials.

Conclusions

Although clinical investigators may consider stopping guidelines to be the responsibility of statisticians, both clinical and statistical expertise are required to address such challenging questions. Careful consideration by clinical investigators and statisticians will help to specify better stopping guidelines in the protocol for effectiveness trials, including those that may need especially careful consideration for special populations (e.g., pregnant women or children) or disorders (e.g., cancer). These guidelines can then be reviewed by the data monitoring committee, and any areas of disagreement can be discussed and addressed before starting the trial. As the methods of clinical trials evolve over time, greater input from clinicians and patients will also be needed to promote progressively better informed, better justified, more useful, and more broadly acceptable stopping guidelines.

Author details
[1]Center for Clinical Research and Evidence-Based Medicine, Department of Pediatrics, University of Texas Health Science Center at Houston, 6431 Fannin St, MSB 2.106, Houston, TX 77030, USA. [2]RTI International, Durham, NC 27709, USA. [3]University of Rochester School of Medicine and Dentistry, Rochester, NY 14627, USA. [4]Department of Pediatrics, University of Iowa, Iowa City, IA 52242, USA. [5]Social, Statistical and Environmental Sciences Unit, RTI International, Durham, NC 27709, USA.

References

1. Pocock SJ. When to stop a clinical trial. BMJ. 1992;305:235–40.
2. Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, et al. Issues in data monitoring and interim analysis of trials. Health Technol Assess. 2005;9:1–238. iii–iv.
3. Bassler D, Montori VM, Briel M, Glasziou P, Guyatt G. Early stopping of randomized clinical trials for overt efficacy is problematic. J Clin Epidemiol. 2008;61:241–6.
4. Guyatt GH, Briel M, Glasziou P, Bassler D, Montori VM. Problems of stopping trials early. BMJ. 2012;344:e3863.
5. Zannad F, Gattis Stough W, McMurray JJV, Remme WJ, Pitt B, Borer JS, et al. When to stop a clinical trial early for benefit: lessons learned and future approaches. Circ Heart Fail. 2012;5:294–302.
6. Pocock SJ. Current controversies in data monitoring for clinical trials. Clin Trials. 2006;3:513–21.
7. Pocock SJ. Interim analyses for randomized clinical trials: the group sequential approach. Biometrics. 1982;38:153–62.
8. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35:549–56.
9. Geller NL, Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. Biometrics. 1987;43:213–23.
10. DeMets DL, Pocock SJ, Julian DG. The agonising negative trend in monitoring of clinical trials. Lancet. 1999;354:1983–8.
11. Lan KK, Wittes J. The B-value: a tool for monitoring data. Biometrics. 1988;44:579–85.
12. Ellenberg SS, Fleming TR, DeMets DL. Data monitoring committees in clinical trials: a practical perspective. Chichester: Wiley; 2003.
13. U.S. Food and Drug Administration. ICH guidance documents. 2013. http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/GuidancesInformationSheetsandNotices/ucm219488.htm. Accessed 4 Apr 2016.
14. Fletcher RH, Fletcher SW. Clinical epidemiology: the essentials. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2005.
15. Sinclair JC. Weighing risks and benefits in treating the individual patient. Clin Perinatol. 2003;30:251–68.
16. Marcucci M, Sinclair JC. A generalised model for individualising a treatment recommendation based on group-level evidence from randomised clinical trials. BMJ Open. 2013;3:e003143.
17. Pocock SJ. When (not) to stop a clinical trial for benefit. JAMA. 2005;294:2228–30.
18. McAlister FA, Straus SE, Guyatt GH, Haynes RB. Users' guides to the medical literature: XX. Integrating research evidence with the care of the individual patient. Evidence-Based Medicine Working Group. JAMA. 2000;283:2829–36.
19. Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength. J Clin Epidemiol. 2013;66:726–35.
20. Vandvik PO, Brandt L, Alonso-Coello P, Treweek S, Akl EA, Kristiansen A, et al. Creating clinical practice guidelines we can trust, use, and share: a new era is imminent. Chest. 2013;144:381–9.
21. Patient-Centered Outcomes Research Institute. www.pcori.org. Accessed 4 Apr 2016.
22. Fayers PM, Ashby D, Parmar MK. Tutorial in biostatistics: Bayesian data monitoring in clinical trials. Stat Med. 1997;16:1413–30.
23. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. BMJ. 1999;319:508–12.
24. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov. 2006;5:27–36.
25. Harrell FE, Shih YC. Using full probability models to compute probabilities of actual interest to decision makers. Int J Technol Assess Health Care. 2001;17:17–26.
26. Parmar MK, Griffiths GO, Spiegelhalter DJ, Souhami RL, Altman DG, van der Scheuren E, et al. Monitoring of large randomised clinical trials: a new approach with Bayesian methods. Lancet. 2001;358:375–81.
27. Wijeysundera DN, Austin PC, Hux JE, Beattie WS, Laupacis A. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. J Clin Epidemiol. 2009;62:13–21. e5.
28. Chalmers I, Sinclair JC. Promoting perinatal health: is it time for a change of emphasis in research? Early Hum Dev. 1985;10:171–91.
29. Wittes J. Stopping a trial early — and then what? Clin Trials. 2012;9:714–20.
30. Hughes MD, Pocock SJ. Stopping rules and estimation problems in clinical trials. Stat Med. 1988;7:1231–42.
31. 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;303:1180–7.

32. U.S. Preventive Services Task Force. Grade definitions. 2014. http://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions. Accessed 4 Apr 2016.
33. DeMets DL, Friedman LM, Furberg CD. Data Monitoring in clinical trials. New York: Springer; 2006.
34. Piantadosi S. Clinical trials: a methodologic perspective. 2nd ed. Hoboken: Wiley; 2005.
35. Demets DL. Futility approaches to interim monitoring by data monitoring committees. Clin Trials. 2006;3:522–9.
36. Spiegelhalter DJ. Bayesian approaches to clinical trials and health care evaluation. Chichester: Wiley; 2004.
37. Djulbegovic B, Hozo I. When should potentially false research findings be considered acceptable? PLoS Med. 2007;4:e26.
38. Saville BR, Connor JT, Ayers GD, Alvarez J. The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. Clin Trials. 2014;11:485–93.
39. Vickers AJ, Kattan MW, Sargent D. Method for evaluating prediction models that apply the results of randomized trials to individual patients. Trials. 2007; 8:14.