

Letter from the editor

Interim analysis: its uses and limitations

The ethics of clinical trials are of paramount concern to investigators and participants — the desire to learn about the benefits of new therapies by testing drugs in human studies must be counterbalanced by a commitment to minimise harm to participants, limit resource expenditures, and maximise the benefit of the trial to the community. Because of the need to adhere to these principles, independent data monitoring committees (DMCs) often preview the accumulating data in a clinical trial, and sometimes recommend actions to study investigators following the review of early data — a process known as interim analysis. In many instances, pre-planned interim analyses are built into the protocols of clinical trials with stopping rules (statistical criteria and plans for stopping the trial) should results suggest impressive benefit or untoward harms.

The recent controversy surrounding the anti-diabetic drug rosiglitazone has highlighted the challenges of interim analyses. Prompted by the publication of a meta-analysis of trials reporting an increase in the risk of myocardial infarction and cardiovascular death with rosiglitazone,[1] the authors of the RECORD study — a large RCT comparing the cardiovascular effects of rosiglitazone with other anti-diabetic regimens in people with type 2 diabetes — released the results of an unplanned interim analysis.[2]

In this editorial, we review the uses and limitations of interim analyses in clinical trials (which we summarise in table 1) and discuss the debate generated by the results of the RECORD interim analysis.

Protocol modifications

Early data review may precipitate protocol modification (see table 1). On the basis of these analyses, investigators may decide to collect additional data, increase the sample size, or extend the length of follow-up — the latter two modifications will enhance the statistical power of the study to detect a treatment effect when the event rate is lower than originally anticipated.

The Caffeine for Apnea of Prematurity (CAP) trial, a multicentre RCT that compared caffeine with placebo for the treatment of apnoea in preterm infants,[3] provides an example of the uses of interim analysis. The external DMC reviewed the study data every 4 to 6 months during the enrolment phase of the trial. The use of drug or surgical therapy to close a persistent ductus arteriosus (PDA) was documented but was not a prespecified outcome. The DMC noted a higher incidence of PDA in the placebo group and, 21 months after the start of enrolment, recommended to the steering committee that it collect additional data regarding all infants whose PDA was closed surgically. An analysis at the end of the trial revealed that infants in the caffeine group were significantly less likely to undergo therapy (in particular, surgery) to close a PDA than were infants in the placebo group. These findings prompted additional ongoing studies clarifying the effect of caffeine on ductal closure.

Another example of the value of interim analysis comes from the CAPRIE trial — a randomised, blinded trial of clopidogrel versus aspirin in people at risk of ischaemic events — in which the investigators extended enrolment because overall event rates were found to be lower than expected in an interim review blinded to allocation.[4] Initially, the investigators designed the trial to include 15,000 people; in the end, they randomized 19,185 participants. The primary end point of the study was a composite of ischaemic stroke, myocardial infarction, or vascular death, which proved less common in the clopidogrel group than in the aspirin group with a P value that just crossed the conventional level of statistical

significance. Without the increase in sample size, the trial would have had inadequate power to detect the relatively small effects in which the investigators were interested. This provides a good example of useful protocol modifications based on an analysis of overall events without reference to differential event rates in treatment and control groups.

Although extending the duration of a trial because of the results of an interim analysis may be useful, keeping the accumulating results blinded is a challenge, as some approaches used to decide whether to extend a trial (for example, conditional power concept, which is the probability that the trial will ultimately show statistical significance in its primary efficacy end point conditional on the observed data) require knowledge of outcomes in both the intervention and control groups. When investigators modify a trial protocol (for example, altering planned sample size, length of follow-up, or outcome measures) in response to an interim analysis on the basis of some prespecified significance level, these changes may introduce bias. To avoid compromising the methodology of the study, investigators should make decisions about extending the study length on the basis of *pooled* event rates in intervention and control groups.[4]

Early termination

If early results suggest that a treatment under study is particularly beneficial or harmful, the investigators may decide to terminate the study earlier than planned (see table 1). Sometimes trials are stopped on the basis of “futility”, when interim results indicate that it is unlikely that results at the scheduled end of the trial will demonstrate a significant effect. We have recently published detailed discussions of the problems inherent in stopping trials early, and will review them here briefly.[5] [6]

In an open-label RCT that assessed different ways of sterilising the catheter lumen, people receiving long-term haemodialysis were assigned to receive two different antibiotic lock solutions (gentamicin or minocycline) or the control solution of heparin.[7] Participants were followed up until the trial end point of catheter-related bacteraemia (CRB). Although interim analysis was primarily performed to assess safety, the trial was terminated early based on an apparent large treatment effect: 7 of 20 people in the heparin group (4.0 events/1,000 catheter days), 1 of 21 people in the minocycline group (0.4 events/1,000 catheter days), and none of 20 people in the gentamicin group developed bacteraemia. Results of the interim analysis were statistically significant and the authors concluded that antibiotic lock solutions are superior to the standard heparin lock alone in the prevention of CRBs.

There are, however, problems associated with stopping RCTs early for benefit — no matter what statistical procedure was used to justify termination of the trial, the estimates of treatment effect may be biased. Bias arises because random fluctuations towards greater treatment effects may result in early termination. In other words, the effects of the estimated treatment may be exaggerated because statistical stopping rules are prone to stop a trial in a “random high”. If the decision to stop the trial does result from observing the apparent benefit of treatment at a “random high”, the resulting estimate of the treatment effect will be misleading. The likelihood of bias is greater when the number of events is small — as it was in the case of the study on catheter lock solutions,[7] in which the magnitude of the treatment effect is likely too good to be true.[8]

When an RCT is stopped early for apparent benefit, the potential overestimation of treatment effects, lack of precision consequent on relatively few events, and paucity of information regarding other important clinical outcomes, can make it difficult for clinicians and future patients to draw reliable inferences about the benefits and risks of an intervention. The resulting uncertainties will compromise wise decision-making. Furthermore, stopping trials early for apparent benefit is particularly hazardous because such trials are often published in high-impact journals and receive considerable attention.[8]

This may result in the quick approval and dissemination of a treatment on the basis of a biased estimate of effect size.

Ethical concerns make stopping trials early for safety more complex than stopping early for efficacy. Most would agree that a lesser degree of evidence is required to declare a harmful trend than to declare a beneficial effect. These less stringent thresholds to stop trials early for harm reflect a preference for forgoing potentially useful interventions rather than exposing patients to harm. Inferences regarding harm are, however, equally susceptible to the biases associated with stopping early for benefit.

Interim analyses and stopping rules for futility can reduce time, money, and effort spent on clinical trials that are unlikely to lead to significant results. These resources can be used for alternative studies asking more promising research questions. However, early stopping typically results in wide confidence intervals, and therefore leaves the primary study question unanswered. Because of the so-called 'regression to the truth effect' — in which early random over- or underestimates of effect move closer to the truth as the data accumulate — RCTs stopped early for futility will, on average, underestimate treatment effects (RCTs stopped early for efficacy will, for the same reason, on average, overestimate the treatment effect). Although there are fundamental differences in the implications of stopping a study early for efficacy compared with futility, the false estimation of treatment effect potentially resulting from studies stopped early for futility might still be problematic in meta-analyses not mindful of this problem.

Conducting and reporting of interim analyses

As we have pointed out, actions following even properly conducted interim analyses may introduce bias in the results of a trial, and subsequently weaken confidence in its conclusions. When the conduct of interim analyses is less rigorous, the dangers of introducing bias increase further. Therefore, investigators should carefully plan the timing, frequency, methods, and potential consequences of interim analyses, and describe these in study protocols; trial reports should describe these processes and the decisions that followed. However, interim analyses are often reported insufficiently or not at all. In a cross-sectional study on the use of DMCs in published reports of RCTs, only 120 of 662 (18%) RCTs reported planned interim analyses.[9] A systematic review of RCTs stopped early because of benefit found that only 67 of 143 (47%) RCTs reported the three key methodological elements that are used to appraise the appropriateness of the interim analyses and the decision to stop the trial early: planned sample size; number of interim analyses performed and the interim analysis after which the RCT was stopped; and stopping rules.[8] A systematic survey on RCTs stopped early for harm in HIV/AIDS reported that 1 of 10 (10%) RCTs reported planned stopping rules.[10]

Publication of interim results prior to completion of study

Sometimes, investigators feel compelled to publish the results of their interim analyses while the trial is ongoing, as was the case for the RECORD study discussed above. In light of the safety concerns raised about the trial drug rosiglitazone in the meta-analysis, the authors of the RECORD study interim analysis, Home and colleagues, believed that "...the current totality of evidence needs to be made available".[2] They expected the publication to protect the study's integrity by avoiding an increase in dropout rates and potential biases in reporting events that may have occurred because of the publicity surrounding the meta-analysis results. After following patients for an average of 3.75 years (the planned mean follow-up of the trial was 6 years), the interim analysis found a low number of events overall, resulting in wide confidence intervals, low statistical power, and no significant difference between the intervention and control groups in the primary outcome of time to first hospitalisation for a cardiovascular event or death from cardiovascular causes (hazard ratio [HR] for rosiglitazone versus other regimens: 1.08; 95% CI 0.89 to 1.31). When end points pending adjudication were included, the 95% confidence interval around the point estimate was consistent with as much as a 7% decrease in cardiovascular risk and as much as a 32% increase in risk with rosiglitazone regimens (HR 1.11; 95% CI 0.93 to 1.32).[2]

The debate that followed the publication of these results highlights the difficulties associated with the interpretation of such analyses, even among sophisticated scientists. An editorial in the *New England Journal of Medicine* by Drazen, Morrissey, and Curfman points out that, because of incomplete follow-up and low event rates, the analysis is underpowered and therefore its results are inconclusive. However, they support the publication of this interim analysis, emphasising that, "...although there may be uncertainty about a drug's safety, there should be no uncertainty about the need for open and honest disclosure."^[11] Other editorials express concern about the safety of rosiglitazone. In his commentary in the *New England Journal of Medicine*,^[12] David Nathan points out that, despite being underpowered, the results of the interim analysis suggest a trend for adverse cardiovascular outcomes in the rosiglitazone group. Nathan argues that, in light of these results, the RECORD study is unlikely to ever establish a cardiovascular benefit for rosiglitazone. He asks whether physicians should feel comfortable using rosiglitazone and concludes that the answer should be no.^[12] By contrast, the authors of the interim report emphasise that the study's DMC has recommended continuation of the trial.^[2]

These challenges and uncertainties are daunting for participants in the trials and for patients in the community. RECORD investigators may be able to determine the extent to which they achieved the desired effect in reporting the interim analysis, as the rates of trial withdrawal and crossover to the control arm become clear. Large withdrawal rates could complicate the interpretation of the final trial results and weaken inferences about the risks and benefits of rosiglitazone.

Conclusions

Actions following interim analyses can bias the results of a trial and subsequently weaken confidence in the conclusions drawn. In particular, a major limitation of interim analyses leading to early stopping for efficacy is their potential for overestimating treatment effects. Interim analyses should be carefully planned in advance and executed according to prespecified rules. Details of interim analyses and stopping rules should be part of study protocols and reports. We recommend circumspection in the planning and conduct of interim analyses, and a high level of scepticism in interpretation of their results. As demonstrated by the publication of the interim analysis of the RECORD study, physicians, patients, and trial participants might have substantial difficulties interpreting results of interim analyses of ongoing studies, and translating them into clinical practice.

Table 1: Uses and limitations of interim analyses, and their potential consequences

| Consequence of interim analysis | Uses | Limitations |
|-------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Protocol modifications | Might help to answer the primary study question and to generate new hypotheses | Requires extra methodological and statistical safeguards |
| Early termination for efficacy | Might speed up dissemination of effective treatments Might fulfil an ethical obligation to protect trial participants | Estimate of treatment effect is imprecise and will, on average, be falsely inflated Information on secondary outcomes diminishes True benefit-risk ratio might be difficult to assess |
| Early termination for harm | Protection of trial participants Avoidance of further patient exposure to harmful treatments | Risks abandonment of a treatment that might be beneficial Information on primary and secondary outcomes diminishes |
| Early termination for futility | Saves resources that could be used on more promising research | Leaves primary study question unanswered Information on secondary outcomes diminishes Estimate of treatment effect is imprecise and might be biased downward Interpretation of trial might become more complex |
| Publication of interim results while trial is ongoing | Open disclosure of total available evidence Might protect the trial's integrity | Might create uncertainty Might lead to a loss of power of the ongoing study Might compromise the trial's integrity |

Dirk Bassler MD, MSc

Department of Neonatology, University Children's Hospital, Tuebingen, Germany

Victor M. Montori MD, MSc

Knowledge and Encounter Research Unit, Division of Endocrinology and Internal Medicine, Mayo Clinic College of Medicine, Rochester, Minnesota, USA.

Gordon Guyatt MD, MSc

CLARITY Research Group, Department of Clinical Epidemiology and Biostatistics, Health Sciences Centre Room 2C12, McMaster University, Hamilton, Ontario, Canada.

guyatt@mcmaster.ca

References

1. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007;356:2457–2471. Epub 2007 May 21. Erratum in: *N Engl J Med* 2007;357:100.
2. Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone Evaluated for Cardiovascular Outcomes — An Interim Analysis *N Engl J Med* 2007 Jun 5. [Epub ahead of print]
3. Schmidt B, Roberts RS, Davis P, et al; Caffeine for Apnea of Prematurity Trial Group. Caffeine therapy for apnea of prematurity. *N Engl J Med* 2006;354:2112–2121.
4. CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet* 1996;348:1329–1339.
5. Bassler D, Montori VM, Briel M, et al. Early stopping of randomised clinical trials for overt efficacy is problematic. *J Clin Epidemiol*, in press.
6. Mueller PS, Montori VM, Bassler D, et al. Ethical issues in stopping randomized trials early because of apparent benefit. *Ann Intern Med* 2007;146:878–881.
7. Nori US, Manoharan A, Yee J, et al. Comparison of low-dose gentamicin with minocycline as catheter lock solutions in the prevention of catheter-related bacteremia. *Am J Kidney Dis* 2006;48:596–605.
8. Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005;294:2203–2209.
9. Sydes MR, Altman DG, Babiker AB, et al; DAMOCLES Group. Reported use of data monitoring committees in the main published reports of randomized controlled trials: a cross-sectional study. *Clin Trials* 2004;1:48–59.
10. Mills E, Cooper C, Wu P, Rachlis B, Singh S, Guyatt GH. Randomized trials stopped early for harm in HIV/AIDS: a systematic survey. *HIV Clin Trials* 2006;7:24–33.
11. Drazen JM, Morrissey S, Curfman GD. Rosiglitazone — continued uncertainty about safety. *N Engl J Med* 2007;357:63–64. Epub 2007, Jun 5.
12. Nathan DM. Rosiglitazone and cardiotoxicity — weighing the evidence. *N Engl J Med* 2007;357:64–66. Epub 2007, Jun 5.