

## Can we trust in trials stopped early for benefit?

Giovanni Casazza · Francesco Casella ·  
Gruppo di Autoformazione Metodologica (GrAM)

Received: 14 September 2012 / Accepted: 10 October 2012 / Published online: 31 October 2012  
© SIMI 2012

In a multicenter, randomized, placebo-controlled, double-blind clinical trial, Elmunzer et al. [1] assigned patients at elevated risk for post- endoscopic retrograde cholangiopancreatography (ERCP) pancreatitis to receive a single dose of rectal indomethacin or a placebo immediately after ERCP.

According to the authors' summary, prophylactic rectal indomethacin significantly reduced the incidence of post-ERCP pancreatitis (PEP) in patients at elevated risk for this complication. The trial has been stopped early after an interim analysis.

Is it reasonable to trust the results of this early stopped trial showing the benefit of indomethacin in preventing PEP?

A randomized clinical trial (RCT) is designed and conducted to provide the answer to a relevant clinical question. The protocol of an RCT should report several details on how the researchers are going to conduct the trial. First of all, the aim of the study has to be clear. Secondly, the outcomes (primary and secondary) should be predetermined as the statistical methods used to analyze them. Next, another item to be specified in the protocol is the number of patients to be included in the study. When determining the sample size, investigators should clarify some details about the scientific hypothesis underlying the clinical question they want to answer.

Large RCTs are periodically monitored as patients are enrolled, and interim analyses are performed on collected data. Sometimes, the decision of stopping a trial early is adopted as a consequence of the results of an interim analysis. There are several reasons for the early termination of a RCT. First, a trial may be stopped early for harm, if significantly worse outcomes are observed among patients receiving the experimental treatment. Another reason for early termination of a trial may be futility, when it is highly unlikely that the trial will accrue the planned number of patients, or if the interim analysis shows that it is extremely unlikely that any benefit will be seen if the study is continued. Finally, trials may be stopped for a significant benefit in the experimental arm (trial stopped early for benefit). In this case, patients receiving the non-experimental treatment may be allowed to “crossover” and receive the beneficial treatment.

In these papers we will only focus on the trials that have been stopped early for benefit.

There are several reasons for stopping an RCT for benefit. First of all, the investigators may feel ethically obligated to stop a trial showing an apparent benefit of a study treatment. An individual ethical issue may drive the decision to stop a trial before the end of the randomization of all the planned patients, since it's considered unethical to deny a study participant an effective treatment. This issue may be compelling for investigators and patients who may be allowed to “crossover”, and receive the more beneficial treatment after an interim analysis shows an apparent benefit of the experimental intervention.

Beyond the individual ethical issues, there are some collective interests leading to the decision to stop a trial for benefit. If a new treatment is more effective than the older one, it's a collective interest that study results spread as quickly as possible, making the treatment available for all

---

G. Casazza (✉)  
Dipartimento di Scienze Biomediche e Cliniche “L. Sacco”,  
Università degli Studi di Milano, Via G.B. Grassi, 74,  
20157 Milan, Italy  
e-mail: giovanni.casazza@unimi.it

F. Casella  
Medicina Interna III, Dipartimento di Medicina,  
Ospedale “L. Sacco”, Università degli Studi di Milano,  
Milan, Italy

the patients as soon as possible. Trials stopped early may prove the efficacy of a treatment some years before the enrollment of the last planned patient.

Furthermore, if the large benefit in the experimental arm is believed to have satisfied the study question, research sources may be invested in other issues, a consideration of concern for the research community and funding agencies.

However, investigators, funding agencies and journals have additional interests to stop a trial. In particular, funding agencies are interested in stopping trials to reduce the cost of trials, while journals have interest in publishing the apparently exciting findings of an early terminated study. All these interests may affect the decision to stop a trial, but it may be inappropriate [2].

On the other hand, the scientific community requires evidence based on a large number of patients to definitively prove the effectiveness of a treatment. Well designed and conducted large clinical trials not stopped early are warranted. For these reasons, RCTs should not be inappropriately stopped. To guarantee a safe and effective conduct within a clinical study, all trials should include an external monitoring committee called Data and Safety Monitoring Committee (DSMC). This independent committee may decide to stop a trial when significant benefits or risks are observed. The decision to stop a RCT is mainly adopted on the results of interim analyses, performed on the basis of prespecified stopping rules. There are several statistical stopping rules that may be used to limit the impact of bias due to early termination on the results of the trial. All of them set a boundary in terms of risk of type 1 error (alpha) lower than 5 % (assuming that  $\alpha = 0.05$  is used in sample size calculation) [3].

In the paper of Elmunzer et al., the outcome was the development of pancreatitis. The study was sized to detect a

50 % reduction in the incidence of pancreatitis (from 10 % in the placebo group to 5 % in the indomethacin group) with an 80 % power and a two-sided alpha of 5 %. So that, the estimated number of patients to enroll was 948 (474 per group). Two interim analyses were planned: one after the enrollment of 400 patients, and the second after the enrollment of 600 patients. The  $p$  value cut-offs for early termination were fixed "... on the basis of the O'Brien–Fleming approach and the Lan–DeMets alpha spending function" [1]. However, the authors do not report the  $p$  value for the second interim analysis (0.005 for the first interim and 0.041 for the final). Anyway, it would be a value between 0.005 and 0.041. The trial was stopped after the enrollment of the first 600 patients, since the second interim analysis showed an apparent benefit in the indomethacin group.

Several risks of RCT stopped for clinical benefit are addressed in recent studies. In 2005, Montori and colleagues [4] published a systematic review of RCT stopped early for benefit. They observe limited reporting of critical features specific to the decision to stop the trial. A significant proportion of these studies do not specify whether a statistical approach to monitor the trial is used while only a minority of them report key methodological elements such as planned sample size, the number of interim analyses and the stopping rules. The lack of information about the decision to stop early is a weakness of the single RCT, and may be overcome by a higher degree of transparency with respect to the number of interim analyses carried out and the stopping rules.

Recent studies observe that truncated RCTs have different biases leading to implausible large treatment effects and misleading estimates of the benefit [5]. Use of well defined statistical procedures may obviate the problem of

#### Commonly used early termination rules

O'Brien and Fleming rule: requires very strong evidence (very low  $p$  value) at each interim analysis to decide to stop the trial. If three analyses are planned (two interim + one final) the trial is stopped, if  $p$  is lower than 0.0005 in the first interim analysis, 0.014 in the second interim and 0.045 in the last analysis.

As it can be seen from the above numbers, the  $p$  value required to stop the trial increases as the trial reaches the planned size. In the last analysis (planned sample size), the  $p$  value for statistical significance approaches the nominal 0.05.

Pocock rule: a constant  $p$  value (to ensure an overall  $\alpha=0.05$ ) is used at each analysis, including the last one. If three analyses are planned, a  $p$  value of 0.0221 is required for statistical significance.

Peto-Haybittle rule: a constant  $p$  value of less than 0.001 is required to stop the trial at each analysis regardless the number of planned interim analysis. At the final analysis the usual 0.05 level is used.

Lan and DeMets method: a more complex but flexible method that takes into account the timing of the interim analysis, and allows the researcher to partition the overall 0.05 alpha value between the prespecified planned analyses.

multiple repeated interim analyses at the same unadjusted level of significance. Furthermore, they may reduce the biases associated to RCTs stopped early by having much more stringent significance levels in the early interim analyses. This issue is emphasized by Korn et al. [6], who suggest that early stopping of trials may be reasonable when the study is well planned and stopping rules are defined. However, even predefined statistical stopping rules may be insufficient to prevent the risk of overestimating treatment benefit of RCTs, stopped early for benefit. Bias arises because large random fluctuations of the estimated treatment effect can occur, particularly early in the progress of a trial [7]. Bassler et al. observe that the risk of overestimating treatment benefit for truncated RCTs is increased in studies with fewer than 500 events. On the contrary, the methodological quality and the presence of defined statistical stopping rules fail to predict the magnitude of the bias of RCTs stopped early for benefit [5].

Another important issue is the choice of the end point used. It is questionable to stop a trial showing positive treatment effects for a surrogate end point.

In conclusion, although it is understandable that investigators and DSMC may feel ethically obligated to protect study participants, they have an ethical obligation to society by helping the scientific community and patients to know which is the best treatment. The risk of truncated trials to overestimate treatment effect on the endpoint that resulted in trial termination may weaken study validity, and thus endanger the wider community to whom the results will be applied.

For these reasons, DSMC should consider the consequences of early termination of a RCT. First of all, the decision to stop a trial may be adopted only if the timing of the interim analyses and the statistical stopping rules has been previously planned in the protocol. Furthermore, investigators and DSMC should not solely use *p*-values to decide on early termination of a study. Other considerations should affect the decision to stop a trial such as the choice of the primary outcome, the number of patients having an outcome event, and the number of interim analyses conducted prior to decision, and their results.

We believe that the reliability of a clinical trial is based on the publication of the protocol ([www.clinicaltrials.gov](http://www.clinicaltrials.gov)) of the study in which all the key points of the RCT (end-points, sample size, interim analyses) should be well described. Furthermore, an independent DSMC should ensure a safe conduct of RCT according to prespecified

rules. Finally, although we recognize the possibility of biased results in RCT that has been stopped early for benefit, we believe that early termination may be reasonable when the trial is well planned, stopping statistical rules are defined, and a sufficient number of events have occurred.

### Bottom line for clinicians

1. When interpreting the results of an RCT stopped early for benefit, you should carefully read the methods section to find out if the decision-making process of early termination is well described, and the statistical rules used to stop the study are appropriate (when it is published on [www.clinicaltrials.gov](http://www.clinicaltrials.gov)).
2. Even if the statistical methods used to monitor the trial are well described and are appropriate, you should remember that the risk of overestimation of the effect may be present in a RCT stopped early for benefit (benefit could be apparent).
3. When interpreting guidelines and systematic reviews, you should always consider if their findings are based on the results of truncated RCTs.

**Conflict of interest** None.

### References

1. Elmunzer BJ, Scheiman JM, Lehman GA et al (2012) A randomized trial of rectal indomethacin to prevent post-ERCP pancreatitis. *N Engl J Med* 366(15):1414–1422
2. Mueller PS, Montori VM, Bassler D et al (2007) Ethical issues in stopping randomized trials early because of apparent benefit. *Ann Intern Med* 146(12):878–881
3. Pocock SJ (1992) When to stop a clinical trial. *BMJ* 305(6847):235–240
4. Montori VM, Devereaux PJ, Adhikari NK et al (2005) Randomized trials stopped early for benefit: a systematic review. *JAMA* 294:2203–2209
5. Bassler D, Briel M, Montori VM, Lane M et al (2010) Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA* 303(12):1180–1187
6. Korn EL, Freidlin B, Mooney M (2009) Stopping or reporting early for positive results in randomized clinical trials: the National Cancer Institute Cooperative Group experience from 1990 to 2005. *J Clin Oncol* 27(10):1712–1721
7. Schulz KF, Grimes DA (2005) Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet* 365(9471):1657–1661