THE DANGERS OF STOPPING TRIALS EARLY

When interim analyses of randomised trials suggest large beneficial treatment effects, investigators sometimes terminate trials earlier than planned. **Gordon H Guyatt and colleagues** show how this practice can have far reaching and harmful consequences

In a seminal simulation study published in 1989, Pocock and Hughes showed that randomised control trials stopped early for benefit will, on average, overestimate treatment effects.[1] Since then, the warning implicit in this simulation study has been largely ignored.

Fifteen years later, we reported a systematic survey which showed that trials stopped early for benefit—which we will refer to as truncated trials—yield treatment effects that are often not credible (relative risk reductions over 47% in half, over 70% in a quarter), and that the apparent overestimates were larger in smaller trials.[2] We subsequently compared effect estimates from all the truncated trials we could identify that had been included in systematic reviews and meta-analyses with the results of non-truncated trials in those same meta-analyses. We found, on average, substantially larger effects in the truncated trials (ratio of relative risks in truncated versus non-truncated of 0.71). Again, we showed an association with the size of the truncated trial: large overestimates were common when the total number of events was less than 200; smaller but important overestimates occurred with 200 to 500 events; and trials with over 500 events showed small overestimates.[3]

Guidance must include a high level of scepticism regarding the findings of trials stopped early for benefit, particularly when those trials are relatively small and replication is limited or absent

The results of simulation studies and systematic surveys of truncated trials therefore show that when true underlying treatment effects are modest—as is usually the case—small trials that are stopped early with few events will result in large overestimates. Larger trials will still, on average, overestimate effects, and these overestimates may also lead to important spurious inferences. Uncritical belief in truncated trials will often, therefore, be misleading—and sometimes very misleading.

The tendency for truncated trials to overestimate treatment effects is particularly dangerous because their apparently compelling results often prompt publication in prominent journals,[2 3] rapid dissemination in media, and speedy incorporation into practice guidelines and quality assurance initiatives. Below we review three instances in which truncated trials have provided misleading estimates of treatment effect and the response of the clinical community possibly resulted in harm to patients.

β blockers in non-cardiac surgery

In 1999 a clinical trial of bisoprolol in patients with vascular disease having non-cardiac surgery with a planned sample size of 266 stopped early after enrolling 112 patients.[4] Two of 59 patients in the bisoprolol group and 18 of 53 in the control group had experienced a composite endpoint event (cardiac death or myocardial infarction). The authors reported a 91% reduction in relative risk with a 95% confidence interval of 63% to 98%.[4] In 2001 a prominent opinion piece recommended β blockers for all high risk patients having non-cardiac surgery, and in 2002 the first authoritative clinical practice guideline recommended β blockers for such patients.[5 6] In 2001, a US quality assurance initiative identified the practice as an opportunity for improving safety.[7]

Although the enthusiastic reception of the results almost stifled subsequent trials, in 2008 a systematic review and meta-analysis, including over 12 000 patients having non-cardiac surgery, documented a 35% reduction in the odds of non-fatal myocardial infarction (95% CI 21% to 46%), a twofold increase in non-fatal strokes (odds ratio 2.1, 27 to 3.68), and a possible increase in all cause mortality (1.20, 0.95 to 1.51).[8]

Despite the results of the systematic review, subsequent guidelines published in 2009 and 2012 continued to recommend β blockers, sometimes with great enthusiasm.[9] Enthusiasts for β blockers suggest that lower doses and ensuring an early start to treatment can prevent the increase in stroke seen in the overall population. The enthusiasts may be right, but there is limited evidence supporting the claim, and if the results of pooled analyses are correct, recommending β blockers in patients having non-cardiac surgery is continuing to cause disabling strokes.

Another explanation for β blockers continuing to be recommended is cognitive dissonance[10]—after prolonged advocacy, it is painful to acknowledge that the policy may result in a large number of patients having a disabling stroke. Those producing, and profiting from, β blockers may also have encouraged their continued perioperative use.

Intensive insulin therapy in critically ill patients

Cognitive dissonance may also help explain responses to the unfolding story of intensive insulin therapy in critically ill patients. In 2001, a single centre randomised trial of an intensive insulin regimen in critically ill patients with raised serum glucose reported a 42% relative risk reduction in mortality (95% confidence interval 22% to 62%). The authors used a liberal stopping threshold (P=0.01) and frequent looks at the data, strategies they said were "designed to allow early termination of the study."[11] The results were, again, met with enthusiasm and rapidly incorporated into practice guidelines, with recommendations published as early as 2003 for an upper limit of glucose of ≤8.3 mmol/L.[12 13] Numerous protocols for achieving upper limits of ≤8.3 mmol/L were also published.

Fortunately, the investigators' decision to stop early did not stifle subsequent research. A systematic review published in 2008 summarised the results of subsequent studies, which refuted the lower mortality with intensive insulin therapy and established an increased

risk of hypoglycaemia.[14] These findings were confirmed in a later systematic review including additional studies. Nevertheless, several guideline groups continue to advocate limits of ≤8.3 mmol/L. These recommendations contrast with those of guidelines that fully account for the results of more recent studies, which recommend a range of 7.8-10 mmol/L.[15]

Activated protein C

The latest example of the phenomenon of misleading results from truncated trials concerns the use of recombinant human activated protein C (rhAPC) in critically ill patients with sepsis. The original trial, published in 2001, was stopped early after the second interim analysis because of an apparent difference in mortality.[16] In 2004 the Surviving Sepsis Campaign, a global initiative to improve management, recommended use of the drug as part of a "bundle" of interventions in sepsis.[17] A subsequent trial, published in 2005,[18] reinforced initial concerns about increased risk of bleeding with rhAPC and raised serious questions about the apparent mortality reduction in the original study. Nevertheless, the 2008 iteration of the Surviving Sepsis guidelines, mirrored in 2009 by another guideline,[19] continued to recommend rhAPC. After further discouraging results, the drug was withdrawn from the market last year, providing no further opportunity for guideline panels to drag their feet in altering recommendations to reflect the latest evidence.

Stopping the rush to judgment

Simulations show that a systematic review of a series of adequately powered trials with similar stopping rules, some of which stop early for benefit and most of which do not, will not appreciably overestimate treatment effects. These simulations, however, assume that the results of one trial will not influence how, or whether, another trial is undertaken.

This assumption is unlikely to reflect the real world. If a trial that by chance overestimates treatment effects, and therefore stops early for benefit, is one of the first, the correcting trials that would bring a pooled estimate towards the truth may never be conducted. Indeed, if the investigators' judgment to stop the trial early—that it is no longer ethical for patients not to receive the experimental intervention—is sound, the correcting trials should never be undertaken.

Overestimation of effects from stopped early trials is therefore an extreme example of two related phenomena: large effects tend not to be replicated, and results from early randomised trials tend to overestimate effects. Increasingly, methodologists are producing guidance for clinicians and guideline developers on how to interpret clinical trial evidence. Such guidance must include a high level of scepticism regarding the findings of trials stopped early for benefit, particularly when those trials are relatively small and replication is limited or absent.

The stories in this article illustrate the linked phenomena of publication of stopped early trials with dramatic results in high profile journals, their rapid and perhaps uncritical uptake by the media and guideline panels, and the experts' understandable disinclination to reverse previous recommendations in the face of new data. Humans tend to seek confirming evidence for previous beliefs and to devalue new evidence.[20] Discomfort with the possibility of having made widely followed recommendations that did more harm than good is natural. It may be equally natural to persuade yourself of the limitations of disconfirming evidence.

Furthermore, the continued use of drugs—particularly if they are expensive and yield large profits—is in the interest of those who produce them. The drug industry is extremely effective in influencing the behaviour of clinicians, and guideline panellists sometimes also have substantial conflicts of interest. Thus, industry influence may partly explain continued recommendations in the face of contradictory evidence.

One solution to this problem would be for investigators to refrain from stopping their trials early. Indeed, we may ask whether trials should ever stop early for apparent benefit. The justification for stopping early is that it is no

bmj.com
Feature: Leaping to conclusions
(*BMJ* 2008;336:1213)

longer ethical to randomise patients to not receive the experimental treatment. If, after a trial stops early, other investigators launch new trials of the apparently beneficial intervention, the original trial authors' judgment was premature. That was the case in the three examples we have presented here. The standard for persuading the entire clinical community that further investigation is not ethical—that is, the appropriate standard for stopping early—should be extremely stringent, both in terms of the magnitude of the evidence and the plausibility of the result. Such stringent criteria are unlikely to be met before 500 events have accumulated.[3]

While awaiting the uniform application of such a cautious approach, opinion leaders and guideline panels should ensure that when the evidence base is modest and comes largely from truncated trials their recommendations and evidence summaries reflect the uncertainty in the evidence and that the effects are likely to be overestimated. This will decrease the likelihood of further possibly harmful, widely promulgated, and inappropriately persisting recommendations.

Gordon H Guyatt professor, Departments of Clinical Epidemiology and Biostatistics and Medicine, McMaster University, Hamilton, Ontario, Canada L8N 3Z5

Matthias Briel assistant professor, Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Basel, Switzerland

Paul Glasziou professor, Centre for Research in Evidence Based Practice, Bond University, Gold Coast, Australia

Dirk Bassler professor, Center for Pediatric Clinical Studies and Department of Neonatology, University Children's Hospital Tuebingen, Tuebingen, Germany

Victor M Montori professor, Departments of Medicine (Knowledge and Evaluation Research Unit) and Health Sciences Research (Health Care and Policy Research), Center for the Science of Healthcare Delivery, Mayo Clinic, Rochester, Minnesota, USA

Correspondence to: G H Guyatt **guyatt@mcmaster.ca**

Accepted: 13 April 2012

Contributors and sources: The authors are clinician methodologists who, over the past 12 years, have led an international network of researchers conducting empirical work on clinical trials stopped early for benefit. The network's work and discussions within it support the thoughts expressed in this article. GHG and VMM drafted the manuscript; MB, PG, and DB made critical revisions. GHG is guarantor.

Competing interests: None declared.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Control Clin Trials* 1989;10(suppl 4):209-21S.
2 Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005;294:2203-9.
3 Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA* 2010;303:1180-7.
4 Poldermans D, Boersma E, Bax JJ, Thomson IR, van de Ven LL, Blankensteijn JD, et al. The effect of bisoprolol on perioperative mortality and myocardial infarction in high-risk patients undergoing vascular surgery. Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echocardiography Study Group. *N Engl J Med* 1999;341:1789-94.
5 Fleisher LA, Eagle KA. Clinical practice. Lowering cardiac risk in noncardiac surgery. *N Engl J Med* 2001;345:1677-82.
6 Eagle KA, Berger PB, Calkins H, Chaitman BR, Ewy GA, Fleischmann KE, et al. ACC/AHA guideline update for perioperative cardiovascular evaluation for noncardiac surgery—executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to update the 1996 guidelines on perioperative cardiovascular evaluation for noncardiac surgery). *Anesth Analg* 2002;94:1052-64.
7 Shojania K, Duncan B, McDonald K, Wachter R. Health care safer: a critical analysis of patient safety practices: evidence report/technology assessment No 43. Agency for Healthcare Research and Quality, 2001.
8 Bangalore S, Wetterslev J, Pranesh S, Sawhney S, Gluud C, Messerli FH. Perioperative beta blockers in patients having non-cardiac surgery: a meta-analysis. *Lancet* 2008;372:1962-76.
9 Poldermans D, Bax JJ, Boersma E, De Hert S, Eeckhout E, Fowkes G, et al. Guidelines for pre-operative cardiac risk assessment and perioperative cardiac management in non-cardiac surgery. *Eur Heart J* 2009;30:2769-812.
10 Festinger L. Cognitive dissonance. *Sci Am* 1962;207:93-102.
11 Van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, et al. Intensive insulin therapy in the critically ill patients. *N Engl J Med* 2001;345:1359-67.
12 Heyland DK, Dhaliwal R, Drover JW, Gramlich L, Dodek P. Canadian clinical practice guidelines for nutrition support in mechanically ventilated, critically ill adult patients. *J Parenter Enteral Nutr* 2003;27:355-73.
13 Garber AJ, Moghissi ES, Bransome ED Jr, Clark NG, Clement S, Cobin RH, et al. American College of Endocrinology position statement on inpatient diabetes and metabolic control. *Endocr Pract* 2004;10(suppl 2):4-9.
14 Wiener RS, Wiener DC, Larson RJ. Benefits and risks of tight glucose control in critically ill adults: a meta-analysis. *JAMA* 2008;300:933-44.
15 Stapleton R, Heyland D, Wilson K. Glycemic control and intensive insulin therapy in critical illness. In: Basow D, ed. *UpToDate*. Wolters Kluwer, 2012.
16 Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med* 2001;344:699-709.
17 Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, Cohen J, et al. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Crit Care Med* 2004;32:858-73.
18 Abraham E, Laterre PF, Garg R, Levy H, Talwar D, Trzaskoma BL, et al. Drotrecogin alfa (activated) for adults with severe sepsis and a low risk of death. *N Engl J Med* 2005;353:1332-41.
19 Sepsis treatment in the emergency department: an evidence-based review. *Medscape Emergency Medicine* 2009.
20 Nickerson R. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol* 1998;2:175-220.

Cite this as: *BMJ* 2012;344:e3863

BMJ.COM BLOGS **Emma Rourke**

Junior doctors don't put patients' lives at risk


JOHN COLE/SPL

For newly qualified doctors across the country, the jubilation associated with passing finals gives way to the incipient dread of the first day in a new job, and the knowledge that very soon patients' lives will be in their hands.

It's well known that if you want to avoid those taking their first professional steps on the wards you should avoid becoming unwell in the first week of August.

It is not difficult to find newspaper articles decrying junior doctors and the training they receive. If the national broadsheet media are to be believed, there exists a breed of "Incompetent junior doctors putting patients' lives at risk" in what was, just last week, coined the "NHS killing season."

A recent systematic review of the literature from 2004-11 (van der Leeuw et al, *BMC Medicine* 2012;10:65) sought to identify correlations between aspects of residency training and patient outcomes.

The article itself refers to residents, who are foundation and core or specialist trainee equivalents in the United States. The literature searched was international and limited only to English language publications by a lack of availability of such articles in other languages.

Many of the study's findings are unsurprising. For example, evidence suggests that residents become more efficient in practical procedures as their training progresses. Those towards the end of their training tend to score more highly on patient satisfaction and have better patient outcomes than those just starting out, and trainees appear to benefit from having more structured training. Teaching hospitals had better patient outcomes than non-teaching hospitals, particularly in surgical cases.

The conclusions reached should reassure all concerned that the care they receive from doctors still in training is not only safe, but comparable with care provided by their more senior counterparts. This reassurance comes with a few necessary conditions, however. Junior doctors must be provided with the required supervision and appropriate additional time to achieve the same level of outcome as their more experienced seniors.

This message needs to be heard, as the headlines casting their competency in doubt are damaging not only to the doctor-patient relationship, but also to the prospects of these junior doctors in terms of training opportunities and building experience.

Emma Rourke is a *BMJ* Clegg scholar and an intercalating medical student at Newcastle University

Read this blog in full and other blogs at bmj.com/blogs