Putting results of a clinical trial into perspective

Nicholas R Wilcken, Val J Gebski, Rhana Pike and Anthony C Keech

Authors of articles reporting clinical trials should discuss the results of their trial in the context of existing evidence. This approach is encouraged by the CONSORT statement (Box 1).[1] A trial's contribution to the evidence can be pinpointed by seeking answers to questions such as:

- How do the results apply in a broader clinical context and patient population?
- Do they confirm or contradict the existing evidence?
- Do they confirm current clinical practice or suggest that a change may be needed?
- Do they point to the next question that should be studied?

A complete picture of the trial results includes how the findings apply to other patients, how they guide or change practice, how they relate to prior knowledge (possibly by a systematic review of the evidence), and any new questions raised.

1 CONSORT checklist of items to include when reporting a randomised trial[1]

Section and topic	Item no.	Descriptor
Discussion		
Overall evidence	22	General interpretation of the results in the context of current evidence ◆

What do the results mean in real-world clinical practice?

Are the trial participants representative of the patients being considered for treatment?

Trials usually focus on single interventions, but clinical practice is complex, and the features of the clinical practice environment — intercurrent illnesses, use of other drugs, mood, compliance, and access to services — need to be taken into account. This can limit our understanding of how widely the results can be safely applied. For example, trials aiming to prevent stroke using antithrombotic therapies among patients with atrial fibrillation have recruited as few as 20% of eligible patients — often excluding older patients, women and people with previous cerebrovascular disease,[2] leading to uncertainty about the net benefit of such treatment in these groups.

Does the study question relate to the patients being considered for treatment?

It is worth reflecting on the question addressed by the study, as results may be misconstrued. One of the Women's Health Initiative (WHI) group's large studies of hormone replacement therapy (HRT) for postmenopausal women serves as an example.[3] The study was undertaken to determine whether, compared with placebo, long-term combined oestrogen and progestin would reduce cardiovascular events. It was already known that short-term HRT relieves symptoms of menopause, such as hot flushes. It was also known that long-term HRT increases the risk of breast cancer, but it was hoped that this risk might be offset by better cardiovascular outcomes. However, the trial of combined HRT, over an average of 5.2 years (the trial was stopped early), showed an increase in breast cancer *and* cardiovascular events with HRT use.[3]

When the WHI reported the results,[3] widespread confusion followed, because many people interpreted these findings as relating to short-term use of HRT to relieve menopausal symptoms. Many women stopped using short-term *and* long-term HRT.

The WHI trial was well designed and might, of itself, be regarded as sufficient evidence for clinicians to no longer consider HRT to prevent cardiovascular disease. The research question was not "What are the effects of short-term use of HRT in newly menopausal women with significant symptoms?"; the safety of *short-term* HRT for control of menopausal symptoms is still unresolved.[4]

Randomised trial or systematic review?

Individual randomised trials and systematic reviews (including meta-analyses) are similar but not identical means of arriving at health care evidence, and each has its place. They have different strengths within the broad common purpose of establishing the efficacy (or otherwise) of interventions.

Thus, a highly specific clinical question (Is X mg of a therapy better than Y mg after a heart attack?) would probably best be answered by a single appropriately powered study. But even where relatively large trials have been done, treatment effects can be modest, and a meta-analysis of several trial results may be a useful way of showing that there is statistically robust evidence to support a change in clinical practice. For example, before a large-scale review of early breast cancer trials,[5] it was known that postoperative tamoxifen is effective in preventing recurrence of cancer in postmenopausal women, but it was unclear whether this applied to premenopausal women. The review, which included an updated meta-analysis of all available trials, was able to show a clear benefit ($2P < 0.00001$) in the premenopausal group. The same dataset definitively answered a subsidiary question: survival was longer after 5 years of treatment than 2 years of treatment. These two observations changed clinical practice, and thousands of lives were potentially saved. Meta-analyses may further refine conclusions by also considering variation in the treatment dosages.[6,7]

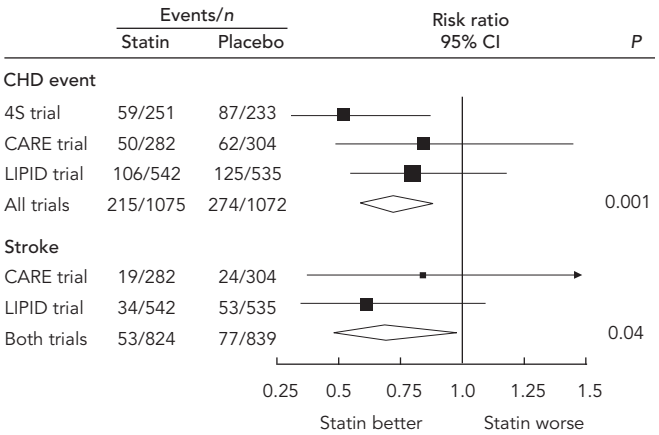
Randomised trial with a meta-analysis

Presenting a meta-analysis as part of the study report helps to show the consistency of the new results with other existing evidence. A forest plot of the results with the published trial evidence will help readers judge the consistency of the results of individual trials and the reliability of the accumulated evidence (Box 2).[8]

Do the results confirm current clinical practice or suggest a change?

Trial results that confirm results of previous studies (particularly in a different patient population) might justify a change in

2 Forest plot of a meta-analysis of a new study (LIPID) with previous studies (4S, CARE) of statin therapy for preventing further cardiovascular events in people with diabetes and existing coronary heart disease[8]



Combining results from these trials leads to a statistically significant finding of a beneficial treatment effect for statins. For stroke, statin therapy has no significant treatment effect in the individual trials; when data for all patients are combined, a reduction in risk with statin therapy is seen. The black squares are proportional to the number of events (here, CHD events or strokes) in the analysis. CHD event = death from coronary heart disease or non-fatal myocardial infarction. 4S = Scandinavian Simvastatin Survival Study (*Arch Intern Med* 1999; 159: 2661-2667). CARE = Cholesterol and Recurrent Events (*Circulation* 1998; 98: 2513-2519). LIPID = Long-Term Intervention with Pravastatin in Ischemic Disease (*N Engl J Med* 1998; 339: 1349-1357). ◆

3 Various trial result scenarios and possible implications for clinical practice

Trial outcome	Population studied	Implications of the results in clinical practice
Benefit	First studied	Is there sufficient evidence to change practice or are confirmatory trials needed? The trial report should propose and discuss these questions.
Confirms previous results	Same as previous population	Change practice (if therapy is cost-effective)
Confirms previous results	New population	Is there enough evidence or are more data needed to extend the treatment indication to the new population?
Apparently contradictory results, compared with previous studies	Same as previous population	Is there an explanation for the result? Is there evidence of significant heterogeneity between the trials?
Does not confirm previous results	New population	Is the difference in effect of treatment between populations biologically plausible? How strong is the evidence from the trials? Are more trials needed before the evidence is enough to support different guidelines in different populations?

clinical practice or even discontinuing further trials of the same clinical question.

For example, lipid-lowering statins given to patients after myocardial infarction are known (from trial evidence) to reduce subsequent cardiovascular events in a general, predominantly male population. A trial substudy testing statins in women also shows possible efficacy.[9] Given what was known previously, the burden of proof may be lower, such that this trial may be sufficient to affect clinical practice.

On the other hand, if the findings of the current study contradict those of previous studies, possible explanations for the discrepancy should be acknowledged and discussed. A change in clinical practice would only be advocated if the results of the new trial significantly outweighed previous evidence in terms of design, adequacy of follow-up, and statistical strength (Box 3).

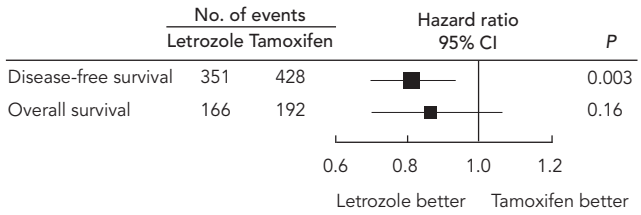
There will always be some legitimate disagreement about the overall significance of individual trial results; one person's "definitive" results will be another's "interesting", "thought-provoking" or "hypothesis-generating" results. Extensive extrapolation may lead to erroneous conclusions; too little or slow extrapolation may retard improvements in clinical practice. However, there are some helpful guidelines, illustrated by the following examples.

Look for biological plausibility: A randomised placebo-controlled trial of folate supplementation during pregnancy confirmed benefits in neonatal outcomes. After decades of follow-up, an unexpected increase in mortality from breast cancer was found. This could be related to the folate therapy, but the result may be a chance finding, and would not be considered biologically plausible at our current level of understanding. In this scenario, women should probably be advised to maintain adequate folate intake around the time of conception, unless further evidence about breast cancer risk emerges.[10]

Interpret the statistical findings for clinical relevance: If multiple outcomes in a single trial are similarly affected by the treatment, some effects may be significant and some non-significant simply because of the differing numbers of events and by the play of chance. For example, a trial in early breast cancer reported that treatment with an aromatase inhibitor resulted in fewer recurrences of the cancer (measured as the disease-free survival period) than treatment with tamoxifen ($P = 0.003$).[11] There was no statistically significant treatment effect for overall survival ($P = 0.16$). At first glance, this might be interpreted as statistical inconsistency across the two outcomes (Box 4). However, the confidence interval for the effect of treatment on mortality extends to plausible worthwhile clinical benefits, so it would be incorrect to rule out a useful survival benefit. Given that that there were relatively few cancer recurrences, a significant survival advantage might still emerge with longer follow-up. A reasonable conclusion would be that "there were fewer deaths in the aromatase inhibitor arm of the trial, but this difference was not beyond the play of chance".

Efficacy is not the only issue: Adverse effects are always important, but take on added significance when efficacy is modest or when we are treating essentially well patients to modify the risk of some future adverse event. Lipid-lowering drugs to prevent heart attacks, blood pressure drugs to prevent strokes, adjuvant chemotherapy to reduce the risk of cancer recurrence, and tamoxifen to prevent a first breast cancer are all examples where large trials have shown modest efficacy.

4 Forest plot showing hazard ratios for disease-free survival (time to recurrence of breast cancer) and overall survival for an aromatase inhibitor, letrozole (4003 patients), compared with tamoxifen treatment (4007 patients)[11]



The median follow-up was about 2 years. The graph shows that disease-free survival was better in the letrozole group. Fewer women died in the letrozole group, but overall survival appears not to differ between the treatment groups. The black squares are proportional to the number of events (here, recurrences and deaths) in the trial. ◆

Drug efficacy is usually more carefully measured and reported than toxicity or quality of life and, perhaps understandably, is more readily trumpeted. Nevertheless, once we have established the biological effect of a drug and how confident we are about this, we should then consider at what cost this was achieved. This might involve formal and complex analyses, such as health economics studies and synthesis of quality-of-life data, but in the shorter term we should also be prepared to make a rough and ready judgement.

Financial cost, extra time spent in hospital or doctors' rooms, extra blood tests and x-rays, as well as adverse effects, all count as costs against which to judge any improvements in clinical outcomes. The patient may also have strong views that need to be taken into account. Cancer patients, when asked, will often accept significant inconvenience and drug toxicity for surprisingly small benefits.[12]

Do the results point to the future or new questions that should be studied?

Randomised trials are the critical units of evidence when the average benefits and harms of interventions are considered. In addition to a statement of the generalisability of the findings,[13] the overall context of the trial results must be incorporated into the discussion. The authors' pragmatic conclusions on the immediate implications for everyday clinical practice (or regulatory decisions), as well as what might constitute the next scientific or clinical question to arise from the current findings, are integral to a report of new evidence.

Competing interests

None identified.

Author details

Nicholas R Wilcken, MB BS, PhD, FRACP, Coordinating Editor, Cochrane Breast Cancer Group
Val J Gebski, BA, MStat, Principal Research Fellow
Rhana Pike, MA, ELS, Publications Officer
Anthony C Keech, MB BS, FRACP, MSc(Epi), Deputy Director
NHMRC Clinical Trials Centre, University of Sydney, Sydney, NSW.
Correspondence: tony@ctc.usyd.edu.au

References

1 Altman DG, Schulz KF, Moher D, et al, for the CONSORT Group. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134: 663-694.
2 Evans A, Kalra L. Are the results of randomized controlled trials on anticoagulation in patients with atrial fibrillation generalizable to clinical practice? *Arch Intern Med* 2001; 161: 1443-1447.
3 WHI Investigators. Risks and benefits of estrogen and progestin in healthy postmenopausal women. *JAMA* 2002; 288: 321-333.
4 HRT: what are women (and their doctors) to do [editorial]? *Lancet* 2004; 346: 2069-2070.
5 Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998; 351: 1451-1467.
6 Gebski V, Lagleva M, Keech A, et al. Survival effects of postmastectomy adjuvant radiation therapy using biologically equivalent doses: a clinical perspective. *J Natl Cancer Inst* 2006; 98: 26-38.
7 Cholesterol Treatment Trialists' (CTT) Collaborators. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins. *Lancet* 2005; 366: 1267-1278.
8 Keech A, Colquhoun D, Best J, et al, for the LIPID Study Group. Secondary prevention of cardiovascular events with long-term pravastatin in patients with diabetes or impaired fasting glucose: results from the LIPID trial. *Diabetes Care* 2003; 26: 2713-2721.
9 Hague W, Forder P, Simes J, et al; LIPID Investigators. Effect of pravastatin on cardiovascular events and mortality in 1516 women with coronary heart disease: results from the Long-Term Intervention with Pravastatin in Ischemic Disease (LIPID) Study. *Am Heart J* 2003; 145: 643-651.
10 Charles DHM, Ness AR, Campbell D, et al. Folic acid supplements in pregnancy and birth outcome: re-analysis of a large randomised controlled trial and update of Cochrane review. *Paediatr Perinat Epidemiol* 2005; 19: 112-124.
11 Breast International Group (BIG) 1-98 Collaborative Group. A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. *N Engl J Med* 2005; 353: 2747-2757.
12 Simes RJ, Coates AS. Patient preferences for adjuvant chemotherapy of early breast cancer: how much benefit is needed? *J Natl Cancer Inst Monogr* 2001; 30: 146-152.
13 Seale JP, Gebski VJ, Keech AC. Generalising the results of trials to clinical practice. *Med J Aust* 2004; 181: 558-560.

(Received 18 Dec 2006, accepted 13 Feb 2007) ❑