Interpreting the results of a clinical trial

Anthony C Keech, Rhana Pike, Renee E Granger and Val J Gebski

In preparing an article reporting a clinical trial, the authors are expected to provide a reasoned interpretation of the results and place them into a broader clinical context. "Interpretation" (Item 20 of the CONSORT statement) refers to how the authors account for their results (Box 1).[1]

Although it has been suggested that authors often have a vested interest in their data and may be biased towards a positive result,[2] they nevertheless have first-hand experience of the design and conduct of the trial, and can offer a unique insight into interpretation of the results.

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

Section and topic	Item no.	Descriptor
Discussion		
Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes ◆

Elements of an interpretation

Authors are generally encouraged to summarise the extent to which the results and findings are consistent with their original hypothesis,[3] and to comment on the robustness of the results for drawing conclusions and making recommendations. This should involve discussion of:

- the suitability of the study design to answer the questions examined;
- the ultimate quality of the trial as conducted;
- the extent to which missing follow-up or imputed data contributed to the reported results; and
- the potential influence of any protocol violations or other biases that may have affected the clinical experiment.[4]

Beyond this, specifically addressing the key aspects of internal validity (such as the fairness of the comparison of treatment groups in the trial) will help the reader assess the results. For example, the reader's confidence in the results is increased by reassurance about the adequacy of randomisation,[5] consistency across treatment arms of the methods used to measure outcomes,[6] and similar levels of background care in the treatment arms.

Additionally, if results are derived from multiple comparisons, the dangers of over-interpretation of a variety of endpoints should be acknowledged.[7,8]

Even if a study has strong internal validity, the results may be surprising or unexpected. Therefore, the plausibility of the results in relation to expectations, and speculation as to possible mechanisms of action of the intervention should be discussed.[7] The sensitivity of the findings to any departures from the assumptions in the design (eg, compliance levels, unblinding, losses to follow-up, and missing data) should be mentioned.[4] Any other limitations or drawbacks of the study design or conduct should be acknowledged, and their influence, or lack of influence, on the outcomes should be argued. This will help the reader to compare these results with other relevant findings.

The strength of the findings (shown by *P* values and confidence intervals around the estimates) signifies their robustness. The size of the estimates of effect and their plausible variability (such as the risk reduction or hazard ratio and confidence intervals) show the potential importance of the intervention in clinical use.

After validity has been discussed, the potential ramifications of the results are usually presented. These include the value of the intervention beyond the trial, including the likely generalisability of the findings,[9] the balance of benefits and harms,[10] and any potential changes to clinical practice that may be appropriate. How consistent the results are with other findings, and how biologically and clinically plausible the interpretation is, will influence this discussion.

The results may raise new questions directing further research. These might arise from the findings for the main outcome, from a subset of patients, or from ancillary analyses. As the authors have an intimate knowledge of the study and the data, their views on the direction of such research may carry weight.

Recommended structure of a Discussion

We recommend an ordered structure for the Discussion section (Box 2).

Statement of the findings

A simple declaration of the meaning of the results should introduce the authors' interpretation. For example, the following sentence introduces the Discussion section of the LIPID study report:[11]

> Our results provide strong evidence that lowering cholesterol levels with pravastatin in patients with a broad range of initial cholesterol levels and a history of myocardial infarction or unstable angina reduces the risk of death from CHD, cardiovascular disease, and all causes combined.

What follows extends this discussion with a brief statement of which patient groups benefited, and the extent and nature of these benefits.

Strengths and weaknesses

The strengths of the trial may include its representative sample, its rigorous design and its clinical relevance.

The account of the weaknesses should aim to explain any flaws in the study identified by the authors, and outline the attempts made to minimise and compensate for these limitations. Identifying weaknesses in the study and discussing their likely influences will help readers appreciate the limitations of interpreting the study results. Discussion of weaknesses should include methodological aspects (eg,

2 Suggested framework for the Discussion section

- A brief statement of the findings;
- Strengths and weaknesses (limitations) of the study, and methods used to minimise and compensate for the limitations;
- Possible mechanisms of action of the intervention, and explanations of these mechanisms;
- Comparison with relevant findings from other published studies;
- Clinical and research implications of the work, as appropriate. ◆

possible biases, the meaning of imprecision in the findings, and the number of multiple comparisons) as well as clinical aspects, such as any problem in translating statistical results to clinical importance.

An example is the study comparing hot water immersion with ice packs to relieve the pain of bluebottle (*Physalia* jellyfish) stings in participants recruited from beach first aid facilities.[12] Hot water immersion for 20 minutes, unlike ice, was highly effective. The Discussion described study weaknesses, such as bias: there was a possibility that, with simultaneous recruitment of family members, treatments could have been allocated after randomisation on the basis of severity.

> We suspect in some cases when two or three patients were simultaneously recruited (often one parent consenting for multiple children), the research assistants may have allocated hot water treatment to the more severe stings once the envelopes were open. However, this was likely to be rare, and a post-hoc analysis using simulations of matched treatment subgroups still showed a highly significant outcome at 20 minutes.

The Discussion also dealt with the subjectivity of pain and the problems of choosing how to measure it.

> The measurement of pain is problematic because it is subjective and is influenced by numerous factors. However, pain is the most important and distressing effect of bluebottle stings, so it was essential that we establish the effect of treatment on pain. The VAS has become a standard tool for the measurement of pain in research, and has been validated in numerous settings.

Mechanisms and explanations

It is important that the authors consider all possible mechanisms underlying the results and explain how they might relate to the outcomes.

For example, in the bluebottle study:[12]

> It might be argued that the hot water immersion may be a symptomatic treatment for jellyfish stings, rather than providing definitive treatment by inactivating venom . . . We demonstrated a time-dependent effect of hot water immersion, with a barely significant effect at 10 minutes and a highly significant effect at 20 minutes. In addition, pain did not recur. This leads us to suggest that the mechanism of reducing pain by heat treatment is inactivation of venom.

However, unlikely or implausible hypothetical mechanisms should not be proposed merely so that they can be disproven.

Relation to other studies

All clinical trials start from a background of previous work. Authors should indicate where results extend, agree with or differ from those of other studies (a forest plot may help readers with interpretation). If there are differences, are these related to differences in methods or the characteristics of participants? An important finding of the Women's Contraceptive and Reproductive Experiences (CARE) Study was new evidence on breast cancer risk:[13]

> In conclusion, high parity and early age at first birth were associated with a reduction in risk only for ER+PR+ tumours. Breastfeeding was associated with a reduction in risk for both ER+PR+ and ER–PR– tumours. Combined with previous research, this suggests that parity and age at first birth act through different mechanisms than breastfeeding. All reproductive factors showed similar associations with both ductal, ductolobular and lobular tumours, suggesting that these tumours have similar aetiologies.

Implications for clinical practice and future research

Readers may not have the authors' background and experience in the research area. Their own interpretations are aided by the authors' commentary, which may include how far the results can be applied in different clinical situations. For example, in the Heart Protection Study:[14]

> As people with blood creatinine concentrations above 200 μmol/L were excluded from the present study, further large trials are required to determine prospectively whether statin therapy can prevent clinically relevant changes in renal function among people at particular risk of developing end-stage kidney disease.

Interpretations, not just interpretation

Most clinical trial reports for publication draw together the expertise and interests of several authors, and the Discussion section is where their views are most likely to diverge. Authors bring different perspectives to interpreting the results.[15] The Discussion needs to reflect a consensus view of all the contributors.

Competing interests

None identified.

Author details

Anthony C Keech, FRACP, MSc(Epi), Deputy Director
Rhana Pike, MA, ELS, Publications Officer
Renee E Granger, BA, BSc(Hons), Medical Writer–Research Assistant
Val J Gebski, BA, MStat, Principal Research Fellow
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 Montori VM, Jaeschke R, Schunemann HJ, et al. Users' guide to detecting misleading claims in clinical research reports. *BMJ* 2004; 329: 1093-1096.
3 Gebski V, Marschner I, Keech AC. Specifying objectives and outcomes for clinical trials. *Med J Aust* 2002; 176: 491-492.
4 Burgess DC, Gebski VJ, Keech AC. Baseline data in clinical trials. *Med J Aust* 2003; 179: 105-107.
5 Beller EM, Gebski V, Keech AC. Randomisation in clinical trials. *Med J Aust* 2002; 177: 565-567.
6 Kirby A, Gebski V, Keech AC. Determining the sample size in a clinical trial. *Med J Aust* 2002; 177: 256-257.
7 Cook DI, Gebski VJ, Keech AC. Subgroup analysis in clinical trials. *Med J Aust* 2004; 180: 289-291.
8 Simes RJ, Gebski VJ, Keech AC. Subgroup analysis: application to individual patient decisions. *Med J Aust* 2004; 180: 467-469.
9 Seale JP, Gebski VJ, Keech AC. Generalising the results of trials to clinical practice. *Med J Aust* 2004; 181: 558-560.
10 Keech AC, Wonders SM, Cook DI, Gebski VJ. Balancing the outcomes: reporting adverse events. *Med J Aust* 2004; 181: 215-218.
11 Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. *N Engl J Med* 1998; 339: 1349-1357.
12 Loten C, Stokes B, Worsley D, et al. A randomised controlled trial of hot water (45° C) immersion versus ice packs for pain relief in bluebottle stings. *Med J Aust* 2006; 184: 329-333.
13 Ursin G, Bernstein L, Lord SJ, et al. Reproductive factors and subtypes of breast cancer defined by hormone receptor and histology. *Br J Cancer* 2005; 93: 364-371.
14 Collins R, Armitage J, Parish S, et al; Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol-lowering with simvastatin in 5963 people with diabetes: a randomised placebo-controlled trial. *Lancet* 2003; 361: 2005-2016.
15 Horton R. The hidden research paper. *JAMA* 2002; 287: 2775-2778.

(Received 15 Dec 2006, accepted 31 Jan 2007) ❑