

RESEARCH METHODOLOGY

Heed the data when results are not significant

Unexpected non-significant results from randomised trials can be difficult to accept.

Catherine Hewitt, Natasha Mitchell, and David Torgerson find that some authors continue to support interventions despite evidence that they might be harmful

When randomised controlled trials show a difference that is not statistically significant there is a risk of interpretive bias.¹ Interpretive bias occurs when authors and readers overemphasise or underemphasise results. For example, authors may claim that the non-significant result is due to lack of power rather than lack of effect, using terms such as borderline significance² or stating that no firm conclusions can be drawn because of the modest sample size.³ In contrast, if the study shows a non-significant effect that opposes the study hypothesis, it may be downplayed by emphasising the results are not statistically significant. We investigated the problem of interpretive bias in a sample of recently published trials with findings that did not support the study hypothesis.

Why interpretive bias occurs

A non-significant difference between two groups in a randomised controlled trial may have several explanations. The observed difference may be real and the study is underpowered or the observed difference may occur simply by chance. Bias can also produce a non-significant difference, but we will not include this in our discussion below.

Trialists are rarely neutral about their research. If they are testing a novel intervention they usually suspect that it is effective otherwise they could not convince themselves, their peers, or research funders that it is worth evaluating. This lack of equipoise, however, can affect the way they interpret negative results. They have often invested a large amount of intellectual capital in developing the treatment under evaluation. Naturally, therefore, it is difficult to accept that it may be ineffective.

A trial with statistically significant negative results should, generally, overwhelm any preconceptions and prejudices of the trialists. However, negative results that are not statistically significant are more likely to be affected by preconceived notions of effectiveness, resulting in interpretive bias. This interpretive bias may lead authors to continue to recommend interventions that should be withdrawn.

Extent of problem

To assess the effect of interpretive bias in trials we hand searched studies published in the *BMJ* during 2002 to 2006 for trials that showed a non-significant difference in the opposite direction of that hypothesised. Two researchers (CEH, NM) identified the papers with a P value above 0.05 and below 0.3 on the primary

Catherine E Hewitt research fellow

Natasha Mitchell research fellow

David J Torgerson director, York Trials Unit, Department of Health Sciences, University of York, York YO10 5DD

Correspondence to: D J Torgerson djt6@york.ac.uk

Accepted: 16 September 2007

outcome, which they agreed between them. Our choice of the limits for the P value was arbitrary and was driven by our decision to identify trials where there was an unexpected difference that could potentially be important and not be statistically significant because of lack of statistical power (type II error).

The decision to use a P value of 0.05 or a 95% confidence interval to determine statistical significance is arbitrary but widely accepted.⁴ Ideally, we should judge the findings of a study not only on its statistical significance but in terms of its relative harms and benefits. Statistical significance is important, however, to guide us in the interpretation of a study's results.

Table 1 | Trials with negative non-significant results published in *BMJ*, 2002-6

Author	Trial	Main outcome	Results (P value)	95%, 67%, and 51% CI*
Henderson et al ⁵	Sex education for 13-15 year olds	Pregnancy termination rate	16% relative increase in terminations, P=0.26	-11 to 42 3 to 29 6 to 26
Ciliberto et al ⁶	Antioxidant supplementation in preventing kwashiorkor	Prevalence of kwashiorkor	Increase, 3.3% v 1.9%, relative risk 1.70, P=0.06	0.98 to 2.42 1.29 to 2.23 1.40 to 2.06
Laurant et al ⁷	Effect of nurse practitioners on general practitioners' workload	No of contacts with general practitioners during surgery hours	Increased by 4.5/week in intervention group; unchanged in control group (z=-1.90, P=0.057)	Unable to calculate
Jespersen et al 2006 ⁸	Short term clarithromycin for patients with stable coronary heart disease	All cause mortality or non-fatal cardiac outcomes	Increase: 15.8% v 13.8%, hazard ratio 1.15, P=0.08	0.99 to 1.34 1.07 to 1.24 1.09 to 1.21
Watson et al ⁹	Safety advice and equipment to reduce unintentional injuries in children under 5 living in deprived areas	Medically attended injuries	Increase: 40.5% v 37.5%, odds ratio 1.14, P=0.09	0.98 to 1.50 1.06 to 1.23 1.08 to 1.20
Dodd et al ¹⁰	Oral misoprostol for induction of labour at term	Vaginal birth not achieved in 24 hours	Increase: 46% v 41.2%, relative risk 1.12, P=0.134	0.95 to 1.32 1.03 to 1.22 1.06 to 1.19
Sanders et al ¹¹	Lidocaine spray to reduce perineal pain during spontaneous vaginal delivery	Pain during delivery on scale of 0 to 100 (100 = worst pain possible)	Increase: 77 v 72, P=0.14	-1.7 to 11.20 1.58 to 8.02 2.51 to 7.09

*67% and 51% confidence intervals were calculated from the data presented in the articles not from the original raw data.

Table 2 | Interpretation of trials with non-significant negative results published in BMJ

Author	Intervention	Effect on main outcome (P value)	Authors' comment	Our view	
				Recommendation	Reason
Henderson et al ⁵	Sex education for 13-15 year olds	Increase in terminations (P=0.26)	High quality sex education should be continued	Sex education as evaluated in this trial should be stopped	Potential harm and more expensive
Ciliberto et al ⁶	Antioxidant supplementation to prevent kwashiorkor	Increase in disease (P=0.06)	Supplementation with these antioxidants was not associated with better growth or less fever, cough, or diarrhoea	Supplements should not be given	Increased risk of harm
Laurant et al ⁷	Adding nurse practitioners to general practice team	Increase in general practitioner consultations (P=0.057)	Nurse practitioners did not reduce general practitioners' workload	Unsure	Increase in workload, which may benefit patients
Jespersen et al ⁸	Short term clarithromycin for patients with stable coronary heart disease	Increase in mortality or cardiac non-fatal outcomes (P=0.08)	Clarithromycin may cause increased mortality	Drug should not be given	Increased harm and cost
Watson et al ⁹	Safety advice and equipment to reduce injuries among children under 5 living in deprived areas	Increase in medically attended injuries (P=0.09)	Advice that includes the offer of free home safety equipment, fitted free of charge, can improve safety practices of families living in deprived areas for up to two years	Intervention should not be given	Increased risk of harm and cost
Dodd et al ¹⁰	Oral misoprostol for induction of labour at term	Increase risk of labour >24 hours (P=0.134)	No significant difference between oral misoprostol and vaginal dinoprostone gel Women preferred oral treatment	Unsure	Increased risk of longer labour but reduced risk of caesarean section
Sanders et al ¹¹	Lidocaine spray to reduce perineal pain during spontaneous vaginal delivery	Increased pain (P=0.14)	Perineal analgesia during second stage labour was acceptable to women and midwives Lidocaine spray had no noticeable effect on perineal pain during spontaneous vaginal delivery	Lidocaine spray should not be used	Increase in pain scores for women receiving intervention

We found 17 papers where there was a difference between the two groups and this difference had a P value of between 0.05 and 0.30. Of these 17 trials, seven (table 1) showed differences in the opposite direction to that specified by the hypothesis.

We calculated three confidence intervals for each identified trial: 95%, 67%, and 51%. We chose 67% as this is half of 95% (that is, the z value for the 67% confidence interval is about half the z value for the 95% interval) and 51% because this range shows where, more often than not, the true treatment estimate will lie. Obviously, each value within the confidence interval is not equally plausible. Values that are close to the point estimate are more likely to correspond to the true value than estimates towards the extreme of the confidence interval.

We used the information in the box in each paper entitled "What this study adds" to determine whether the authors recommended the intervention. We then assessed the data in the paper and used the three confidence intervals to make our recommendation. The authors seem to recommend that the intervention should or could be used in four studies (table 2). We disagreed with this conclusion for three of these studies and were unsure for the other one, as discussed below.

Sex education programme for 13-15 year olds

Twenty five schools in Scotland were randomised to receive either normal sex education or an enhanced package.⁵ The trial was powered to show a 33% reduction in termination rates and had over 99% follow-up after 4.5 years. The intervention schools had an increase of 15.7 terminations per 1000 compared with the control schools (P=0.26). Although the 95% confidence intervals did not exclude an 11% decrease in

terminations, they included a 42% increase in terminations. The 67% confidence intervals did not pass through zero, thus on balance the intervention was more likely to be associated with an increase in terminations than a decrease. The cost of the intervention was up to 45 times greater than usual sex education.

To support use of the intervention the authors refer to an earlier report that "pupils and teachers preferred the SHARE programme . . . It also increased pupils' knowledge of sexual health . . . and had a small but beneficial effect on beliefs about alternatives to sexual intercourse and intentions to resist unwanted sexual activities and to discuss condoms with partners." Although the authors admit that the programme "was not more effective than conventional provision," they do not discuss the possibility that the increase in termination rates might be real and that the programme should be withdrawn until further research supported its implementation. Indeed, the Scottish Executive supports its use in Scottish schools.

Providing free child safety equipment to prevent injuries

A total of 3428 families were randomised to provide 80% power to show a 10% reduction in medically attended injuries.⁹ Free safety equipment was offered to families living in deprived areas along with advice from health visitors. Data on injuries attended in primary care were available for >80% of participants and secondary care >92%. There was an increased risk of having medically attended injuries in the intervention group (P=0.08). The 67% confidence intervals suggested that on balance the most likely value for the true effect is to increase the risk of injuries. The intervention is associated with increased cost and increased risk.

SUMMARY POINTS

Bias can occur when interpreting randomised controlled trials that produce unexpected results that are not statistically significant. Some authors seem to support interventions despite evidence that they might be ineffective. Authors should be careful when they interpret non-significant negative results.

Despite this, the authors seem to use proxy measures of outcome as justification for the intervention: “Our findings in relation to safety practices and degrees of satisfaction are encouraging for safety equipment schemes such as those organised by SureStart.” The authors also note that it was unlikely that intervention would not reduce injury rates because “several observational studies have shown a lower risk of injury among people with a range of safety practices.” Observational studies are potentially biased, which is one of the main reasons we do randomised trials. It is, therefore, surprising to seek reassurance from non-randomised data when a randomised trial shows the “wrong” result. The authors suggest that bias could have been introduced because of differential raised parental awareness, although they acknowledge that the intervention could have increased injury through the process of risk compensation.

Oral misoprostol for induction of labour

In this trial, 741 pregnant women with an indication for prostaglandin induction of labour were randomised to oral misoprostol or vaginal dinoprostone gel.¹⁰ The trial was powered to show a 30% difference in vaginal birth after 24 hours. Follow-up rates were 100% in both groups, allocated treatment adherence was greater than 99%. 46% of women in the oral misoprostol group did not achieve a vaginal birth within 24 hours compared with 41% of the vaginal dinoprostone group. The 95% confidence intervals suggested, at best, the intervention could be associated with 0.95 relative risk improvement.

The authors stated that there was no difference between the two treatments but women preferred oral treatment. However, the 67% confidence interval was significant, suggesting that oral treatment increased the risk of delayed vaginal birth. We could not make a definite recommendation because the risk of caesarean section was reduced for the intervention group (0.82, $P=0.13$), and the 67% confidence interval (0.73 to 0.91) on this outcome favours the intervention.

Lidocaine spray to reduce pain during vaginal delivery

This trial randomised 185 women to receive a topically applied anaesthetic spray or placebo.¹¹ The primary outcome was pain during delivery. Follow-up was 100% at delivery. The pain on delivery was increased by 4.8 points in the intervention group, although the 95% confidence intervals suggested that it could reduce pain by 1.7 points or increase it by 11.2 points. The 67% interval suggested that the true difference was an increase in pain. An adjusted analysis suggested a bigger difference in pain scores. Therefore, this intervention should not be used.

Acting on evidence

Randomised trials are usually considered the best method of establishing effectiveness. All of the trials we identified were well designed and powered to test a two tailed hypothesis, which by implication accepts that the intervention could cause benefit or harm. The results on

proxy measures or from observational studies cannot justify ignoring the main results of the trial.

The use of measures of uncertainty, such as confidence intervals, inform the need for further research not necessarily policy decisions. The Scottish Executive implemented the sex education programme described above on the basis of proxy markers of effect. The main follow-up has been completed. The decision should now be made on a combination of effectiveness and costs. We know the point estimate favours the control group, and we know that on balance when we examine both the 67% and 51% confidence intervals that the likely true estimate of effect is an increase in terminations, and finally we know the costs also favour the control group. The logical interpretation, therefore, of this evidence is to withdraw this programme until further research shows another sex education programme is effective at reducing unwanted pregnancies. A similar argument applies to the accident prevention programme.

Journal editors, readers, and authors need to listen to the data presented in the paper. Sometimes the data speak clearly. Often, however, the data speak more softly and we must be more careful in our interpretation. Journal editors and peer reviewers have an important role in making sure that authors do not make recommendations that are not supported by the data presented.

We thank Nick Freemantle, Cathie Sudlow, and BMJ editors for helpful comments.

Contributors and sources: DJT has published widely on the design of randomised controlled trials in health care and education. This article arose from discussions around interpretation bias within a single study. DJT suggested the idea of the study. CEH and NM identified the studies and extracted the data. All authors interpreted the data. All authors drafted and revised the manuscript critically. All authors give final approval of the version to be published. DJT is guarantor.

Competing interests: None declared.

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

- 1 Kaptchuk TJ. Effect of interpretive bias on research evidence. *BMJ* 2003;326:1453-5.
- 2 Meyer G, Warnke A, Bender R, Muhlhauser I. Effect on hip fractures of increased use of hip protectors in nursing homes: cluster randomised controlled trial. *BMJ* 2003;326:76-80.
- 3 Craig TKJ, Garety P, Power P, Rahaman N, Colbert S, Fornells-Ambrojo M, et al. The Lambeth Early Onset (LEO) Team: randomised controlled trial of the effectiveness of specialised care for early psychosis. *BMJ* 2004;329:1067-71.
- 4 Sterne JA, Davey Smith G. Sifting the evidence—what’s wrong with significance tests? *BMJ* 2001;322:226-31.
- 5 Henderson M, Wight D, Raab GM, Abraham C, Parkes A, Scott S, et al. Impact of a theoretically based sex education programme (SHARE) delivered by teachers on NHS registered conceptions and terminations: final results of cluster randomised trial. *BMJ* 2007;334:1333-8.
- 6 Ciliberto H, Ciliberto M, Briend A, Ashorn P, Bier D, Manary M. Antioxidant supplementation for the prevention of kwashiorkor in Malawian children: randomised, double blind, placebo controlled trial. *BMJ* 2005;330:1109-13.
- 7 Laurant MGH, Hermens RMPG, Braspenning JCC, Sibbald B, Grol RPTM. Impact of nurse practitioners on workload of general practitioners: randomised controlled trial. *BMJ* 2004;328:927-32.
- 8 Jespersen CM, Als-Nielsen B, Damgaard M, Hansen JF, Hansen S, Helø OH, et al. Randomised placebo controlled multicentre trial to assess short term clarithromycin for patients with stable coronary heart disease: CLARICOR trial. *BMJ* 2006;332:22-7.
- 9 Watson M, Kendrick D, Coupland C, Woods A, Futers D, Robinson J. Providing child safety equipment to prevent injuries: randomised controlled trial. *BMJ* 2005;330:178-83.
- 10 Dodd JM, Crowther CA, Robinson JS. Oral misoprostol for induction of labour at term: randomised controlled trial. *BMJ* 2006;332:509-13.
- 11 Sanders J, Campbell R, Peters TJ. Effectiveness and acceptability of lidocaine spray in reducing perineal pain during spontaneous vaginal delivery: randomised controlled trial. *BMJ* 2006;333:117-21.