

- 9 Cox DR. Regression models and life tables. *J R Stat Soc B* 1972;34: 187-220.
- 10 Ismail AA, O'Neill TW, Cockerill W, Finn JD, Cannata JB, Hoszowski K, et al. Validity of self-report of fractures: results from a prospective study in men and women across Europe. *Osteoporos Int* 2000;11:248-54.
- 11 Nevitt MC, Cummings SR, Browner WS, Seeley DG, Cauley JA, Vogt TM, et al. The accuracy of self-report of fractures in elderly women: evidence from a prospective study. *Am J Epidemiol* 1992;135:490-9.
- 12 Gillespie WJ, Henry DA, O'Connell DL, Robertson J. Vitamin D and vitamin D analogues for preventing fractures associated with involutional and post-menopausal osteoporosis. *Cochrane Database Syst Rev* 2002;(3):CD000227.
- 13 Lips P, Graafmans WC, Ooms ME, Bezemer PD, Bouter LM. Vitamin D supplementation and fracture incidence in elderly persons. A randomized, placebo-controlled clinical trial. *Ann Intern Med* 1996;124:400-6.
- 14 Heikinheimo RJ, Inkovaara JA, Harju EJ, Haavisto MV, Kaarela RH, Kataja JM, et al. Annual injection of vitamin D and fractures of aged bones. *Calcif Tissue Int* 1992;51:105-10.
- 15 Scragg R, Jackson R, Holdaway IM, Lim T, Beaglehole R. Myocardial infarction is inversely associated with plasma 25-hydroxyvitamin D3 levels: a community based study. *Int J Epidemiol* 1990;19:559-62.
- 16 Garland CF, Garland FC, Gorham ED. Calcium and vitamin D: their potential roles in colon and breast cancer prevention. *Ann N Y Acad Sci* 1999;889:107-19.
- 17 Chapuy MC, Preziosi P, Maamer M, Arnaud S, Galan P, Hercberg S, et al. Prevalence of vitamin D insufficiency in an adult normal population. *Osteoporos Int* 1997;7:439-43.
- 18 Barrett-Connor E, Gore R, Browner WS, Cummings SR. Prevention of osteoporotic hip fracture: global versus high-risk strategies. *Osteoporos Int* 1998;8:S2-7.
- 19 Rose G. *Strategy of preventive medicine*. Oxford: Oxford University Press, 1992.
- 20 Torgerson D, Donaldson C, Reid D. Using economics to prioritize research: a case study of randomized trials for the prevention of hip fractures due to osteoporosis. *J Health Serv Res Policy* 1996;1:141-6.

(Accepted 10 December 2002)



This is an abridged version; the full version is on bmj.com

Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses

Fujian Song, Douglas G Altman, Anne-Marie Glenny, Jonathan J Deeks

Department of Public Health and Epidemiology, University of Birmingham, Birmingham B15 2TT

Fujian Song
senior research fellow

Centre for Statistics in Medicine, Institute of Health Sciences, Oxford OX3 7LF

Douglas G Altman
professor of statistics in medicine

Jonathan J Deeks
senior medical statistician

Cochrane Oral Health Group, University Dental Hospital of Manchester, Manchester M15 6FH

Anne-Marie Glenny
lecturer in evidence based oral health care

Correspondence to: F Song
f.song@bham.ac.uk

BMJ 2003;326:472-5



Details of methods and a worked example, references for 28 systematic reviews, and three tables are on bmj.com

Abstract

Objective To determine the validity of adjusted indirect comparisons by using data from published meta-analyses of randomised trials.

Design Direct comparison of different interventions in randomised trials and adjusted indirect comparison in which two interventions were compared through their relative effect versus a common comparator. The discrepancy between the direct and adjusted indirect comparison was measured by the difference between the two estimates.

Data sources Database of abstracts of reviews of effectiveness (1994-8), the Cochrane database of systematic reviews, Medline, and references of retrieved articles.

Results 44 published meta-analyses (from 28 systematic reviews) provided sufficient data. In most cases, results of adjusted indirect comparisons were not significantly different from those of direct comparisons. A significant discrepancy ($P < 0.05$) was observed in three of the 44 comparisons between the direct and the adjusted indirect estimates. There was a moderate agreement between the statistical conclusions from the direct and adjusted indirect comparisons (κ 0.51). The direction of discrepancy between the two estimates was inconsistent.

Conclusions Adjusted indirect comparisons usually but not always agree with the results of head to head randomised trials. When there is no or insufficient direct evidence from randomised trials, the adjusted indirect comparison may provide useful or supplementary information on the relative efficacy of competing interventions. The validity of the adjusted indirect comparisons depends on the internal validity and similarity of the included trials.

Introduction

Well designed randomised controlled trials generally provide the most valid evidence of relative efficacy of competing interventions in health care and minimise

the possibility of selection bias.¹ However, many competing interventions have not been compared directly (head to head) in randomised trials. Even when different interventions have been compared directly, such evidence is often limited and insufficient. Because of the lack of direct evidence, indirect comparisons have been recommended² and used for evaluating the efficacy of alternative interventions (Glenny AM, et al, international society of technology assessment in health care, The Hague, 2000). There are concerns that indirect comparisons may be subject to greater bias than direct comparisons and may overestimate the efficacy of interventions.³ Empirical evidence is required to assess the validity of indirect comparisons.

We previously examined the validity of indirect comparisons using examples in a systematic review of antimicrobial prophylaxis in colorectal surgery.⁴ We found some discrepancies between the results of direct and indirect comparisons, depending on which indirect method was used. The results of the study, however, were based on only one topic and the findings may not be generalisable. We therefore used a sample of 44 comparisons of different interventions from 28 systematic reviews to provide stronger evidence about the validity of indirect comparisons.

Methods

To identify relevant meta-analyses of randomised controlled trials, we searched the database of abstracts of reviews of effectiveness (1994-8), the Cochrane database of systematic reviews (Issue 3, 2000), Medline, and references of retrieved articles. Our two inclusion criteria were that competing interventions could be compared both directly and indirectly and that the same trial data had not been used in both the direct and indirect comparison.

Comparison methods

The relative efficacy in each meta-analysis was measured by using mean difference for continuous data and log relative risk for binary data. We use two

comparative methods: the direct (head to head) comparison and the adjusted indirect comparison. See webextra for more details about the statistical methods used and a worked example.

For the direct comparisons, comparison of the result of group B with the result of group C within a randomised controlled trial gave an estimate of the efficacy of intervention B versus C. We used the method suggested by Bucher et al for adjusted indirect comparisons.³

Measures of discrepancy

The discrepancy between the direct estimate (T_{BC}) and the adjusted indirect estimate (T^*_{BC}) was measured by the difference (Δ) between the two estimates:

$$\Delta = T_{BC} - T^*_{BC}$$

We then calculated the standard error and 95% confidence intervals around this difference. In addition, we categorised the results of meta-analyses as non-significant ($P > 0.05$) or significant ($P \leq 0.05$). The significant effect can be further separated according to whether intervention B was less or more effective than intervention C. The degree of agreement in statistical conclusions between the direct and indirect method was assessed by a weighted κ .⁵

Results

We identified 28 systematic reviews in which both the direct and indirect comparison of competing interventions could be conducted, although indirect comparison was not explicitly used in many of these meta-analyses. Some systematic reviews assessed more than two active interventions, and a total of 44 comparisons (see webextra table A) could be made by using data from the 28 systematic reviews.^{w1-w28}

Figure 1 summarises the discrepancies between the direct and the adjusted indirect estimates. There was significant discrepancy ($P < 0.05$) in three of the 44 comparisons—that is, the 95% confidence interval did not include zero. The relative efficacy of an intervention was equally likely to be overestimated or underestimated by the indirect comparison compared with the results of the direct comparison.

There was a moderate agreement in statistical conclusions between the direct and the adjusted indirect method (weighted κ 0.53) (table). In terms of statistical conclusions, 32 of the 44 indirect estimates fell within the same categories as the direct estimates. According to the direct comparisons, 19 of the 44 comparisons suggested a significant difference ($P < 0.05$) between competing interventions. Compared with direct estimates, the adjusted indirect estimates were less likely to suggest a significant difference. Ten of the 19 significant direct estimates became non-significant in the adjusted indirect comparison, while only two of the 25 non-significant direct estimates were significant in the adjusted indirect comparison.

Discussion

The 44 meta-analyses in 28 systematic reviews included in this study covered a wide range of medical topics. The categories of patients included those with an increased risk of vascular occlusion, HIV infection, viral hepatitis C, gastro-oesophageal reflux disease, postoperative pain, heart failure, dyspepsia, and

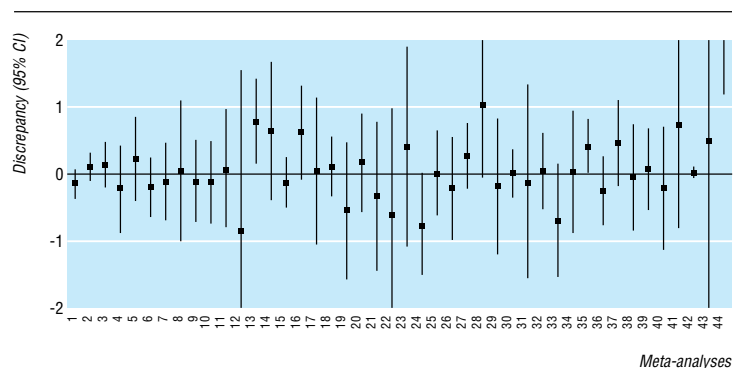


Fig 1 Discrepancy between direct and adjusted indirect comparison defined as difference in estimated log relative risk (meta-analyses 1-39) or difference in estimated standardised mean difference (meta-analysis 40) or difference in estimated mean difference (meta-analyses 41-44): empirical evidence from 44 published meta-analyses (see webextra table A)

cigarette smoking. The results of adjusted indirect comparisons were usually similar to those of direct comparisons. There were a few significant discrepancies between the direct and the indirect estimates, although the direction of discrepancy was unpredictable. These findings are similar to (but more convincing than) those of our previous study of antibiotic prophylaxis in colorectal surgery.⁴

Discrepancies between the direct and the adjusted indirect estimate may be due to random errors. Partly because of the wide confidence interval provided by the adjusted indirect comparison, significant discrepancies between the direct and the adjusted indirect estimate were infrequent (3/44). Results that were significant when we used the direct comparison often became non-significant in the adjusted indirect comparison (table).

The internal validity of trials involved in the adjusted indirect comparison should be examined because biases in trials will inevitably affect the validity of the adjusted indirect comparison. In addition, for the adjusted indirect comparison to be valid, the key assumption is that the relative efficacy of an intervention is consistent in patients across different trials. That is, the estimated relative efficacy should be generalisable. Generalisability (external validity) of trial results is often questionable, however, because of restricted inclusion criteria, exclusion of patients, and differences in the settings where trials were carried out.⁶

Of the 44 comparisons, three showed significant discrepancy ($P < 0.05$) between the direct and the adjusted indirect estimate. In two cases the discrepancies seem to have no clinical importance as both the direct and the adjusted indirect estimates were in the same direction.^{w6 w17} However, the discrepancy between the direct and the adjusted indirect estimate was

Methods of comparison and number of significant findings* in 44 meta-analyses of competing interventions. Weighted κ 0.53 for agreement between direct and adjusted indirect estimate

Direct estimate	Adjusted indirect estimate		
	Significant effect (-) (n=6)	Non-significant effect (n=33)	Significant effect (+) (n=5)
Significant effect (-) (n=8)	5	3	0
Non-significant effect (n=25)	1	23	1
Significant effect (+) (n=11)	0	7	4

*Non-significant effect: difference between intervention groups is non-significant ($P > 0.05$); significant effect ($P \leq 0.05$) is separated according to whether intervention A is less (-) or more effective (+) than intervention B.

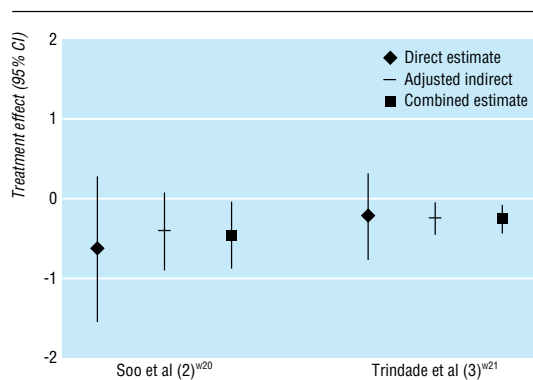


Fig 2 Combination of direct and adjusted indirect estimates in two meta-analyses

clinically important in another case: a comparison between paracetamol plus codeine and paracetamol alone in patients with pain after surgery.^{w28} A close examination of this example showed that the discrepancy could be explained by different doses of paracetamol and codeine used in trials for the indirect comparison. See bmj.com for details.

When is the adjusted indirect comparison useful?

When there is no direct evidence, the adjusted indirect method may be useful to estimate the relative efficacy of competing interventions. Empirical evidence presented here indicates that in most cases results of adjusted indirect comparisons are not significantly different from those of direct comparisons.

When direct evidence is available but inadequate, the adjusted indirect comparison may provide supplementary information.⁷ Sixteen of the 44 direct comparisons in this paper were based on one randomised trial while the adjusted indirect comparisons were based on a median of 19 trials (range 2-86). Such a large amount of data available for adjusted indirect comparisons could usefully strengthen conclusions based on direct comparisons, especially when there are concerns about the methodological quality of a single randomised trial.

Results of the direct and the adjusted indirect comparison could be quantitatively combined to increase statistical power or precision when there is no

important discrepancy between the two estimates. The non-significant effect estimated by the direct comparison may become significant when the direct and the adjusted indirect estimate are combined, as happened in two of the 44 comparisons (fig 2). In each case, the change was because of an increased amount of information. Equally, a significant relative effect estimated by a direct comparison can become non-significant when direct and adjusted indirect estimates are combined.

It will be a matter of judgment whether and how to take account of indirect evidence. It is not desirable to base such decisions on whether or not the difference between the two estimates is significant, although this is the easiest approach. A more constructive approach would be to base the decision on the similarity of the participants in the different trials and the comparability of the interventions.

Some authors have used a naive (unadjusted) indirect comparison, in which results of individual arms between different trials were compared as if they were from a single trial (Glenny AM, et al, international society of technology assessment in health care). Simulation studies and empirical evidence (not shown in this paper) indicate that the naive indirect comparison is liable to bias and produces overprecise estimates (Altman DG, et al, third symposium on systematic reviews: beyond the basics, Oxford, 2000). The naive indirect comparison should be avoided whenever possible.

Direct estimates from randomised trials may not always be reliable

Direct evidence from randomised trials is generally regarded to be the best, but such evidence may sometimes be flawed. Observed discrepancies between the direct and the adjusted indirect comparison may be partly due to deficiencies in the trials making a direct comparison or those contributing to the adjusted indirect comparison, or both.

Conclusions

When there is no direct randomised evidence, the adjusted indirect method may provide useful information about relative efficacy of competing interventions. When direct randomised evidence is available but not sufficient, the direct and the adjusted indirect estimate could be combined to obtain a more precise estimate. The internal validity and similarity of all the trials involved should always be carefully examined to investigate potential causes of discrepancy between the direct and the adjusted indirect estimate. A discrepancy may be due to differences in patients, interventions, and other trial characteristics including the possibility of methodological flaws in some trials.

We thank the secretarial and information support by staff at the NHS Centre for Reviews and Dissemination, University of York. We thank W Y Zhang, Aston University, for sending us the data used in a systematic review. Any errors are the responsibility of the authors.

Contributors: See bmj.com

Funding: This paper is based on a project (96/51/99) funded by the NHS R&D Health Technology Assessment Programme, UK. DGA is supported by Cancer Research UK. The guarantor accepts full responsibility for the conduct of the study, had access to the data, and controlled the decision to publish.

Competing interests: None declared.

What is already known on this topic

Many competing interventions have not been compared in randomised trials

Indirect comparison of competing interventions has been carried out in systematic reviews, often implicitly

Indirect comparison adjusted by a common control can partially take account of prognostic characteristics of patients in different trials

What this study adds

Results of adjusted indirect comparison usually, but not always, agree with those of head to head randomised trials

The validity of adjusted indirect comparisons depends on the internal validity and similarity of the trials involved

- 1 Pocock SJ. *Clinical trials: a practical approach*. New York: John Wiley, 1996.
- 2 McAlister F, Laupacis A, Wells G, Sackett D. Users' guides to the medical literature: XIX. Applying clinical trial results B. Guidelines for determining whether a drug is exerting (more than) a class effect. *JAMA* 1999;282:1371-7.
- 3 Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50:683-91.
- 4 Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. *Contr Clin Trials* 2000;21:488-97.
- 5 Altman DG. *Practical statistics for medical research*. London: Chapman Hall, 1991.
- 6 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.
- 7 Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;15:2733-49.

(Accepted 16 January 2003)

Survey of claims of no effect in abstracts of Cochrane reviews

Phil Alderson, Iain Chalmers

It is never correct to claim that treatments have no effect or that there is no difference in the effects of treatments. It is impossible to prove a negative or that two treatments have the same effect. There will always be some uncertainty surrounding estimates of treatment effects, and a small difference can never be excluded.¹

Claims of no effect or no difference may mean that patients continue to be denied or exposed to interventions with important effects, either beneficial or harmful. They may also suggest that further research is unnecessary, so delaying satisfactory estimates of treatment effects.

The impossibility of proving no effect or no difference should be distinguished from the concept used for equivalence trials, where bounds are set on the differences that are deemed practically important. An analysis of 45 reports of trials purporting to test equivalence found that only a quarter set boundaries on their concept of equivalence.² Given the rationale for avoiding claims of no effect or difference, such claims should be infrequent. We measured their prevalence in abstracts of systematic reviews published in the *Cochrane Database of Systematic Reviews*.

Methods and results

We downloaded the abstract for each of 989 complete reviews in the *Cochrane Database of Systematic Reviews* in issue 1, 2001, and 80 reviews published for the first time in issue 2, 2001. We extracted the sections headed Main Results and Reviewers' Conclusions. One of us (IC) read these sections, looking for claims of no effect. Only those reviews that stated "there was no difference" or "there was no effect" without any qualification about clinical or statistical significance were classified as claiming no effect or difference. The process was repeated for 143 Cochrane abstracts published for the first time in issues 1 and 2 of 2002, except that PA assessed them. Both authors then reviewed all the abstracts of reviews identified as claiming no effect or difference to agree on categorisation.

Claims of no effect or difference were made in 240 (22.5%) abstracts published in the 2001 issues of the *Cochrane Database of Systematic Reviews* and in 19 (13.3%) abstracts first published in 2002. The difference in proportions was -9.2%, 95% confidence interval -2.2% to -14.5%.

Comment

Inappropriate claims of no effect or no difference occurred in about a fifth of abstracts of Cochrane reviews. These claims may have been due to careless wording rather than a mistaken belief that no effect or difference had been shown. It is encouraging that these errors seem to be decreasing.

The decrease may reflect our unconscious use of more lenient evaluations in more recent publications or be due to a higher proportion of significant differences being detected. Alternatively, those reporting and editing Cochrane reviews may have become more aware of the inappropriateness of using no effect or difference because of recent initiatives aimed at improving the quality of these reviews. If so, we hope that there will soon be no Cochrane reviews making claims of no effect or difference and that this will not result in an increase in vague wording. Acceptable phrases include "no significant differences were detected" and "there is insufficient evidence either to support or to refute."

We decided to concentrate on identifying errors in abstracts of Cochrane reviews because they are widely available. We have reported the errors to the relevant editorial teams. Inappropriate wording may have been used in other parts of the review, however, and we urge readers to promote improvements by using the electronic feedback system in the *Cochrane Database of Systematic Reviews*.

We thank Mike Clarke, Paul Garner, Peter Gøtzsche, Philippa Middleton, Jim Neilson, Andy Smith, and Liz Wager for comments on earlier drafts of this manuscript. The views expressed in this article represent those of the authors and are not necessarily the views or the official policy of the Cochrane Collaboration.

Contributors: The study was conceived and conducted by both authors, and both drafted the manuscript. PA is the guarantor.

Funding: The work was carried out as part of the programme of the UK Cochrane Centre, which is funded by the NHS Research and Development Programme. The guarantor accepts full responsibility for the conduct of the study, had access to the data, and controlled the decision to publish.

Competing interests: None declared.

- 1 Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.
- 2 Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: are they supported by the evidence? *Ann Intern Med* 2000;132:715-22.

(Accepted 23 December 2002)

UK Cochrane Centre, NHS Research and Development Programme, Oxford OX2 7LG
Phil Alderson
associate director (training)
Iain Chalmers
former director

Correspondence to:
P Alderson
palderson@cochrane.co.uk

BMJ 2003;326:475