

- 3 Nashan B, Moore R, Amlot P, Schmidt AG, Abeywickrama K, Soullou JP, et al. Randomised trial of basiliximab versus placebo for control of acute cellular rejection in renal allograft recipients. *Lancet* 1997;350:1193-8.
- 4 Eriksson BI, Ekman S, Kalebo P, Zachrisson B, Bach D, Close P, et al. Prevention of deep-vein thrombosis after total hip replacement: direct thrombin inhibition with recombinant hirudin, CGP 39393. *Lancet* 1996; 347:635-9.
- 5 Goldman SC, Holcenberg JS, Finklestein JZ, Hutchinson R, Kreissman S, Johnson FL, et al. A randomized comparison between rasburicase and allopurinol in children with lymphoma or leukemia at high risk for tumor lysis. *Blood* 2001;97:2998-3003.
- 6 IFNB Multiple Sclerosis Study Group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. *Neurology* 1993;43:655-61.
- 7 Garattini S. Are me-too drugs justified? *J Nephrol* 1997;10:283-94.
- 8 Perez EA, Rodeheffer R. Clinical cardiac tolerability of trastuzumab. *J Clin Oncol* 2004;22:322-9.
- 9 Uppenkamp M, Engert A, Diehl V, Bunjes D, Huhn D, Brittinger G. Monoclonal antibody therapy with CAMPATH-1H in patients with relapsed high- and low-grade non-Hodgkin's lymphomas: a multicenter phase I/II study. *Ann Hematol* 2002;81:26-32.
- 10 Winkler U, Jensen M, Manzke O, Schulz H, Diehl V, Engert A. Cytokine-release syndrome in patients with B-cell chronic lymphocytic leukemia and high lymphocyte counts after treatment with an anti-CD20 monoclonal antibody (rituximab, IDEC-C2B8). *Blood* 1999;94:2217-24.
- 11 Olsen NJ, Stein CM. New drugs for rheumatoid arthritis. *N Engl J Med* 2004;350:2167-79.
- 12 Gardam MA, Keystone EC, Menzies R, Manners S, Skamene E, Long R, et al. Anti-tumour necrosis factor agents and tuberculosis risk: mechanisms of action and clinical management. *Lancet Infect Dis* 2003; 3:148-55.
- 13 International Joint Efficacy Comparison of Thrombolytics Investigators. Randomised, double-blind comparison of reteplase double-bolus administration with streptokinase in acute myocardial infarction (INJECT): trial to investigate equivalence. *Lancet* 1995;346:329-36.
- 14 Assessment of the Safety and Efficacy of a New Thrombolytic (ASSENT-2) Investigators. Single-bolus tenecteplase compared with front-loaded alteplase in acute myocardial infarction: the ASSENT-2 double-blind randomised trial. *Lancet* 1999;354:716-22.

(Accepted 29 June 2005)

Simultaneous comparison of multiple treatments: combining direct and indirect evidence

Deborah M Caldwell, A E Ades, J P T Higgins

How can policy makers decide which of five treatments is the best? Standard meta-analysis provides little help but evidence based decisions are possible

Several possible treatments are often available to treat patients with the same condition. Decisions about optimal care, and the clinical practice guidelines that inform these decisions, rely on evidence based evaluation of the different treatment options.^{1,2} Systematic reviews and meta-analyses of randomised controlled trials are the main sources of evidence. However, most systematic reviews focus on pair-wise, direct comparisons of treatments (often with the comparator being a placebo or control group), which can make it difficult to determine the best treatment. In the absence of a collection of large, high quality, randomised trials comparing all eligible treatments (which is invariably the situation), we have to rely on indirect comparisons of multiple treatments. For example, an indirect estimate of the benefit of A over B can be obtained by comparing trials of A v C with trials of B v C,³⁻⁵ even though indirect comparisons produce relatively imprecise estimates.⁶ We describe comparisons of three or more treatments, based on pair-wise or multi-arm comparative studies, as a multiple treatment comparison evidence structure.

The need to combine direct and indirect evidence

Concerns have been expressed over the use of indirect comparisons of treatments.^{4,5} The Cochrane Collaboration's guidance to authors states that indirect comparisons are not randomised, but are "observational studies across trials, and may suffer the biases of observational studies, for example confounding."⁷ Some investigators believe that indirect comparisons may systematically overestimate the effects of treatments.³ When both indirect and direct comparisons are available, it has been recommended that the two approaches be considered separately and that direct



Angioplasty balloon device used to unblock and widen arteries

comparisons should take precedence as a basis for forming conclusions.^{5,7}

Difficulties arise, however, if the direct evidence is inconclusive but the indirect evidence, either alone or in combination with the direct evidence, is not. Furthermore, this approach becomes increasingly impractical as the number of treatments increases. If five treatments have been compared with each other,

Medical Research Council Health Services Research Collaboration, Department of Social Medicine, University of Bristol, Bristol BSS 2PR
Deborah M Caldwell
research associate
A E Ades
professor of public health science

Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge CB2 2SR
J P T Higgins
senior research associate

Correspondence to: D M Caldwell
d.m.caldwell@bristol.ac.uk

BMJ 2005;331:897-900



Further details of the method are on bmj.com

Table 1 Evidence structure for comparison of multiple treatments used in two meta-analyses: number of randomised controlled trials directly comparing seven treatments for acute myocardial infarction. Ps denote the treatments compared

No of trials	Streptokinase	Alteplase-	Accelerated alteplase	Streptokinase +alteplase	Retepase	Tenecteplase	PCTA
Boland et al¹⁵:							
8	P	P					
1	P		P	P			
1	P			P			
1	P				P		
2			P		P		
1			P			P	
Keeley et al¹⁶:							
8	P						P
3		P					P
11			P				P

PCTA = primary percutaneous transluminal coronary angioplasty.

there are 10 possible direct pair-wise comparisons and 70 indirect comparisons. Keeping this information separate makes little sense, particularly when the entire body of evidence could be summarised in terms of four relative treatment effects.

Suitable statistical methods for comparing multiple treatments that fully respect randomisation have been available for some time.⁸⁻¹¹ They have not been widely used, although applications of these methods have begun to appear in US medical journals¹² and in medical decision making.^{13 14} Below, we provide a worked example of these methods and show their advantages over restricting attention to direct pair-wise comparisons. We also examine some concerns about bias and randomisation.

Thrombolysis and angioplasty after myocardial infarction

Two recent overviews of treatment for acute myocardial infarction show the difficulties of relying on standard pair-wise meta-analysis. Boland and colleagues reviewed 14 randomised controlled trials making two or three way comparisons of six thrombolytic treatments (table 1).¹⁵ Their findings are limited to summary statements on each of the pair-wise comparisons for which direct evidence was available (streptokinase is as effective as non-accelerated alteplase, tenecteplase is as effective as accelerated alteplase, reteplase is at least as effective as streptokinase, etc). Presenting results in this way makes it difficult to draw an overall conclusion about which treatment is best on the chosen outcome, or even to form an internally consistent summary of their relative effects.

At the other extreme, Keeley and colleagues looked at 22 randomised controlled trials that compared primary percutaneous transluminal coronary angioplasty with thrombolytic treatment (streptokinase, alteplase, or accelerated alteplase).¹⁶ They found that primary percutaneous transluminal coronary angioplasty was better, but only by "lumping" the three thrombolytic drugs to form a single comparator. This approach was criticised because the relevant comparison is with the best thrombolytic drug not the average one.^{17 18} The full set of evidence from both reviews is available on bmj.com.

To deal with such multiple treatment evidence structures we need a single statistical analysis providing estimates for all the 21 possible pair-wise comparisons

Table 2 Pair-wise odds ratios between seven treatments for acute myocardial infarction obtained by direct and multiple treatment comparisons with fixed effect and random effects analyses*

Treatment comparison	Fixed effect		Random effects	
	Direct comparisons	Multiple comparison	Direct comparisons	Multiple comparison
Streptokinase v:				
Alteplase	1.00 (0.94 to 1.06)	0.99 (0.94 to 1.06)	0.89 (0.54 to 1.14)	0.96 (0.74 to 1.10)
Accelerated alteplase	0.86 (0.78 to 0.94)	0.86 (0.78 to 0.93)		0.84 (0.68 to 0.99)
Streptokinase+alteplase	0.96 (0.87 to 1.05)	0.96 (0.87 to 1.05)		0.97 (0.75 to 1.25)
Retepase	0.95 (0.79 to 1.12)	0.90 (0.80 to 1.01)		0.88 (0.65 to 1.06)
Tenecteplase		0.86 (0.74 to 1.00)		0.85 (0.57 to 1.17)
PCTA	0.52 (0.36 to 0.73)	0.63 (0.52 to 0.77)	0.49 (0.20 to 0.91)	0.62 (0.47 to 0.77)
Alteplase v:				
Accelerated alteplase		0.86 (0.77 to 0.95)		0.88 (0.70 to 1.19)
Streptokinase+alteplase		0.96 (0.86 to 1.07)		1.02 (0.78 to 1.51)
Retepase		0.90 (0.79 to 1.02)		0.92 (0.70 to 1.24)
Tenecteplase		0.86 (0.73 to 1.01)		0.90 (0.61 to 1.35)
PCTA	0.63 (0.25 to 1.29)	0.64 (0.51 to 0.77)		0.65 (0.49 to 0.86)
Accelerated alteplase v:				
Streptokinase+alteplase	1.12 (1.00 to 1.25)	1.12 (1.01 to 1.24)		1.16 (0.91 to 1.55)
Retepase	1.02 (0.90 to 1.16)	1.05 (0.94 to 1.17)		1.04 (0.81 to 1.28)
Tenecteplase	1.01 (0.88 to 1.14)	1.01 (0.89 to 1.14)		1.01 (0.74 to 1.35)
PCTA	0.81 (0.64 to 1.02)	0.74 (0.61 to 0.89)	0.79 (0.55 to 1.05)	0.73 (0.59 to 0.90)
Streptokinase+alteplase v:				
Retepase		0.94 (0.82 to 1.07)		0.92 (0.62 to 1.19)
Tenecteplase		0.90 (0.76 to 1.05)		0.89 (0.57 to 1.27)
PCTA		0.66 (0.53 to 0.81)		0.64 (0.45 to 0.85)
Retepase v:				
Tenecteplase		0.96 (0.82 to 1.13)		0.98 (0.68 to 1.43)
PCTA		0.71 (0.57 to 0.87)		0.71 (0.53 to 0.94)
Tenecteplase v PCTA				
		0.74 (0.58 to 0.92)		0.74 (0.50 to 1.03)

PCTA= primary percutaneous transluminal coronary angioplasty.

* Empty cells represent pair-wise comparisons that have not been evaluated in trials (fixed effect) or for which there are fewer than three trials (random effects).

between seven treatments, and, more importantly, an assessment of which treatment is most likely to have the lowest mortality. The first objective can be achieved by a variety of traditional (frequentist)⁶⁻⁸⁻¹⁰ and bayesian statistical methods,⁸⁻¹¹ all of which are based on a logistic regression model.

Calculation of the probability that each treatment is best on some chosen outcome requires simulation-based methods. We have used an existing bayesian Markov chain Monte Carlo method,⁸ adapted to apply to any connected network set of treatment comparisons. Vague prior distributions are used for comparisons of treatments, so the findings should be close to those obtained with frequentist methods.¹⁹ Further technical details of the method, including computer programs that are applicable to a wide range of evidence structures, are available on bmj.com.

Multiple comparisons versus pair-wise meta-analysis

Table 2 compares results from a single analysis comparing multiple treatments with those from standard pair-wise meta-analyses of direct comparisons. The multiple treatment analysis provides a full set of odds ratios for all the 21 comparisons, effectively combining all the direct and indirect evidence for each comparison. Direct data exist for only 10 pair-wise comparisons, and for four of these comparisons only one randomised controlled trial is available. When direct evidence is available, it agrees with the results obtained by combining all the available evidence, and the multiple treatment analysis tends to produce narrower confidence intervals.

In Table 3 we present estimates of absolute risk of mortality for each treatment, along with the estimated probability that each treatment is best. The results decisively confirm Keeley and colleagues' conclusion that primary percutaneous transluminal coronary angioplasty is the most effective treatment but by using a method that retains the identity of each treatment.¹⁶ Note that, although on the direct evidence primary percutaneous transluminal coronary angioplasty is not significantly better than accelerated alteplase (fixed effect odds ratio = 0.81, 95% confidence interval 0.64 to 1.02),¹⁷ in the multiple treatment comparison analysis it clearly is (0.74, 0.61 to 0.89). The random effects analysis produces an even greater narrowing of the confidence interval. The combination of all available data in this way allows relevant treatments to be compared and leads to more precise conclusions.

This statistical analysis can be embedded in a decision analysis by including data on costs and total life years gained for each treatment. The same methods can then be extended to calculate the probability that each treatment is the most cost effective.¹³⁻¹⁴

Bias, randomisation, and generalisability

What assumptions are being made in this analysis? One way to conceptualise this is to imagine that all 36 trials had examined all seven treatments but that in each trial results for all but two or three treatments had been lost at random. The key assumption for the fixed effect analysis is that the relative effect of one treatment compared with another is the same across the entire set of trials.³⁻⁴ This means that the true odds ratio comparing A

Table 3 Percentage mortality at 35 days and the probability that each treatment is best (lowest mortality) in multiple treatment comparison analysis*

	Fixed effect model		Random effects model	
	35 day Mortality %	Probability best	35 day Mortality %	Probability best
Streptokinase	6.7	0	6.8	0
Alteplase	6.7	0	6.5	0.003
Accelerated alteplase	5.8	0	5.8	0.001
Streptokinase + alteplase	6.5	0	6.6	0.002
Retepase	6.1	0	6.0	0.01
Tenecteplase	5.8	0.004	5.8	0.03
Percutaneous transluminal coronary angioplasty	4.4	0.995	4.3	0.95

*Absolute mortality is based on the average mortality with streptokinase in the 19 randomised controlled trials that included it (see bmj.com for further details).

with B in trials of A v B is exactly the same as the true odds ratio for A v B in the A v C, B v C, and indeed E v F trials, even though A and B were not included in those studies. In a random effects model, where it is assumed that the odds ratios in each trial are different but from a single common distribution, the assumption is that this common distribution is the same across all sets of trials.

These assumptions are remarkably similar to those that underlie a standard pair-wise meta-analysis. The only additional assumption is that the similarity of the relative effects of treatment holds across the entire set of trials, irrespective of which treatments were actually evaluated. Indeed, one way to question this assumption is to imagine all trials had compared the same two treatments and to judge whether they are sufficiently similar to be combined in a meta-analysis.

It may be helpful to consider how the target population relates to the patient groups in the studies. If we are attempting to find the single best treatment with a view to recommending it to all the various patient groups represented in previous trials, a single combined analysis is clearly indicated. If our target is a specific subgroup of patients who were represented in studies around a specific set of treatment comparisons, it may be more appropriate to consider restricting the analysis to a subset of the trials.

The assumptions behind comparisons of multiple treatments are unlikely to be statistically verifiable, and it seems reasonable to rely on expert clinical and epidemiological judgment, both for multiple treatment comparison and for standard pair-wise meta-analysis. This is, after all, the criticism that clinicians often make of meta-analyses—that like is not being compared with like. Multiple treatment meta-analysis is no different from pair-wise meta-analysis in requiring such a judgment. Poor judgments in both cases may induce heterogeneity of effects. Poor judgments in meta-analyses of multiple treatment comparisons may, in common with subgroup analyses and meta-regression of pair-wise comparisons, lead to confounding (if, for example, trials of A v B and A v C are systematically different from trials of B v C). However, bias would not be expected generally to operate in any particular direction.

A further assumption relates to scale of measurement. In common with studies based on indirect comparisons,³⁻⁵⁻²⁰ multiple comparison models are based on the assumption that treatment effects add together so that the relative effect of A v C can be predicted from the effects of A v B and B v C. This assumes that the appropriate measure of effect (log odds ratios, relative risk, or risk difference) has been chosen.²¹

Summary points

Healthcare decisions often involve choosing from a selection of treatment options

Most systematic reviews and meta-analyses focus on pair-wise comparisons, forcing reliance on indirect comparisons

Statistical methods for comparing multiple treatments that combine direct and indirect evidence in a single analysis are available

These methods make the similar assumptions to standard pair-wise meta-analyses but require that they hold over the entire set of trials

Multiple treatment comparisons should be more frequently used to inform healthcare decisions

Both of the pair-wise meta-analyses described earlier were built on the above assumption.^{15 16} In our multiple analysis we made the further assumption that the two sets of studies are homogeneous regarding the relative effects of all the treatments, or (in the random effects analysis) that the relative effects are from the same common distribution. This was generally supported by the similarity between the direct evidence and the estimates obtained by combining direct and indirect evidence. This observation has been made in previous analyses.⁴

Another common feature of our proposed multiple comparison methods and standard pair-wise meta-analyses is a lack of any assumptions about baseline risks across studies. Thus, our analyses are based only on randomised comparisons. Methods that compare patient outcomes in one treatment arm of one trial directly with those in a treatment arm in another trial break randomisation and have been rightly criticised.⁶

Can these methods be used routinely?

The precepts of systematic review enjoin us to put all the available evidence together, to avoid selection bias, increase precision, and examine generalisability.²² A unified, coherent analysis can be achieved only by analysing the entire collection of relevant randomised controlled trials while respecting randomisation. Rather than asking whether analyses comparing multiple treatments should be used routinely, it is more appropriate to ask whether they can be avoided. Methods for comparing multiple treatments will have increasing scope as new treatments proliferate. Most manufacturers' trials, aimed at obtaining licenses for new products, tend to use current or older treatments as comparators. Thus the head to head comparisons of new treatments that are most useful to clinical practice are not available. The need for integrated analyses to inform technology appraisals and clinical guidelines in these circumstances is a recurring theme at the National Institute for Clinical Excellence.¹ Meta-analyses comparing multiple treatments are feasible and should be considered as the bedrock for decisions when several treatments are available.

Common sense is obviously required when applying these methods. Investigators must consider whether it is meaningful to generalise over the entire set of studies.

Particular caution should be exercised in combining contemporary trials with historical trials, since earlier trials may include more severely ill participants. If efficacy depends on baseline risk,²³ a multiple treatment comparison that does not take this into account will be biased, just as a pair-wise meta-analysis would be.

No statistical method is a panacea. Other models making weaker assumptions remain to be explored. More advanced methods that allow for random errors in the generalisability assumption may also prove valuable.^{10 12} This is an area of active research, and we look forward to a period of extensive development and, most importantly, increasing application of these methods.

Contributors and sources: DMC is working on the role of multiple treatment comparisons in decision analysis. AEA is leading a programme on methods for multi-parameter evidence synthesis. JPTH developed the statistical method on which the current analyses are based. DMC wrote the first draft and carried out the statistical analyses. AEA supervised the analyses and is the guarantor. AEA, JPTH, and DMC contributed to the final draft. Competing interests: None declared.

- 1 Rawlins M. In pursuit of quality: the National Institute for Clinical Evidence. *Lancet* 1999;353:1079-82.
- 2 Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Potential benefits, limitations, and harms of clinical guidelines. *BMJ* 1999;318:527-30.
- 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, Altman D, Glenny MA, Deeks J. Validity of indirect comparison for estimating efficacy of competing interventions: evidence from published meta-analyses. *BMJ* 2003;326:472-6.
- 5 Yazdanpanah Y, Sissoko D, Egger M, Mouton Y, Zwahlen M, Chene G. Clinical efficacy of antiretroviral combination therapy based on protease inhibitors or non-nucleoside analogue reverse transcriptase inhibitors: indirect comparison of controlled trials. *BMJ* 2004;328:249-53.
- 6 Glenny AM, Altman DG, Song F, Sakarovich C, Deeks JJ, D'Amico R, et al. Indirect comparisons of competing interventions. *Health Technol Assess* 2005;9(26):1-148.
- 7 Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions* 4.2.4. In: *Cochrane library*. Issue 2. Chichester: John Wiley, 2005.
- 8 Higgins JPT, Whitehead J. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;15:2733-49.
- 9 Hasselblad V. Meta-analysis of multi-treatment studies. *Med Decis Making* 1998;18:37-43.
- 10 Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;21:2313-24.
- 11 Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;23:3105-24.
- 12 Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, et al. Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis. *JAMA* 2003;289:2534-44.
- 13 Wilby J, Kainth A, McDaid D, McIntosh H, Golder S, O'Meara S, et al. A rapid and systematic review of the clinical effectiveness, tolerability and cost effectiveness of newer drugs for epilepsy in adults: a systematic review and economic evaluation. *Health Technol Assess* 2005;9(15):1-157.
- 14 Bridle C, Palmer S, Bagnall AM, Darba J, Duffy S, Sculpher M, et al. A rapid and systematic review of the clinical and cost-effectiveness of newer drugs for treatment of mania associated with bipolar affective disorder. *Health Technol Assess* 2004;8(19):1-187.
- 15 Boland A, Dunder Y, Bagust A, Haycox A, Hill R, Mujica Mota R, et al. Early thrombolysis for the treatment of acute myocardial infarction: a systematic review and economic evaluation. *Health Technol Assess* 2003;7(15):1-136.
- 16 Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. *Lancet* 2003;361:13-20.
- 17 Auer J, Melandri G, Keeley EC. Primary angioplasty or thrombolysis for acute myocardial infarction [letter]? *Lancet* 2003;361:965-8.
- 18 Fresco C, French J, Buchan I, Keeley EC. Primary angioplasty or thrombolysis for acute myocardial infarction [letter]? *Lancet* 2003;361:1303-5.
- 19 Spiegelhalter DJ, Myles JP, Jones DR, Abrams K. An introduction to Bayesian methods in health technology assessment. *BMJ* 1999;319:508-12.
- 20 Bucher HC, Griffith L, Guyatt GH, Opravil M. Meta-analysis of prophylactic treatments against *Pneumocystis carinii* pneumonia and toxoplasma encephalitis in HIV-infected patients. *J AIDS Hum Retrovir* 1997;15:104-14.
- 21 Deeks JJ. Issues on the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002;21:1575-600.
- 22 Mulrow CD. Rationale for systematic reviews. In: Chalmers I, Altman DG, eds. *Systematic reviews*. London: BMJ Publishing, 1995:1-8.
- 23 Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Stat Med* 1997;16:2741-58. (Accepted 29 June 2005)