# RESEARCH

# Multiplicity of data in trial reports and the reliability of meta-analyses: empirical study

OPEN ACCESS

Britta Tendal *research fellow*[1], Eveline Nüesch *research fellow*[2][3], Julian P T Higgins *senior statistician*[4], Peter Jüni *head of division*[2][3], Peter C Gøtzsche *professor and director*[1]

[1]Nordic Cochrane Centre, Rigshospitalet, Blegdamsvej 9, DK-2100, Copenhagen Ø, Denmark; [2]Institute of Social and Preventive Medicine, University of Bern, Switzerland; [3]CTU Bern, Bern University Hospital, Switzerland; [4]MRC Biostatistics Unit, Cambridge, UK

## Abstract

**Objectives** To examine the extent of multiplicity of data in trial reports and to assess the impact of multiplicity on meta-analysis results.

**Design** Empirical study on a cohort of Cochrane systematic reviews.

**Data sources** All Cochrane systematic reviews published from issue 3 in 2006 to issue 2 in 2007 that presented a result as a standardised mean difference (SMD). We retrieved trial reports contributing to the first SMD result in each review, and downloaded review protocols. We used these SMDs to identify a specific outcome for each meta-analysis from its protocol.

**Review methods** Reviews were eligible if SMD results were based on two to ten randomised trials and if protocols described the outcome. We excluded reviews if they only presented results of subgroup analyses. Based on review protocols and index outcomes, two observers independently extracted the data necessary to calculate SMDs from the original trial reports for any intervention group, time point, or outcome measure compatible with the protocol. From the extracted data, we used Monte Carlo simulations to calculate all possible SMDs for every meta-analysis.

**Results** We identified 19 eligible meta-analyses (including 83 trials). Published review protocols often lacked information about which data to choose. Twenty-four (29%) trials reported data for multiple intervention groups, 30 (36%) reported data for multiple time points, and 29 (35%) reported the index outcome measured on multiple scales. In 18 meta-analyses, we found multiplicity of data in at least one trial report; the median difference between the smallest and largest SMD results within a meta-analysis was 0.40 standard deviation units (range 0.04 to 0.91).

**Conclusions** Multiplicity of data can affect the findings of systematic reviews and meta-analyses. To reduce the risk of bias, reviews and meta-analyses should comply with prespecified protocols that clearly identify time points, intervention groups, and scales of interest.

## Introduction

Meta-analyses of randomised clinical trials are crucial for making evidence based decisions. However, trial reports often present the same data in multiple forms when reporting different intervention groups, time points, and outcome measures.[1] Although this multiplicity has always been a challenge in meta-analyses, its potential as a source of bias has received little attention.

The choice of the outcome of interest to include in systematic reviews is generally based on clinical judgment. However, since a fundamentally similar outcome might be measured on different scales, standardisation to a common scale is therefore required before the outcome can be combined in the meta-analysis. This standardisation is typically achieved by calculating the standardised mean difference (SMD) for each trial, which is the difference in means between the two groups, divided by the pooled standard deviation of the measurements.[2] By this transformation, the outcome becomes dimensionless and the scales are comparable, because the results are expressed in standard deviation units. For example, a meta-analysis addressing pain might include trials measuring pain on a visual analogue scale and trials using a five point numerical rating scale. Combining these outcomes on different scales potentially adds a layer of multiplicity, because the outcome of interest might be measured on more than one scale not only across trials but also within the same trial. Multiplicity of data in trial reports might lead to biased decisions about which data to include in meta-analyses and hence threaten the validity of their results. In this study, we empirically assessed whether selecting between multiple time points, scales, and treatment groups affected SMD results in a randomly selected sample of Cochrane reviews.

Correspondence to: B Tendal bt@cochrane.dk

## Methods

### Data source and selection

We included all Cochrane systematic reviews published in the Cochrane Library over 1 year (between issue 3 in 2006 and issue 2 in 2007) that presented a result as an SMD. For every review, we retrieved reports of all randomised trials that contributed to the first SMD result, and downloaded the latest protocols for all reviews in June 2007. Reviews were eligible if the SMD result was based on two to ten randomised trials and if the review protocol described the outcome. We excluded reviews if they only presented results of subgroup analyses.

We defined the index SMD result as the first pooled SMD result presented in the abstract or in the main body of text of the review that was not based on a subgroup analysis. We used index SMD results to identify a specific outcome for each meta-analysis from its protocol. To ensure that the review authors had not received additional outcome data from the authors of relevant trials, we only considered the first SMD result that was based exclusively on published data.

Based on the published protocol of each review, two observers (BT, EN) independently extracted all data from the original trial reports that could be used to calculate the SMD for the outcome that met our inclusion criteria. From each trial report, we extracted data for all experimental or control groups, time points, and measurement scales, provided that they were compatible with the definitions in the review protocol. If any required data were unavailable, we made approximations as previously described.[3] We did not include interim analyses. Disagreements were resolved by discussion. We did not contact trial authors for unpublished data. Selection of reviews and trials and the extraction of data from trial reports were prespecified (protocol available on request).

### Review methods

We used Monte Carlo simulations to determine the variation in meta-analysis results from different SMD estimates, calculated from multiple time points, intervention groups, and measurement scales. We also used this simulation to estimate the overall impact of multiplicity. During each simulation, we randomly sampled one SMD and the corresponding standard error for each component trial in a specific meta-analysis. We used sampling with replacement from the population of all possible SMDs caused by multiplicity, and selected one SMD per trial. We then used this randomly sampled SMD and the corresponding standard error for fixed or random effects meta-analysis (as originally done in the published reviews), and calculated a pooled SMD for each meta-analysis. We repeated this process 10 000 times—that is, we undertook each meta-analysis 10 000 times, with a random selection of one SMD per trial each time. We then examined the distribution of pooled SMDs in histograms.

To estimate the impact of a single source of multiplicity (intervention groups, time points, measurement scales), we allowed only one source of multiplicity to vary at a time when randomly sampling SMDs for each trial. We standardised the other sources of multiplicity at prespecified standard values (group: pooled groups, time point: post treatment values, scale: first scale mentioned in text). For example, the analysis of multiplicity from different scales was based on post treatment values and pooled groups (if there were several possible groups). We would then randomly sample the values of the different scales for this time point and these groups to calculate the pooled SMD results. We expressed the variability of SMD results due to multiplicity as the difference between the smallest and largest pooled SMD results obtained from the Monte Carlo simulations. Only meta-analyses including trials with multiplicity contributed to these analyses. Finally, we compared the median pooled SMD from the Monte Carlo simulations to the index SMD that was published in the Cochrane review using a paired Wilcoxon test.

## Results

Figure 1⇓ shows the flowchart for the selection of meta-analyses. The 19 eligible meta-analyses included 83 trials that contributed to our study.[4-22] Table 1⇓ shows the characteristics of included reviews, which addressed various condition types: psychiatric (eight reviews), musculoskeletal (two), neurological (two), gynaecological (one), hepatological (one), respiratory (one), and other (four). We studied psychological interventions in 10 meta-analyses, pharmacological interventions in four, physical interventions in three, and other interventions in two (exercise and humidified air). The index outcomes analysed in the 19 meta-analyses were diverse: pain in three, another symptom in 13, and other outcomes in three.

### Information in review protocols

Table 2⇓ shows the level of information given in the review protocols. The protocols did not contain any information about which scales should be preferred. Eight protocols gave information about which time point or period to select, but only one gave enough information to avoid multiplicity, because the time point relevant for the selected index outcome was post-treatment, meaning that the data were collected by the end of treatment. A typical statement, which allowed for a potentially biased choice regarding the selection of a time point, was: "All outcomes were reported for the short term (up to 12 weeks), medium term (13 to 26 weeks), and long term (more than 26 weeks)."[7] Another review about humidified air for treating croup[15] stated: "The outcomes will be separately recorded for the week following treatment." The selected outcome in this particular review was croup symptom score and none of the three included trials ran for this length of time, but reported symptoms 20 min to 12 hours after the intervention. Eighteen protocols described which type of control group to select but none reported any hierarchy among similar control groups or any intention to combine such groups.

Furukawa and colleagues provided an example of a protocol with many possible intervention or control groups.[8] The authors aimed to compare combined psychotherapy and pharmacotherapy with psychotherapy or pharmacotherapy alone. They defined psychotherapy broadly, as "any other psychological approach."[8] The pooled index SMD was based on seven trials, three of which had more than one possible intervention group.[23-25] For three trials[26-28] with only one intervention group, each contained three groups that could be used as control groups: one receiving pharmacotherapy only, one receiving psychotherapy only, and one receiving psychotherapy plus placebo.

### Observed multiplicity in trial reports

Table 3⇓ presents the extent of multiplicity observed in the eligible reviews. Table 4⇓ gives an example of multiple eligible measurement scales, showing the different scales possible in the meta-analysis by Hunot and colleagues.[10]

## Observed multiplicity in meta-analyses

In 11 (58%) meta-analyses, we identified at least one trial that provided data for more than one intervention or control group. Thirteen (68%) meta-analyses included at least one trial that reported more than one eligible time point and 12 (63%) included at least one trial that reported the index outcome using more than one eligible measurement scale. We identified one meta-analysis without multiplicity, because all three included trials only reported data of one intervention and control group, one eligible time point, and one measurement scale for the index outcome.[18]

## Effects of multiplicity on meta-analysis results

Figure 2⇓ shows the distributions of possible pooled SMDs in each meta-analysis, after we randomly selected one possible SMD result per trial. Any type of multiplicity of data in the included trials affected pooled SMD results in 17 (89%) of 19 meta-analyses. The remaining two meta-analyses were not affected, because one study did not have multiple data in the trial reports[18] and the observed multiplicity in another had no effect on the pooled SMD results.[7] In one study, the Monte Carlo distributions do not include the published SMD, because the review authors used changes instead of end of follow-up values to calculate the SMD.[18]

In all 11 (58%) meta-analyses including at least one trial with more than one experimental or control group, we found variability in the pooled SMD results due to this type of multiplicity. In 12 (63%) meta-analyses, we found variability in the pooled SMD results due to multiplicity of data regarding time points (figure 2). In one meta-analysis with two trials that reported more than one eligible time point, we did not find multiplicity due to these different time points.[7] In ten (53%) meta-analyses, we found variability in pooled SMD results from trial data of multiple measurement scales used for the index outcome. In two meta-analyses, one trial in each meta-analysis reported data for more than one measurement scale for the index outcome, but this multiplicity did not affect the pooled SMD results.[6] [22] In 12 (63%) reviews, the published pooled SMDs were more favourable for the experimental intervention than the median pooled SMD from the simulations (P=0.49).

Table 5⇓ presents the variability of pooled SMD results according to different sources of multiplicity (that is, groups, time points, or scales). Eighteen meta-analyses included trials with multiple data for at least one source. In these 18 meta-analyses, the treatment effect from multiplicity of data varied greatly (median difference between the smallest and largest SMDs within the same meta-analysis, 0.40 standard deviation units, range 0.04 to 0.91).

## Discussion

In 18 of the 19 meta-analyses in our study, we found multiplicity of data in at least one trial report within each meta-analysis, which frequently resulted in substantial variation in the pooled SMD results. The impact of multiple data in trial reports regarding intervention groups, time points, or measurement scales on meta-analysis results varied greatly across meta-analyses, ranging from almost no effect (0.04 standard deviation units) to a substantial one (0.91 standard deviation units, corresponding to a large treatment effect),[29] with a median difference of 0.40 standard deviation units. We also estimated the effect of the individual sources of multiplicity, holding the other sources constant.

## Example of potential implications of multiplicity of data

Table 6⇓ provides an example of data from trials investigating the effects of pharmacotherapy on anxiety levels. Depending on which time point is examined, the effect of pharmacotherapy varied widely from week to week within the individual trials. When we randomly selected one time point for each trial, SMDs varied from −0.76 (indicating a large benefit) to 0.05 (indicating little effect). For example, in the Fineberg 2005 trial, there was a large difference in the treatment effect from weeks 8 to 16.

If a meta-analysis were to pick only the most favourable trial results, its result would be biased and overly optimistic, which might affect clinical judgment about whether to use a particular treatment. Therefore, if the protocol does not state any prespecified time points for the meta-analyses, the meta-analysts might make data driven decisions based on the trial results as a whole. In the example in table 6, one could argue two strategies: to include the latest time point from all trials, or to use the length of the shortest trials and extract time points from the other trials that match this time point best. Another solution would be to include all time points in one analysis, similar to an analysis of repetitive measures in an individual trial (see below).

## Strengths and limitations of study

Our selection of Cochrane reviews in this study was random, and the variability of the SMD results did not seem related to particular types of interventions or outcomes. To estimate the impact of multiplicity on meta-analysis results, we randomly selected one SMD per trial from a pool of eligible SMDs with equal probability and used these to calculate pooled SMDs for each meta-analysis. However, in practice, implicit rules regarding data extraction might apply within specialties. For example, one scale might be more commonly used than others, for example, Hamilton's depression scale. Such implicit hierarchy of scales would be expected to reduce the multiplicity, but should be made explicit in protocols for systematic reviews.

Our results are transparent because we only included published results. Therefore, we probably underestimated the true level of multiplicity, since selective reporting of outcomes in trials is common.[30-33] Positive, significant results are more likely to be published than non-significant results.[34] Alternatively, if our random selection of SMDs for the meta-analyses did not reflect how review authors typically select in practice, we might have overestimated the observed effects of multiplicity.

Our study was possible because authors of Cochrane reviews must publish their protocols before they undertake and publish the review. We believe that most non-Cochrane meta-analyses do not have available protocols,[35] and therefore the scope for multiplicity is probably greater than in Cochrane reviews. Although we examined three common sources of multiplicity of data in trial reports, there are other types of multiple data in trial reports—for example, different types of analysis such as intention to treat and per protocol analyses. Review authors might also be influenced by how many and which outcomes to select according to how favourable results appear to be in the published trial reports. The effect of the selection of scales, time points, and control groups has not been systematically assessed in any of the published Cochrane reviews.

The extent of multiplicity of data identified in trial reports is a function of the information provided in the review protocols: we would expect a poorly specified outcome to increase multiplicity. Data extraction for a meta-analysis depends on the information given by trial reports, and therefore it cannot be

fully specified in advance without knowledge of the included trials. However, to minimise data driven selection of time points, measurement scales, or intervention groups, researchers should specify these decisions at the protocol stage. If amendments to the protocol are indicated, they should be reported transparently.[36] [37]

## Comparison with other studies

To our knowledge, our study is the first to show empirically the extent to which multiplicity of data can compromise the reliability of meta-analysis results. We have previously reported results from an observer agreement study of ten meta-analyses included in the present study.[2] We found that disagreements in observers were common and often large, mainly because of: the different choices of groups, time points, scales, and calculations; different decisions on the inclusion or exclusion of particular trials; and data extraction errors.[2] Bender and colleagues describe the problem of multiple comparisons in systematic reviews.[1] They identified common reasons for multiplicity in reviews, but did not estimate the impact on the meta-analysis results.[1] In our study, we included meta-analyses of SMDs, which could be associated with multiplicity of data because of the use of different measurement scales in included trials. However, multiplicity of data due to selection of time points and groups is not unique to SMD results and could apply to other effect measures, such as binary outcomes.

## Possible approaches to minimise bias due to multiplicity of data

One approach to dealing with multiplicity in systematic reviews is to extract, analyse, and report all data available for intervention groups, time points, and measurement scales. However, this method could lead to considerable problems at interpretation, in view of the potential discrepancies between different scales or time points. As with the repetitive measures in an individual trial, all available time points reported in included trials could be analysed in a single meta-analysis while fully accounting for the correlation of repetitive measurements within a trial.[38] Alternatively, assessments from different scales measuring similar concepts could be analysed in a single multivariate model, similar to the use of bivariate models in diagnostic research.[39] Although the first approach of including repetitive assessments in a single analysis could be easily understandable, many readers could find the second approach difficult to understand.

Another approach could be to provide detailed protocols for systematic reviews with clearly specified time points, scales, and groups. Protocols should also include explicit and transparent hierarchies of each source of data, or strategies to combine sources (for example, if there are several control groups). Clinical judgment will be important here. Ideally, the choice of time points and scales should be evidence based, but empirical evidence for the most interesting time points and a hierarchy of scales according to their validity and responsiveness are rarely available. In addition, it is difficult to foresee everything at the protocol stage, and the scope, methodological quality, and quality of reporting of included studies might require subsequent modifications.[40] Only Cochrane reviews are formally required to have a published protocol; however, and only about 10% of non-Cochrane reviews explicitly state a formal protocol.[37] Protocol amendments could affect the results and conclusions of systematic reviews and should be made only after careful consideration and be reported transparently.[36] [37] Furthermore, the reporting of the methods and the results in

meta-analyses must clearly explain how the results were achieved and how any multiplicity of data was handled.[36]

## Conclusions

Multiplicity of data in trial reports and review protocols lacking a detailed specification of eligible time points, scales, and treatment groups can lead to substantial variability in meta-analysis results. Authors of systematic reviews should anticipate and consider the multiplicity of data in trial reports when writing protocols. To enhance reliability of meta-analyses, protocols should clearly define time points to be extracted, provide a hierarchy of scales, clearly define eligible treatment and control groups, and present strategies for handling multiplicity of data.

1   Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, et al. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008;61:857-65.
2   Tendal B, Higgins JP, Juni P, Hrobjartsson A, Trelle S, Nuesch E, et al. Disagreements in meta-analyses using outcomes measured on continuous or rating scales: observer agreement study. *BMJ* 2009;339:b3128.
3   Reichenbach S, Sterchi R, Scherer M, Trelle S, Burgi E, Burgi U, et al. Meta-analysis: chondroitin for osteoarthritis of the knee or hip. *Ann Intern Med* 2007;146:580-90.
4   Abbass AA, Hancock JT, Henderson J, Kisely S. Short-term psychodynamic psychotherapies for common mental disorders. *Cochrane Database Syst Rev* 2006;4:CD004687.
5   Afolabi BB, Lesi FE, Merah NA. Regional versus general anaesthesia for caesarean section. *Cochrane Database Syst Rev* 2006;4:CD004350.
6   Ahmad G, Duffy J, Watson AJ. Pain relief in hysterosalpingography. *Cochrane Database Syst Rev* 2007;2:CD006106.
7   Buckley LA, Pettit T, Adams CE. Supportive therapy for schizophrenia. *Cochrane Database Syst Rev* 2007;3:CD004716.
8   Furukawa TA, Watanabe N, Churchill R. Combined psychotherapy plus antidepressants for panic disorder with or without agoraphobia. *Cochrane Database Syst Rev* 2007;1:CD004364.
9   Gava I, Barbui C, Aguglia E, Carlino D, Churchill R, De Vanna M, et al. Psychological treatments versus treatment as usual for obsessive compulsive disorder (OCD). *Cochrane Database Syst Rev* 2007;2:CD005333.
10  Hunot V, Churchill R, Silva de Lima M, Teixeira V. Psychological therapies for generalised anxiety disorder. *Cochrane Database Syst Rev* 2007;1:CD001848.
11  Ipser JC, Carey P, Dhansay Y, Fakier N, Seedat S, Stein DJ. Pharmacotherapy augmentation strategies in treatment-resistant anxiety disorders. *Cochrane Database Syst Rev* 2006;4:CD005473.
12  Larun L, Nordheim LV, Ekeland E, Hagen KB, Heian F. Exercise in prevention and treatment of anxiety and depression among children and young people. *Cochrane Database Syst Rev* 2006;3:CD004691.
13  Martinez Devesa P, Waddell A, Perera R, Theodoulou M. Cognitive behavioural therapy for tinnitus. *Cochrane Database Syst Rev* 2007;1:CD005233.
14  Mistiaen P, Poot E. Telephone follow-up, initiated by a hospital-based health professional, for postdischarge problems in patients discharged from hospital to home. *Cochrane Database Syst Rev* 2006;4:CD004510.
15  Moore M, Little P. Humidified air inhalation for treating croup. *Cochrane Database Syst Rev* 2006;3:CD002870.

**What is already known on this topic**

Considerable observer variation exists in data extraction, which can be attributed to different choices and errors in the data extraction

The extent to which multiplicity of data in trial reports can compromise the reliability of meta-analysis results is unknown

**What this study adds**

Multiplicity of data in trial reports is substantial and has an important effect on meta-analyses results

16  Mytton J, DiGuiseppi C, Gough D, Taylor R, Logan S. School-based secondary prevention programmes for preventing violence. *Cochrane Database Syst Rev* 2006;3:CD004606.

17  O'Kearney RT, Anstey KJ, von Sanden C. Behavioural and cognitive behavioural therapy for obsessive compulsive disorder in children and adolescents. *Cochrane Database Syst Rev* 2006;4:CD004856.

18  Orlando R, Azzalini L, Orando S, Lirussi F. Bile acids for non-alcoholic fatty liver disease and/or steatohepatitis. *Cochrane Database Syst Rev* 2007;1:CD005160.

19  Trinh KV, Graham N, Gross AR, Goldsmith CH, Wang E, Cameron ID, et al. Acupuncture for neck disorders. *Cochrane Database Syst Rev* 2006;3:CD004870.

20  Uman LS, Chambers CT, McGrath PJ, Kisely S. Psychological interventions for needle-related procedural pain and distress in children and adolescents. *Cochrane Database Syst Rev* 2006;4:CD005179.

21  Woodford H, Price C. EMG biofeedback for the recovery of motor function after stroke. *Cochrane Database Syst Rev* 2007;2:CD004585.

22  Yousefi-Nooraie R, Schonstein E, Heidari K, Rashidian A, Akbari-Kamrani M, Irani S, et al. Low level laser therapy for nonspecific low-back pain. *Cochrane Database Syst Rev* 2007;2:CD005107.

23  Fahy TJ, O'Rourke D, Brophy J, Schazmann W, Sciascia S. The Galway study of panic disorder. I: Clomipramine and lofepramine in DSM III-R panic disorder: a placebo controlled trial. *J Affect Disord* 1992;25:63-75.

24  Mavissakalian M, Michelson L. Two-year follow-up of exposure and imipramine treatment of agoraphobia. *Am J Psychiatry* 1986;143:1106-12.

25  Zitrin CM, Klein DF, Woerner MG, Ross DC. Treatment of phobias. I. Comparison of imipramine hydrochloride and placebo. *Arch Gen Psychiatry* 1983;40:125-38.

26  Barlow DH, Gorman JM, Shear MK, Woods SW. Cognitive-behavioral therapy, imipramine, or their combination for panic disorder: a randomized controlled trial. *JAMA* 2000;283:2529-36.

27  De Beurs E, van Balkom AJ, Lange A, Koele P, van Dyck R. Treatment of panic disorder with agoraphobia: comparison of uvoxamine, placebo, and psychological panic management combined with exposure and of exposure in vivo alone. *Am J Psychiatry* 1995;152:683-91.

28  Sharp DM, Power KG. Fluvoxamine, placebo and cognitive behaviour therapy used alone and in combination in the treatment of panic disorder and agoraphobia. *J Anxiety Dis* 1996;10:219-42.

29  Cohen J. *Statistical power analysis for the behavioral sciences* . 2nd ed. Lawrence Earlbaum, 1988.

30  Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330:753.

31  Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.

32  Chan AW, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.

33  Vedula SS, Bero L, Scherer RW, Dickersin K. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *N Engl J Med* 2009;361:1963-71.

34  Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009;1:MR000006.

35  Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4:e78.

36  Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.

37  Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.

38  Wandel S, Juni P, Tendal B, Nuesch E, Villiger P, Welton N, et al. Effects of glucosamine, chondroitin or placebo for osteoarthritis of the hip or knee: network meta-analysis. *BMJ* 2010;341:c4675.

39  Harbord RM, Whiting P, Sterne JA, Egger M, Deeks JJ, Shang A, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J Clin Epidemiol* 2008;61:1095-103.

40  Juni P, Egger M. PRISMAtic reporting of systematic reviews and meta-analyses. *Lancet* 2009;374:1221-3.

Cite this as: *BMJ* 2011;343:d4829

# Tables

**Table 1| Characteristics of included systematic reviews**

| Author | Outcome | Condition | Intervention | Group |
|---|---|---|---|---|
| Mytton et al[16] | School responses | Aggression/violence | Violence prevention programme | Cochrane Injuries Group |
| Afolabi et al[5] | Neonatal neurological and adaptive score | Caesarean section | Epidural | Cochrane Pregnancy and Childbirth Group |
| O'Kearney et al[17] | Depression | Obsessive compulsive disorder | Behavioural therapy or cognitive behavioural therapy | Cochrane Depression, Anxiety, and Neurosis Group |
| Buckley and Pettit[7] | General functioning score | Schizophrenia | Supportive therapy | Cochrane Schizophrenia Group |
| Abbass et al[4] | Anxiety/depression | Common mental disorders | Psychotherapy | Cochrane Depression, Anxiety, and Neurosis Group |
| Orlando et al[18] | Radiological response | Non-alcoholic fatty liver disease | Bile acids | Cochrane Hepato-Biliary Group |
| Mistaen and Poot[14] | Patient disease knowledge or symptom management | Post discharge problem | Telephone follow-up | Cochrane Consumers and Communication Group |
| Moore and Little[15] | Symptom score | Croup | Humidified air | Cochrane Acute Respiratory Infections Group |
| Yousefi-Nooraie et al[22] | Low back related disability | Low back pain | Low level laser treatment | Cochrane Back Group |
| Trinh et al[19] | Pain | Neck disorder | Acupuncture | Cochrane Back Group |
| Martinez Devesa et al[13] | Subjective tinnitus loudness | Tinnitus | Cognitive behavioural therapy | Cochrane Ear, Nose, and Throat Disorders Group |
| Ahmad et al[6] | Pain | Hysterosalpingography (tubal patency) | Analgesic | Cochrane Menstrual Disorders and Subfertility Group |
| Woodford and Price[21] | Range of movement | Stroke | EMG biofeedback | Cochrane Stroke Group |
| Larun et al[12] | Anxiety | Anxiety | Exercise | Cochrane Depression, Anxiety, and Neurosis Group |
| Gava et al[9] | Symptom level | Obsessive compulsive disorder | Psychological treatment | Cochrane Depression, Anxiety, and Neurosis Group |
| Furukawa et al[8] | Global judgment | Panic disorders | Combined treatment (psychotherapy and antidepressant) | Cochrane Depression, Anxiety, and Neurosis Group |
| Ipser et al[11] | Symptom severity scale | Treatment-resistant anxiety disorders | Pharmacotherapeutic augmentation | Cochrane Depression, Anxiety, and Neurosis Group |
| Uman et al[20] | Pain | Needle-related procedural pain and distress | Psychological interventions | Cochrane Pain, Palliative and Supportive Care Group |
| Hunot et al[10] | Worry/fear symptoms | Generalised anxiety disorder | Psychological therapies | Cochrane Depression, Anxiety, and Neurosis Group |

EMG=electromyography.

Table 2| **Content of the review protocols**

| | Eligible intervention groups | Eligible control groups | Hierarchy of control groups | Eligible time points | Hierarchy of time points | Eligible measuring methods or scales | Hierarchy of measuring methods or scales |
|---|---|---|---|---|---|---|---|
| Mytton et al[16] | Yes | Yes | – | Yes | – | Yes | – |
| Afolabi et al[5] | Yes | Yes | Yes* | – | – | Yes | – |
| O'Kearney et al[17] | Yes | Yes | – | – | – | Yes | – |
| Buckley and Pettit[7] | Yes | Yes | – | Yes | – | Yes | – |
| Abbass et al[4] | Yes | Yes | – | Yes | – | Yes | – |
| Orlando et al[18] | Yes | Yes | – | – | – | Yes | – |
| Mistaen and Poot[14] | Yes | Yes | – | Yes | – | Yes | – |
| Moore and Little[15] | Yes | Yes | Yes* | Yes | – | Yes | – |
| Yousefi-Nooraie et al[22] | Yes | Yes | – | – | – | Yes | – |
| Trinh et al[19] | Yes | Yes | – | – | – | Yes | – |
| Martinez Devesa et al[13] | Yes | Yes | – | – | – | Yes | – |
| Ahmad et al[6] | Yes | Yes | – | Yes | – | Yes | – |
| Woodford and Price[21] | Yes | – | – | – | – | Yes | – |
| Larun et al[12] | Yes | Yes | – | – | – | Yes | – |
| Gava et al[9] | Yes | Yes | – | – | – | Yes | – |
| Furukawa et al[8] | Yes | Yes | – | Yes | – | Yes | – |
| Ipser et al[11] | Yes | Yes | – | – | – | Yes | – |
| Uman et al[20] | Yes | Yes | – | – | – | Yes | – |
| Hunot et al[10] | Yes | Yes | – | Yes | Yes | Yes | – |

*Only one possible control group stated.

**Table 3| Observed multiplicity of data from trials in meta-analyses**

| | | No (%) of trials with multiplicity of data | | | |
|---|---|---|---|---|---|
| | No of trials included | Any of the three sources | Intervention groups | Time points | Measurement scales |
| Mytton et al[16] | 2 | 1 (50) | 1 (50) | 0 | 1 (50) |
| Afolabi et al[5] | 2 | 1 (50) | 0 | 1 (50) | 0 |
| O'Kearney et al[17] | 2 | 1 (50) | 1 (50 | 0 | 0 |
| Buckley and Pettit[7] | 2 | 2 (100) | 0 | 2 (100) | 0 |
| Abbass et al[4] | 2 | 2 (100) | 0 | 1 (50) | 2 (100) |
| Orlando et al[18] | 3 | 0 | 0 | 0 | 0 |
| Mistaen and Poot[14] | 3 | 1 (33) | 1 (33) | 0 | 0 |
| Moore and Little[15] | 3 | 2 (67) | 0 | 2 (67) | 0 |
| Yousefi-Nooraie et al[22] | 3 | 2 (67) | 0 | 1 (33) | 1 (33) |
| Trinh et al[19] | 3 | 3 (100) | 0 | 3 (100) | 1 (33) |
| Martinez Devesa et al[13] | 4 | 4 (100) | 3 (75) | 3 (75) | 1 (25) |
| Ahmad et al.[6] | 5 | 3 (60) | 1 (20) | 2 (40) | 1 (20) |
| Woodford and Price[21] | 5 | 4 (80) | 2 (40) | 2 (40) | 3 (60) |
| Larun et al[12]* | 5 | 4 (80) | 1 (20) | 3 (60) | 2 (40) |
| Gava et al[9] | 7 | 6 (86) | 4 (57) | 3 (43) | 5 (71) |
| Furukawa et al[8] | 7 | 6 (86) | 6 (86) | 2 (29) | 4 (57) |
| Ipser et al[11] | 7 | 6 (86) | 0 (0) | 5 (71) | 3 (43) |
| Uman et al[20] | 9 | 2 (22) | 2 (22) | 0 | 0 |
| Hunot et al[10] | 9 | 5 (56) | 2 (22) | 0 | 5 (56) |
| All reviews | 83 | 55 (66) | 24 (29) | 30 (36) | 29 (35) |

*One trial was excluded because of lack of data in the trial reports.

**Table 4| Possible measurement scales to include in the meta-analysis by Hunot and colleagues10**

| Trial | Scale |
|---|---|
| Akkerman 2001 | Penn state worry questionnaire |
| | Worry scale |
| Barlow 1992 | Worry scale |
| | Percentage of an average day during past month that patient reported worrying |
| | Fear questionnaire |
| Dugas 2003 | Penn state worry questionnaire |
| Ladouceur 2000 | Penn state worry questionnaire |
| | Worry and anxiety questionnaire (only six GAD somatic symptom items included) |
| Mohlman 2003a | Worry composite: standardised combined scores on Penn state worry questionnaire and Spielberger state-trait anxiety inventory scales |
| Mohlman 2003b | Anxiety and worry composite: standardised combined scores on Beck anxiety inventory, symptom checklist anxiety, and Penn state worry questionnaire scales |
| Stanley 2003 | Penn state worry questionnaire |
| Wetherell 2003 | Percentage of an average day during past month that patient reported worrying (question in ADIS-IV (DiNarelo 1994)) |
| | Penn state worry questionnaire |
| Woodward 1980 | Fear thermometer (assessment of overall subjective anxiety on 20 cm unmarked scale which patients marked to show anxiety in the past week overall) |
| | Fear survey schedule intensity subscale |
| | Fear survey schedule severity subscale |

GAD=generalised anxiety disorder.

**Table 5| Variability in meta-analyses results**

| Source of multiplicity | No (%) of meta-analyses with multiplicity of data (n=19) | SMD range within the same meta-analysis (median, range) |
|---|---|---|
| Intervention groups | 11 (58) | 0.09 (0.01 to 0.43) |
| Time points | 13 (68) | 0.19 (0.03 to 0.82) |
| Measurement scales | 12 (63) | 0.23 (0.01 to 0.45) |
| Any source | 18 (95) | 0.40 (0.04 to 0.91) |

**Table 6| Potential implications of multiple time points on treatment effect in a meta-analysis11**

| Timepoint (weeks) | Atamaca 2002 | Carey 2005 | Denys 2004 | Erzegovesi 2004 | Fineberg 2005 | Hollander 2003 | McDougle 2000 |
|---|---|---|---|---|---|---|---|
| 1 | – | – | – | – | – | – | 0.44 (−0.25 to 1.14) |
| 2 | −0.03 (−0.79 to 0.73) | −0.27 (−0.89 to 0.34) | – | – | – | – | 0.10 (−0.59 to 0.78) |
| 3 | – | – | – | – | – | – | −0.30 (−0.99 to 0.39) |
| 4 | −0.40 (−1.16 to 0.37) | −0.15 (−0.76 to 0.47) | – | – | −0.17 (−1.03 to 0.69) | – | −0.02 (−0.70 to 0.67) |
| 5 | – | – | – | – | – | – | −0.66 (−1.36 to 0.05) |
| 6 | −0.90 (−1.69 to −0.10) | −0.27 (−0.89 to 0.34) | – | – | – | – | −0.92 (−1.65 to −0.20) |
| 8 | −2.12 (−3.08 to −1.17) | – | −0.82 (−1.47 to −0.18) | – | 0.12 (−0.74 to 0.98) | −0.61 (−1.65 to 0.42) | – |
| 12 | – | – | – | – | −0.15 (−1.01 to 0.71) | – | – |
| 15 | – | – | – | −0.52 (−1.41 to 0.38) | – | – | – |
| 16 | – | – | – | – | −0.27 (−1.13 to 0.59) | – | – |
| 18 | – | – | – | −1.64 (−2.66 to −0.61) | – | – | – |

Data are standardised mean difference (95% CI).

# Figures



```
┌────────────────────────────────────────────────────┐
│     Potentially eligible reviews (n=32)            │
└────────────────────────────────────────────────────┘
        │
        │   ┌────────────────────────────────────────────────────┐
        │   │ Excluded reviews (n=13):                           │
        │   │   No SMD result based exclusively on randomised    │
        │   │    trials (n=6)                                    │
        ├──▶│ No pooled SMD result (n=2)                         │
        │   │ Only SMD results based on unpublished data (n=2)   │
        │   │ Only subgroup result (n=1)                         │
        │   │ No available review protocol (n=1)                 │
        │   │ Outcome not available in protocol (n=1)            │
        │   └────────────────────────────────────────────────────┘
        ▼
┌────────────────────────────────────────────────────┐
│            Eligible reviews (n=19)                 │
└────────────────────────────────────────────────────┘
                    │
                    ▼
┌────────────────────────────────────────────────────┐
│  1 meta-analysis selected from every review (n=19) │
└────────────────────────────────────────────────────┘
```
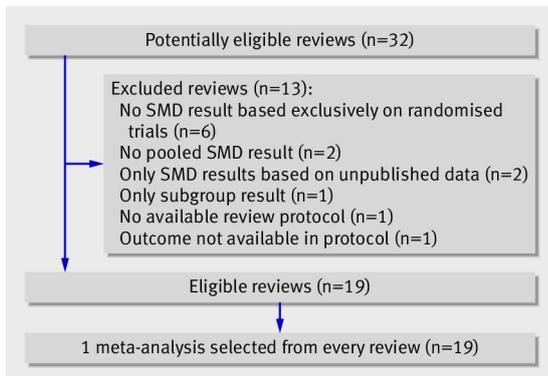
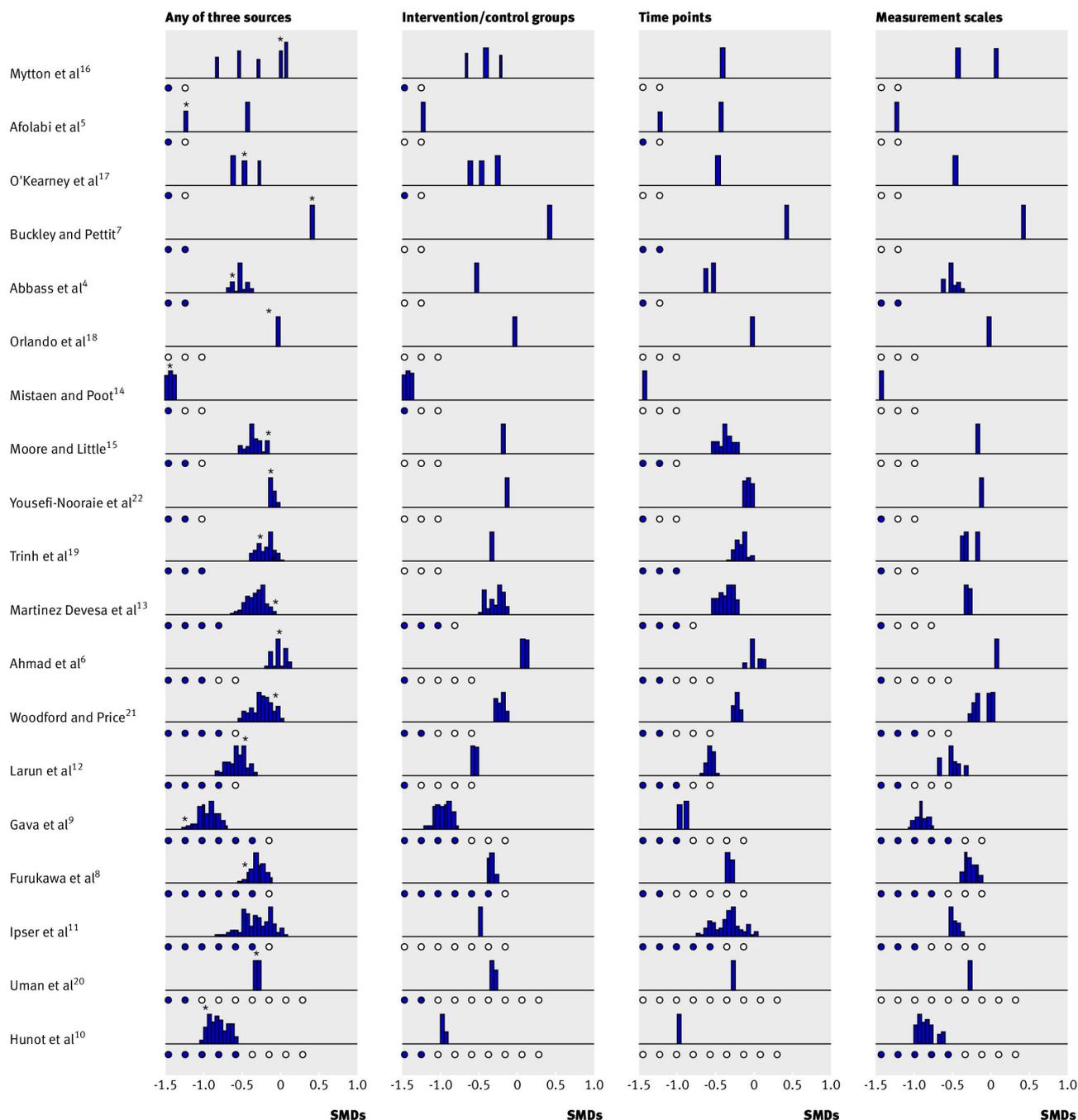**Fig 1** Flowchart for selection of meta-analyses

**Fig 2** Monte Carlo distributions of possible pooled SMDs in each meta-analysis. Dots=number of trials included. Open dots=trials without multiplicity of data. Filled dots=trials with multiplicity of data. Stars=published pooled SMDs. Meta-analyses are ordered according to the number of trials included. Negative SMDs on y axis indicate experimental intervention has more beneficial effect than control intervention