Analysis of quality of interventions in systematic reviews

Robert D Herbert, Kari Bø

The quality of interventions can affect the results of clinical trials. Reviews of complex interventions need to take this into account

Complex health interventions, such as surgery or physiotherapy, can be administered well or badly. Variation in the quality of administration of interventions may explain some of the variability in estimates of effects between trials in systematic reviews. We argue that systematic reviews of complex interventions should assess the quality of interventions, and we suggest how to make such assessments.

Best evidence

Randomised controlled trials provide the best test of the efficacy of preventive or therapeutic interventions because they can separate the effects of the intervention from those of extraneous factors such as natural recovery and statistical regression. When more than one trial has examined a particular intervention, systematic reviews potentially provide the best summaries of the available evidence.¹

Systematic reviewers can summarise findings of randomised trials using an impressionistic approach (qualitative synthesis) or they can produce quantitative syntheses by statistically combining the results from several studies (meta-analysis). Regardless of the method used, most systematic reviewers seek to reduce data from clinical trials into simple statements about treatment. Systematic reviews that provide succinct statements about the effects of an intervention are particularly useful to clinicians. But simple summaries are possible only when the studies address similar questions in similar ways.

Heterogeneity

We can differentiate several types of heterogeneity between trials. Clinical heterogeneity can be identified before analysis of data. It may be due to variations across trials in the inclusion and exclusion criteria, the way in which the intervention is administered or the outcomes that are measured. Statistical heterogeneity occurs when estimates of effects of interventions vary substantially across trials. Unlike clinical heterogeneity, statistical heterogeneity becomes apparent only after estimates of effects have been obtained. Sometimes clinical heterogeneity seems to produce statistical heterogeneity: trials with different types of participants, interventions, or outcomes show different effects. When statistical heterogeneity exists,

Effect of training pelvic floor muscle on urinary incontinence in pregnancy. Pooled estimate for all trials with sufficient data (top) and excluding a trial without supervised exercise (bottom)

review summaries become more complicated because the conclusions are uncertain or conditional.

Effect of quality

A potential source of clinical heterogeneity is variation between trials in the way in which the intervention is delivered. This problem is least likely in reviews of simple interventions, particularly some drugs. Dose finding trials are required before drugs can be used in clinical trials, so the optimal dose is usually known and administration is consistent across trials. In contrast, complex, multifaceted interventions are likely to be administered in different ways in different settings. Some examples of complex interventions relevant to our discipline (physiotherapy) are back schools,² programmes to prevent falls,³ and functional restoration programmes for injured workers.⁴ Other complex interventions include education programmes⁵ and most surgical procedures.⁶ These interventions may be administered in quite different ways across trials, and it is reasonable to expect that the way in which they are administered could influence their effectiveness. Heterogeneity of effects may occur because, all else being equal, well planned, intensive, competently administered interventions are likely to be more effective than low intensity interventions that are poorly planned and administered.

Exercise, a component of many physiotherapy interventions, provides a case in point. Exercise

Centre for Evidence-Based Physiotherapy, University of Sydney, PO Box 170, Lidcombe, NSW 1825, Australia

Robert D Herbert
associate professor

Norwegian School of Sport Sciences, Department of Sport Medicine, Oslo, Norway

Kari Bø
professor

Correspondence to:
R Herbert
R.Herbert@fhs.usyd.edu.au

BMJ 2005;331:507-9

science provides us with some insights into how the design of an exercise programme influences physiological responses to exercise. A dose-response relation exists, as with drugs. Physiological responses to exercise are determined by the mode, frequency, intensity, and duration of exercise.⁷ In practice, the dose is also determined by adherence to the exercise programme. The effects of exercise programmes observed in clinical trials are likely to vary because trials use different training doses and inspire different levels of adherence.⁸

An example of statistical heterogeneity that could be attributed to trials administering the interventions differently comes from studies of the effects of training the pelvic floor muscle. We identified four randomised trials of the effects of pelvic floor training to prevent urinary incontinence during pregnancy.^{9–12} Three presented enough data to permit meta-analysis.^{10–12} The studies were heterogeneous with respect to intervention. Two showed significant and clinically important effects of antenatal training,^{11 12} whereas one study reported non-significant and clinically trivial effects.¹⁰ In the two trials with positive effects, training was supervised regularly by a physiotherapist, whereas in the study with negative effects women saw the physiotherapist only once.

The pooled estimate of effect obtained from a meta-analysis of all three trials did not show an effect of pelvic floor training on risk of urinary incontinence (odds ratio 0.67, 95% confidence interval 0.39 to 1.16; figure). When we excluded the large trial of a low intensity intervention¹⁰ from the meta-analysis, we found a clinically worthwhile effect of antenatal training (0.50, 0.34 to 0.75). The largest trial may have reported a smaller effect because of its size. Resource limitations often mean that large trials provide less intensive interventions, and in large trials it may be logistically difficult to provide well supervised interventions. Yet large trials are most heavily weighted in meta-analyses. If large studies with less intense interventions show smaller effects, they will tend to dilute the effects of smaller studies that show larger effects. An uncritical synthesis of these data suggests that the intervention is ineffective, but a more accurate interpretation may be that the intervention is effective only if administered intensively.

The differences in conclusions reached by analyses that do and do not consider the quality of intervention are likely to be clinically important. In our opinion, clinicians, providers of health care and patients should be more impressed by systematic reviews that explicitly consider the quality of interventions, particularly when the interventions are complex. Systematic reviews of complex interventions should routinely examine the quality of interventions.

Assessing quality

How should the quality of interventions be assessed, when should reviewers suspect that statistical heterogeneity is due to the quality of interventions, and how should trials with poor quality interventions be dealt with in systematic reviews? These questions go to the heart of some of the most difficult methodological issues in systematic reviews.

Formal assessment of the quality of the intervention in systematic reviews is potentially problematic. It will always be difficult to assess the quality of the intervention before it is known if the intervention is effective. However, the findings of other research can sometimes guide considerations of quality. When other research strongly suggests an intervention should be administered in a particular way, that research can be used as a benchmark for assessing the quality of intervention. For example, studies of the dose-responsiveness of strength training clearly indicate that strength training programmes produce the greatest increases in muscle strength when the training load is high.⁷ So systematic reviews of interventions designed to increase muscle strength should assess whether the training load was adequate.

Assessment of the quality of the intervention relies on sufficient detail in trial reports, but many reports provide only superficial descriptions of complex interventions. The CONSORT statement on reporting of clinical trials recommends that reports of clinical trials include “precise details of the interventions intended for each group and how and when they were actually administered.”¹³ Interventions should be described in sufficient detail to enable readers to assess if the intervention was administered well.

In some reviews of complex interventions it may be possible to formally incorporate considerations of the quality of the interventions into the analysis. At the simplest level, the systematic reviewer could separately analyse trials with high and low quality interventions—that is, stratify by intervention quality. A more sophisticated approach is to quantify interactions between measures of the quality of an intervention and its effect with meta-regression techniques.¹⁴

Whatever approach is used, the analysis is prone to statistical errors that are similar to the errors that arise in subgroup analysis of clinical trials.^{15–17} Stratified analyses are more likely to miss real effects within strata (because each stratum is smaller than the set of all trials). They are also exposed to an increased risk of finding spurious effects within strata (because the risk of falsely significant findings accrues with each test of each stratum) and, consequently, they risk finding spurious differences in effects across strata.¹⁷ Several strategies can be used to minimise statistical errors. Most importantly, the strata that are to be tested should be specified in the review protocol before the review is started and the analysis should focus on the differences between effects across strata—that is, the analysis should examine interactions between the quality and effect of an intervention; claims of stratum specific effects should not be based on tests of effects within strata.

Even when reviewers take care to minimise statistical errors, the effects of intervention quality may still generate false conclusions. One reason is that such analyses almost always rely on comparisons between studies—that is, they compare the effects of an intervention in trials in which the intervention was administered well with those in trials in which the intervention was administered poorly. Such comparisons are prone to confounding because studies with high quality interventions may differ from studies with low quality interventions in other respects, such as the types of participants or their methodological quality. In the pelvic floor muscle example above, the study by

Summary points

Systematic reviews of complex interventions should consider the quality of interventions in individual trials

Quality can be assessed when other research provides clear indications of how interventions should be administered

Assessment of effects of intervention quality can be built into analyses of the effects of intervention

Such analyses should be specified in the review protocol and should focus on interactions between the quality and the effects of the intervention

Hughes et al¹⁰ provided the lowest intensity of intervention, but there were other differences between this study and the other two: the women who participated in the study by Reilly et al¹¹ had no symptoms at baseline, and fewer of the women in the study by Mørkved et al¹² had symptoms of urinary incontinence at baseline than in the study by Hughes et al. These factors, rather than the quality of intervention, might have produced the apparent differences in effect. Analyses based on comparisons between studies should always be considered exploratory rather than definitive.

Conclusions

The QUORUM statement on reporting of meta-analyses of randomised trials exhorts reviewers to describe details of the intervention provided in each trial.¹⁸ In our opinion reviewers of complex interventions should go one step further: they should explicitly assess the quality of interventions. If reviewers describe interventions but do not evaluate the quality of interventions, they abdicate judgments about the quality of intervention to readers who may not have access to the complete reports of individual trials in the review.

Contributors and sources: This article arose from discussions in a workshop on meta-analysis conducted by RDH and KB in Oslo in January 2004. Both RDH and KB are physiotherapists, teachers, and researchers with experience in the design and analysis of randomised trials and systematic reviews. KB had the idea for the article. KB and RDH wrote the article. RDH performed the meta-analyses. Both authors have approved the final version of the article. RDH will act as the guarantor.

Competing interests: None declared.

- 1 Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing Group, 2001.
- 2 Heymans MW, van Tulder MW, Esmail R, Bombardier C, Koes BW. Back schools for non-specific low-back pain. *Cochrane Database Syst Rev* 2004;2:CD000261.
- 3 Gillespie LD, Gillespie WJ, Robertson MC, Lamb SE, Cumming RG, Rowe BH. Interventions for preventing falls in elderly people. *Cochrane Database Syst Rev* 2003;4:CD000340.
- 4 Schonstein E, Kenny DT, Keating J, Koes BW. Work conditioning, work hardening and functional restoration for workers with back and neck pain. *Cochrane Database Syst Rev* 2002;4:CD001822.
- 5 Wells HM, Sin D, Kaufman BJ, Spooner CH, Rowe BH. Education interventions for adults who attend the emergency room for acute asthma. *Cochrane Database Syst Rev* 2003;2:CD003000.
- 6 Black NA, Downs SH. The effectiveness of surgery for stress urinary incontinence in women: a systematic review. *Br J Urology* 1996;78:497-510.
- 7 Kraemer WJ, Adams K, Cafarelli E, Dudley GA, Dooly C, Feigenbaum MS, et al. Progression models in resistance training for healthy adults. *Med Sci Sports Exerc* 2002;34:364-80.
- 8 Kettunen JA, Kujala UM. Exercise therapy for people with rheumatoid arthritis and osteoarthritis. *Scand J Sci Med Sport* 2002;14:138-42.
- 9 Sampselle CM, Miller JM, Mims BL, DeLancey J, Ashton-Miller JA, Antonakos CL. Effect of pelvic muscle exercise on transient incontinence during pregnancy and after birth. *Obstet Gynecol* 1998;91:406-12.
- 10 Hughes P, Jackson S, Smith P, Abrams P. Can antenatal pelvic floor exercises prevent postnatal incontinence. *Neurourol Urodyn* 2001;20:447-8.
- 11 Reilly E, Freeman R, Waterfield M, Waterfield A, Steggles P, Pedlar F. Prevention of postpartum stress incontinence in primigravidae with increased bladder neck mobility: a randomised controlled trial of antenatal pelvic floor exercises. *BJOG* 2002;109:68-76.
- 12 Mørkved S, Bø K, Schei B, Salvesen K. Pelvic floor muscle training during pregnancy to prevent urinary incontinence. A single blind randomized controlled trial. *Obstet Gynecol* 2003;101:313-9.
- 13 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Med Res Methodol* 2001;1:2. www.biomedcentral.com/1471-2288/1/2 (accessed 30 June 2005).
- 14 Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21:1559-73.
- 15 Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomised clinical trials. *JAMA* 1991;266:93-8.
- 16 Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002;21:2917-30.
- 17 Brookes ST, Whitely E, Egger M, Davey Smith G, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004;57:229-36.
- 18 Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORUM statement. *Lancet* 1999;354:1896-900.

(Accepted 17 May 2005)

Submitting articles to the *BMJ*

We are now inviting all authors who want to submit a paper to the *BMJ* to do so via the web (<http://submit.bmj.com>).

We have introduced Benchpress, our web based manuscript tracking system, with the aim of streamlining our processes and providing better, quicker information for authors, reviewers, and editors.

Benchpress is a website where authors deposit their manuscripts and editors go to read them and record their decisions. Reviewers' details are also held on the system, and when asked to review a paper reviewers will be invited to access the site to see the relevant paper. The system is secure, protected by passwords, so that authors see only their own papers and reviewers see only those they are meant to.

For authors in particular the system offers several benefits. The system provides all our guidance and forms and allows authors to

suggest reviewers for their paper—something we'd like to encourage. Authors get an immediate acknowledgement that their submission has been received, and they can watch the progress of their manuscript. The record of their submission, including editors' and reviewers' reports, remains on the system for future reference.

Anyone with an internet connection and a web browser can use the system.

The system itself offers extensive help, and the *BMJ* Online Submission Team office is geared up to help authors and reviewers if they get stuck.

Benchpress is accessed via <http://submit.bmj.com> or via a link from *bmj.com*