

BMJ

Systematic reviews in health care: Assessing the quality of controlled clinical trials

Peter Jüni, Douglas G Altman and Matthias Egger

BMJ 2001;323:42-46
doi:10.1136/bmj.323.7303.42

Updated information and services can be found at:
<http://bmj.com/cgi/content/full/323/7303/42>

These include:

References

This article cites 27 articles, 13 of which can be accessed free at:
<http://bmj.com/cgi/content/full/323/7303/42#BIBL>

77 online articles that cite this article can be accessed at:
<http://bmj.com/cgi/content/full/323/7303/42#otherarticles>

Rapid responses

3 rapid responses have been posted to this article, which you can access for free at:
<http://bmj.com/cgi/content/full/323/7303/42#responses>

You can respond to this article at:
<http://bmj.com/cgi/eletter-submit/323/7303/42>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top left of the article

Topic collections

Articles on similar topics can be found in the following collections

[Systematic reviews \(incl meta-analyses\): descriptions](#) (133 articles)
[Systematic reviews \(incl meta-analyses\): examples](#) (352 articles)

Notes

To order reprints follow the "Request Permissions" link in the navigation box

To subscribe to *BMJ* go to:
<http://resources.bmj.com/bmj/subscribers>

- 7 Working Party of the Royal College of Physicians. Improving communication between doctors and patients. *J R Coll Physicians Lond* 1997;31:258-9.
- 8 Meryn S. Improving doctor-patient communication. *BMJ* 1998;316:1922.
- 9 Coiera E, Tombs V. Communication behaviours in a hospital setting: an observational study. *BMJ* 1998;316:673-6.
- 10 Neil-Dwyer N, Lang D. Brain attack—aneurysmal subarachnoid haemorrhage: death due to delayed diagnosis. *J R Coll Physicians Lond* 1997;31:49-52.
- 11 Van Gijn J. Slip-ups in diagnosis of subarachnoid haemorrhage. *Lancet* 1997;349:1492.
- 12 South Thames A & E Speciality Sub Committee Audit Group. Brain attack—how good is the early management of subarachnoid haemorrhage in accident and emergency departments? *J Accid Emerg Med* 2000;17:176-9.
- 13 Cooke MW, Wilson S, Bridge P. Questionnaires of an accident and emergency department: are they reproducible? *J Accid Emerg Med* 2000;17:355-6.
- 14 Espinosa JA, Nolan TW. Reducing errors made by emergency physicians in interpreting radiographs: a longitudinal study. *BMJ* 2000;320:737-40.
- 15 *A first class service—quality in the new NHS*. London: Department of health, 1998. (Government white paper.)
- 16 National Audit Office. *NHS accident and emergency departments in England*. London: HMSO, 1992.
- 17 Wyatt JP, Henry J, Beard D. The association between seniority of accident and emergency doctor and outcome following trauma. *Injury* 1999;30:165-8.
- 18 Cooke MW, Kelly C, Khattab A, Lendrum K, Morrell R, Rubython EJ, et al. Accident and emergency 24 hour senior cover—a necessity or a luxury? *J Accid Emerg Med* 1998;15:181-4. (Accepted 23 April 2001)

Systematic reviews in health care

Assessing the quality of controlled clinical trials

Peter Jüni, Douglas G Altman, Matthias Egger

This is the first in a series of four articles

Department of Social and Preventive Medicine, University of Bern, Bern, 3012 Switzerland
Peter Jüni
research fellow

Imperial Cancer Research Fund Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, Oxford OX3 7LF
Douglas G Altman
professor of statistics in medicine

Medical Research Council Health Services Research Collaboration, Department of Social Medicine, University of Bristol, Bristol BS8 2PR

Matthias Egger
senior lecturer in epidemiology and public health medicine

Correspondence to: M Egger
m.egger@bristol.ac.uk

Series editor: Matthias Egger

BMJ 2001;323:42-6

The quality of controlled trials is of obvious relevance to systematic reviews. If the “raw material” is flawed then the conclusions of systematic reviews cannot be trusted. Many reviewers formally assess the quality of primary trials by following the recommendations of the Cochrane Collaboration and other experts.^{1 2} However, the methodology for both the assessment of quality and its incorporation into systematic reviews and meta-analysis are a matter of ongoing debate.³⁻⁵ In this article we discuss the concept of study quality and the methods used to assess quality.

Quality is a multidimensional concept, which could relate to the design, conduct, and analysis of a trial, its clinical relevance, or quality of reporting.⁶ The validity of the findings generated by a study clearly is an important dimension of quality. In the 1950s the social scientist Campbell proposed a useful distinction between internal and external validity (see box below).^{7 8} Internal

Components of internal and external validity of controlled clinical trials

Internal validity—extent to which systematic error (bias) is minimised in clinical trials

- Selection bias: biased allocation to comparison groups
- Performance bias: unequal provision of care apart from treatment under evaluation
- Detection bias: biased assessment of outcome
- Attrition bias: biased occurrence and handling of deviations from protocol and loss to follow up

External validity—extent to which results of trials provide a correct basis for generalisation to other circumstances

- Patients: age, sex, severity of disease and risk factors, comorbidity
- Treatment regimens: dosage, timing and route of administration, type of treatment within a class of treatments, concomitant treatments
- Settings: level of care (primary to tertiary) and experience and specialisation of care provider
- Modalities of outcomes: type or definition of outcomes and duration of follow up

Summary points

Empirical studies show that inadequate quality of trials may distort the results from systematic reviews and meta-analyses

The influence of the quality of included studies should routinely be examined in systematic reviews and meta-analyses

The use of summary scores from quality scales is problematic—it is preferable to examine the influence of key components of methodological quality individually

Based on empirical evidence and theoretical considerations, the generation and concealment of the allocation sequence, blinding, and handling of patient attrition in the analysis should always be assessed

validity implies that the differences observed between groups of patients allocated to different interventions may, apart from random error, be attributed to the treatment under investigation. In contrast, external validity, or generalisability, is the extent to which the results of a study provide a correct basis for generalisations to other circumstances. In itself, there is no external validity. The term is only meaningful with regard to specified “external” conditions, such as other patient populations or treatment regimens. Internal validity is a prerequisite for external validity: the results of a flawed trial are invalid, and the question of its external validity becomes redundant.

Dimensions of internal validity

Internal validity is threatened by bias, “any process at any stage of inference tending to produce results that differ systematically from the true values.”⁹ In clinical trials, biases fall into four categories: selection bias, performance bias, detection bias, and attrition bias (box).

The two interrelated steps of randomisation

Generation of allocation sequences

- Adequate if sequences are suitable to prevent selection bias: random numbers generated by computer, table of random numbers, drawing of lots or envelopes, tossing a coin, shuffling cards, throwing dice, etc
- Inadequate if sequences could be related to prognosis and thus introduce selection bias: case record number; date of birth; day, month, or year of admission; etc

Concealment of allocation sequences

- Adequate if patients and investigators enrolling patients cannot foresee assignment: a priori numbered or coded drug containers of identical appearance prepared by an independent pharmacy; central randomisation (performed at a site remote from the trial's location); sequentially numbered, sealed, opaque envelopes; etc
- Inadequate if patients and investigators enrolling patients can foresee assignments and thus introduce selection bias: procedures based on inadequate generation of allocation sequences, open allocation schedule, alternation and other unsealed or non-opaque envelopes, etc

Selection bias

The aim of randomisation is the creation of groups that are comparable for any known or unknown potential confounding factors.¹⁰ Success depends on two interrelated procedures (see box above).¹¹ Firstly, an allocation sequence that is suitable to prevent selection bias must be generated—for example, by using a computer algorithm, tossing a coin, or throwing a dice. Secondly, this sequence must be concealed from investigators enrolling patients. Knowledge of assignments—for example, from a table of random numbers posted on a bulletin board—can cause selective enrolment of patients on the basis of prognostic factors.¹² Patients who would have been assigned to a treatment deemed to be “inappropriate” may be rejected, and some patients may be deliberately directed to the “appropriate” treatment.¹³ Deciphering of allocation schedules may occur even if concealment was attempted. For example, envelopes may be opened or held against a bright light to reveal the contents.¹⁴

Performance bias and detection bias

Performance bias occurs if additional treatment interventions are provided preferentially to one group. Blinding of patients and care providers prevents this type of bias and also safeguards against differences in placebo responses between the groups. Detection bias arises if the knowledge of patient assignment influences the assessment of outcome.¹⁵ This is avoided by the blinding of those assessing outcomes—for example, patients, care providers, radiologists, or end point review committees (box).

Attrition bias

Deviations from protocol and loss to follow up often lead to the exclusion of patients after they have been allocated to treatment groups, which may introduce attrition bias. Possible deviations from protocol include the violation of eligibility criteria and non-adherence to treatments. Loss to follow up refers to patients

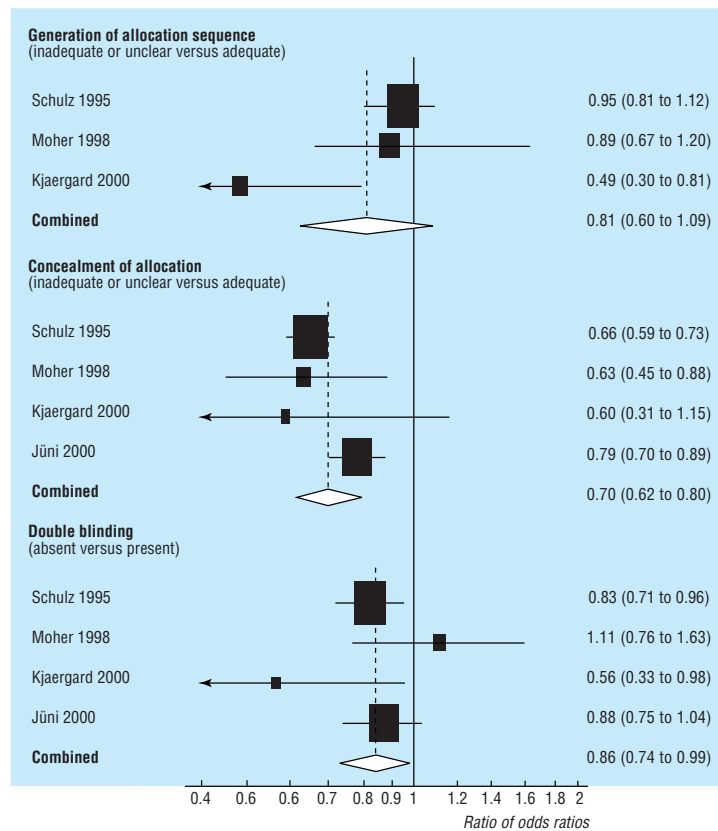
becoming unavailable for examinations at some stage during the study period because they refuse to participate further (also called drop outs), cannot be contacted, or clinical decisions are made to stop the assigned interventions.

Patients excluded after allocation are unlikely to be representative of patients remaining in the study. For example, patients may not be available for follow up because they have an acute exacerbation of their illness or severe side effects.¹⁶ Patients not adhering to treatments generally differ in respects that are related to prognosis.¹⁷ All randomised patients should therefore be included in the analysis and kept in their original groups, regardless of their adherence to the study protocol. In other words the analysis should be performed according to the intention to treat principle, thus avoiding selection bias.^{16 18} This implies that the primary outcome was recorded for all randomised patients at the prespecified times throughout the follow up period.¹⁹ If the end point of interest is mortality from all causes this can be established most of the time. It may, however, be impossible retrospectively to ascertain other binary or continuous outcomes, and some patients may therefore have to be excluded from the analysis. In this case the proportion of patients not included in the analysis must be reported and the possibility of attrition bias discussed.

Empirical evidence of bias

Numerous case studies show that the biases described above do occur in practice, distorting the results of clinical trials.⁵ The authors are aware of four methodological studies that have gauged their relative importance in a large number of clinical trials while avoiding confounding by disease or intervention.^{20–23} The figure shows a meta-analysis of the results from these studies. Inadequate or unclear concealment of treatment allocation was associated with an exaggeration of treatment effects in all four studies. Odds ratios from trials with inadequate or unclear concealment were on average 30% lower (more beneficial) than those from trials with adequate methodology (combined ratio of odds ratios 0.70, 95% confidence interval 0.62 to 0.80). The inappropriate generation of allocation sequences was assessed in three studies only and was not consistently associated with treatment effects, although an effect was evident in the study from Denmark (figure).^{20 21 23} Interestingly, when only trials with adequate concealment of allocation were analysed in Schulz et al's study, those with an inadequate generation of allocation sequences did yield inflated treatment effects.²⁰ This indicates that if assignments are predictable some deciphering can occur, even with adequate concealment. On the other hand, the generation of unbiased sequences is probably irrelevant if the sequences are not concealed from those involved in the enrolment of patients.¹³

Results for double blinding were more heterogeneous: the two larger studies^{20 22} found that estimates were on average moderately biased in open trials, whereas one of the two smaller studies showed no effect,²¹ and the other showed substantial bias associated with lack of double blinding (figure).²³ To some extent the importance of blinding depends on the outcomes assessed. In some situations—for example, when exam-



Meta-analysis of four empirical studies relating key aspects of methodological quality of controlled trials to their effect estimates. Meta-analysis was by random effects model. Size of squares is proportional to inverse of variance of estimate

ing the effect of an intervention on overall mortality—blinding of outcome assessment is irrelevant. Differences in the type of outcomes examined could thus explain the discrepancy between the studies.

Furthermore, investigators' understanding of who exactly should be blinded in double blind trials varies,²⁴ and this may also introduce heterogeneity. Two studies addressed attrition bias but used different definitions. Schulz et al compared trials that reported exclusions with trials that either explicitly reported no exclusions or gave the impression that no exclusions had taken place.²⁰ In contrast, Kjaergard et al compared trials that reported adequately on attrition (independent of whether exclusions occurred) to trials with inadequate reporting.²³ Schulz et al found little difference in effect estimates (ratio of odds ratios 1.07, 95% confidence interval 0.94 to 1.21) whereas Kjaergard et al found a trend towards larger effect estimates in trials with adequate reporting (ratio of odds ratios 1.50, 0.80 to 2.78).^{20, 23} The methods used to assess attrition were unsatisfactory in both of these studies. Future research in this area should distinguish between quality of reporting and methodological quality and consider that some exclusions and losses to follow up may be unavoidable whereas others are clearly inappropriate.

Dimensions of external validity

External validity relates to the applicability of the results of a study to other "populations, settings,

treatment variables, and measurement variables".⁸ External validity is a matter of judgment, which depends on the characteristics of the patients included in the trial, the setting, the treatment regimens, and the outcomes assessed (box).⁸ In recent years large meta-analyses based on data from individual patients have shown that important differences in treatment effects may exist between patient groups and settings. For example, antihypertensive treatment reduces total mortality in middle aged patients with hypertension, but this may not be the case in elderly people.²⁵ The benefits of fibrinolytic treatment in suspected acute myocardial infarction has been shown to decrease linearly with the delay between the start of symptoms and the initiation of treatment.²⁶ In trials of cholesterol lowering drugs the benefits of a reduction in non-fatal myocardial infarction and mortality due to coronary heart disease depends on the reduction in total cholesterol concentration and the duration of follow up.²⁷ At the very least, therefore, assessment of a trial's applicability requires adequate information about the characteristics of the participants.

Quality of reporting

The assessment of the methodological quality of a trial is intertwined with the quality of reporting—that is, the extent to which a report provides information about the design, conduct, and analysis of the trial.⁴ Reports often omit important methodological details. For example, only 1 of 122 randomised trials of selective serotonin reuptake inhibitors specified the method of randomisation.²⁸ A widely used approach to this problem is to assume that the quality was inadequate unless the information to the contrary is provided (the "guilty until proved innocent" approach). This is often justified because faulty reporting generally reflects faulty methods.^{20, 29} A well conducted but badly reported trial will, however, be misclassified. An alternative approach is to explicitly assess the quality of the reporting rather than the adequacy of the methods. This is also problematic because a biased but well reported trial will receive full credit.³⁰ The adoption of guidelines on the reporting of clinical trials has recently improved this situation for several journals,^{31, 32} but deficiencies in reporting will continue to be confused with deficiencies in design, conduct, and analysis.

Assessing trial quality

How the quality of trials should be assessed is being debated. Quality scales combine information on several features in a single numerical value, whereas the component approach examines key dimensions individually, without calculation of a score. Moher et al reviewed the use of quality scores in systematic reviews published in medical journals and the Cochrane database of systematic reviews.³³ Trial quality was assessed in 78 (38%) of the 204 reviews from journals, of which 20 (26%) used components and 52 (67%) used scales. By contrast, all 36 reviews from the database assessed quality, of which 33 (92%) used components and none used scales.

Scales vary considerably in dimensions covered and complexity.⁴ Many scales include items for which there is little evidence that they are related to the inter-

nal validity of a trial. For example, a widely used instrument includes items related to the presentation of data and the organisation of the trial.³⁴ Unsurprisingly, different scales can lead to discordant results. This was shown in a study in which 25 different scales were used to assess 17 trials comparing low molecular weight heparin with standard heparin for thromboprophylaxis.⁵ With some scales, the relative risks of the “high quality” trials were close to unity and not statistically significant, indicating that low molecular weight heparin was not superior to standard heparin, whereas the “low quality” trials assessed by these scales showed better protection with the low molecular weight heparin. With other scales the opposite was the case: high quality trials indicated that low molecular weight heparin was superior to standard heparin, whereas low quality trials found no significant difference.⁵

When the association of effect estimates with quality scores is examined, interpretation of results is difficult. In the absence of an association there are three possible explanations³⁵: there is no association with any of the components; there are associations with one or several components, but these components have so little weight that the effects are lost in the summary score; or there are associations with two or more components, but these cancel out so that no association is found with the overall score. On the other hand, if treatment effects do vary with quality scores then investigators will have to identify the component or components that are responsible for this association to interpret this finding.

The analysis of individual components of trial quality overcomes many of the shortcomings of composite scores. The component approach takes into account that the importance of individual quality domains, and the direction of potential biases associated with these domains, varies between the contexts in which trials are performed.

Incorporating study quality into meta-analysis

It makes intuitive sense to take into account information on the quality of studies when doing systematic reviews. One approach is to exclude trials that fail to meet some standard of quality. This may often be justified but could exclude studies that might contribute valid information. It may therefore be prudent to exclude only trials with gross deficiencies in design—for example, those that clearly failed to study comparable groups. The possible influence of study quality on effect estimates should, however, always be examined in a given set of included studies. Several approaches have been proposed for this purpose.

Quality as a weight in statistical pooling

The most radical approach is to directly incorporate information on study quality as weighting factors in the analysis. Study weights can be multiplied by quality scores, thus increasing the weight of trials deemed to be of high quality and decreasing the weight of those of low quality.^{3 21} A trial with a quality score of 40 out of 100 will thus get the same weight in the analysis as a trial with half the amount of information but a quality score of 80.



MARK OLDROYD

Weighting by quality scores is problematic for several reasons. As mentioned, the choice of the scale influences the weight of individual studies in the analysis, and the combined effect estimate and its confidence interval therefore depend on the scale. However, there is no reason why study quality should modify the precision of estimates. Poor studies are still included. Thus any bias associated with poor methodology is only reduced, not removed. Including both good and poor studies may also increase heterogeneity of estimated effects across trials and may reduce the credibility of a systematic review. The incorporation of quality scores as weights lacks statistical or empirical justification.³

Sensitivity analysis

The robustness of the findings of a meta-analysis to different assumptions should always be examined in a thorough sensitivity analysis. An assessment of the influence of methodological quality should be part of this process. Simple stratified analyses and meta-regression models are useful for exploring associations between treatment effects and study characteristics. Quality summary scores or categorical data on individual components can be used for this purpose. For the reasons discussed the authors recommend that sensitivity analysis should be based on the components of study quality that are considered important in the context of a given meta-analysis. Other approaches, such as plotting effect estimates against quality scores or performing cumulative meta-analysis in order of quality, are also affected by the problems surrounding composite scales.^{3 36}

Conclusions

There is ample evidence that many trials are methodologically weak and increasing evidence that deficiencies translate into biased findings of systematic reviews. The assessment of the methodological quality of controlled trials and the conduct of sensitivity analyses should therefore be considered routine procedures in systematic reviews and meta-analysis. Although composite quality scales may provide a useful overall assessment when comparing populations of trials, such scales should generally not be used to identify trials of apparent low quality or high quality in a given systematic review. Rather, the relevant methodological aspects should be identified a priori and assessed individually. This should include the generation and concealment of treatment allocation,

blinding, and handling of attrition in the analysis. Other ways of investigating and dealing with bias in systematic reviews will be discussed and illustrated later in this series.³⁷

We thank Ken Schulz and Lise Kjaergard for unpublished data and Iain Chalmers for useful comments on an earlier version of this paper.

Funding: PJ is supported by the Swiss National Science Foundation. The work on trial quality in Bristol was supported by the NHS Research and Development Programme.

Competing interests: None declared.

- Clarke M, Oxman AD, eds. *Cochrane reviewers' handbook 4.0*. In: Cochrane Collaboration. *Cochrane Library*. Oxford: Update Software, 1999.
- Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epidemiol* 1995;48:167-71.
- Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbé KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992;45:255-65.
- Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clin Trials* 1995;16:62-73.
- Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trial for meta-analysis. *JAMA* 1999;282:1054-60.
- Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*, 2nd ed. London: BMJ Books, 2001.
- Campbell DT. Factors relevant to the validity of experiments in social settings. *Psychol Bull* 1957;54:297-312.
- Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research on teaching. In: Gage NL, ed. *Handbook of research on teaching*. Chicago: Rand McNally, 1963:171-246.
- Murphy EA. *The logic of medicine*. Baltimore: Johns Hopkins University Press, 1976.
- Altman DG, Bland JM. Treatment allocation in controlled trials: why randomise? *BMJ* 1999;318:1209.
- Altman DG. Randomisation. Essential for reducing bias. *BMJ* 1991;302:1481-2.
- Keirse MJ. Amniotomy or oxytocin for induction of labor. Re-analysis of a randomized controlled trial. *Acta Obstet Gynecol Scand* 1988;67:731-5.
- Schulz KF. Randomised trials, human nature, and reporting guidelines. *Lancet* 1996;348:596-8.
- Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995;274:1456-8.
- Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 1994;44:16-20.
- Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979;301:1410-2.
- Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the CDP. *N Engl J Med* 1980;303:1038-41.

- May GS, Demets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation* 1981;64:669-73.
- Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319:670-4.
- Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
- Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-13.
- Jüni P, Tallon D, Egger M. 'Garbage in - garbage out? Assessment of the quality of controlled trials in meta-analyses published in leading journals. *Proceedings of the 3rd symposium on systematic reviews: beyond the basics*, St Catherine's College, Oxford. Oxford: Centre for Statistics in Medicine, 2000:19.
- Kjaergard LL, Villumsen J, Gluud C. Quality of randomised clinical trials affects estimates of intervention efficacy. *Proceedings of the 7th Cochrane colloquium*. Universita S.Tommasso D'Aquino, Rome. Milan: Centro Cochrane Italiano, 1999:57 (poster B10).
- Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Mouton VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;285:2000-3.
- Gueyffier F, Bulpitt C, Boissel JP, Schron E, Ekblom T, Fagard R, et al. Antihypertensive drugs in very old people: a subgroup meta-analysis of randomised controlled trials. *Lancet* 1999;353:796.
- Fibrinolytic Therapy Trialists' (FTT) Collaborative Group. Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. *Lancet* 1994;343:311-22.
- Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Stat Methods Med Res* 1993;2:173-92.
- Hotopf M, Lewis G, Normand C. Putting trials on trial—the costs and consequences of small trials in depression: a systematic review of methodology. *J Epidemiol Community Health* 1997;51:354-8.
- Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986;4:942-51.
- Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 1995;48:71-9.
- Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials. *JAMA* 2001;285:1987-91.
- Egger M, Jüni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;285:1996-9.
- Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technol Assess* 1999;3(12).
- Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, Reitman D, et al. A method for assessing the quality of a randomized control trial. *Controlled Clin Trials* 1981;2:31-49.
- Greenland S. Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994;140:300-2.
- Linde K, Scholz M, Ramirez G, Clausius N, Melchart D, Jonas WB. Impact of study quality on outcome in placebo-controlled trials of homeopathy. *J Clin Epidemiol* 1999;52:631-6.
- Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001 (in press).



Systematic Reviews in Health Care: Meta-analysis in Context can be purchased through the BMJ Bookshop (www.bmjbookshop.com); further information and updates for the book are available on www.systematicreviews.com

A memorable patient A record follow up

On 5 September 1944 when I was a medical officer at the 59th British Military Hospital, then in Fano on the Adriatic coast of Italy, a nun from a small Roman Catholic hospital came to ask for help. A 16 year old girl had been rescued after spending several hours trapped in the rubble of her home in Gemmano, a small town about 20 miles away; it had been virtually destroyed by bombing. Now, after this length of time, I can't remember whether it was by the Germans or Allies.

The girl's father, who was the mayor of the town, had escaped with superficial injuries. Her mother had bled to death from multiple injuries including a severed foot. Now the daughter was lying seemingly near to death in the semi-derelict hospital where there was no doctor and only minimal equipment. She had obviously lost a great deal of blood; her right femur was shattered; and her extensive wounds were infected.

The hospital was fairly quiet at the time. During several visits two of us were able to clean the wounds using pentothal anaesthetic. We also managed to find some blood and over the next week or two we gave her about two pints, though I can no longer remember how we got the blood or the other details of her treatment (for example, did we have any penicillin?). When

eventually the military hospital received orders to move on, we were greatly relieved to see that the girl looked as if she would recover.

I received letters from her and her father in 1946 but after that, nothing. The image of that poor girl has always haunted me, and I often talked to family and friends about her. In February this year, my daughter, who has lived in Florence for 28 years and speaks fluent Italian, decided to try to find the girl. By using the internet she discovered that my patient was still alive and living in Treviso; she was now the widow of the orthopaedic surgeon who had looked after her when she was transferred to a hospital in Rome. She was planning to travel through Florence to visit her daughter in Rome at a time when we had already arranged to stay with our daughter over Easter.

So she came to lunch: a sprightly and articulate 74 year old who had some difficulty walking because of a shortened right leg resulting from injuries received 57 years ago. I wonder if the length of follow up could be a record.

N C Mond *retired general practitioner, Oxfordshire*