Meta-analysis
Unresolved issues and future developments

George Davey Smith, Matthias Egger

Since its recent introduction into clinical epidemiology, meta-analysis has established itself as an influential branch of biostatistics. Several books have focused mainly or entirely on meta-analysis in medicine,[1-5] and the latest editions of relevant textbooks generally include a section on meta-analysis.[6-9] Computer software entirely devoted to meta-analysis has been developed, and meta-analytic procedures have been introduced in general statistical software packages. We will soon be providing an overview of software packages on the *BMJ*'s website.[10] Several unresolved issues concerning meta-analysis remain, and in this final article of our series we address some of the topics that are likely to feature in future discussions of the appropriate practice and domain of meta-analysis.

Summary points

- Meta-analysis has established itself as an important technique in clinical epidemiology, but several issues remain unresolved
- The inclusion of unpublished, non-peer reviewed data can be problematic, particularly if these data come from interested sources, such as the pharmaceutical industry
- Individual patient data are often required to address important questions, but the mechanisms to facilitate increasing availability of trial data for meta-analysis are lacking
- The clinical application of results from meta-analyses to the individual patient often remains a difficult matter of judgment
- The Cochrane Collaboration will have an important role in future developments in the field of systematic reviews and meta-analyses

This is the last in a series of six articles examining the procedures in conducting reliable meta-analysis in medical research

Department of Social Medicine, University of Bristol, Bristol BS8 2PR
George Davey Smith, *professor of clinical epidemiology*
Matthias Egger, *reader in social medicine and epidemiology*
Correspondence to: Professor Davey Smith
zetkin@bristol.ac.uk

BMJ 1998;316:221–5

Should unpublished data be included in meta-analyses?

Publication bias, discussed in detail in a previous article,[11] is a major threat to the validity of meta-analysis. Obtaining and including data from unpublished studies seems to be the obvious way of avoiding this problem. Including data from unpublished studies can itself introduce bias, however. Even after extensive consultation with the research community, unpublished studies may remain hidden. The unpublished studies that can be located may thus be an unrepresentative sample of unpublished studies. Whether bias is reduced or increased by including unpublished studies cannot formally be assessed as it is impossible to be certain that all unpublished studies have been located. A further problem relates to the willingness of investigators of located unpublished studies to provide data. This may depend on the findings of the study, more favourable results being provided more readily. This could again bias the findings of a meta-analysis.

An analysis of 150 meta-analyses published between 1988 and 1991 showed that most meta-analysts had searched for unpublished material, although such data were located and included in only 31% of meta-analyses.[12] A questionnaire assessing the attitudes towards inclusion of unpublished data was sent to the authors of these reports and to the editors of the journals that had published them: 78% of meta-analysts supported the use of unpublished material, compared with only 47% of journal editors.[12] This lack of support by some editors is on the grounds that the data have not been peer reviewed. The refereeing process, however, has not always been a successful way of ensuring that published results are valid.[13 14] On the other hand, meta-analyses of unpublished data from interested sources is clearly of concern. Such unchallengeable data have been produced in circumstances in which an obvious commercial interest exists (box 1 gives an example).

The most satisfactory approach to the inclusion of unpublished data in meta-analyses is to carry out an extensive search for such data and obtain them if possible. The analysis should then be performed with and without the unpublished data, as a form of sensitivity analysis. If the conclusions are altered through the inclusion or exclusion of such data, the results of either approach should be treated cautiously.

Subjectivity in data analysis and reporting

Using published results exclusively can introduce biases other than those of publication bias. The choice of the outcome that is reported can be influenced by the results: the outcome with the most favourable findings will generally be reported. An example of how published results can be misleading comes from two separate analyses of a double blind, placebo controlled trial assessing the efficacy of amoxycillin in children with non-suppurative otitis media.[17 18] Opposite conclusions were reached, mainly because different weight was given to the various outcome measures assessed in the study. This disagreement was conducted in the public arena, as it was accompanied by accusations of impropriety against the team producing the findings favourable to amoxycillin. The leader of this team had received considerable funding, both in research grants and as personal honorarium, from the manufacturers of amoxycillin.[19] This is a good example of how reliance on the data chosen to be presented by the investigators can lead to

Box 1: Controversy over selective serotonin uptake inhibitors and depression

- Selective serotonin uptake inhibitors are widely used for the treatment of depression, although their clinical advantages over the much less expensive tricyclic antidepressants have not been well established.
- In their meta-analysis Song et al used the dropout rate among randomised controlled trial participants taking selective serotonin uptake inhibitors and those taking conventional antidepressants as an indicator of therapeutic success[15]: patients who stop taking their treatment because of inefficacy or side effects are the ones who are not benefiting, and thus the class of drug with the lower dropout rate can be considered the one with the more favourable effects.
- There was little difference between selective serotonin uptake inhibitors and the other—usually tricyclic—antidepressants. In response to this analysis, Nakielny (for Lilly Industries, the manufacturers of fluoxetine) presented a meta-analysis of 14 investigational studies of new drugs which they stated included every study completed by December 1990.[16] This included what were called "unpublished data on file." The pooled dropout rates calculated by Nakielny differed markedly from the literature based analysis.

		Fluoxetine		Tricyclic antidepressant		
	No of trials	No of patients	Drop out rate (%)	No of patients	Drop out rate (%)	P value
Song et al[15]	18*	913	34.5	916	36.7	0.4
Nakielny[16]	14	781	36.5	788	47.5	<0.0001

*References 6, 12-15, 18, 29, 31, 33-35, 44, 47, 63, 65-67, 69 in Song et al.[15]

Lilly Industries claimed that its analysis was not "subject to biases introduced by selective publication and literature searches," but this is difficult to assess if the trials included represent unpublished data "on file." To make such data available in the future is one of the major challenges facing meta-analysts and the promoters of systematic reviews and evidence based medicine.

distortion.[20] This has probably been a frequent source of bias, which only in rare occasions becomes common knowledge. With improving standards of clinical trial reporting[21] subjectivity in data analysis should become less common in the future.

Individual patient data or summary statistics—which should be included in a meta-analysis?

Meta-analyses that have been entirely dependent on summary data obtained from published reports of clinical trials have provided robust indicators of treatment outcomes. Such analyses have been described as meta-analyses of the literature.[22] If a researcher is interested in outcomes in different groups, however, the analysis will be made difficult if the various trials do not report data accordingly. For example, a literature based meta-analysis of the effect of drug treatment of hypertension in elderly people[23] was obliged to use a definition of "elderly" that included the participants aged 60 or over from some studies and those aged 65 or over from others. Also, because many trials failed to report age stratified data, less than half of the potential trials could be included in the analysis. This could create serious bias, as the decision of investigators to publish age stratified data may have been dependent on results.

Supplementary data from individual trials are increasingly being obtained for meta-analyses. For example, by obtaining data on mortality from coronary heart disease according to grouped follow up periods from the original investigators of cholesterol lowering trials, Law et al were able to show that the reduction in risk of coronary heart disease consequent on cholesterol lowering increased with the duration of treatment.[23] Several collaborative groups have assembled data on each participant within the separate trials. This greatly increases flexibility when defining groups within the different trials for subgroup analyses, and also allows use of data on the exact time to the event for each participant. For example, the Fibrinolytic Therapy Trialists' Collaborative Group investigated the effect of thrombolysis after myocardial infarction according to *(a)* the electrocardiographic abnormalities of patients at entry to the study; *(b)* the time at which treatment was received after onset of symptoms; *(c)* the age and sex of the patients; and *(d)* the presence or absence of various comorbid conditions.[24] This permits comparisons that retain the advantage of the original randomisation to be made, with the proviso that the separate trials did not necessarily use stratified randomisation according to these characteristics. Box 2 presents a further example.

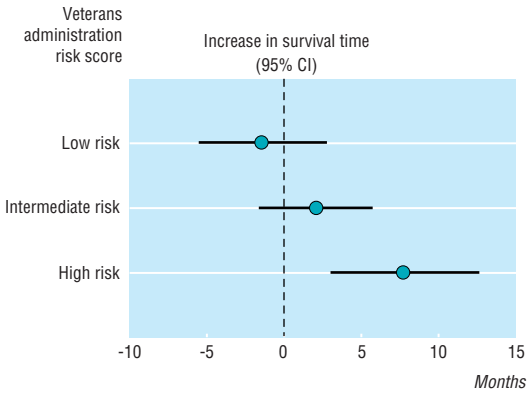
Obtaining individual patient data has advantages beyond the ability to perform standardised subgroup analyses.[27] Contact with individual investigators can help to identify further trials—published and unpublished—which the meta-analysts had missed. It may be possible to identify deviations from protocols in the trials—for example, participants who were included even though they did not satisfy entry criteria. Incorrect analyses—for example, deviation from "intention to treat" analysis, the presence of unreported dropouts, and simple oversights—may be identified. Outcome measures can be better standardised across the trials, which will counteract the tendency of researchers to publish the results only in terms of the most striking effect on a particular outcome. Additional follow up data can also be obtained, as for some trials the period of randomised comparison continues beyond the initial publication, but only the published data are publicly available.

Value of "failed" meta-analyses

In some cases a conclusive meta-analysis may not be possible if methodological standards are to be maintained. In such "failed" meta analyses[28] the treatment methods, concurrent treatment, length of follow up, characteristics of the study participants, or end points that were measured might be too varied to allow for the sensible combination of results. A meta-analysis exclusively based on a small number of

Box 2: Coronary artery bypass graft surgery and survival: meta-analysis using individual patient data

- It has long been accepted that coronary artery bypass graft surgery provides effective relief from angina pectoris and that it prolongs survival in high risk patients with left main artery disease
- The effect of such surgery on survival in other categories of patients with coronary heart disease, however, remains controversial
- A meta-analysis of trials compared coronary artery bypass graft surgery with conventional treatment in patients with stable coronary heart disease.[25] The graft surgery overall was associated with a significant reduction of mortality—for example, at five years 10.2% *v* 15.8%, P = 0.0001)
- For this meta-analysis the individual patient data made it possible to perform several subgroup analyses. For example, by using a modification of the veterans administration risk score[26] (which is based on the presence of class III or IV angina, ST depression at rest, history of hypertension, and history of myocardial infarction) the relation of benefit with the level of risk could be explored. No benefit was evident in the third at lowest risk, which was characterised by a relatively low five year mortality of 5.5%. Conversely, benefit was present for groups of patients at higher risk of death (figure). This information is crucial to the clinical application of the results from meta-analyses, indicating that targeting coronary artery bypass graft surgery at high risk individuals would be an efficient way of using limited resources
- This example illustrates how important information can be derived from risk stratification based on individual patient data

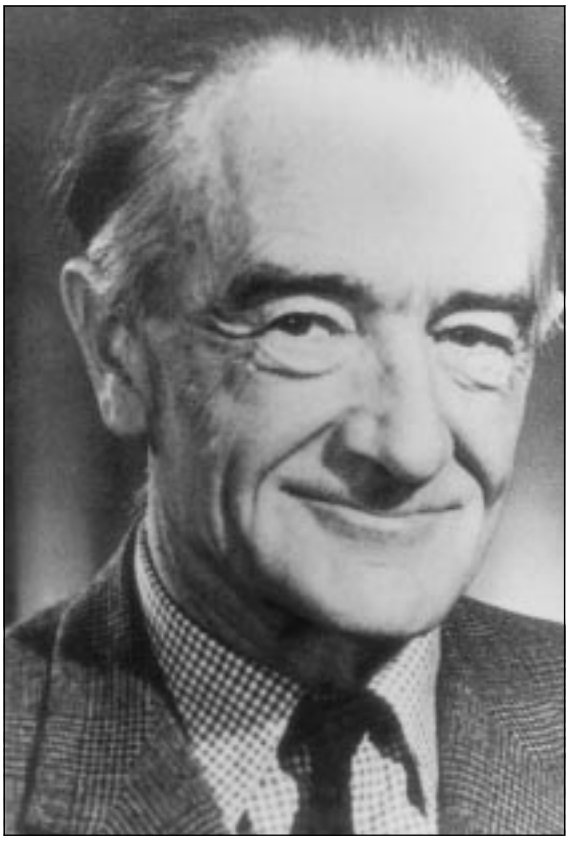


trials will often have to be inconclusive, even if the combined estimate of effect is significant.[11 29]

The *Cochrane Database of Systematic Reviews* contains many examples of treatment interventions for which, the reviewers thought, meta-analysis had failed to produce a conclusive answer. For example, the review on thrombolysis in acute ischaemic stroke, published in the second issue of the *Cochrane Library* in 1996,[30] stated: "… the data so far are scant, and quite insufficient to make any definite conclusion about the benefit or otherwise of thrombolysis to treat acute ischaemic stroke." Additional trials have since been published, and an updated version of the same review (issue 2, 1997) concluded that, although more research was needed, clear evidence existed for a substantial excess risk of intracranial haemorrhage and early death with high doses of thrombolytic drugs. Clearly stating and showing the inadequacy of existing evidence should serve as a stimulus for conducting the appropriate and necessary trials.

The Cochrane Collaboration

The dissemination of failed reviews is an important task, which is neglected by traditional journals. The examples mentioned above illustrate that this is increasingly being taken on by the Cochrane Collaboration, along with the dissemination of many other, conclusive reviews. This international group, named after Archie Cochrane, is a unique initiative in the evaluation of healthcare interventions. In his seminal book *Effectiveness and Efficiency: Random Reflections on Health Services*, published in 1972, Cochrane forcefully argued that the healthcare resources should be used to provide equitably those interventions that have been shown in well designed studies to be effective.[31] The collaboration's effort to prepare, disseminate, and continuously update systematic reviews of controlled trials is an essential, and timely, step towards achieving this goal. The Cochrane Collaboration will have an important role in future developments in the field of systematic reviews and meta-analysis. The collaboration's working groups are addressing many of the currently unresolved issues, including, for example, the approach to observational data and data from evaluations of screening and diagnostic tests. Ways of improving the applicability of reviews, discussed below, and of strengthening the involvement of consumer representatives, are also being studied.



Archie Cochrane (1909-88), the pioneer in health services research whose visions are at the heart of the Cochrane Collaboration

Clinical application of results from meta-analyses

Single large trials showing beneficial effects of treatments do influence medical practice, whereas meta-analyses of smaller studies have generally had limited impact. For example, the use of thrombolysis to reduce mortality from myocardial infarction increased only after publication of two large trials in the late 1980s,[32 33] although the same reduction in mortality had already been shown in 1982 in a meta-analysis of eight smaller studies[34] and again in a 1985 meta-analysis.[35] The increase in the use of thrombolysis is in line with the recommendations made in authoritative reviews and textbooks. Only after publication of the first trial by the Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico was thrombolysis increasingly recommended as routine treatment after myocardial infarction.[36] The 1982 meta-analysis has received only 150 citations over the 14 years since its publication in the *New England Journal of Medicine* (the medical journal with the highest impact factor), whereas the 1985 meta-analysis has received about 350 citations, the same as those received by a small, inconclusive trial that was published in the same year.[37] The two large trials, however, have received several thousand citations over a shorter period. Clearly, meta-analyses, even when conclusive, currently receive less attention than the trials which they pool, and this is presumably reflected in a smaller degree of influence on clinical practice.

Clinicians want results from clinical research that can usefully inform their clinical practice. Perhaps meta-analyses are seen as not providing information beyond the effect of treatment on a hypothetical "average" patient. The confidence interval, often narrow in meta-analysis, reflects how certain one can be about the size of the overall effect seen in a population. Of more relevance to the clinician, however, is how certain one can be about the effect in his or her particular patient. Although the overall effect will generally provide the best available estimate, the uncertainty with respect to a particular patient will always be greater than with respect to the overall patient group. This is because, in the same way as the effect under examination varies between the component studies in the meta-analysis, the effect further varies between different patients.[38]

Many clinical opinion leaders simply do not trust the results from meta-analyses. This could be seen as a cautious attitude to a relatively new technique, which is justified considering the existence of misleading meta-analyses.[29 39] The emergence of the "professional meta-analyst" [40] moving monthly from issue to issue, happily engaging in areas outside their domain of primary expertise, filling the pages of the medical journals, and sometimes viewed as lackeys for governmental agencies with a cost cutting agenda, has certainly not helped here. We believe that with improved methodological standards that routinely involve thorough sensitivity analyses, confidence in the results from meta-analyses will gradually grow. Although knowledge of the accumulated evidence from clinical trials should certainly provide a strong guide for practice, it is appropriate that features of the particular clinical situation should also be incorporated into the decision making process. The failure to recognise that the world is characterised as much by difference as similarity, which may be lost to those faced by numbers not patients, has on occasion led to overconfident assertions from practitioners of meta-analysis, which have understandably antagonised clinicians. Retaining a degree of humility in the face of the diversity of humanity served by medicine, and thus admitting to greater uncertainty than may be wished, will in the end prove the best way of furthering the goals of meta-analysis and the practice of evidence-based medicine to which it contributes.


MARK HUDSON

Outlook

In this series we outlined and illustrated the principles, strengths, and weaknesses of meta-analysis. We believe that this technique is clearly superior to the narrative approach to reviewing medical research. In addition to providing a precise estimate of the overall treatment effect in some instances, appropriate examination of heterogeneity across individual studies can produce clinically useful information with which to guide rational and cost effective treatment decisions. Both the uncritical synthesis of data from observational studies and the unconsidered synthesis of disparate results from randomised controlled trials threaten to damage the reputation of meta-analysis.

Some of the shortcomings of meta-analysis, however, are a consequence of a more general failing with respect to the dissemination of research findings. Currently this process is highly dependent on the publication of study results in peer reviewed, English-language journals. Considerations regarding publication and location biases have shown that this can result in a selected portion of all the evidence becoming available for systematic review. This is clearly unsatisfactory and can misdirect clinical practice, whether or not a formal meta-analysis is performed. Meta-analyses based on individual patient data have

shown that making such data available can contribute valuable and clinically relevant information that could not be obtained from published sources. Mechanisms to facilitate such collaborative analyses and to ensure wide accessibility of results from clinical research, including results kept as "data on file" by the pharmaceutical industry, must be developed further. The technological barriers to worldwide data exchange and collaboration are tumbling down—we can only hope that the remaining barriers, rooted in customary practice, political agendas, and commercial interests, will swiftly fall too.

The department of social medicine at the University of Bristol is part of the Medical Research Council's health services research collaboration.

Funding: ME was supported by the Swiss National Science Foundation.

1 Petitti DB. *Meta-analysis, decision analysis, and cost-effectiveness analysis. Methods for quantitative synthesis in medicine.* New York: Oxford University Press, 1994.
2 Warren KS, Mosteller F, eds. *Doing more good than harm. The evaluation of health care interventions.* New York: New York Academy of Sciences, 1993.
3 Cooper H, Hedges LV, eds. *The handbook of research synthesis.* New York: Russell Sage Foundation, 1994.
4 Hedges LV, Olkin I. *Statistical methods for meta-analysis.* Boston: Academic Press, 1985.
5 Eddy DM, Hasselblad V, Shachter R. *Meta-analysis by the confidence profile method. The statistical synthesis of evidence.* Boston: Academic Press, 1992.
6 Armitage P, Berry G. *Statistical methods in medical research.* Oxford: Blackwell Science, 1994.
7 Altman DG. *Practical statistics for medical research.* London: Chapman and Hall, 1991.
8 Lilienfeld DE, Stolley PD. *Foundations of epidemiology.* New York: Oxford University Press, 1994.
9 Davey Smith G, Egger M, Phillips AN. Meta-analysis and data synthesis in medical research. In: Detels R, Holland WW, McEwen J, Omenn GS, eds. *Oxford textbook of public health.* New York: Oxford University Press (in press).
10 Egger M, Davey Smith G. Meta-analysis software (http.www.bmj.com).
11 Egger M, Davey Smith G. Meta-analysis: bias in location and selection of studies. *BMJ* 1998;316:61-6.
12 Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, et al. Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 1993;269:2749-53.
13 Rennie D. Problems in peer review and fraud: cleave ever to the sunnier side of doubt. In: *Balancing act. Essays to honour Stephen Lock.* London: Keynes Press, 1991.
14 Nylenna M, Riis P, Karlsson Y. Multiple blinded reviews of the same two manuscripts: effects of referee characteristics and publication language. *JAMA* 1994;272:149-51.
15 Song F, Freemantle N, Sheldon TA, House A, Watson P, Long A, et al. Selective serotonin reuptake inhibitors: meta-analysis of efficacy and acceptability. *BMJ* 1993;306:683-7.
16 Nakielny J. Effective and acceptable treatment for depression. *BMJ* 1993;306:1125.
17 Mandel EH, Rockette HE, Bluestone CD, Paradise JL, Nozza RJ. Efficacy of amoxicillin with and without decongestant-antihistamine for otitis media with effusion in children. *New Engl J Med* 1987;316:432-7.
18 Cantekin EI, McGuire TW, Griffith TL. Antimicrobial therapy for otitits media with effusion ("secretory" otitits media). *JAMA* 1991;266:3309-17.
19 Rennie D. The Cantekin affair. *JAMA* 1991;266:3333-7.
20 Subjectivity in data analysis [editorial]. *Lancet* 1991;337:401-2.
21 Begg CB, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-9.
22 Stewart LA, Parmar MKB. Meta-analysis of the literaure or of individual patient data: is there a difference? *Lancet* 1993;341:418-22.
23 Law MR, Wald NJ, Thompson SG. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease. *BMJ* 1993;308:367-72.
24 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.
25 Yusuf S, Zucker D, Peduzzi P, Fisher LD, Takaro T, Kennedy JW, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Graft Surgery Trialists Collaboration. *Lancet* 1994;344:563-70.
26 Takaro T, Hultgren HN, Lipton MJ, Detre KM. The VA cooperative randomized study of surgery for coronary arterial occlusive disease II: subgroup with significant left main lesions. *Circulation* 1976;54(suppl 3):107-17.
27 Clarke MJ, Stewart LA. Systematic reviews: obtaining data from randomised controlled trials: how much do we need for reliable and informative meta-analyses? *BMJ* 1994;309:1007-10.
28 Naylor CD. The case for failed meta-analyses. *J Clin Eval Clin Pract* 1995;1:127-30.
29 Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. *BMJ*;315:629-34.
30 Wardlaw J, Yamaguchi T, del Zoppo G, Hacke W. Thrombolysis in acute ischaemic stroke. In: Warlow C, van Gijn J, Sandercock P, eds. *Stroke Module of the Cochrane Database of Systematic Reviews, [updated 03 June 1996]. Available in The Cochrane Library [database on disk and CD ROM]. The Cochrane Collaboration; Issue 2.* Oxford: Update Software, 1996.
31 Cochrane A. *Effectiveness and efficiency. Random reflections on health services.* London: Nuffield Provincial Hospital Trust, 1972.
32 Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986;397-402.
33 ISIS-2 Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988;ii:349-60.
34 Stampfer MJ, Goldhaber SZ, Yusuf S, Peto R, Hennekens CH. Effect of intravenous streptokinase on acute myocardial infarction. Pooled results from randomized trials. *New Engl J Med* 1982;307:1180-2.
35 Yusuf S, Collins R, Peto R, Furberg C, Stampfer MJ, Goldhaber SZ, et al. Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction: overview of results on mortality, reinfarction and side-effects from 33 randomized controlled trials. *Eur Heart J* 1985;6:556-85.
36 Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *JAMA* 1992;268:240-8.
37 Simoons ML, Serruys PW, van den Brand M, Bar F, de Zwaan C, Res J, et al. Improved survival after early thrombolysis in acute myocardial infarction. *Lancet* 1985;ii:578-82.
38 Gelber RD, Goldhirsch A. From the overview to the patient: how to interpret meta-analysis data. *Recent Results in Cancer Research* 1993;127:167-76.
39 Egger M, Davey Smith G. Misleading meta-analysis. Lessons from "an effective, safe, simple" intervention that wasn't. *BMJ* 1995;310:752-4.
40 Rosendaal FR. The emergence of a new species: the professional meta-analyst. *J Clin Epidemiol* 1994;47:1325-6.

Correction

Papers that report diagnostic or screening tests

In the issue of 11 October 1997 (p 942) we published a correction to this article by Trisha Greenhalgh (30 August, p 540-3). The definition of the negative likelihood ratio was misleading in both the original article and the correction. The negative likelihood ratio is defined by the equation (1 – sensitivity)/specificity, as stated in the original article, and addresses the question, "How much more or less likely is a negative result to be found in a person with, as opposed to without, the condition?"

In addition, an author's error occurred in the box on p 542 in the same article; the fraction for the negative predictive value should have been 966/987 = 97.8% [not 966/973 = 97.8%].

One hundred years ago

Cookery as a branch of medical study

The medical faculty of the State University of Minnesota has decided to add a new course to the medical studies of that institution. As soon as the new term begins the senior class will have to take up the study of cooking. On the catalogue this study will be designated "Practical Dietetics." The students will have to go into the "laboratory" and make soups, teas, gravies, farinas, and a host of other dishes for the sick and convalescent. Is the time coming, when the overburdened medical student will be required to "take up the study" of cutlery (to be designated "Practical Scalpelography") because he may have occasion to use knives; of plumbing because it may fall to his lot to inspect drains; of bedmaking, sweeping, and dusting that he may more efficiently direct the practice of these mysteries in the sick room; of hardware manufacture that he may be able to pass a sound judgment on the quality of his gallipots?
(*BMJ* 1898;i:1475)