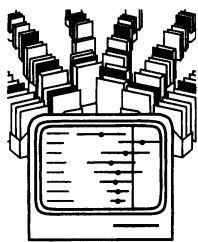


## Why sources of heterogeneity in meta-analysis should be investigated

Simon G Thompson



*This paper was presented at a meeting on Systematic Reviews organised jointly by the BMJ and the UK Cochrane Centre and held in London in July 1993; it is the last in this series*

Although meta-analysis is now well established as a method of reviewing evidence, an uncritical use of the technique can be very misleading. One common problem is the failure to investigate appropriately the sources of heterogeneity, in particular the clinical differences between the studies included. This paper distinguishes between the concepts of clinical and statistical heterogeneity and exemplifies the importance of investigating heterogeneity by using published meta-analyses of epidemiological studies of serum cholesterol concentration and clinical trials of its reduction. Although not without some dangers of speculative conclusions, prompted by overzealous inspection of the data to hand, a sensible investigation of sources of heterogeneity should increase both the scientific and the clinical relevance of the results of meta-analyses.

The purpose of a meta-analysis of a set of clinical trials is rather different from the specific aims of an individual trial. For example, a particular clinical trial investigating the effect of serum cholesterol reduction on the risk of ischaemic heart disease tests a particular treatment regimen, given for a specified duration to participants fulfilling certain selection criteria, using a particular definition of outcome measures. The purpose of a meta-analysis of cholesterol lowering trials is broader—that is, to estimate the extent to which serum cholesterol reduction, achieved by a variety of means, generally influences the risk of ischaemic heart disease. A meta-analysis also attempts to gain greater objectivity, generalisability, and precision by including all the available evidence from randomised trials that pertain to the issue.<sup>1</sup> Because of the broader aims of a meta-analysis, the trials included usually encompass a substantial variety of specific treatment regimens, types of patients, and outcomes. In this paper I argue that the influence of these clinical differences between trials, or clinical heterogeneity, on the overall results needs to be explored carefully.

The paper starts by clarifying the relation between clinical heterogeneity and statistical heterogeneity. It then gives examples of meta-analyses of both observational epidemiological studies of serum cholesterol concentration and clinical trials of its reduction in which exploration of heterogeneity was important in the overall conclusions reached. The discussion addresses the dangers of post hoc exploration of results and consequent overinterpretation.

### Clinical and statistical heterogeneity

To make the concepts clear, it is useful to focus on a meta-analysis where heterogeneity was found to be a problem in its interpretation. Figure 1 shows the results of 19 randomised trials investigating the use of endoscopic sclerotherapy for reducing mortality in the primary treatment of cirrhotic patients with oesophageal varices.<sup>2</sup> As is usual, the results of each trial are shown as odds ratios and 95% confidence intervals, with odds ratios less than unity representing a beneficial effect of sclerotherapy. As is noted in the original paper, the trials differed considerably in their patient selection, baseline disease severity, endoscopic tech-

nique, management of intermediate outcomes such as variceal bleeding, and duration of follow up.<sup>2</sup> So in this meta-analysis, as in almost all, there is extensive clinical heterogeneity. There were also methodological differences in the mechanism of randomisation and in the extent and handling of withdrawals and losses to follow up.

It would not be surprising, therefore, to find that the results of these trials were to some degree incompatible with one another, even when expressed on an odds ratio scale. Such incompatibility in the quantitative results is termed statistical heterogeneity. Statistical heterogeneity may be caused by known clinical differences between trials or by methodological differences, or it may be related to unknown or unrecorded trial characteristics. In assessing the direct evidence of statistical heterogeneity, the imprecision in the estimate of the odds ratio from each trial, as expressed by the confidence intervals in figure 1, has to be taken into account. The statistical question is then whether there is greater variation between the results of the trials than is compatible with the play of chance. As might be surmised from inspection of figure 1, the published test of statistical heterogeneity yielded a highly significant result ( $\chi^2_{18}=43$ ,  $P<0.001$ ), giving very substantial evidence of statistical heterogeneity.<sup>2</sup> (For the interpretation of such tests, it is useful to know that a  $\chi^2$  statistic has on average a value equal to its degrees of freedom, so a result of  $\chi^2_{18}=18.0$  would give no evidence of heterogeneity; values much larger, such as that observed for the sclerotherapy trials, give small P values and provide evidence of statistical heterogeneity.)

The existence of clinical heterogeneity would be expected to lead to at least some degree of statistical

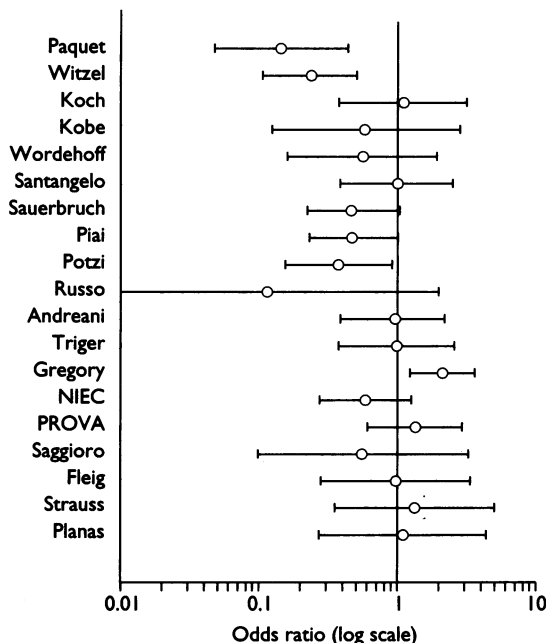


FIG 1—Odds ratios of death (and 95% confidence intervals) in 19 trials of sclerotherapy. Odds ratios less than unity represent beneficial effects of sclerotherapy. Trials are identified by principal author, referenced by Pagliaro et al<sup>2</sup>

Medical Statistics Unit,  
London School of Hygiene  
and Tropical Medicine,  
London WC1E 7HT  
Simon G Thompson, reader  
in medical statistics

BMJ 1994;309:1351-5

heterogeneity in the results. In the example of the sclerotherapy trials, the evidence for statistical heterogeneity is substantial. In many meta-analyses, however, statistical evidence for heterogeneity will be lacking and the test of heterogeneity will be non-significant. Yet this cannot be interpreted as evidence of homogeneity (that is, total consistency) of the results of all the trials included. This is not only because a non-significant test can never be interpreted as direct evidence in favour of the null hypothesis (of total consistency),<sup>3</sup> but in particular because such tests of heterogeneity have low power and may fail to detect as statistically significant even a moderate degree of genuine heterogeneity.<sup>4,5</sup>

We would of course be somewhat happier to ignore the problems of clinical heterogeneity in the interpretation of the results if direct evidence of statistical heterogeneity is lacking, and more inclined to try to understand the reasons for any heterogeneity for which the evidence is more convincing. However, the extent of statistical heterogeneity, which can be quantified,<sup>6</sup> is more important than the evidence for its existence. The guiding principle should be to investigate the influences of the specific clinical differences between studies rather than to rely on an overall statistical test of heterogeneity. This then focuses attention on particular contrasts between the trials included, which will be more powerful at detecting genuine differences—and clinically and scientifically more relevant to the overall conclusions. For example, in the sclerotherapy trials, the underlying disease severity as evidenced by the rate of bleeding varices was discussed as being potentially related to the efficacy of sclerotherapy observed.<sup>2</sup>

The most important conclusion of a meta-analysis is usually the quantitative summary of the results—for example, in terms of an overall odds ratio and 95% confidence interval. For the sclerotherapy trials, the overall odds ratio for death was given as 0.76 with a 95% confidence interval of 0.61 to 0.94.<sup>2</sup> A naive interpretation of this would be that sclerotherapy convincingly decreased the risk of death, with an odds reduction of around 25%. But what are the implications of clinical and statistical heterogeneity in the interpretation of this result? Given the clinical heterogeneity, we do not know to which endoscopic technique, to which selection of patients, or in conjunction with what ancillary clinical management such a conclusion is supposed to refer. It is some sort of “average” statement that is not easy to interpret quantitatively in relation to the benefits that might accrue from the use of a specific clinical protocol. In this particular case the evidence for statistical heterogeneity is also overwhelming and this, as stated in the original meta-analysis,<sup>2</sup> introduces even more doubt about the interpretation of any one overall estimate of effect. Even if we accept that some sort of average or typical<sup>7</sup> effect is being estimated, the confidence interval given is too narrow in terms of extrapolating the results to future trials or patients, since the extra variability between the results of the different trials is ignored.<sup>5</sup>

The clinical and scientific answer to such problems is that meta-analyses should incorporate a careful investigation of potential sources of heterogeneity. Three examples of the benefits of applying such an approach in published meta-analyses are now given. An obvious example is provided by the relation of serum cholesterol concentration and the risk of ischaemic heart disease in clinical trials; and a more speculative example is the relation of serum cholesterol concentration to the risk of cancer.

### Serum cholesterol concentration and risk of ischaemic heart disease

An extreme example of heterogeneity is evident in a recent review of the 10 largest prospective studies of serum cholesterol concentration and the risk of ischaemic heart disease in men, which included data on 19 000 myocardial infarctions or deaths from ischaemic heart disease.<sup>8</sup> Here the purpose was to summarise the magnitude of the relation between serum cholesterol and risk of ischaemic heart disease in order to estimate the long term benefit that might be expected to accrue from reduction in serum cholesterol concentrations.

Figure 2 shows the results from the 10 prospective studies. These are expressed as proportionate reductions in risk associated with a reduction in serum cholesterol of 0.6 mmol/l (about 10% of average levels in Western countries), having been derived from the apparently log-linear associations of risk of ischaemic heart disease with serum cholesterol concentration. They also take into account the underestimation of the relation of risk of ischaemic heart disease that results from the fact that a single measurement of serum cholesterol is an imprecise estimate of long term level, sometimes termed regression dilution bias.<sup>9</sup> Although all of the 10 studies showed that cholesterol reduction was associated with a reduction in the risk of ischaemic heart disease, they differed substantially in the estimated magnitude of this effect. This is clear from figure 2, and an extreme value for an overall test of heterogeneity ( $\chi^2=127$ ,  $P < 0.001$ ) is obtained. This shows that simply combining the results of these studies into one overall estimate is misleading; an understanding of the reasons for the heterogeneity is necessary.

The most obvious cause of the heterogeneity relates to the ages of the participants, or more particularly the average age of experiencing coronary events during follow up, since it is well known that the relative risk association of ischaemic heart disease with a given serum cholesterol increment declines with advancing age.<sup>10,11</sup> The data from the 10 studies were therefore divided, as far as was possible from published and unpublished information, into groups according to age at entry.<sup>8</sup> This yielded 26 substudies, the results of which were plotted against the average age of experiencing a coronary event (fig 3). The percentage reduction in ischaemic heart disease clearly decreases with age. This relation could be summarised with a quadratic regression on age, appropriately weighted to take account of the different precisions of each estimate. It was concluded that a decrease in cholesterol concentration of 0.6 mmol/l was associated with a decrease in risk of ischaemic heart disease of 54% at age 40, 39% at age 50, 27% at age 60, 20% at age 70, and 19% at age 80. In fact, there remains considerable evidence of heterogeneity in figure 3 even from this summary of

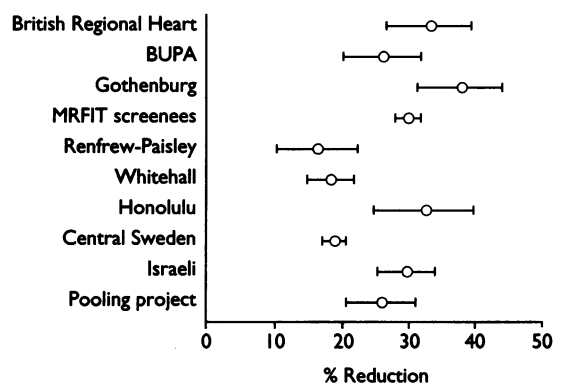


FIG 2—Percentage reduction in risk of ischaemic heart disease (and 95% confidence intervals) associated with 0.6 mmol/l serum cholesterol reduction in 10 prospective studies of men. Studies referenced by Law *et al*<sup>8</sup>

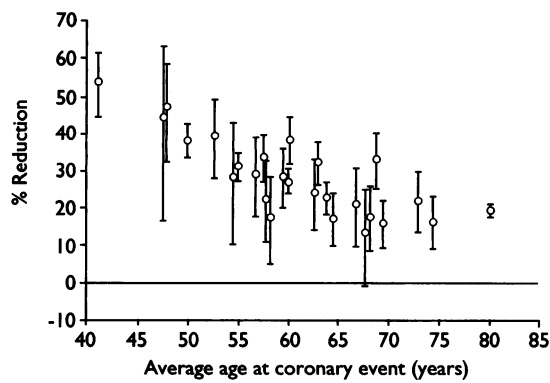


FIG 3—Percentage reduction in risk of ischaemic heart disease (and 95% confidence intervals) associated with 0.6 mmol/l serum cholesterol reduction, according to age at experiencing a coronary event

results ( $\chi^2_{23}=45$ ,  $P=0.005$ ), but it is far less extreme than the heterogeneity evident before age was considered (figure 2).

The effect on the conclusions brought about by considering age are of course crucial—for example, in considering the impact of cholesterol reduction in the population. The proportionate reductions in the risk of ischaemic heart disease associated with reduction in serum cholesterol are strongly related to age. The large proportionate reductions in early middle age cannot be extrapolated to old ages, at which more modest proportionate reductions are evident.

#### Serum cholesterol reduction and risk of ischaemic heart disease

The randomised controlled trials of serum cholesterol reduction have been the subject of a number of recent meta-analyses<sup>8 12 13</sup> and much controversy. In conjunction with the review of the 10 prospective studies just described, the results of 28 randomised trials were summarised in order to quantify the observed effect of serum cholesterol reduction on the risk of ischaemic heart disease in the short term, the trials having an average duration of about five years.<sup>8</sup> There was considerable clinical heterogeneity between the trials in the interventions tested (different drugs, different diets, and in one case surgical intervention using partial ileal bypass grafting), in the duration of the trials (0.3 to 10 years), in the average extent of serum cholesterol reduction achieved (0.3 to 1.5 mmol/l), and in the selection criteria for the patients such as pre-existing disease (for example, primary or secondary prevention trials) and level of serum cholesterol concentration at entry. As before it would seem likely that these substantial clinical differences would lead to some heterogeneity in the observed results.

Conventional meta-analysis diagrams such as figure 1 are not very useful for investigating heterogeneity. A better diagram for this purpose was proposed by Galbraith<sup>14</sup> and is shown for the risk of ischaemic heart disease in figure 4. For each trial the ratio of the log odds ratio to its standard error (the Z statistic) is plotted against the reciprocal of the standard error. Hence the least precise results from small trials appear towards the left of the figure and results from the largest trials appear towards the right. An overall (log) odds ratio is represented by the slope of the solid line through the origin in the figure. The dotted lines are positioned two units above and below the solid line and delimit an area in which, in the absence of statistical heterogeneity, the great majority (that is, about 95%) of the trial results would be expected to lie. It is thus interesting to note the characteristics of those trials that lie near or outside these dotted lines. For example, in figure 4 there are two dietary trials that lie above the

upper line and showed apparently adverse effects of serum cholesterol reduction on the risk of ischaemic heart disease. One of these trials achieved only a very small cholesterol reduction; the other had a particularly short duration.<sup>15</sup> Conversely the surgical trial, below the bottom dotted line and showing a large reduction in the risk of ischaemic heart disease, was both the longest trial and the one that achieved the greatest cholesterol reduction.<sup>15</sup> These observations add weight to the need to investigate heterogeneity of results according to extent and duration of cholesterol reduction.

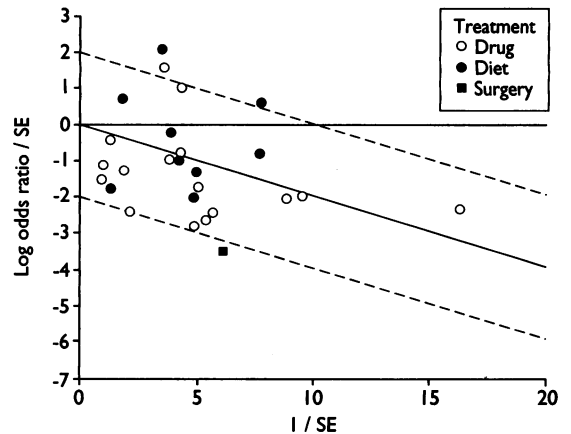


FIG 4—Galbraith plot of odds ratios of ischaemic heart disease in 28 trials of serum cholesterol reduction (see text for explanation). Two trials were omitted because of no events in one group

Figure 5 shows the results according to average extent of cholesterol reduction achieved. There is very strong evidence ( $P<0.001$ ) that the proportionate reduction in the risk of ischaemic heart disease increases with the extent of average cholesterol reduction.<sup>15</sup> A suitable summary of the trial results, represented by the sloping line in figure 5, is that the risk of ischaemic heart disease is reduced by an estimated 18% (95% confidence interval 13% to 22%) for each 0.6 mmol/l reduction in serum cholesterol concentration. Obtaining data subdivided by time since randomisation<sup>8</sup> to investigate the effect of duration was also very informative (fig 6). Whereas the reduction in ischaemic heart disease risk in the first two years was rather limited, the reductions thereafter were around 25% per 0.6 mmol/l reduction. After extent and duration of cholesterol reduction were allowed for in this way, the evidence for further heterogeneity of the results from the differential trials was limited ( $P=0.11$ ). In particular there was no evidence of further differences in the results between the drug and the dietary trials or between the primary prevention and the secondary prevention trials.<sup>8 15</sup>

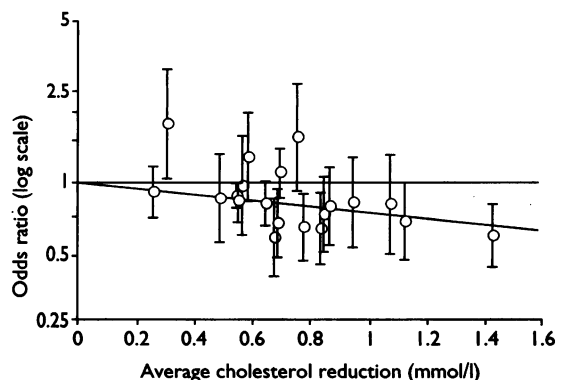


FIG 5—Odds ratios of ischaemic heart disease (and 95% confidence intervals) according to the average extent of serum cholesterol reduction achieved in each of 28 trials. Overall summary of results is indicated by sloping line. Results of the nine smallest trials have been combined

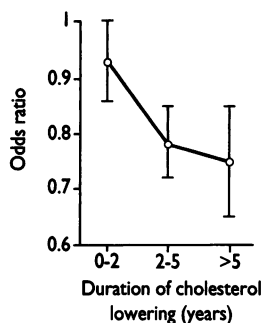


FIG 6—Odds ratios of ischaemic heart disease (and 95% confidence intervals) per 0.6 mmol/l serum cholesterol reduction in 28 trials, according to the duration of cholesterol lowering

This investigation of heterogeneity was also crucial to the conclusions reached. The analysis showed that the percentage reduction in the risk of ischaemic heart disease depends both on the extent and the duration of cholesterol reduction. Meta-analyses ignoring these factors<sup>12,13</sup> may well be misleading. It also seems that these factors are more important determinants of the proportionate reduction in ischaemic heart disease than the mode of intervention or the underlying risk of the patient. Patients at high risk of ischaemic heart disease of course have most to gain from cholesterol reduction in absolute terms for ischaemic heart disease and in both proportionate and absolute terms for all cause mortality.<sup>12</sup> Investigation of treatment benefits according to the underlying risk of the patient is one particular aspect of heterogeneity.<sup>16</sup> However, analyses that simply relate the event rate in the treated group (or the odds ratio of treated subjects to controls) to the event rate in the control group—using regression, for example—need very careful interpretation because of the problems induced by regression to the mean.<sup>17</sup>

### Serum cholesterol concentration and the risk of cancer

An association between low serum cholesterol concentrations and increased risk of cancer has been identified in a number of epidemiological prospective studies, and in 1991 a meta-analysis of the results from the 33 available prospective studies was published.<sup>18</sup> Because preclinical cancer lowers serum cholesterol, attention focused on cancers diagnosed at least two years and cancer deaths occurring at least five years after cholesterol measurement. Here these results for men (table) are discussed. The relation between cancer risk and serum cholesterol was summarised as the mean cholesterol in those subsequently developing cancer minus the mean in those who did not. Hence a negative mean difference in cholesterol corresponds to an association of low cholesterol levels with an increased risk of cancer. The overall mean difference for all the 33 studies was indeed negative,  $-0.04$  mmol/l in the table. This is significant ( $P < 0.001$ ) but small, being equivalent to about a 15% increase in the lowest fifth of the distribution of cholesterol levels relative to all the remainder of the distribution. Of interest here is that there was some evidence of statistical heterogeneity between the results of the different studies ( $\chi^2_{32}=53$ ,  $P=0.01$ ; table).

Investigation of possible sources of heterogeneity revealed that the predominant socioeconomic status of the men recruited seemed to be important (table). The association between low cholesterol and increased risk of cancer seemed most pronounced in studies of men with predominantly low socioeconomic status, moderate in studies of mixed populations, and absent or even reversed in the studies of men with high socioeconomic status. After this division of studies according to socioeconomic status, the heterogeneity was substantially less ( $\chi^2_{30}=37$ ,  $P=0.18$ ). Thus socio-

economic grouping seemed to explain a substantial part of the original heterogeneity of results.

Another subdivision considered was that according to cancer site. Where data were available, results within studies were separated into lung cancers and other cancers (table). Lung cancers accounted for most of the overall association with serum cholesterol concentration and showed similar heterogeneity according to socioeconomic status as described above. The association for other cancers was less, not statistically significant, and showed less evidence of heterogeneity. This suggests that a factor particularly related to lung cancer, presumably cigarette smoking, is involved in the explanation of these results.

Although there may be other explanations, these findings with respect to socioeconomic status and lung cancer suggest an explanation in terms of confounding by the intensity of cigarette smoking.<sup>18</sup> For example, more intensive smoking among poorer people who may have lower serum cholesterol concentrations could produce the observed results. Such an explanation requires confirmation, but the heterogeneity is important in that it tends to argue against a conclusion that low cholesterol concentrations are a direct cause of cancer.

### Discussion

As meta-analysis becomes widely used as a technique for reviewing scientific evidence, an overly simplistic approach to its implementation needs to be avoided. A failure to investigate potential sources of clinical heterogeneity is one aspect of this. As shown in the above examples, such investigation can importantly affect the overall conclusions to be drawn, as well as the clinical implications of the review. Therefore the issues of clinical and statistical heterogeneity and how to approach them need emphasis in written guidelines and in the computer software currently being developed for conducting meta-analyses.<sup>19</sup>

Discussion of heterogeneity in meta-analysis affects whether it is reasonable to believe in one overall estimate that applies to all the studies encompassed, implied by the so called fixed effect method of statistical analysis.<sup>3</sup> Undue reliance may have been put on this approach in the past, causing overly simplistic and overly dogmatic interpretation.<sup>5</sup> Although the so called random effects method of analysis<sup>6</sup> may be useful when statistical heterogeneity is present but cannot be obviously explained by clinical differences, the main focus should be on trying to understand any sources of heterogeneity that are present. In practice, however, there may be no great difference between those who advocate a fixed effect approach<sup>7</sup> and those who are more doubtful<sup>5,20,21</sup> when it comes to undertaking particular meta-analyses. For example, the recent large scale overview of early breast cancer treatment, carried out ostensibly with a fixed effect approach, includes an appropriate investigation of heterogeneity according to type and duration of treatment, dose of drug, use of concomitant therapy, age, nodal status, oestrogen receptor status, and outcome (recurrence or death).<sup>22</sup> Likewise, extensive investigation of heterogeneity was undertaken in the recent overview of antiplatelet therapy.<sup>23</sup>

Considerable dangers of overinterpretation can, however, be induced by attempting to investigate heterogeneity, since such investigations are usually inspired, at least to some extent, by looking at the results to hand. Moreover, apparent (even statistically significant) heterogeneity may always be due to chance, and searching for its causes would then be misleading. The problem is akin to that of subgroup analyses within an individual clinical trial.<sup>24</sup> However,

Prospective studies of serum cholesterol concentration and the risk of cancer in men<sup>18</sup>: cancers diagnosed at least two years and cancer deaths occurring at least five years after cholesterol measurement

	No of cancer cases	No of studies	Mean (SE) difference in serum cholesterol (mmol/l)*	Heterogeneity
All studies:				
Overall	12 516	33	-0.041 (0.009)	$\chi^2=53$ , df=32, P=0.01
Socioeconomic status:				} $\chi^2=37$ , df=30, P=0.18
High	619	4	+0.032 (0.048)	
Mixed	10 378	20	-0.030 (0.010)	
Low	1 519	9	-0.130 (0.025)	
Studies with lung cancer data:				} $\chi^2=40$ , df=18, P=0.002
All cancers	8 062	19	-0.043 (0.012)	
Lung cancers	2 239	19	-0.101 (0.022)	
Cancers other than lung	5 823	19	-0.023 (0.014)	$\chi^2=32$ , df=18, P=0.02

\*Mean cholesterol in those who subsequently developed cancers minus mean in those who did not.

the degree of clinical heterogeneity across different clinical trials is greater than that within individual trials and represents a more serious problem. Guidelines for deciding whether to believe results that stem from investigation of heterogeneity depend on, for example, the magnitude and statistical significance of the differences identified, the extent to which the potential sources of heterogeneity had been specified in advance, and indirect evidence and biological considerations which support the investigation.<sup>25</sup>

These problems in meta-analysis are greatest when there are many clinical differences but only a small number of trials available. In such situations there may be several alternative explanations of statistical heterogeneity, and ideas about sources of heterogeneity can be considered only as hypotheses for evaluation in future studies. Some of these problems may be more satisfactorily approached by basing meta-analyses on the individual patient data from each trial<sup>26</sup> rather than their summary results, so that divisions according to patients' characteristics can be made within trials and these results combined across trials.

Although clinical causes of heterogeneity have been focused on in this paper, it is important to recognise that there are other potential causes. Statistical heterogeneity may be caused by publication bias<sup>27</sup> (in that among small trials those with dramatic results may be preferentially published), by defects of methodological quality,<sup>28</sup> or even by early termination of clinical trials for ethical reasons.<sup>29</sup> For example, poor methodological quality was of concern in the meta-analysis of sclerotherapy trials discussed at the beginning of this paper. Statistical heterogeneity may also be induced by using an inappropriate scale for measuring treatment effects—for example, using absolute rather than relative differences.

Despite the laudable attempts to achieve objectivity in reviewing scientific data, considerable subjective judgment is necessary in carrying out meta-analyses. These judgments include those about which studies are "relevant" and which studies are methodologically sound enough to be included, as well as the issue of whether and how to investigate sources of heterogeneity. Such scientific judgments are as necessary in meta-analysis as they are in other forms of medical research, and skills in recognising appropriate analyses and dismissing overly speculative interpretations need to be developed. However, in many meta-analyses heterogeneity can and should be investigated so as to increase the clinical relevance of the conclusions drawn and the scientific understanding of the studies reviewed.

I thank Peter England, Rebecca Hardy, Iain Chalmers, and Douglas Altman for their constructive criticisms of a previous version of this paper.

- 1 Dickersin K, Berlin J. Meta-analysis: state-of-the-science. *Epidemiol Rev* 1992;14:154-76.
- 2 Pagliaro L, D'Amico G, Sorensen TIA, Lebrech D, Burroughs AK, Morabito A, et al. Prevention of first bleeding in cirrhosis: a meta-analysis of randomised trials of non-surgical treatment. *Ann Intern Med* 1992;117:59-70.
- 3 Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991:167-70.
- 4 Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomised clinical trials. *Statist Med* 1991;10:1665-77.
- 5 Thompson SG, Pocock SJ. Can meta-analyses be trusted? *Lancet* 1991;338:1127-30.
- 6 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986;7:177-88.
- 7 Peto R. Why do we need systematic overviews of randomised trials? *Statist Med* 1987;6:233-40.
- 8 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* 1994;308:367-73.
- 9 MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, et al. Blood pressure, stroke, and coronary heart disease. Part I, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990;335:765-74.
- 10 Manolio TA, Pearson TA, Wenger NK, Barrett-Connor E, Payne GH, Harlan WR. Cholesterol and heart disease in older persons and women: review of an NHLBI workshop. *Ann Epidemiol* 1992;2:161-76.
- 11 Shipley MJ, Pocock SJ, Marmot MG. Does plasma cholesterol concentration predict mortality from coronary heart disease in elderly people? 18 year follow-up in Whitehall study. *BMJ* 1991;303:89-92.
- 12 Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992;305:15-9.
- 13 Davey Smith G, Song F, Sheldon T. Cholesterol lowering and mortality: the importance of considering initial level of risk. *BMJ* 1993;306:1367-73.
- 14 Galbraith RF. A note on the graphical presentation of estimated odds ratios from several clinical trials. *Statist Med* 1988;7:889-94.
- 15 Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Statist Meth Med Res* 1993;2:173-92.
- 16 Brand R, Kratt H. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. *Statist Med* 1992;11:2077-82.
- 17 Senn S. Importance of trends in the interpretation of an overall odds ratio in a meta-analysis of clinical trials. *Statist Med* 1994;13:293-6.
- 18 Law MR, Thompson SG. Low serum cholesterol and the risk of cancer: an analysis of the published prospective studies. *Cancer Causes Control* 1991;2:253-61.
- 19 Oxman A. Preparing and maintaining systematic reviews. In: Sackett D, ed. *Cochrane Collaboration Handbook*. Section VI. Oxford: Cochrane Collaboration, 1994.
- 20 Meier P. Meta-analysis of clinical trials as a scientific discipline [commentary]. *Statist Med* 1987;6:329-31.
- 21 Bailey KR. Inter-study differences: how should they influence the interpretation and analysis of results? *Statist Med* 1987;6:351-8.
- 22 Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. *Lancet* 1992;339:1-15, 71-85.
- 23 Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy. I. Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ* 1994;308:81-106.
- 24 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.
- 25 Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992;116:78-84.
- 26 Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993;341:418-22.
- 27 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72.
- 28 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodologic quality are associated with estimates of treatment effects in 250 controlled trials from 33 meta-analyses. *JAMA* (in press).
- 29 Hughes MD, Freedman LS, Pocock SJ. The impact of stopping rules on heterogeneity of results in overviews of clinical trials. *Biometrics* 1992;48:41-53.

## ANY QUESTIONS

*What is the risk in France of incurring non-respiratory tuberculosis from consuming unpasteurised milk? Can we assume that those who are tuberculin positive or have been vaccinated with BCG are protected?*

The incidence of infection with *Mycobacterium bovis* in cattle in France is unknown. It is not possible, therefore, to assess the risk of contracting non-respiratory (extra-pulmonary) tuberculosis from unpasteurised milk. But if it is assumed that the position is similar to that in Britain then the risk is negligible.

Positivity on skin testing for tuberculosis is not a measure of protection. It merely indicates that the person has been exposed at some time to the tubercle bacillus or a closely related organism. There have been no studies of the protective efficacy of BCG against *M bovis* infection. But as BCG was originally derived from a virulent strain of *M bovis* its protective efficacy against this species would probably be the same as its efficacy against *M tuberculosis* (roughly 77% in Britain).—P A JENKINS, head of the Public Health Laboratory Service's *Mycobacterium Reference Unit*, Cardiff