Treating Individuals 4

Can meta-analysis help target interventions at individuals most likely to benefit?

Simon G Thompson, Julian P T Higgins

Lancet 2005; 365: 341–46

MRC Biostatistics Unit, Institute of Public Health, Cambridge CB2 2SR, UK
(Prof S G Thompson DSc, J P T Higgins PhD)

Correspondence to:
Prof Simon G Thompson
simon.thompson@mrc-bsu.cam.ac.uk

Meta-analyses of randomised trials aim to summarise the effects of interventions across many patients, and can seem remote from the clinical issue of how individual patients should be treated and which patient groups will benefit the most from treatment. One method that attempts to address this point entails relating the overall effect in every trial to summaries of patient characteristics. This is called meta-regression. The interpretation of such analyses is not straightforward, however, because of a combination of confounding and other biases. Much more useful is to compare the outcomes for patient subgroups within trials and combine these results across trials. Unfortunately this method is rarely possible using published information, so analyses of individual patient data from trials are necessary. Also, although meta-analyses generally summarise an intervention's effect as a relative risk reduction, the groups of patients with the greatest absolute risk reduction have the most to gain.

Introduction

Systematic reviews of health-care interventions are an attempt to collate information from all relevant studies and, if deemed appropriate, combine their results using meta-analysis.[1] This process inevitably brings together studies that are diverse in their designs (in terms of outcomes assessed and length of follow-up, for example), in the specific interventions used (method, intensity, and duration), and in the types of patients studied (demographic and clinical characteristics). Thus the results, based on such a broad range of evidence, can seem remote from the issue of how to treat individual patients, and even somewhat irrelevant to clinical practice.[2] Nevertheless it is incontrovertible that treatment decisions should be based on evidence when it exists, and that good quality systematic reviews provide an essential mechanism in reviewing available evidence.[3] The issue is how best to bridge the gap between evidence based on many patients and making decisions about treating individuals.

The larger randomised trials are, the less their results will be subject to chance. Many patients are needed to distinguish true treatment benefits that are clinically important, but moderate in size, from chance effects.[4] Increasing numbers of patients by combining results across trials provides a principal rationale for meta-analysis.[3] At the other extreme, n-of-1 trials attempt to isolate effective treatments for a particular individual;[5] however, such trials can only be undertaken in specific clinical situations, for example, for treatments to relieve symptoms in chronic disorders, and do not provide evidence about medical policy that can be generalised to new patients. In between these extremes lies the aim of targeting interventions by identifying subgroups of patients most likely to benefit. Subgroup analyses within a clinical trial investigate the effects of an intervention for specific groups of patients—eg, defined by their clinical characteristics—in an attempt to refine how the treatment might best be used in practice.[6] Such analyses are, however, inevitably plagued by chance effects—both wider confidence intervals due to the fewer patients involved, leading to more uncertain inferences, and false positive results arising from the multiplicity of subgroups typically investigated.[7]

Comparing patient subgroups within a meta-analysis might help to ameliorate the tension between decision making in clinical medicine and overall statements of evidence in systematic reviews. Researchers have suggested that meta-analysis should go beyond estimating one overall effect,[8,9] although this expansion has drawbacks.[10] One aim of meta-analysis should be to estimate how treatment effectiveness varies according to patients' characteristics.[11] In this article, we discuss the extent to which this aim is achievable, and investigate whether we can progress beyond the general statement that meta-analytic conclusions should be borne in mind in clinical decision making. In doing so, we need to distinguish the relative risk reductions usually summarised in meta-analyses from their implications for absolute risks, which describe how much patients benefit.

Conventional meta-analysis

To focus the discussion, we introduce a specific example. The effectiveness of platelet glycoprotein IIb/IIIa inhibitors (PGIs) in acute coronary syndromes (non-Q-wave infarction and unstable angina) has received much attention, being the subject of a Health Technology Assessment review,[12] National Institute for Clinical Excellence guidance,[13] and a Cochrane systematic review.[14] Although PGIs reduce the risk of death and myocardial infarction in patients undergoing percutaneous coronary intervention,[14] their role in acute coronary syndromes in which coronary revascularisation is not planned is more uncertain. We focus on this issue by undertaking a meta-analysis of six large randomised trials (each >1000 patients) reviewed by Boersma and colleagues.[15]

The six trials (PRISM, PRISM-PLUS, PARAGON, PURSUIT, GUSTO IV-ACS, and PARAGON-B)[16–21] included 31 402 patients with unstable angina or myocardial infarction without persistent ST-segment elevation, who were not routinely scheduled for early revascularisation. PGIs were given intravenously (bolus plus infusion) and compared with placebo or control (aspirin or heparin). The PGI drugs used varied between trials (abciximab, eptifibatide, lamifiban, and tirofiban), as did the doses and durations of infusion (24 to 120 h). Myocardial infarction was defined objectively in every study, with slightly varying criteria for cardiac enzyme concentrations. We consider the risks of death and myocardial infarction up to 30 days after randomisation, since all trials reported results at this timepoint. A total of 3530 events occurred, an average risk of 11%.

The meta-analysis yields an overall odds ratio of 0·91 (95% CI 0·85–0·98, p=0·02, figure 1). Since relative risks and odds ratios are similar for risks up to about 20%, this result corresponds to a 9% reduction in risk from the use of PGIs. The analysis is based on an assumption of a common effect across all trials. The failure to show direct statistical evidence against this assumption (the test for heterogeneity is not significant, p=0·33), does not however mean that the underlying odds ratio in every trial is in fact the same. The test for heterogeneity lacks statistical power[22] and cannot distinguish true differences between the results in the different trials from chance effects. Indeed, in view of the clinical diversity of the trials, an assumption of a common effect is highly implausible. From a clinical standpoint, it would be convenient to assume that the 9% reduction in risk applies to all patients with acute coronary syndromes. Unfortunately, this assumption would be unjustified, not only because the test for heterogeneity lacks power but also because the test only addresses differences between trials (for example because of varying treatment protocols) rather than potential differences between patients with varying characteristics.

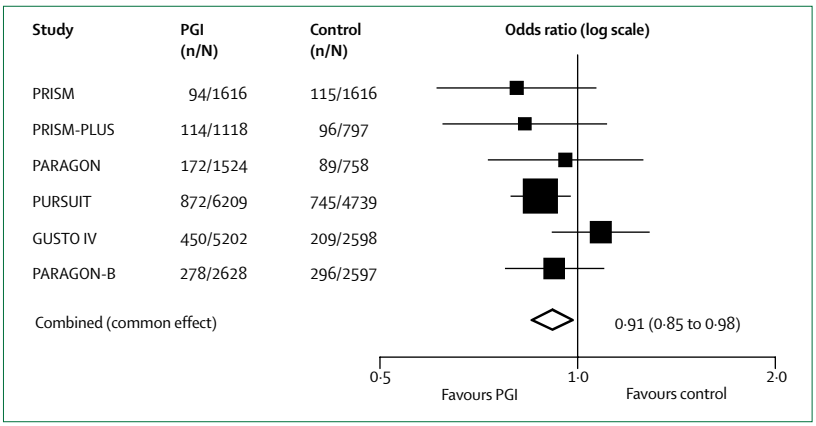

Figure 1: **Meta-analysis of six trials of platelet glycoprotein IIb/IIIa inhibitors (odds ratios [95% CI])**
n=number of deaths and myocardial infarctions up to 30 days after patients were randomly assigned treatment or control. N=total number of patients.

Absolute risks are more informative for clinical decision making than are relative measures such as the relative risk or odds ratio.[23] The absolute risk difference estimates the risk reduction that is expected on average for every patient. For the six PGI trials, a meta-analysis of the differences in the proportions of patients dying or having a myocardial infarction within 30 days yields an overall absolute risk difference of 0·89% (95% CI 0·17–1·60) in favour of PGIs. This finding corresponds to an expected number needed to treat to prevent one event of 112 (95% CI 63–590). In this case, we might argue that the absolute risk differences are as likely to be consistent between the trials as are the odds ratios, since the period of follow-up for the outcome considered is identical. In general, and especially for trials with different follow-up periods, the relative risk or odds ratio is probably more consistent between trials and patients.[24] Extrapolation to a specific patient group then involves applying the relative risk reduction from the meta-analysis to the group's baseline level of risk.[25–27]

Conventional meta-analysis thus does not effectively identify groups of patients who might benefit most from an intervention. To do this, extent of treatment benefit should be related to patients' characteristics. On the basis of published data, one way to link benefit to characteristics is to relate the treatment effect in every trial to some average characteristic of the patients in that trial (such as mean age or proportion of women). Such an analysis is called meta-regression,[28] which, although straightforward to do, is subject to substantial difficulties in interpretation.

Meta-regression

Meta-regression aims to relate the treatment effects recorded in different trials to the overall characteristics of those trials. We will consider the example of whether the effectiveness of PGIs is different between men and women. The basic characteristics of patients recruited into trials are usually reported fully in publications. For example, we can relate the odds ratio noted in every trial to the proportion of women in that study (figure 2). Meta-regression assesses the strength of the relation between the two. In this case, the estimated log odds ratio for the effect of PGIs is estimated to increase by 0·044 (SE 0·024, p=0·06) for every 1% rise in the proportion of women. This result could be taken to imply that the odds ratio for women is 81-times that in men, corresponding to studies of 100% women and 0% women respectively (exponential of 100×0·044=81). This conclusion is clearly totally implausible, even though the relation is of borderline significance.

First, we note some technical issues about undertaking meta-regression, since they are sometimes incorrectly done.[29] Odds ratios or relative risks are usually log-transformed because they can more justifiably be regarded as normally distributed. The regression also has to be weighted, taking into account not only the precision

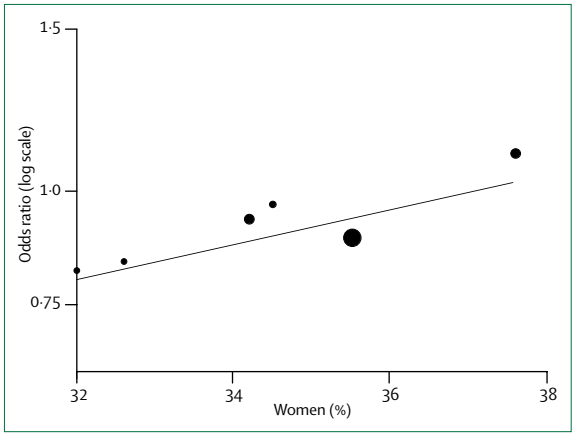

Figure 2: **Meta-regression relation of log odds ratios across trials**
Size of every circle is proportional to the precision of each log odds ratio estimate.

of every trial's result (as shown by the size of the circles in figure 2) but also the extent of residual differences between their results not attributable to the characteristic being considered.[30] Statistical software to do such analyses is now widely available.[31] To investigate directly whether the treatment effect (odds ratio or absolute risk difference, for example) varies with the baseline risk of the patients in the different trials is tempting. However, the baseline risk can usually only be measured by the reported risk in the control groups, which directly enters the calculation of the treatment effect. Therefore, simple meta-regression can give biased results.[32] Although more appropriate statistical methods can be used for such analyses,[33] results are not necessarily robust.

Even if meta-regression is undertaken correctly from a technical point of view, relations with averages of patients' characteristics are potentially misleading. First, meta-regression describes observational relations across studies, which are subject to confounding by other characteristics that vary between the trials. Even though every trial is randomised, meta-regression is only the study of the epidemiology of trials[34] and relations may well not be causal. For example, many (and probably more important) characteristics vary across the PGI trials than merely the proportion of women, but which might be correlated with it. Thus, the relation of the odds ratio with the proportion of women could be attributable to other factors. A second difficulty is the limited range of characteristics when they are averaged over all the patients in a trial. For example, across the six PGI trials the proportions of women only vary from 32% to 38%, whereas individual patients are either 0% or 100% female! Similarly, mean ages vary from 62 to 66 years, whereas the individual patients' ages have a much wider spread (typically 45–85 years), suggesting that little statistical power exists to detect relations by meta-regression.[35] In the extreme case, in which the proportions of women are identical in every trial, there is no possibility of detecting a relation. One final difficulty is that, in any systematic review, many characteristics (of trials or their patients) could be investigated by meta-regression, but there are usually only a few trials.[36] This fact leads to the likelihood of data dredging and the reporting in publications of only significant findings, which are, therefore, likely to be false positives—misleading for both clinical practice and future research.[30]

Thus, there are many reasons why meta-regression should be avoided. Its use should be restricted to investigation of differences between trials that relate to trial features (such as treatment regimen) and patients' characteristics that vary substantially across trials and not within trials, when these features have been prespecified and many trials are available.[36] Although data for average patients' characteristics are usually available, and meta-regression is easily undertaken, this does not mean that findings can be reliably interpreted. To provide a way of investigating patients' characteristics we have to move away from looking at relations across trials, to inspection of relations within trials. So we need to compare subgroups of patients within every trial (eg, men *vs* women) and then combine these results over trials.

Meta-analysis of subgroup differences

Most large trials report whether certain baseline patients' characteristics are effect modifiers, that is whether the treatment effect varies according to these characteristics. For example, treatment effects in men and women, or by age group, are calculated and presented separately. Evidence for differential effects should be assessed by a statistical test of interaction.[37] In principle, we should be able to extract this information from every trial—eg, calculate the difference in (log) odds ratios between subgroups in every trial and undertake a meta-analysis of these differences across all trials to assess the evidence for, and extent of, any overall treatment interaction. However, to undertake this analysis, sufficient numerical information must be present in the publication, enabling calculation of both the log odds ratios within subgroups and their SEs, which is uncommon in practice.

In the six PGI trials,[16–21] to what extent is such within-trial information on the sex difference in treatment effect available? Only one trial gave the basic numerical information from which the relevant quantities could be calculated for men and women. Another did not present any subgroup findings, but the remaining four provided some information for men and women separately. However, these data were presented as diagrams rather than tables, with commentary in the text restricted to the issue of significance of interactions. For characteristics other than sex, greater problems arose. For continuous characteristics, such as age, or more complex categorical variables, such as ECG findings, different groupings were used in the trial reports. Moreover, the variables chosen for presentation were not consistent between trials. For example, of the five trials presenting any subgroup findings, only two presented subgroup results by weight,

another two by smoking category, and one by previous use of β blockers. A concern here is that the selection of variables for presentation might have been determined by the results obtained, leading to false positive interactions. A final difficulty was that the outcome chosen for detailed subgroup analysis varied substantially between the trials: different combinations of death, myocardial infarction, refractory ischaemia, and unstable angina were used, at timepoints ranging from 2 days to 6 months.

The PGI trials example is no doubt typical. Even if one extracted as much information as possible out of the subsidiary publications from every trial, we would be unlikely to resolve the difficulties. Without sufficient and consistent information from every trial, a meta-analysis of within-trial subgroup findings, which would overcome some of the worst drawbacks of meta-regression, cannot be undertaken. Two solutions can be considered. The first is that tabulated information in a consistent format is requested from every set of trialists. Such a requirement might be successful, but sharing of individual patient data from different trials is usually more fruitful. This has many additional advantages[38] including the possibility to check basic data and analyses, to improve consistency between trials (for example in terms of definition of outcomes), to undertake extra analyses, and to consider confounding of subgroup effects by other individual characteristics. We now describe an individual patient data meta-analysis in the context of the PGI trials.

Individual patient data meta-analysis

The trialists from the six large PGI trials undertook a collaborative project in which data for every patient were collated centrally and analysed.[15] Again we focus on the 30-day outcome of death or myocardial infarction. A simple meta-analysis based on these data gave the overall odds ratio noted before, that is 0·91 (95% CI 0·85–0·98). Now, however, the results by subgroup could be extracted in a consistent manner. For example, by logistic regression,[15] the estimated odds ratio was 0·81 for men (a beneficial effect of PGIs) and 1·15 for women (an apparent adverse effect) as shown in figure 3. This differential treatment effect was highly significant (p<0·0001, a test of interaction based on a meta-analysis of within-trial differences). The investigators also

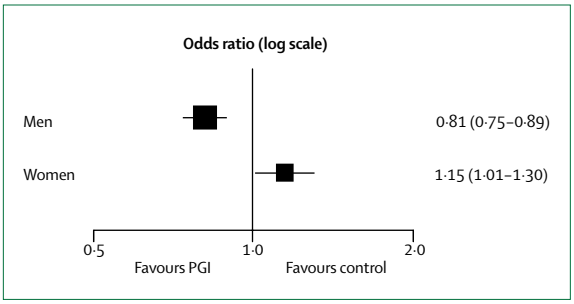

Figure 3: Odds ratios (95% CI) of death or myocardial infarction in men and women based on individual patient data

reported whether 12 other baseline characteristics modified the overall odds ratio. None was as convincing as the sex difference, but there was some evidence that the benefit of PGIs was greater in younger people (p=0·10) and in those without ST-segment depression (p=0·06).

How should such interactions, based on individual patients' data, be judged? One consideration is the extent to which the results are compatible with chance, which depends not only on the p value for the interaction test but also on how many characteristics have been investigated (which might be more than the number reported). With a simple adjustment for multiple testing,[39] one might reasonably regard the sex difference in the PGI trials as most unlikely to have arisen by chance, but judge that the differential effects by age and ST-segment depression are unconvincing because at least 13 characteristics have been investigated. The magnitude of the sex difference might be exaggerated merely because it was the most extreme among many interactions investigated. A second issue is the extent to which the findings are biologically plausible, although such arguments are prone to post-hoc speculation. Some researchers argue that qualitative interactions (treatment effects in opposite directions) are intrinsically implausible. A third point is whether the relation revealed might be attributable to other characteristics. In the case of the sex difference for the effect of PGIs, one relevant consideration is the concentration of troponin, a marker of the extent of myocardial damage. Boersma and colleagues[15] argue that, since men generally had higher concentrations of troponin than women, a sex difference might be caused by differential effects of PGIs in those with different levels of myocardial damage. However, men also had other characteristics that differed from women, such as age and prevalence of a history of myocardial infarction and diabetes.

With individual patients' data we can, in principle, investigate this type of confounding. By adjustment simultaneously for the potential confounding variables, one can see whether the sex difference becomes compatible with chance. In the case of the PGI trials, this possibility was not the case, with one exception. When adjusting for baseline troponin concentration, the sex difference was no longer evident.[15] Although such findings are observational in nature, and can be subject to residual confounding and measurement error,[40] in this case troponin might be the more important moderator of the effect of PGIs than sex per se. However, troponin data were only available for 35% of the entire population, and when restricting the analysis to this subgroup the unadjusted sex difference was no longer evident. Had full data for troponin been available from the trials, the confounding of sex and troponin concentration could have been fully addressed. With incomplete data available, the answer remains uncertain. This difficulty is typical of other individual patient data meta-analyses,

when different baseline data have been obtained in the included trials.

The technical demands of individual patient data meta-analyses are substantially greater than are those of meta-analysis or meta-regression.[41] Indeed, statistical methods need to be developed in this area. In general, we would suggest that the estimated relations between the extent of treatment benefit and patients' characteristics are derived only from within-trial information, so that confounding because of differences between trials is avoided. As discussed above, such confounding is one issue that affects meta-regression. To avoid this problem, interaction effects (eg, the difference between men and women) are calculated in every trial separately and then combined over trials. When some studies contain only women or only men, they contribute no within-trial information to the evidence about a sex difference in treatment effect and so would be omitted from the analysis. Such a method is designed to reduce bias, at the expense of losing some precision. Some multilevel model methods of analysis do not clearly separate within and between trial information, and can consequently result in misleading conclusions.[42]

The conclusions from the PGI trials' individual patient data meta-analysis might reasonably be that the proportionate risk reduction for men of 19% seems to apply reasonably uniformly across patient subgroups. For women, the results do not seem encouraging, with a reported 15% increase in risk. Whether this finding is attributable to an intrinsic difference between the sexes, or to the generally less severe myocardial damage in women, remains uncertain.

Discussion

Identification of patient groups who benefit most from an intervention is never going to be easy, since it is a task for which enormous quantities of randomised evidence are necessary. Even in large trials, apparent subgroup differences can result merely from chance. Meta-analyses of large trials based on individual patient data allow subgroups to be contrasted within trials, and for these results to be combined across trials, producing more reliable evidence. Individual patient data also allow investigation of whether treatment interactions associated with one clinical characteristic are potentially confounded by another. Attempts to target treatments by meta-regression of overall trial results and averages of patients' characteristics are generally misleading. For example, in the PGI trials, a completely unrealistic estimate of the sex difference in treatment effect was obtained from a meta-regression across trials.

Clinical decisions for the individual patient, and medical policy decisions, have always to be made, at least to some degree, on the basis of incomplete or insufficient evidence. When we do not have evidence about treatment effects in specific subgroups of patients, these decisions have to be made with evidence about overall effectiveness. We should only make different decisions for specific patient groups when strong evidence supporting these decisions becomes available. For policy decisions at a national level, cost-effectiveness has to be considered in addition to clinical effectiveness.[43] A treatment should be targeted at those for whom it is most cost effective, ideally individuals who get the greatest clinical benefit with the least use of medical resources. Generally, we are far from having sufficient information to make policy decisions on this basis.

Evidence of generally consistent relative risk reductions has been striking in some meta-analyses. For example, antiplatelet treatment produced about a 25% relative reduction in risk of serious vascular events across a wide range of patient groups,[44] and fibrinolytic treatment after myocardial infarction showed about a 20% reduction in mortality.[45] In such situations, the benefit for specific patient groups depends crucially on their baseline risk; those with low baseline risk have little to gain, but those with high baseline risk have much to gain. There are, however, some exceptions to consistent relative effects: the PGI trials apparently showed beneficial effects in men and adverse effects in women; for antiplatelet treatment, there was a lesser proportionate reduction in the risk of serious vascular events in patients with acute stroke than in other high risk groups;[44] for endarterectomy in patients with carotid stenosis, benefit was restricted to those with at least 70% stenosis.[46]

Even when relative risk reductions are used as the summary of effects in individual patient data meta-analysis, absolute risk reductions should be explicitly estimated for specific patient groups.[44] Estimation of the baseline risks for different patient groups is conventionally not part of meta-analyses. Nevertheless, it can be done, with either data from the trials themselves[46] or from external observational studies.[47] Such analyses would add to the clinical usefulness of the usual meta-analytic summaries of relative risk reductions.[48] Individual patient data meta-analysis should be seen as more than merely a gold standard method for doing simple meta-analysis. For example, patient subgroups can be consistently defined and systematically contrasted for evidence of possible differential treatment effects. Pre-specification of a limited number of such patient subgroups can help guard against the risk of false positive results. Individual patient data meta-analyses would then be explicitly designed to directly address the best targeting of interventions.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgments

We thank Eric Boersma, Colin Baigent, John Danesh, and Shah Ebrahim for their helpful comments on an earlier version of this review. Both authors are funded by the UK Medical Research Council, which was not otherwise involved with the production of this review.

References

1 Egger M, Davey Smith G, Altman DG. Systematic reviews in health care: meta-analysis in context. London: BMJ Books, 2001.

2 Wittes RE. Problems in the medical interpretation of overviews. *Stat Med* 1987; **6:** 269–80.
3 Mulrow CD. Rationale for systematic reviews. *BMJ* 1994; **309:** 597–99.
4 Peto R. Why do we need systematic overviews of randomised trials? *Stat Med* 1987; **6:** 233–44.
5 Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy: randomized trials in individual patients. *N Engl J Med* 1986; **314:** 889–92.
6 Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995; **345:** 1616–19.
7 Oxman AD, Guyatt GH. A consumers' guide to subgroup analyses. *Ann Intern Med* 1992; **116:** 78–84.
8 Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994; **309:** 1351–55.
9 Berlin JA. Benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *Am J Epidemiol* 1995; **142:** 383–87.
10 Davey Smith G, Egger M, Phillips AN. Meta-analysis: beyond the grand mean? *BMJ* 1997; **315:** 1610–14.
11 Lau J, Ioannidis JPA, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet* 1998; **351:** 123–27.
12 McDonagh MS, Bachmann LM, Golder S, Kleijnen J, Ter Riet G. A rapid and systematic review of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists in the medical management of unstable angina. *Health Technol Assess* 2000; **4:** 1–95.
13 Guidance on the use of glycoprotein IIb/IIIa inhibitors in the treatment of acute coronary syndromes (NICE Technology Appraisal Guidance No 12). London: National Institute for Clinical Excellence, 2000.
14 Bosch X, Marrugat J. Platelet glycoprotein IIb/IIIa blockers for percutaneous coronary revascularization, and unstable angina and non-ST-segment elevation myocardial infarction (Cochrane Review). *Cochrane Database Syst Rev* 2001; **4:** CD002130.
15 Boersma E, Harrington RA, Moliterno DJ, et al. Platelet glycoprotein IIb/IIIa inhibitors in acute coronary syndromes: a meta-analysis of all major randomised clinical trials. *Lancet* 2002; **359:** 189–98.
16 Platelet Receptor Inhibition in Ischemic Syndrome Management (PRISM) Study Investigators. A comparison of aspirin plus tirofiban with aspirin plus heparin for unstable angina. *N Engl J Med* 1998; **338:** 1498–505.
17 The Platelet Receptor Inhibition in Ischemic Syndrome Management in Patients Limited by Unstable Signs and Symptoms (PRISM-PLUS) Study Investigators. Inhibition of the platelet glycoprotein IIb/IIIa receptor with tirofiban in unstable angina and non-Q-wave myocardial infarction. *N Engl J Med* 1998; **338:** 1488–97.
18 The PARAGON Investigators. International, randomized, controlled trial of lamifiban (a platelet glycoprotein IIb/IIIa inhibitor), heparin, or both in unstable angina. *Circulation* 1998; **97:** 2386–95.
19 The PURSUIT Trial Investigators. Inhibition of platelet glycoprotein IIb/IIIa with eptifibatide in patients with acute coronary syndromes. *N Engl J Med* 1998; **339:** 436–43.
20 GUSTO IV-ACS Investigators. Effect of glycoprotein IIb/IIIa receptor blocker abciximab on outcome in patients with acute coronary syndromes without early coronary revascularisation: the GUSTO IV-ACS randomised trial. *Lancet* 2001; **357:** 1915–24.
21 Global Organization Network (PARAGON-B) Investigators. The platelet IIb/IIIa antagonist for the reduction of acute coronary syndrome events. Randomized, placebo-controlled trials of titrated intravenous lamifiban and acute coronary syndromes. *Circulation* 2002; **105:** 316–21.
22 Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Stat Med* 1998; **17:** 841–56.
23 Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! *Evidence-based Medicine* 1996; **1:** 164–66.
24 Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002; **21:** 1575–600.
25 Bailey KR. Generalizing the results of randomized clinical trials. *Control Clin Trials* 1994; **15:** 15–23
26 Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses: sometimes informative, usually misleading. *BMJ* 1999; **318:** 1548–51.
27 Glasziou PP, Irwig LM. An evidence based approach to individualizing treatment. *BMJ* 1995; **311:** 1356–59.
28 Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Stat Med* 1995; **14:** 395–411.
29 Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999; **18:** 2693–708.
30 Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; **21:** 1559–73.
31 Sharp S. Meta-analysis regression. *Stata Technical Bulletin* 1998; **42:** 16–22.
32 Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. *BMJ* 1996; **313:** 735–38.
33 Sharp SJ, Thompson SG. Analysing the relationship between treatment benefit and underlying risk in meta-analysis: comparison and development of approaches. *Stat Med* 2000; **19:** 3251–74.
34 Sterne JA, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002; **21:** 1513–24.
35 Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol* 2002; **55:** 86–94.
36 Higgins J, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy* 2002; **7:** 51–61.
37 Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; **355:** 1064–69.
38 Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Stat Med* 1995; **14:** 2057–79.
39 Altman DG. Practical statistics for medical research. London: Chapman and Hall, 1991.
40 Phillips AN, Davey Smith G. The design of prospective epidemiological studies: more subjects or better measurements? *J Clin Epidemiol* 1993; **46:** 1203–11.
41 Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Stat Med* 2001; **20:** 2219–41.
42 Thompson SG, Turner RM, Warn DE. Multilevel models for meta-analysis, and their application to absolute risk differences. *Stat Methods Med Res* 2001; **10:** 375–92.
43 National Institute for Clinical Excellence. Guide to the methods of technology appraisal. http://www.nice.org.uk/ (accessed August, 2003).
44 Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 2002; **324:** 71–86.
45 Fibrinolytic Therapy Trialists' 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.
46 Rothwell PM, Warlow CP. Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. *Lancet* 1999; **353:** 2105–10.
47 Yusuf S, Zucker D, Peduzzi P, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. *Lancet* 1994; **344:** 563–70.
48 Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994; **343:** 1209–11.