International Journal of Technology Assessment in Health Care, 15:2 (1999), 304–315.

DERIVING TREATMENT RECOMMENDATIONS FROM EVIDENCE WITHIN RANDOMIZED TRIALS

The Role and Limitation of Meta-Analysis

Nick Freemantle
James Mason
University of York

Martin Eccles
University of Newcastle upon Tyne

Abstract

Meta-analysis is commonly used in reviews of the effectiveness of medical technologies, but this approach has not been used in direct support of guidelines development groups. This paper describes the approach of the North of England Guidelines Development Project in describing the evidence using meta-analyses that were conducted explicitly to address questions on the choice of therapy raised by the guidelines development groups. Particular emphasis is placed on the context within which the contributing trials were conducted and the extent to which systematic differences between trials (heterogeneity) was observed, described, and explained. There is a trade-off between internal and external validity for different metrics when presenting the results of trials. More interpretable metrics, such as risk differences or weighted mean differences, are confounded by study design issues and strong assumptions. More robust measures such as odds ratios or standardized weighted mean differences are difficult to interpret physically. Individual patient data may prove particularly helpful in addressing pivotal questions on the magnitude of effects of interventions, though accessing and reanalyzing these data requires a substantial investment in time and other resources.

Keywords: Meta-Analysis, Evidence-based Guidelines, Treatment Recommendations, Randomized Controlled Trials

The development of a systematic approach to abstracting evidence from clinical trials, coupled with the availability of a range of methods for the statistical pooling

The North of England Guidelines Development Project was funded for this work under the Prescribing Research Initiative by the Research and Development Directorate of the Department of Health. Methodological development work on the translation of research results into practice was funded by the NHS Standing Group on Health Technology Assessment. We are grateful to Anne Burton for her administrative support in developing the guidelines and for her help in preparing the paper.

of data, has led to an explosion in the number of meta-analyses in the published literature. Although these may provide more precise and valid estimates of treatment effect, they should not be thought of as an end in themselves. In order to maximize their usefulness, it is important that they are viewed as a method of summarizing evidence within a process aimed at delivering treatment or deriving policy recommendations. Viewed in this light, meta-analysis becomes an important tool, informing a broader guidelines development process (16), although, to date, no U.K. guidelines have made extensive use of this approach.

This paper describes the methodological issues addressed by the North of England Guidelines Development Project while developing evidence-based guidelines for primary care in four clinical areas (10;29;30;31;32). These were angiotensin-converting enzyme (ACE)-inhibitors in the primary care treatment of adults with symptomatic heart failure; aspirin for the secondary prophylaxis of vascular disease in primary care; the choice of antidepressants for depression in primary care; and nonsteroidal anti-inflammatory drugs (NSAIDs) versus basic analgesia in the treatment of osteoarthritis. The general principles used in the process are addressed elsewhere (10). The purpose of this paper is to examine in more detail the approaches to formal summary of the evidence base (meta-analysis) that arose during the period of guideline development across four major clinical areas.

The following sections examine four main questions:

1. Establishing what a treatment can achieve;
2. Fixed versus random effects and the importance of exploring heterogeneity;
3. Choice of metric to maximize interpretation; and
4. Estimating the magnitude of effects.

ESTABLISHING WHAT A TREATMENT CAN ACHIEVE

Within each of the guidelines, there was a need to establish what the various drugs could achieve and under what circumstances. Construct validity describes the extent to which an experiment may lead to an understanding of the action and interaction of a treatment effect (7), which is distinct from internal validity, which describes the extent to which a trial measures what it is intended to measure. Frequently, there is a trade-off in trials between high construct validity and generalizability, as the degree of experimental control required to ensure the former (such as double-blinding or extra investigation) often leads to important differences from the realities of routine care (12).

Evidence-based guideline recommendations require information on what interventions can do, and the extent to which potential benefits may be realized in routine clinical practice. Phase III drug trials, typically double-blind and carefully controlled, comparing an intervention with a standard comparator and/or placebo, tend to have high construct and internal validity by virtue of their design, and thus provide useful information on what may be achieved in what are often near ideal conditions. However, large, pragmatically designed, "real world," trials are required to provide estimates of the extent to which potential effects identified in phase III trials may be realized in routine practice (46). Such real world trials are frequently not available.

In three of the four clinical areas examined, there were pre-existing meta-analyses that adequately described the underlying construct and could be updated with additional trials identified after their publication (3;4;14;42). These analyses

provided strong evidence that ACE inhibitors could be effective in the treatment of heart failure (14), that antiplatelet therapy could be effective in reducing the risk of thrombosis in patients at raised vascular risk (4), and that there are no large differences in the effectiveness or tolerability of different antidepressants (3;42). Although this information was important, it fell short of what was required for deriving recommendations on which patients should be treated and how.

Although up to 123 phase III randomized trials were included in analyses comparing newer and older antidepressants in the antidepressant guideline, only one real world trial was found (40); this study was conducted in a population that was not directly relevant for U.K. primary care. The trial also had a number of methodological weaknesses (12). The importance of such considerations of design and setting was highlighted by the dichotomy between patient characteristics in these trials. For example, the phase III trials required subjects' condition to be both severe and chronic, typically requiring a minimum of 18 points on the Hamilton Depression Rating Scale for inclusion, and for this condition to be continued for a number of days of single-blind placebo "wash out" treatment. Entry requirement to the real world trial was based on the physician's decision to initiate drug treatment for depression. At entry, the average patient in that trial had a Hamilton Depression Rating Scale score of only 13 points. It is not clear how to interpret these data, since U.K. primary care patients treated for depression are likely to be less severe than the selected groups in the phase III trials, but the context of care is clearly quite different from that in which patients are treated in the single flawed U.S. trial.

In contrast, generating evidence-based recommendations on the use of ACE inhibitors in heart failure was relatively straightforward since a single large trial included patients of similar severity, demonstrated, for example, through similarities in hospitalization rates, as patients in U.K. primary care (41).

In summary, clinical trials frequently feature an explanatory design geared to give a precise estimate of small changes in treatment effect. Meta-analyses of such trials are necessary to provide a robust overview of treatment efficacy, but provide only part of the information required to guide treatment decisions for individual patients or set general guidelines. In the four areas we examined, at least, none adequately addressed issues of external validity nor described treatment effects in ways that may aid interpretation, a point that is described in some detail below. Although helpful, the manner in which available meta-analyses have been conducted was not directly suited to the development of evidence-based clinical practice guidelines, and further work is needed.

FIXED VERSUS RANDOM EFFECTS: EXPLORING HETEROGENEITY

The choice of model for meta-analysis remains controversial (44), with advocates of both fixed (15) and random effects (26) in the literature. Fixed effects models, such as the Mantel-Haenzsel variance, assume a single underlying treatment effect (38), while random effects models assume a distribution (usually Gaussian) of treatment effects and that studies available are a representative sample of all studies (9).

Although random effects models frequently provide the most appropriate approach for synthesizing data from trials with apparently heterogeneous effects, they are hard to interpret clinically. Some exploration of the potential causes of heterogeneity can be helpful (45). It is sometimes referred to as subgroup analysis

(33); however, the exploration of heterogeneity in meta-analyses differs from subgroup analysis in individual trials with randomization within trials being preserved, although subsequent groupings of studies are observational since studies will differ on a range of characteristics. Firm conclusions may be drawn where different estimates of effect are found through grouping trials on the basis of important characteristics derived prior to randomization. Caution is required when some characteristic of trials after randomization is used to categorize trials, because bias may be introduced, notably through regression to the mean (39), but also through data-driven analyses.

In the guidelines on the use of ACE inhibitors in the treatment of heart failure, 15 trials were located that addressed the question of whether patients soon after myocardial infarction benefitted from treatment with ACE inhibitors (1;2,6;11;13;17;20;21;23;24;25;28;35;36;43). This analysis involved nearly 10,000 deaths in more than 100,000 patients randomized to receive an ACE inhibitor or control treatment after myocardial infarction, and there was evidence of a reduction in overall mortality from treatment (odds ratio, 0.91; 95% CI, 0.87 to 0.95) (37;38). However, substantial heterogeneity was found ($Q = 24.1$; $df = 14$; $p = .045$) (9). Applying a random effects model incorporated this heterogeneity into the overall estimate of treatment effect (9), providing an estimate of the reduction of the odds of death of 0.88 (95% CI, 0.82 to 0.96).

Dividing trials of patients with recent myocardial infarction into two groupings on the basis of whether an explicit sign of left ventricular dysfunction was an entry criterion provided clear insights into the likely effects of ACE inhibition (Figure 1). For the 11 trials that did not require left ventricular dysfunction as an inclusion criterion, the odds ratio for the reduction in mortality was 0.94 (95% CI, 0.89 to 0.98), while for patients in the four trials with signs of left ventricular dysfunction as an entry criterion, the odds ratio was 0.74 (95% CI, 0.66 to 0.83). The test for heterogeneity, although weak in situations where data were sparse, provided support for this analysis in the unusual circumstance of considerable replication within trials, giving an estimate of heterogeneity for the studies that did not specify left ventricular dysfunction of $Q = 9.73$; $df = 10$; $p = .46$, and for those specifying left ventricular dysfunction of $Q = 0.75$; $df = 3$; $p = .86$.

As about 40% of patients develop left ventricular dysfunction after myocardial infarction, it appears that benefit is concentrated in this group. No good rationale for the expectation that ACE inhibitors may improve survival in patients without left ventricular dysfunction was located to support this conclusion. The overall reduction in mortality for patients with recent myocardial infarction and evidence of left ventricular dysfunction was similar to that found in placebo controlled trials of patients with heart failure but who had not recently experienced an infarct (odds ratio, 0.76; 95% CI, 0.67 to 0.86): $Q = 32.52$; $df = 39$; $p = .76$). Thus, exploring heterogeneity using the approach suggested by Thompson (45) enabled a further understanding of the likely benefits of ACE inhibitors in specific patient groups, extending evidence-based recommendations to those who had recently experienced a myocardial infarction.

CHOICE OF METRIC TO MAXIMIZE INTERPRETATION

The choice of metric used to describe results has substantial impact upon both the interpretability of meta-analyses finding and the robustness of the results.

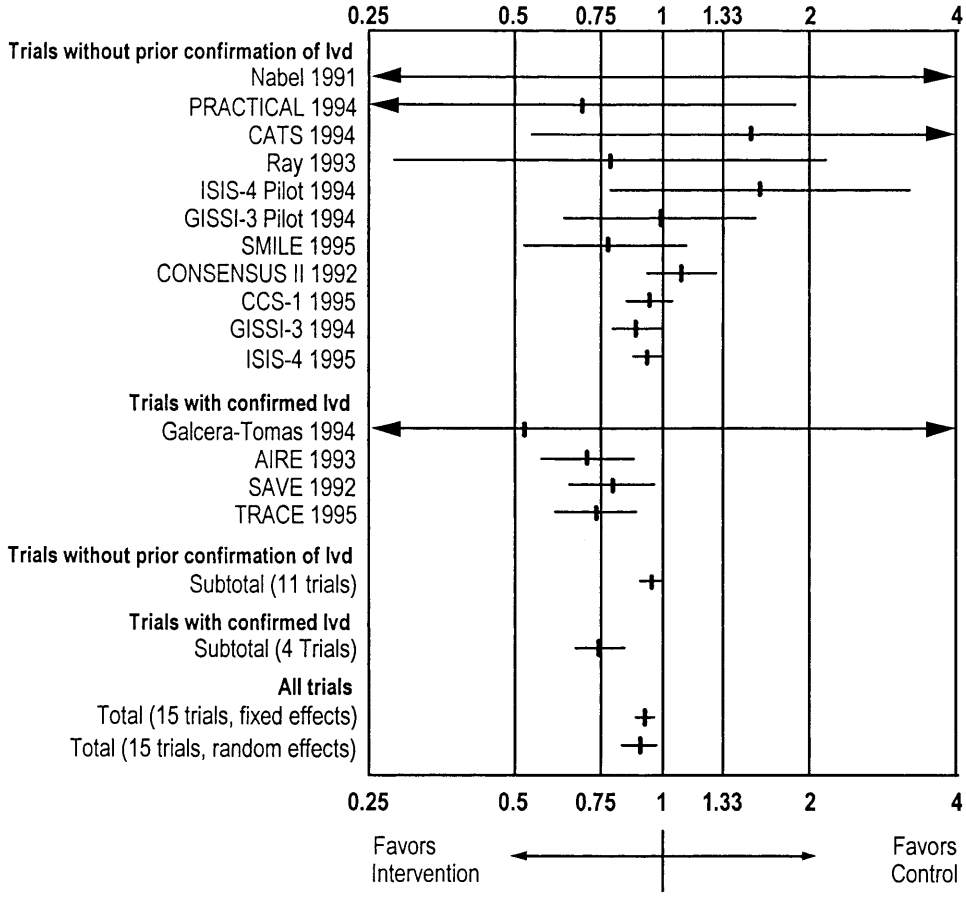

Figure 1. Pooled odds ratio and approximate 95% confidence intervals for trials including patients early post myocardial infarction grouped by left ventricular function as an entry criteria.

Binary Outcomes

For binary outcomes (such as alive/dead), metrics commonly used include odds ratios, risk ratios, and risk reductions. DerSimonian and Laird (9) suggest that odds ratios and risk ratios suffer from a lack of interpretability, and "by far the most intuitively appealing measure for trials of clinical efficacy is the risk difference, since it measures actual gains which can be expected in terms of percentages of patients treated." This observation is reflected in the interest in "numbers needed to treat (NNT)" which is the inverse of the risk difference and is advocated by some as a more interpretable measure (5;8). Although the estimation of crude odds ratios and their confidence intervals have some statistical advantages over risk ratios, it is rare to find qualitative differences between results achieved using these two metrics, and risk ratios are slightly easier to interpret, at least superficially.

Odds ratios and risk ratios are particularly useful when examining the level of effect of a treatment at different levels of absolute risk. A problem with the risk difference is that it is confounded by design effects, such as length of follow up, in a way that a risk or odds ratio is not. Thus, there remain major difficulties in interpreting risk differences and numbers needed to treat, and these are seen clearly

Table 1. Acute MI and Stable Angina: Comparison of Metrics

Trials	Metric	Effect	95% CI		Q
Acute MI	OR	0.71	0.65	0.77	Q = 7.30; *df* = 8; *p* = .51
Stable angina	OR	0.67	0.53	0.85	Q = 2.80; *df* = 5; *p* = .73
Acute MI	RD	−0.038	−0.047	−0.028	Q = 6.12; *df* = 8; *p* = .63
Stable angina	RD	−0.045	−0.071	−0.018	Q = 7.48; *df* = 5; *p* = .19
Acute MI	IRD (month)	−0.033	−0.043	−0.024	Q = 6.44; *df* = 8; *p* = .60
Stable angina	IRD (year)	−0.007	−0.017	0.004	Q = 6.76; *df* = 5; *p* = .24

Abbreviations: MI = myocardial infarction; OR = odds ratio; RD = risk difference; IRD = incidence risk difference.

when addressing the appropriate use of antiplatelet therapy. From nine trials of antiplatelet therapy in patients with acute myocardial infarction, the crude Mantel-Haenszel odds ratio to avoid a vascular event or death is 0.71 (95% CI, 0.65 to 0.77). Similarly, the crude odds ratio from six trials of patients with stable angina is 0.67 (95% CI, 0.53 to 0.85). Estimates of pooled risk difference are also similar (Table 1). These equate to a number needed to treat to avoid a subsequent vascular event or death of 26 for acute myocardial infarction or 22 for stable angina, though these examples include trials ranging in planned follow-up from 1 to 12 months for acute myocardial infarction, and 0.5 to 4.5 years for treatment of stable angina. This renders the concept of the NNT, with its need for a specified time interval, meaningless.

A method for addressing this has been suggested by Ioannidis and colleagues (19), which involves estimating a pooled random effects incidence risk difference, or in other words, adjusting the pooled estimates of risk difference by the time of exposure in some convenient unit (e.g., a month or year). The magnitude of effects found in these analyses are expressed more satisfactorily with this approach, with a time-adjusted risk difference varying from a percent reduction of 3.3% per month (95% CI, 2.4% to 4.3%) for acute myocardial infarction to 0.7% per year (95% CI, 1.7% to 0.04%) for stable angina. These equate to a number needed to treat to avoid a cardiovascular event or death of 30 for acute myocardial infarction from 1 month of treatment, or 143 for stable angina for 1 year of treatment.

The proposed solution, providing an estimate of the crude risk reduction adjusted by time, has the disadvantage that it undermines replication in the trials included, since the denominator becomes the period of time selected rather than the number of patients. Studies that have longer follow-up make a proportional contribution to the results greater than those with shorter follow-up, and the independence assumption may be lost with a number of treatment periods potentially derived from the same patient. However, the trade-off is that the less robust estimator expresses results in a way that can more readily be interpreted.

Continuous Outcomes

The intuitive way to combine differences in continuous measures between treatment and control groups of individuals trials is to pool trial findings to calculate a weighted mean difference (i.e., the differences from each trial are weighted by the inverse of study variance and then a weighted average calculated). The pooled estimate retains the same physical interpretation as the original measurements and is thus relatively easy to interpret clinically for many outcomes, but there are often concerns about the validity of pooling continuous measures where there is a substantial

subjective component involved in measurement. For example, pain scores used in the evaluation of NSAID treatment for osteoarthritis may be inconsistently presented and perceived by different trialists and patient cohorts. More fundamentally, the use of a weighted mean difference requires the assumption that the same underlying population standard deviation applies to the difference of means observed in each study, making any differences between studies attributable solely to measurement error. In practice, there are many reasons why the population standard deviation will not be common across clinical trials: these include differences in patient selection, the context of care, and the mode of intervention. Since variances can only be estimated from the data, it is unclear (particularly when data are sparse) to what extent the assumption of a common population standard deviation between studies may bias the pooled result. As it does not make this assumption, a standardized weighted mean difference may be the most useful measure to establish that an observed effect is unlikely to be due to the play of chance.

Hedges and Olkin (18) describe a robust approach to calculate a standardized weighted mean difference, in which standardized effect sizes are based upon the pooled within-study variance and adjusted for small sample bias. Since there is no correction for small sample bias when calculating weighted mean differences (27), this method may slightly overestimate the magnitude of treatment effect when data are sparse (18), but standard deviation units are difficult to interpret clinically.

Presented alongside the standardized weighted mean difference, an estimate of the weighted mean difference may provide the most interpretable practical estimate of an interventional in practice. Nonetheless, analysts should consider the appropriateness of assumptions (a common underlying variance, consistent use, or measures) when presenting a weighted mean difference. As with binary outcomes, more robust estimation is gained at the expense of simplicity of interpretation.

ESTIMATING THE MAGNITUDE OF EFFECTS

An advantage of the meta-analytic approach for developing evidence-based treatment recommendations is that it provides an objective estimate of effect based upon all the relevant available studies. None of the metrics discussed above are suitable for estimating the total benefit of health care interventions in terms of attributable changes in patient health status. In the development of evidence-based treatment guidelines, estimates of the health gains are required alongside the resource implications and costs of achieving change. Health gains are ideally expressed as years of life saved and adjusted for patient quality of life.

In the ACE inhibitor guideline, the development group was interested in the increased survival attributable to the drug. In the SOLVD treatment trial (41), in which patients were randomized to enalapril (an ACE inhibitor) or control, the relative risk of death was 0.89 (95% CI, 0.80 to 0.98). The risk difference was −0.045 (95% CI, −0.083 to −0.008) over up to 48 months of treatment, which translates to a number needed to treat with enalapril for a single year to prevent one death of about 77. Similarly, the hazard ratio based upon a stratified log rank test (with the 23 treatment centers as strata) is 0.84 (95% CI, 0.74 to 0.96) (34). However, the outcome of interest for the guidelines development group was best expressed in terms of the restricted mean, providing an estimate of the difference between survival curves for the treated and control groups (Figure 2). The health gain can be estimated as equaling 2.44 months of life during the 4-year period of the trial (22). To calculate an accurate estimate of the restricted mean, the original

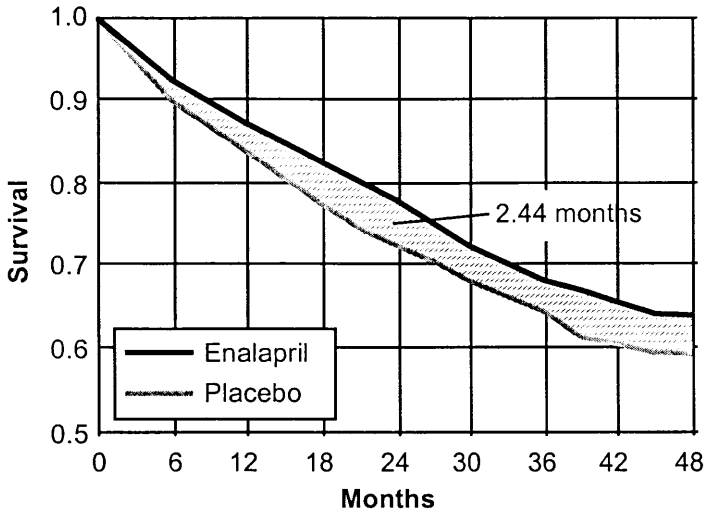

Figure 2. Survival curve derived from the SOLVD treatment trial.

patient data were required — these data are rarely available to guideline developers and even when available, require the investment of considerable time and other resources to re-analyze and describe. Since the treatment curves in the SOLVD trial have not converged (nor are they converging) at the end of follow-up, the restricted mean is clearly an underestimate of the overall treatment effect.

Arguably, the preoccupation with statistical power and significance in trials has led to a reduced emphasis on estimation of the magnitude of important health benefits: information that may be required to assess the value of an intervention to patients. In the case of the SOLVD treatment trial, the best that can be done in the knowledge that ultimately the survival curves must converge is to assume a shape for that convergence. Therefore, if it is assumed that the survival curve converges at the rate it diverged, the average patient in the SOLVD treatment trial derives *at least* a 2.44-month benefit in terms of extended life from ACE inhibition, but this benefit may well be about 5 months or potentially more.

CONCLUSIONS

The experience of the North of England Guidelines Development Group in deriving treatment recommendations based explicitly on information relating to effectiveness and efficiency demonstrated the limitations of many commonly used methods for deriving and describing results. Paradoxically, those methods that were most robust normally provided results that were least interpretable at a practical level. The principles that were used to overcome these limitations are summarized in Table 2.

In general, it is most helpful to use a range of descriptors of effects and to attempt to interpret the information that each of these provide in the light of its methodological limitations. During the development of the clinical practice guidelines, we used the most robust approach available to establish the potential usefulness of an intervention, and then less robust but more interpretable approaches were used to describe the likely impact of these results in practice. Where available, large trials providing survival curves, from which estimates of life-years gained could be derived, proved the most helpful in estimating likely treatment

Table 2. Using Evidence To Support Treatment Guidelines

- Summarize all available relevant evidence, using formal statistical methods to pool the results from similar trials.
- Examine heterogeneity between studies both through the use of random effects approaches to pooling data and through the examination of differences in the patient population included in trials.
- Use the most robust available metrics to describe the underlying effects of treatments and use interpretable approaches to describe the magnitude of treatment effects.
- Where necessary, seek unpublished data or individual patient data from the sponsors of trials to aid adequate interpretation.

benefits, especially when the findings of these trials were supported by broader meta-analyses. Reanalysis of these data required a substantial investment of time and other resources.

Two commonly advocated methods of describing results, risk differences (or the inverse: numbers needed to treat) and weighted mean differences, create difficulties in meta-analysis. Their interpretability may be problematic due to different length of follow-up or severity in trials with binary outcomes, or strong assumptions on the shape of study distributions and small sample bias in meta-analyses of continuous variables. The random effects incidence risk difference proved useful, in conjunction with more traditional metrics, although this approach to pooling data undermines the structure of randomized trials and is not, on its own, robust. Similarly, the weighted mean difference may not be robust on its own, but in addition, was only rarely helpful because of differences in outcome measures used between trials. Nevertheless, this approach has the potential to provide relatively interpretable results alongside the more robust standardized weighted mean difference.

Greater attention should be paid to the interpretation of results from trials, particularly the need to establish the value of an intervention in real world settings and thus guide treatment policy. This change in emphasis may be predicted over the next few years, as the pool of potentially helpful, but resource-intensive, interventions continues to grow, and it is recognized that greater efficiency in the allocation of resources is required if patients are to derive optimum benefit from technological developments in health care.

REFERENCES

1. The Acute Infarction Ramipril Efficacy (AIRE) study investigators. Effect of ramipril on mortality and morbidity of survivors of acute myocardial infarction with clinical evidence of heart failure. *Lancet*, 1993, 342, 821–28.
2. Ambrosioni, E., Borghi, C., & Magnani, B., for the Survival of Myocardial Infarction Long Term Evaluation (SMILE) Investigators. The effect of angiotensin-converting enzyme inhibitor zofenopril on mortality and morbidity after anterior myocardial infarction. *New England Journal of Medicine*, 1995, 332, 80–85.
3. Anderson, I. M., & Tomsenson, B. M. Treatment discontinuation with selective serotonin reuptake inhibitors compared with tricyclic antidepressants: A meta-analysis. *British Medical Journal*, 1995, 310, 1433–38.

4. 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. *British Medical Journal*, 1994, 308, 81–106.
5. Chatellier, G., Zapletal, E., Lemaitre, D., Menard, J., & Degoulet, P. The number needed to treat: A clinically useful monogram in its proper context. *British Medical Journal*, 1996, 312, 426–29.
6. Chinese Cardiac Study Collaborative Group. Oral captopril versus placebo among 13,634 patients with suspected acute myocardial infarction: Interim report from the Chinese Cardiac Study (CCS-1). *Lancet*, 1995, 345, 686–87.
7. Cook, T. D., & Campbell D. T. *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally, 1979.
8. Cook, R. J., & Sackett, D. L. The number needed to treat: A clinically useful measure of treatment effect. *British Medical Journal*, 1995, 310, 452–54.
9. DerSimonian, R., & Laird, N. Meta-analysis in clinical trials. *Controlled Clinical Trials*, 1986, 7, 177–88.
10. Eccles, M., Freemantle, N., & Mason, J. M. Methods of developing guidelines for efficient drug use in primary care: North of England Evidence Based Guidelines Development Project. *British Medical Journal*, 1998, 316, 1232–35.
11. Foy, S. G., Crozier, I. G., Turner, J. G., et al. Comparisons of enalapril versus captopril on left ventricular function and survival three months after acute myocardial infarction (the 'PRACTICAL' study). *American Journal of Cardiology*, 1994, 73, 1180–86.
12. Freemantle, N., & Drummond, M. F. Should clinical trials with concurrent economic analyses be blinded? *JAMA*, 1997, 277, 63–64.
13. Galcera-Tomas, J., Antonio Nuno de la Roas, J., Torres-Martinez, G., et al. Effects of early use of captopril on haemodynamics and short-term ventricular remodelling in acute anterior myocardial infarction. *European Heart Journal*, 1993, 14, 259–66.
14. Garg, R., & Yusuf, S. Overview of randomized trials of ACE inhibitors on mortality and morbidity in patients with heart failure. *JAMA*, 1995, 273, 1450–56.
15. Greenland, S. Invited commentary: A critical look at some popular meta-analytic methods. *American Journal of Epidemiology*, 1994, 140, 290–96.
16. Grimshaw, J. M., Eccles, M. P., & Russell, I. T. Developing clinically valid guidelines. *Journal of Evaluation in Clinical Practice* 1995, 1, 37–48.
17. Gruppo Italiano per lo Studio della Sopravvivenza nell'infarto Miocardico (GISSI-3). GISSI-3: Effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. *Lancet*, 1994, 343, 1115–22.
18. Hedges, L. V., & Olkin, I. *Statistical methods for meta-analysis.* London: Academic Press, 1985.
19. Ioannidis, J. P. A., Cappelleri, J. C., Lau, J., et al. Early or deferred therapy in HIV infected patients without an AIDS defining illness: A meta-analysis. *Annals of Internal Medicine,* 1995, 122, 856–66.
20. ISIS-4 (Fourth International Study of Infarct Survival) Pilot Study Investigators. Randomized control trial of oral captopril, or oral isosorbide mononitrate and of intravenous magnesium sulphate started early in acute myocardial infarction: Safety and haemodynamic effects. *European Heart Journal*, 1994, 15, 608–19.
21. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. ISIS-4: A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet,* 1995, 345, 669–85.
22. Karrison, T. G. Use of Irwin's restricted mean as an index for comparing survival in different treatment groups: interpretation and power considerations. *Controlled Clinical Trials,* 1997, 18, 151–67.
23. Kingma, J. H., van Gilst, W. H., Peels, C. H., et al., for the CATS Investigators. *European Heart Journal,* 1994, 15, 898–907.

24. Kober, L., Torp-Pedersen, C., Carlesen, J. E., et al., for the Trandolapril Cardiac Evaluation (TRACE) Study Group. *New England Journal of Medicine*, 1995, 333, 1670–76.
25. Latini, R., Avanzini, R., de Nicolao, A., Rocchetti, M., and the GISSI-3 Investigators. Effects of lisinopril and nitroglycerin on blood pressure early after myocardial infarction: The GISSI-3 pilot study. *Clinical Pharmacological Therapeutics*, 1994, 56, 680–92.
26. Mosteller, F., & Colditz, G. A. Understanding research synthesis (meta-analysis). *Annual Review of Public Health*, 1996, 17, 1–23.
27. Mulrow, C. D., & Oxman, A. D. (eds). *Cochrane Collaboration Handbook* (updated December 9, 1996). Available in The Cochrane LIbrary (database on disk and CDROM). The Cochrane Collaboration, Issue 1. Oxford: Update Software, 1997. Updated quarterly.
28. Nabel, E. G., Topo, E. J., Galeana, A., et al. A randomized placebo-controlled trial of combined early intravenous captopril and recombinant tissue-type plasminogen activator therapy in acute myocardial infarction. *Journal of the American College of Cardiology,* 1991, 17, 467–73.
29. North of England Guidelines Development Project. *ACE-inhibitors in the primary care management of adults with symptomatic heart failure.* Newcastle upon Tyne: Centre for Health Services Research, 1998.
30. North of England Guidelines Development Project. *Aspirin for the secondary prophylaxis of vascular disease in primary care.* Newcastle upon Tyne: Centre of Health Services Research, 1998.
31. North of England Guidelines Development Project. *Non-steroidal anti-inflammatory drugs (NSAIDs) versus basic analgesia in the treatment of osteoarthritis*. Newcastle upon Tyne: Centre for Health Services Research, 1998.
32. North of England Guidelines Development Project. *The choice of antidepressants for depression in primary care.* Newcastle upon Tyne: Centre for Health Services Research, 1998.
33. Oxman, A. D., & Guyatt, G. H. A consumers guide to subgroup analyses. *Annals of Internal Medicine,* 1992, 116, 78–84.
34. Peto, R., Pike, M. C., & Armitage, P. Design and analysis of randomised clinical trials requiring prolonged observation of each patient, part II: Analysis and examples. *British Journal of Cancer*, 1977, 35, 1–39.
35. Pfeffer, M. A., Branunwald, E., Moy, L. A., et al., for the SAVE Investigators. Effect of captopril on mortality and morbidity in patients with left ventricular dysfunction after myocardial infarction. Results of the Survival and Ventricular Enlargement Trial. *New England Journal of Medicine*, 1992, 327, 669–77.
36. Ray, S., Pye, M., Oldroyd, K. G., et al. Early treatment with captopril after acute myocardial infarction. *British Heart Journal*, 1993, 69, 215–22.
37. Robins, J., Breslow, N., & Greenland, S. Estimators of the Mantel-Haenszel variance consistent in both sparse data and large strata models. *Biometrics*, 1986, 42, 311–23.
38. Rothman, K. J. *Modern epidemiology*. Boston: Little & Brown, 1986.
39. Sharp, S. J., Thompson, S. G., & Altman, D. G. The relation between treatment benefit and underlying risk in meta-analysis. *British Medical Journal*, 1996, 313, 753–58.
40. Simon, G. E., Vonkorff, M., Heilingenstein, J. H., et al. Initial antidepressant choice in primary care: Effectiveness and cost of fluoxetine vs. tricyclic antidepressants. *JAMA*, 1996, 275, 1897–902.
41. SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *New England Journal of Medicine* 1991, 325, 293–302.
42. Song, F., Freemantle, N., Sheldon, T. A., et al. Selective serotonin reuptake inhibitors: Meta-analysis of efficacy and acceptability. *British Medical Journal*, 1993, 306, 683–87.
43. Swedberg, G., Held, P., Kjekshus, J., et al., for the CONSENSUS II Study Group. Effects of the early administration of enalapril on mortality in patients with acute myocardial

infarction. Results of the cooperative New Scandinavian Enalapril Survival Study II (CONSENSUS II). *New England Journal of Medicine,* 1992, 327, 678–84.
44. Thompson, S. G. Controversies in meta-analysis: The case of the trials of serum cholesterol reduction. *Statistical Methods in Medical Research*, 1993, 2, 173–92.
45. Thompson, S. G. Why sources of heterogeneity in meta-analyses should be investigated. *British Medical Journal*, 1994, 309, 1351–55.
46. Yusuf, S., Collins, R., & Peto, R. Why do we need some large, simple randomized trials? *Statistics in Medicine*, 1984, 3, 409–20.