GENERAL CARDIOLOGY

Evaluation of large scale clinical trials and their application to usual practice

Andrew M Tonkin

National Heart Foundation of Australia,
Austin and Repatriation Medical Centre,
Melbourne, Australia

679

"If it were not for the great variability among individuals, Medicine might be a Science, not an Art"—*Sir William Osler, 1882, The Principles and Practice of Medicine*

It is important to apply current best evidence in making decisions about management of individual patients. While the evidence may be derived from basic and applied research, the findings from large scale clinical trials of interventions are the most relevant. However, in many cases there are uncertainties around the effects of treatments and indeed guidelines can "legitimise" these uncertainties by defining boundaries within which decisions are reasonable. Therefore, the appropriate interpretation of clinical trial results is just as important for those who are charged with the development and implementation of guidelines as they are for the clinician in discussing options with individual patients.

Important aspects relating to trial design and interpretation are discussed, using illustrative examples drawn from various fields of cardiovascular medicine.

The science of clinical trial methodology has been discussed in detail elsewhere,[1] and the application of trial results to individual patients considered by other authors,[2] including overviews of trials of many interventions. However, to date, relatively few relating to cardiovascular medicine have been produced through the Cochrane Collaboration (http://www.epi.bris.ac.uk/cochrane.heart.htm).[3]

Rationale for the trial

The background to the clinical trial should be very clearly stated (and read) in the introduction to the paper which reports a trial result, as it will have a major influence on the trial design and hence its results. The intervention should have a sound biologic and/or pathophysiological rationale. The trial will often test the principal mechanism of action of the intervention. However, drugs often have pleiotropic effects and it needs to be borne in mind that the trial will test the particular drug (often in one dose) and not its mechanism(s); indeed, dose–response relations for different effects may vary.

The hypothesis to be tested will often have been generated from a meta-analysis of previous studies in the area. The recently published HOPE study[4] illustrates the manner in which the hypothesis for the trial can be generated from an overview of studies in more restricted patient populations. While meta-analyses may be very useful in defining likely effects in certain subgroups, by and large such overviews should be regarded as hypothesis generating. However, an example of what may be the unique benefits of meta-analyses is the antiplatelet trialists collaboration,[5] following which the more widespread use of aspirin would likely not have been achieved without an overview of many trials which individually were underpowered to show significant benefit.

The cohort of patients: generalisability of results

The main purpose of large scale trials is to cause widespread appropriate change in clinical practice. Typically, controlled clinical trials examine the effects of an intervention which is administered following tightly specified protocols to patients who are selected and generally compliant. This contrasts to the care of unselected patients by usual practices and practitioners.

It follows that it is important that patients recruited to trials closely resemble those in typical practice. Therefore, evaluation of a trial requires consideration of exclusion as well as inclusion criteria, and, if possible, of baseline characteristics of those patients who were "logged" but not recruited. Typically baseline characteristics are presented in the first table in reports of large scale studies. When the intervention modifies a biomedical risk factor, such as in the case of lipid modifying treatment in patients with known coronary artery disease, the trial has most relevance when the cholesterol concentrations of those studied most represent those of usual patients.

It is also important that trials test the particular treatment on a background of usual accepted practice. Indeed when usual care of study patients does not include general advances in treatment, the trial results must be interpreted with a degree of caution. The management of patients with coronary artery disease is an important example. Many large scale trials of different therapeutic approaches do not embrace the contemporary approach which might include more complete use of arterial conduits during bypass surgery, stent deployment during percutaneous coronary intervention, and an aggressive approach to cholesterol lowering treatment as part of medical management.

In cardiovascular trials the elderly and women are often under represented. The incidence of cardiovascular disease, including coronary heart disease and its manifestations, increases greatly with age. Absolute risk is greater in the elderly and failure to include such patients could lead to underestimation of the benefits of intervention. Alternatively, the true effects of treatment in the elderly may be missed because rates of deleterious outcomes

Correspondence to:
Professor Andrew Tonkin, Director of Health, Medical and Scientific Affairs, National Heart Foundation, 411 King Street, West Melbourne, Victoria, Australia 3003
andrew.tonkin@
heartfoundation.com.au

may also be different. Recently published observational data in almost 8000 patients showed that among patients with myocardial infarction receiving thrombolytic treatment, in those over 75 years old who received treatment the mortality rate was 18% in the first month after discharge, compared to 15% in those who did not receive treatment.[6] While some older patients undoubtedly benefit from thrombolytic treatment, others have an increased risk of cerebral haemorrhage and other complications. Comorbidities such as hypertension or previous stroke which may have increased bleeding risk may have been ignored. However, because controlled trials of thrombolysis have been confined to relatively younger patients, a randomised trial of thrombolysis in the "old old" may be appropriate.

The elderly have been notably under represented in trials of treatments for heart failure. Because the average age of patients recruited to heart failure trials is younger than those usually treated, this in turn may also lead to recruitment of fewer females as they develop disease manifestations at an older age.[7] Furthermore, heart failure trials frequently recruit from cardiology departments in the hospital environment, and inclusion criteria may require objective evidence of greater left ventricular dysfunction than is found in usual patients, particularly in the community setting.

Trial design and monitoring

An understanding of the principles and different types of trial design is important. Observational studies are particularly affected by issues of bias and confounding that cast doubts about their validity. Indeed, randomisation is one of the major factors that has increased the relevance of clinical trials. Even then, all attempts must still be made to reduce bias at the time of randomisation. The randomisation process may include stratification for key baseline descriptor(s), but in very large scale studies it is often assumed that baseline risks should be matched between the two groups assigned different therapeutic approaches.

The importance of an adequate (ideally placebo) control group has been demonstrated repeatedly. As one example, without inclusion of a contemporary, placebo group, the important proarrhythmic effect of class 1c antiarrhythmic drugs may not have been recognised in the CAST (cardiac arrhythmia suppression trial) study,[8] as event rates in those randomised to active treatment were similar to those from previous individual patient usage data held by the pharmaceutical company.

A placebo limb may be unethical in certain circumstances—for example, thrombolysis in acute myocardial infarction. In such a context, different "active" treatments should be compared. Because of decreasing mortality rates with general improvements in management, trials which attempt to show superiority of newer agents above standard treatments are increasingly more difficult. The large number

of patients needed can be a major problem. As an extension to this, trials designed to demonstrate "equivalence" with narrow confidence intervals actually require more rather than fewer patients compared with "superiority" studies.[9] Because of this, the latest shift has been to "non-inferiority" trials. Then the clinical value of demonstrating that there is no clinically significant difference in outcomes between a new agent and conventional treatment may lie in the lower cost, greater ease of administration, or greater safety of the new agent. These analyses are often undertaken in conjunction with the main trial.

Other design strategies which may be incorporated to increase power in comparative studies are not only to increase the sample size, but to randomise unevenly by including fewer patients in "control" groups, and to deliberately enrol patients at higher risk so as to increase the number of end points.

A further relatively new development has been the possible use of a "cluster" design which allows randomisation of groups of people. This technique is used when the intervention is administered to and can affect entire clusters of people rather than individuals within the cluster, or when the intervention, although given to individuals, may "contaminate" others in the control group so as to weaken any estimate of treatment difference.[10] The methodology can be particularly applied to studies of methods of care. An example could be a telephone based support system for patients when compared to usual outpatient care.

Factorial design (and simplicity) are other methodological approaches that may increase the efficiency of randomised controlled trials. Factorial design not only allows more than one hypothesis to be tested simultaneously, but allows large scale evaluation of some treatments such as dietary supplements that might not otherwise be possible because of difficulty in attracting the necessary funding.

Very large scale clinical trials should have an independent data and safety monitoring board. Their role should be clearly stated. Typically, the board will operate with pre-specified general stopping rules but they should usually be encouraged not to terminate a trial too early. This is because the reliability of data is greater with an increasing number of end points, perhaps euphemistically termed "regression to the truth". Accordingly, the mathematical functions which determine stopping often require more extreme evidence of effect earlier compared with later in the trial. Particularly, trials should rarely be terminated very early on the basis of "futility" because this deduction is unreliable when there are relatively few end points.

Appropriate end points: clinical relevance

One major end point should be clearly specified and used as the basis for power calculations, the estimate of the "reliability" of the

result. These power calculations should be presented.

All cause mortality is the hardest end point and allows for inaccuracies in the certification of the cause of death.[11] It usually requires inclusion of a very large number of patients in the study. Increasingly, an expanded end point which is a composite of a number of outcomes is the primary end point. An expanded end point could be a composite of cause specific death, related non-fatal events, and perhaps an index of cost benefit such as a measure of hospitalisation. Each component should be biologically plausible and there should be an attempt to minimise any possibility of "double counting".

Care is prudent before there is wide extrapolation from the results of secondary end point data from smaller trials. As an example, data from the ELITE II study[12] failed to confirm a mortality benefit of an angiotensin receptor antagonist compared to an angiotensin converting enzyme (ACE) inhibitor in heart failure patients, although this had previously been demonstrated in the smaller ELITE I study. The primary end point in ELITE I was renal function rather than mortality, but the somewhat dramatic effect on survival had been sufficient to convince a number of regulatory authorities throughout the world to liberalise indications for angiotensin receptor antagonism.

Methods of analysis

Intention to treat analyses are vital to minimise bias and must always be presented. These analyses present outcomes by treatment assigned at the start of the trial, irrespective of whether there is adherence throughout the period of follow up.

However, it is appropriate to examine the data which are presented for the extent of non-adherence to assigned treatment. The reader should ascertain whether or not there was significant "crossover" to the other treatment limb which was being compared. Crossover between assigned treatments can be a particular problem in trials which compare non-pharmacologic interventions and drug treatment. One example involves the trials in patients with unstable angina which have com-

pared outcomes after early coronary angiography and, possibly, revascularisation with a "conservative" approach based on medical treatment. In the TIMI IIIb, VANQWISH, and FRISC II studies, from 14–57% and 48–73%, respectively, of those patients assigned to a conservative therapeutic approach had cardiac catheterisation while an inpatient or within 12 months.[13] These intervention rates translated to revascularisation approaches by 12 months in 33–49% in those assigned initial conservative treatment compared to 44–78% of those assigned to an initial invasive strategy. This made meaningful conclusions concerning the role of early revascularisation very difficult.

The examples also suggest the potential value of additional presentation of "on-treatment" analyses when this is appropriate.

Net benefit: public health impact

Figure 1 shows a schema within which the overall effects of a treatment might be considered.

The distinction between relative and absolute risk (and reduction) is very important. Relative risk is the increase (for a risk factor) or decrease (the typical case for an intervention) in the likelihood of an event compared to a reference group. The odds ratio (OR) is another measure of this, calculated as the ratio of odds (OR = $p \div 1-p$, where p is the probability of the event).

However, it is much more important to examine absolute risks. Absolute risk reduction is the arithmetic difference in rates of outcomes between the experimental and "reference" (control) groups in the trial. The reciprocal of
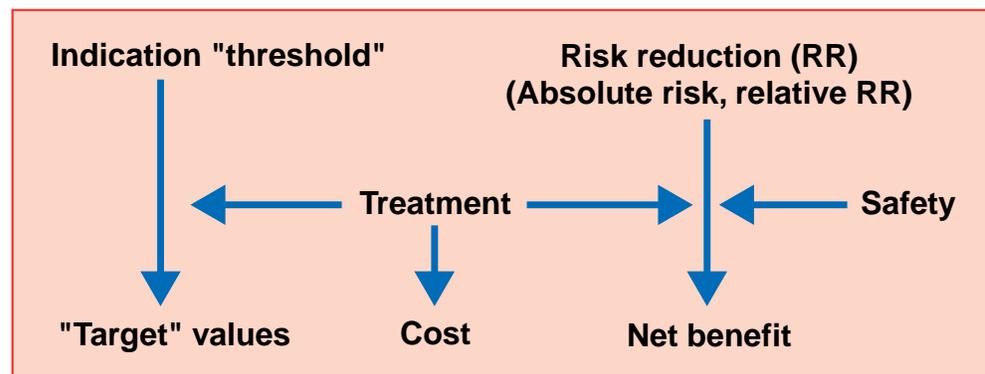

Figure 1. A schema within which to consider aspects of a treatment. Information on many of these can be obtained within the context of a large scale trial.

the absolute risk reduction is the number who would need to be treated to prevent one adverse outcome ("number needed to treat"). This takes into account both the relative risk reduction and underlying risk and is often used to gauge the absolute effect of the intervention being tested. To enable comparisons for chronic treatments, the numbers needed to treat are often estimated for five years of intervention.

The thresholds for initiating treatment should reflect the level of absolute risk at which first, the benefits and hazards of treating outweigh those of not treating, and which secondly, justify the associated costs and inconvenience to the patient. As shown in fig 2, the risk reduction with an effective treatment should increase, somewhat in proportion to the level of risk of the patient cohort. However, the magnitude of any harmful effects is usually independent of the level of risk for the indication for treatment. A net benefit can then be derived as a composite of these considerations of absolute benefit and harm.

Clinicians need to compare the absolute risk of trial patients with their own patient. If the relative risk reduction is anticipated to be the same, the absolute benefit of an intervention is greatest in the patients at highest risk. Such groups could include the elderly or people with diabetes. These considerations can also be relevant when absolute risk rates are greater in clinical practice than in selected patients recruited to the trial.

An example of the logic outlined above can be found in considering the risk and prevention of stroke in patients with chronic non-rheumatic atrial fibrillation.[14] The overall risk is around 5% per annum but this increases with increasing age, recent congestive heart failure, presence of hypertension or diabetes, a history of previous stroke or transient ischaemic attack, and evidence of left atrial enlargement or left ventricular dysfunction on transthoracic echocardiography. In both primary and secondary prevention trials, warfarin has been shown to decrease risk by around two thirds, but from a baseline annual risk of 12% in secondary prevention compared to 5% in primary prevention. The same relative risk reduction results in much greater absolute benefit in those who have had previous events, but bleeding risk is no different in the two scenarios. It should be further noted that the rate of bleeding observed in the trials (0.5–0.8% per annum) is much less than that seen in usual clinical practice (around 5% per annum). Therefore, it is important to assess individual patients carefully for comorbidities which could increase risk of bleeding.

Another example concerns primary prevention of coronary heart disease events with lipid modifying treatment. Absolute risk in individuals with similar cholesterol concentrations depends critically on their age, sex, and levels of other established cardiovascular risk factors. On the basis of consideration of multiple risk factors, groups such as the joint European task force have applied multivaried mathematical modelling to enable prediction of an arbitrary
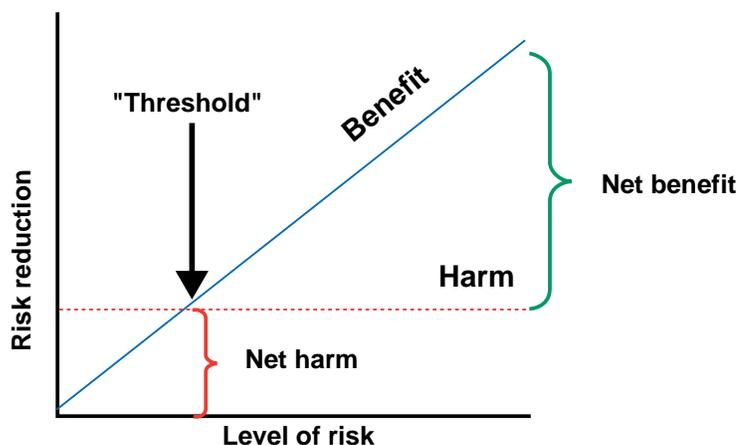
"Threshold" here excludes consideration of costs and inconvenience of therapy

Figure 2. Net benefit is a composite of absolute benefit (which will often vary according to baseline level of risk) and harm (which is often independent of the level of risk).

risk of events over 10 years and to suggest various "thresholds" at which initiation of treatment may be appropriate.[15]

To establish relative public health benefits, often trials are "lumped" to compare the number needed to be treated in different scenarios. However, because baseline risk often varies widely between trials, care is necessary in pooling of data from multiple trials.[16]

Another note of caution concerns the interpretation of safety data. Few clinical trials extend beyond five years because of factors such as investigator and subject fatigue, and the accumulation of crossovers. This time frame may be inadequate to detect some very important adverse affects such as cancer. As a corollary, the risk:benefit ratio may differ at different time points after the initiation of the treatment.

Cost-benefit analyses

A large part of the direct costs associated with cardiovascular disease relate to hospitalisation, and a disproportionate amount is associated with the care of the elderly. Many regulatory authorities now require formal evaluation of cost–benefit of new treatments and these are often conducted within the clinical trial environment. The findings obviously impact on translation of the outcomes of trials to the clinical context and demonstration of important outcomes can be used to justify the more widespread use, albeit with higher initial costs of some treatments. An example is the value of the implantable cardioverter defibrillator in patients at higher risk of "malignant" ventricular arrhythmias.[17]

Surrogate measures

Because studies which compare two active treatments require much higher numbers to ascertain or exclude differences in treatment

682

Evaluation of clinical trials: summary

COMPONENT	ASPECTS
The hypothesis	Biologic rationale, meta-analysis
The cohort	Demographics, background treatments
Trial design and monitoring	Appropriate design, independence
Primary and subsidiary end points	Relevance, plausibility
Analysis	Intention to treat, on treatment
Net benefit	Absolute benefit, safety
Affordability	Cost–benefit considerations
Intermediate outcomes	Validity as "surrogate" measures
Subgroup analyses	Whether prespecified

effects reliably, "surrogate" measures may be reported. Study of intermediate outcomes for "harder" clinical end points should require appropriate scientific data to suggest the relation is truly a mechanistic one. An association, even a compelling epidemiologic relation, does not necessarily imply a causal relation. Particularly, interventions may have multiple mechanisms of action. Indeed, the relative importance of these different effects may vary according to the criteria which are used to define the patient population under study.

An associated question is whether it is valid to use intermediate outcomes as surrogate measures to deduce a "class" effect among drugs which can differ in their pharmacokinetic and pharmacodynamic properties, and in their spectrum of adverse effects. The argument could apply to ACE inhibitors which, because of differences in tissue binding properties, could have different strengths of action on paracrine ACE systems, or to 3-hydroxy-3-methylglutaryl coenzyme A (HMG CoA) reductase inhibitors which might differ not only in the potency of their lipid modifying effects but also of other potentially relevant mechanisms.

As stated earlier, while the approach is conservative, it is usually sensible to regard large scale trials as testing specific treatments (and in the case of drugs, in particular doses) and not their mechanisms of action. However, guidelines for determining whether or not a drug is exerting (more than) a class effect have been published.[18] One very important example of the potential value of intermediate measures and of substudies was that which established the importance of restoration of TIMI III flow through the infarct related artery for preservation of left ventricular function and long term outcome following myocardial infarction.[19]

<h2 style="color:red">Subgroup analyses</h2>

Trials are designed to have sufficient power to reliably test the effect of the intervention in the cohort which is defined by the particular inclusion and exclusion criteria. Results in subgroups of the cohort are nearly always less reliable and frequently over interpreted.

The credibility of a subgroup analysis depends on the size of the subgroup, the biologic plausibility of the analysis, and consistency of effects between different trials. Particularly, unless subgroup analyses have been prespecified (both the patient subgroups and the outcomes of interest) and all such prespecified analyses are presented or at least available, bias is frequent if not inevitable, and the selection of subgroup analyses presented might be viewed as "data dredging".

When subgroup analyses are presented, the appropriate statistical test often examines for evidence of heterogeneity between different subgroups—for example, between sex or age groups, or those with or without particular risk factors.

<h2 style="color:red">Final comments: patient preferences</h2>

Trials of cardiovascular treatments have conclusively shown the efficacy of a wide variety of treatments. Patient outcomes can be improved by appropriate translation of the results of these trials to usual practice.

However, it is appropriate that patients are being further empowered concerning decisions relating to their health. It is worth noting the opinions of different groups in a recent survey to establish a threshold above which it was judged to be appropriate to use antihypertensive drugs. The number of patients with hypertension regarded as appropriate to be treated over five years to save one life was lower for consultants than general practitioners, but particularly much higher among nurses and, notably, the public.[20] In the current environment with increasing application of information management, perhaps in the future the community and patients may also be seeking details relating to trial management and interpretation, rather than taking what to this time has been the relatively passive role of subjects in such trials.

1. **Friedman LM, Furberg C, DeMets DL.** *Fundamentals of clinical trials*, 3rd ed. St Louis: Mosby, 1996.
• *A comprehensive edition devoted to fundamentals of the design, conduct and analysis of clinical trials.*

2. **Dans AL, Dans LF, Guyatt GH,** *et al* for the Evidence-Based Medicine Working Group. User's guides to the medical literature. XIV. How to decide on the applicability of clinical trial results to your patient. *JAMA* 1998;**279**:545–9.
• *Another discussion of the issues—biologic, social and economic, and epidemiologic—that might influence applicability of clinical trial results to an individual patient.*

3. **Cochrane AL.** *Effectiveness and efficiency. Random reflections on health services.* London: Nuffield Provincial Hospitals, Trust, 1972.
• *The Cochrane Collaboration is an international venture which provides overviews of controlled trials for a variety of treatments. It is named after Archie Cochrane, a British epidemiologist who wished people to make more informed decisions about health care, by providing better access to reliable reviews of the relevant evidence.*

4. **The Heart Outcomes Prevention Evaluation Study Investigators.** Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. *N Engl J Med* 2000;**342**:145–53.
• *HOPE is a landmark study which was based on subsidiary analyses of older trials of ACE inhibitors in patients with heart failure or asymptomatic left ventricular dysfunction. It confirmed a reduction in cardiovascular events as the primary end point, but extended this to a wide range of vascular patients, irrespective of their left ventricular function.*

5. **Antiplatelet Trialists' Collaboration.** Collaborative overview of randomised trials of antiplatelet therapy. 1:

prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ* 1994;**308**:81–106.
• *This meta-analysis of antiplatelet treatments enabled conclusions for practice to be drawn which were not otherwise available from individual trials which had been underpowered.*

6. **Thiemann DR, Coresh J, Schulman SP,** *et al.* Lack of benefit for intravenous thrombolysis in patients with myocardial infarction who are older than 75 years. *Circulation* 2000;**101**:2239–46.

7. **Cohen-Solal A, Desnos M, Delahaye F,** *et al* for the Myocardial and Heart Failure Working Group of the French Society of Cardiology, the National College of General Hospital Cardiologists and the French Geriatrics Society. A national survey of heart failure in French hospitals. *Eur Heart J* 2000;**21**:763–9.
• *This observational study characterised heart failure patients in medical and geriatric as well as cardiology departments in 120 French hospitals. Patient characteristics such as age and sex distribution differed significantly from those in patients recruited to recent heart failure trials.*

8. **The Cardiac Arrhythmia Suppression Trial Investigators.** Preliminary report: effects of encainide and flecainide on mortality in a randomised trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989;**321**:406–12.
• *The placebo controlled design of this study allowed recognition of the potential for proarrhythmic effects and worsened outcome with antiarrhythmic drugs. The trial was based on the hypothesis that suppression of ventricular ectopic beats would prevent sudden death, an assumption which has since been shown to be fallacious.*

9. **White HD.** Thrombolytic therapy and equivalence trials. *J Am Coll Cardiol* 1998;**31**:494–6.

10. **Donner A.** Some aspects of the design and analysis of cluster randomised trials. *Appl Stat* 1998;**47**:95–113.

11. **Lauer MS, Blackstone EH, Young JB,** *et al.* Cause of death in clinical research: time for a reassessment? *J Am Coll Cardiol.* 1999;**34**:618–20.

12. **Pitt B, Poole-Wilson PA, Segal R,** *et al.* Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial. The losartan heart failure survival study ELITE II. *Lancet* 2000;**355**:1582–7.

13. **Mehta SR, Eikelboom JW, Yusuf S.** Long-term management of unstable angina and non-Q-wave myocardial infarction. *Eur Heart J* 2000;**2**(suppl E):E6–12.

14. **Atrial Fibrillation Investigators.** Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomised controlled trials. *Arch Intern Med* 1994;**154**:1449–57.

15. **Wood DA, DeBacker G, Faergeman O,** *et al* on behalf of the Task Force, Prevention of Coronary Heart Disease in Clinical Practice. Recommendations of the second joint task force of the European Society of Cardiology, European Atherosclerosis Society and European Society of Hypertension. *Eur Heart J* 1998;**19**:1434–503. *Atherosclerosis* 1998;**140**:199–270. *J Hypertens* 1998;**16**:1407–14.

16. **Smeeth L, Haines A, Ebrahim S.** Numbers needed to treat derived from meta-analyses—sometimes informative, usually misleading. *BMJ* 1999;**318**:1548–51.
• *A critical commentary on the caution needed in interpreting pooled data from trials, in this case specifically, numbers needed to treat. Contains examples from various therapies in cardiovascular disease.*

17. **Roberts PR, Betts TR, Morgan JM.** Cost effectiveness of the implantable cardioverter defibrillator. *Eur Heart J* 2000;**21**:712–19.

18. **McAlister FA, Laupacis A, Wells GA,** *et al* for the Evidence-Based Medicine Working Group. Users' guides to the medical literature: XIX. Applying clinical trial results. B. Guidelines for determining whether a drug is exerting more than a class effect. *JAMA* 1999;**282**:1371–7.
• *This interesting article gives examples and discusses level of evidence for determining whether or not drugs are exhibiting a class effect.*

19. **The GUSTO Angiographic Investigators.** The effects of tissue-plasminogen activator, streptokinase, or both on coronary-artery patency, ventricular function, and survival after acute myocardial infarction. *N Engl J Med* 1993;**329**:1615–22.
• *This substudy of the first GUSTO trial compared patency rates and left ventricular function for different thrombolytic regimens. It illustrates how careful substudies may yield very important data on mechanisms of therapeutic effect.*

20. **Steel N.** Thresholds for taking antihypertensive drugs in different professional and lay groups: questionnaire survey. *BMJ* 2000;**320**:1446–7.

CORRECTION

Non-invasive coronary artery imaging with electron beam computed tomography and magnetic resonance imaging. PJ de Feyter, K Nieman, P van Ooijen, M Oudkerk. *Heart* 2000;**84**:442–8.

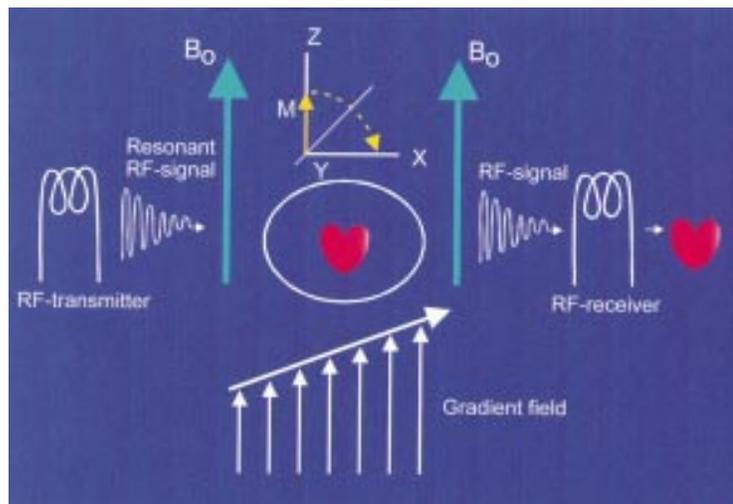In this article fig 2 was incorrect, as certain elements were missing. The correct fig 2 is reproduced here.

Figure 2 The patient (heart) is placed within a strong external magnetic field (Bo). The RF transmitter rotates the net tissue magnetisation in the transverse plane, and after termination, relaxation occurs which emits a signal detected by the RF receiver. The gradient coils produce a supplemental magnetic field gradient to allow precise location of the excited protons. The received signals have certain signal intensity (brightness) and location, both of which are processed to form the desired image.