

REVIEW

Cautionary tales in the clinical interpretation of therapeutic trial reports

I. A. SCOTT^{1,2*} and P. B. GREENBERG^{3,4*}

¹Princess Alexandra Hospital and ²University of Queensland, Brisbane, Queensland and ³Clinical Epidemiology and Health Care Evaluation Unit, Department of General Medicine, Royal Melbourne Hospital and ⁴University of Melbourne, Melbourne, Victoria, Australia

Abstract

Practising clinicians are assailed daily with reports of new therapeutic clinical trials. The evidence-based medicine movement has developed critical appraisal methods for assessing the validity and impact of such studies. However, challenges persist in regards to the appropriate interpretation and application of trial results

within everyday clinical settings. Using selected examples from recently published literature, we illustrate 15 cautionary themes for translating research evidence from therapeutic trials into clinical practice. (Intern Med J 2005; 35: 611–621)

Key words: clinical interpretation, therapy trials, cautions.

INTRODUCTION

The evidence-based medicine (EBM) movement has emphasised the need for critical appraisal of results of randomised clinical trials (RCT) in regards to validity, clinical impact and patient applicability,¹ and has promulgated standardised reporting methods that better highlight strengths and weaknesses of individual trials.²

Despite this, the Evidence-based Medicine Working Group of the Internal Medicine Society of Australia and New Zealand contends,³ as have done others,⁴ that clinicians remain vulnerable to changing their practice on the basis of misleading claims. The reporting of trial methodology is inadequate in more than half of therapy publications,⁵ outcome measures are frequently misreported or omitted,⁶ and conclusions and inferences are often biased and unsubstantiated.⁷ This article discusses, with examples, 15 cautionary themes for lessening the risk of clinical misinterpretation of reports of therapy trials. For each theme we highlight the relevant components of therapeutic trial design: Patient group, Intervention, Comparison and Outcome (PICO).⁸

Beware limited generalisability

Generalisability ('external validity' or 'applicability') refers to how much trial findings can be extrapolated to

the broader population, or to particular patients. This depends on the representativeness of the trial population (**P**) in terms of clinical characteristics (age, sex, disease severity and other comorbidities), setting and concomitant therapies (or cointerventions) (Table 1).⁹ No more than 4%¹⁰–15%¹¹ of screened populations actually achieve randomisation in most trials, with severely ill or elderly patients, those with multisystem disease, and others in whom risks of therapy are considered disproportionately high tending to be excluded.

Unintended harm can result from generalising promising results of trials involving selected patients to unselected populations in the absence of close surveillance or confirmation of observed effects in large trials with broad patient selection criteria.¹²

Example. In the RALES trial, spironolactone use significantly reduced all-cause mortality and rehospitalisation (23% relative risk reduction (RRR) over 2 years) in patients with congestive heart failure (CHF).¹³ However, trial patients had normal renal function and severe left ventricular (LV) dysfunction (ejection fraction <35%). Cohort studies have since revealed an 'indication drift', with spironolactone use in patients with advanced renal failure and/or mild LV dysfunction.¹⁴ In Ontario, a more than threefold post-RALES increase in spironolactone prescribing in CHF patients was correlated with an almost fourfold rise in the numbers of hospitalisations and deaths secondary to hyperkalaemia.¹⁵

Conversely, foregoing administration of therapies to non-trial patients at higher risk than trial patients might also be inappropriate if evidence suggests no increased risk of therapy-induced harm.¹⁶

Example. Early trials of β -blocker therapy following myocardial infarction excluded patients with chronic obstructive pulmonary disease (COPD), diabetes or

Correspondence to: Ian A. Scott, Princess Alexandra Hospital, Ipswich Road, Brisbane 4102, Australia. Email: ian_scott@health.qld.gov.au

Received 20 March 2005; accepted 1 June 2005.

Funding: None

Potential conflicts of interest: None

*Ian A Scott and Peter B Greenberg for the Evidence-based Medicine Working Group of the Internal Medicine Society of Australia and New Zealand.

depression because of concerns about β -blocker-induced bronchospasm, hypoglycaemia unawareness, or severe depression.¹⁷ However, cohort studies revealed that β -blockade in these patient groups failed to show excess toxicity,^{18–20} and, indeed, in patients with COPD or diabetes, demonstrated greater absolute reductions in mortality than in trial patients.²¹

Beware faulty comparators

Wherever possible, experimental therapies should be compared with acceptable current practice or 'usual care' (C) rather than placebo drugs. Usual care therapies should, in turn, reflect conventional dosing schedules that minimise toxicity. Head-to-head trials of different treatments might, in their design, unfairly advantage one drug over its comparator.

Example. High-dose (80mg/day) atorvastatin is heavily promoted as the most appropriate lipid-lowering regimen in patients with coronary artery disease based on two recent trials evaluating long-term cardiovascular end-points.^{22, 23} However, in one trial high dose atorvastatin was compared to a different statin (pravastatin) at a less potent dose (40mg/day),²³ while in the other, high-dose atorvastatin was compared to an unconventionally low dose (10mg/day) of the same drug.²⁴ If both trials had included a comparison of high-dose with intermediate dose (40mg/day) atorvastatin, cholesterol levels and incidence of cardiovascular events in high and intermediate dose atorvastatin groups might have been shown to be the same, and lower than in the pravastatin or low-dose atorvastatin groups. Equivalence of high and intermediate dose atorvastatin, if demonstrated, would obviate the additional cost and risk of toxicity entailed in using high-dose atorvastatin in large numbers of patients.

Beware surrogate end-points

Surrogate end-points are physiological, biochemical, radiological, neuropsychological or other outcome measures, distinct from 'clinically important' end-points (O),

such as death, morbid events, or hospitalisation. When assumed to correlate with improved clinical end-points, changes in surrogates might obviate the need for large trials by occurring more quickly and with greater precision than rare, more remote clinical end-points. However, surrogate-based trials should only influence practice under strictly defined circumstances (Table 2) because of repeated dissonance between their results and those of clinical outcome-based trials.²⁴

Example 1. In an RCT of 502 postinfarct patients with ventricular ectopy on electrocardiogram, patients randomised to encainide or flecainide achieved higher rates of suppression of ectopics than those receiving placebo (79 and 83% vs 37%).²⁵ This reduction in ectopy was assumed to predict a reduced risk of sudden arrhythmic cardiac death. However, in a subsequent RCT of 1498 postinfarct patients, antiarrhythmic treatment compared to placebo increased the risk of cardiac deaths and arrests (relative risk (RR) 2.64 (95% confidence interval (CI) 1.60–4.36)).²⁶

Example 2. In a Cochrane review of 16 trials ($n = 4365$) of the acetylcholinesterase inhibitor donepezil in patients with mild to moderately severe Alzheimer's dementia,²⁷ improvements were noted over 6–12 months in cognitive function, which, although statistically significant, represented marginal clinical gains (mean 1.84 point increase on 30-point Mini-Mental State Examination (MMSE) scale;²⁸ mean 2.92 decrease on 70-point Alzheimer's Disease Assessment Scale–cognitive subscale).²⁹ Modest improvements in global clinical states, activities of daily living (ADL) measures and behaviour were seen, but with non-standardised measures giving heterogeneous results.

Stabilisation or improvement in cognitive and functional performance is assumed to predict less need for institutionalisation, reduced disability, fewer deaths or adverse events, lower carer burden and decreased health-care costs. However, a recent RCT of 565 community-living patients with mild to moderate dementia showed no difference between donepezil and placebo groups for

Table 1 Criteria for assessing generalisability⁹

-
- Similarity of clinical characteristics of trial patients to those encountered in practice
 - Conformity of treatments and background care (doses, durations, follow-up period, care settings etc.) to standard practice patterns
 - Consistency of measured outcomes with conclusions drawn
 - Coverage of all relevant outcomes (adverse events and side-effects)
 - Consideration of the study findings in the context of other available evidence
-

Table 2 Criteria for accepting trials based on surrogate end-points²⁴

-
- Strong, independent and consistent association between surrogate and clinical outcome
 - RCTs involving same drug class and other drug classes show consistent improvement in clinical outcome following improvement in surrogate
 - Treatment effect is large, precise and lasting
 - Likely treatment benefits outweigh potential costs and harms on the basis of high baseline patient risk, patient preferences to avoid target clinical outcome, and absence of alternative therapies
-

RCT, randomised clinical trials.

any of these clinical end-points over 3 years,³⁰ despite changes in MMSE and ADL scores of a magnitude similar to those seen in the Cochrane review.

Beware relative versus absolute effects

The magnitude of effect size, or differences in outcomes (**O**), can be expressed in relative or absolute terms. Doctors, patients and health managers all become more conservative in supporting new therapies when efficacy is expressed as absolute risk reduction (ARR), or number needed to treat (NNT) to prevent an event, as opposed to RRR or odds ratios (OR).^{31–33} Absolute benefit is, in turn, related to the absolute level of baseline risk. Treating patients with moderate to severe hypertension, for example, will prevent more strokes (ARR = 8%; NNT = 12) than treating mild hypertension (ARR = 0.6%; NNT = 166) in the first instance, even though the RRR from treatment is identical (40%) in both groups.³⁴

Example. Of 174 pharmaceutical advertisements from 6 popular Australian medical journals, less than 10% presented any numerical estimate of efficacy, and of these, no more than a quarter reported ARR, NNT and RRR, or provided the data that allowed their calculation.³⁵ Instead, misleading graphical depictions of relative risk,³⁶ combined with visual and linguistic imagery,³⁷ are used to influence clinician practice.

Beware small effect sizes

Outcome differences (**O**) between groups, although reaching statistical significance, might be of minor clinical significance. We caution against an overinterpretation of small to modest effect sizes, especially if new treatments, compared to alternatives, are more costly or might induce more harm.

Example. An Australian trial involving 6083 elderly hypertensive patients found, over 4 years, a RRR of 11% in all cardiovascular events or all-cause mortality favouring ACE inhibitors over diuretics.³⁸ However this RRR was of borderline statistical significance (95% CI –1 to 21%; $P = 0.05$) and, in absolute terms, represented a difference of only 0.4 events per 100 patient years; that is, NNT of 250 to prevent 1 additional event over 1 year. The much larger trial, ALLHAT, ($n = 33\,357$)³⁹ and a

meta-analysis of 42 trials ($n = 192\,478$)⁴⁰ also demonstrated no benefit for ACE inhibitors over diuretics as first-line agents in treating hypertension, including that in elderly patients.

Beware subgroup analyses

Subgroup analyses attempt to identify subsets of patients (**P**) who derive greater or lesser benefit (**O**) from experimental therapies than does the average trial patient. Subgroups can be defined by age, sex, disease severity, comorbidities or cointerventions. However, before basing treatment decisions on results of subgroup analyses, several criteria should be satisfied (Table 3).⁴¹

Example 1. The previously cited hypertension trial reported that, in men, ACE inhibitors significantly reduced the hazard of death or cardiovascular events (RRR 17%; 95% CI 3–29%; $P = 0.02$), whereas in women there was no difference.³⁸ However, this subgroup analysis satisfied only one (within-study comparison) of the seven criteria listed in Table 3, with a 15% probability that the difference in RRR between the sexes could be explained by chance alone. Although conceding that ‘the significance of these secondary results should be judged cautiously’ the authors concluded in their abstract that ‘ACE inhibitors in older subjects, *particularly men*, (italics added) appear to lead to better outcomes than treatment with diuretics’.

Example 2. The PRAISE-1 trial assessed the effects of amlodipine versus placebo in 3000 patients with systolic CHF,⁴² with randomisation stratified according to aetiology: ischaemic versus non-ischaemic. There was a mortality trend favouring amlodipine overall (RRR 16%; 95% CI –2 to 31%; $P = 0.06$), no survival effect in the ischaemic subgroup, but a highly significant survival benefit (RRR 46%; 95% CI 21–63%; $P = 0.004$) in the non-ischaemic subgroup. In contrast, the subsequent PRAISE-2 trial, involving 1652 patients with non-ischaemic heart failure and using a study design identical to that in PRAISE-1, showed no survival difference.⁴³

Beware de-emphasised or unmeasured risk of harm

The negative outcomes of treatment-induced harm might not be reported or even measured, or misreported in ways that mask their clinical impact. We suggest

Table 3 Criteria for accepting results of subgroup analyses⁴¹

-
- Clinically significant: the magnitude of the differences between treatment groups are clinically important and would lead to different clinical decisions for different subgroups
 - Statistically significant: the differences remain statistically significant after formally testing for treatment–subgroup interactions using appropriate statistical methods
 - A priori hypothesis: the hypothesis of subgroup differences preceded rather than followed the analysis (i.e. a priori hypothesis prespecified in the trial protocol, not a discovery made from post-hoc analyses)
 - Limited number of comparisons: the subgroup analysis in question was one of a small number of hypotheses tested to minimise the number of seemingly significant differences (i.e. interactions) that could occur simply by chance
 - Within study comparisons: subgroup differences were suggested by comparisons within studies (i.e. direct comparisons) rather than between studies (indirect comparisons)
 - Reproducibility: the subgroup difference is reproducible in other studies that have adequate power and are of similar design in terms of patient characteristics, cointerventions and outcome measures
 - Supporting evidence: the subgroup difference is biologically compelling; that is, consistent with current understanding of biologic mechanisms of disease
-

caution in accepting results of trials that inadequately report all important adverse effects regarded as plausible on the basis of previous clinical experience, observational data, or postulated mechanisms of drug action.

Example. Two trials assessing effects of cyclooxygenase-2 (COX-2) inhibitors on risk of colonic cancer were suspended when it became apparent that, in comparison to placebo, rofecoxib and celecoxib in moderate to high dosage over a 3-year period, were associated with a two to threefold increase in risk of myocardial infarction, stroke or death.^{44,45} Previous trials comparing COX-2 inhibitors with non-selective non-steroidal anti-inflammatory drugs (NSAIDs) in arthritis, and focusing on differences in short-term gastrointestinal toxicity,^{46,47} had not revealed this cardiovascular risk. However, more complete pooled data from these trials at 12 months follow up clearly showed a 1.3% increase in absolute risk of death, hospitalisation, life-threatening or disabling event with COX-2 inhibitors.⁴⁸ The 6 in 1000 higher incidence of cardiovascular events for rofecoxib cancelled out the 5 in 1000 lower incidence of complicated ulcers.⁴⁹

By December 2000, accumulated evidence had confirmed the more than twofold increased risk of cardiac events secondary to rofecoxib,⁵⁰ and to celecoxib,⁵¹ but it took another 4 years before safety data from the cancer prevention trials led to withdrawal of rofecoxib from the market and a cautionary letter to all practitioners from the makers of celecoxib. During this time, hypertensive patients prescribed COX-2 inhibitors in the USA had higher coronary risk profiles than those prescribed non-selective NSAIDs,⁵¹ further increasing the magnitude of harm. Moreover, a 41% rise in overall use of NSAIDs in 1.3 million elderly patients in Ontario from 1994 to 2002, driven by the use of COX 2 inhibitors, was accompanied by 10% increase in hospitalisation rates for upper gastrointestinal bleeding.⁵²

The detection of rare and delayed, but serious, toxicity is difficult. To be 95% sure of seeing at least one adverse event that occurs once in every given number of patients during the duration of a trial, you need to follow up 3 times that number of patients.⁵³ Given the size and duration of most trials, adverse events occurring less

than 1 in 1000, or which take longer than 6 months to appear, will generally remain undetected.

Nevertheless, randomised trials (and meta-analyses of such trials) provide the most unbiased estimates of excess hazard, and should include all data from non-published trials or submissions to regulatory authorities. In the absence of trial data, premarketing safety data (phase 1 and 2 studies), case-control studies and postmarketing surveillance data must serve to confirm safety of new drugs.^{54,55}

Beware composite end-points

The use of composite end-points (combining multiple end-points into a single outcome measure (**O**)) is becoming popular, especially in cardiovascular trials.⁵⁶ Although composites advantage investigators, they disadvantage clinicians because of their propensity to misinterpretation (Table 4).⁵⁷ When component end-points are not reported separately, clinicians are unable to discern which of those that matter to individual patients have been altered by therapy, or might wrongly believe that all component end-points have undergone changes of similar magnitude and direction.

Example. The UK prospective diabetes study (UKPDS) randomised patients with type 2 diabetes to intensive glycaemic control versus conventional control.⁵⁸ The primary end-point was a composite of 21 end-points: time to (i) first 'diabetes-related end-point' comprising sudden death, death from hyper- or hypoglycaemia, fatal or non-fatal myocardial infarction, angina, heart failure, stroke, renal failure, amputation of at least one digit, vitreous haemorrhage, retinal photocoagulation (end-point added after trial onset), unocular blindness, or cataract extraction, (ii) 'diabetes-related death' from myocardial infarction, stroke, peripheral vascular disease, renal disease, hyperglycaemia or hypoglycaemia and sudden cardiac death or (iii) all-cause death.

Although a significant decrease was seen in the composite end-point (RRR = 12%; 95% CI 1–21%; $P = 0.01$), most of this effect comprised reduction in retinal photocoagulation, with no changes in diabetes-related deaths and all-cause mortality; an observation emphasised in only 1 of 35 reviews of the UKPDS results.⁵⁹

Table 4 Advantages and disadvantages of composite end-points⁵⁷

Advantages

- More events allowing smaller sample size and reduced follow-up period
- Evaluation of effects of treatment on different but related disease processes
 - Greater insights into mechanism of drug action
 - Identification of disease populations that might benefit preferentially in terms of reduction in rates of specific clinical events
 - Aids extrapolation of drug effect from single to multiple disease conditions that share a common pathogenesis

Disadvantages

- Multiple component end-points that individually might be non-significant or moving in opposite directions
- Problematic issues of definition and/or measurement of each end-point both within and between trials
- Equal statistical weighting of component clinical end-points (e.g. mortality is treated statistically the same as recurrent angina)
- Potential for double-counting of component clinical end-points if measured as 'all-events' rather than 'first specified event' at the level of individual patients
- Potential for misinterpretation by clinicians wherein reported benefits can be ascribed to all component end-points, inflating perceptions of efficacy

Beware clinician-driven end-points

The UKPDS trial also raises another issue: the inclusion of clinician-driven end-points as outcomes (**O**). These end-points do not reflect natural history of disease, but embody pragmatic, clinician-instigated therapeutic manoeuvres such as retinal photocoagulation, revascularisation procedures, mechanical ventilation, or initiation of dialysis. Assessment is rarely undertaken of the precision and reproducibility of these different and arbitrarily defined proxy end-points and of the clinical and/or environmental circumstances that trigger them, or whether they correlate with clinical end-points such as death.

In an analysis of 179 comparisons involving composite end-points, the inclusion in 78 of a clinician-driven end-point more than doubled the odds of a result favouring experimental therapy (OR = 2.24; 95% CI 1.15–4.34) after adjusting for differences in event rates, quality of trial design and numbers of component end-points.⁵⁷

Beware discordant or unexpected results for secondary end-points

Interpreting results of a trial or group of trials showing no treatment effect on the primary end-point but an apparent benefit on a secondary end-point is problematic. We recommend caution in accepting discordant or unexpected results for secondary end-points, particularly those analysed post-hoc, and suggest that they be viewed as hypothesis-generating in need of confirmation by additional, independent evidence.

Example. In the ELITE I trial involving 722 elderly patients with heart failure randomised to losartan or captopril, there was no difference in renal function as the primary end-point. However, an unexpected decrease was seen in the secondary end-point of all-cause mortality favouring losartan (4.8 vs 8.7%; RRR = 46%; 95% CI 5–69%; $P = 0.035$).⁶⁰ In response, the larger ($n = 3152$) ELITE II study was undertaken, using virtually the same study design, in which all-cause death, as a primary end-point, was similar for both drugs (11.7% vs 10.4%).⁶¹

Beware conflated trials

We define a conflated trial as the situation where two separate experiments (**I**) and their outcome results (**O**), which may go in opposite directions, are reported as if they arose from a single trial and where aggregated data

may give rise to misleading interpretations. We advise clinicians to be clear on how comparisons were performed, and to ensure that results of all comparisons are presented and any interpretations justified.

Example. In the PROGRESS trial, 6105 patients with previous stroke or transient ischaemic attack were randomly assigned active treatment or placebo, the former comprising perindopril (4 mg/day), with addition of a second agent indapamide at the discretion of treating clinicians.⁶² After mean follow up of 3.9 years, the active treatment group as a whole had significantly fewer strokes (RRR = 28%; 95%CI 17–38%) and major vascular events (RRR = 26%; 95%CI 16–34%) than the placebo group. The results as reported, together with the phrase ‘perindopril-based blood-pressure-lowering regimen’ in the title might invoke the erroneous interpretation that perindopril alone protects against stroke recurrence, with added benefit if combined with indapamide.

However, the PROGRESS study was in fact two trials in one as patients were randomised, according to clinician preference, to: (i) perindopril plus indapamide or double placebo or (ii) perindopril alone or placebo. (Note: no third arm randomising patients to indapamide alone vs placebo.) The separate results for each trial (Table 5) clearly show that perindopril alone had no outcome effect, a result de-emphasised in several interpretations of PROGRESS results recommending perindopril be initiated post-stroke,^{63,64} and indapamide added if possible. Indapamide alone might be superior to perindopril plus indapamide on the basis of a 29% decrease in stroke recurrence with indapamide versus placebo in the PATS trial.⁶⁵

Beware class effect assumptions

Virtually all drug classes include multiple compounds manufactured by different drug companies. We define ‘drug class’ as those drugs that share a similar chemical structure, mechanism of action and biological target (**I**). Because of these similarities, drug classes are generally thought to confer similar pharmacologic effects and clinical outcomes; that is, to show class effects, which often imply, in addition, similar safety profiles.

If a class effect is assumed, then choosing one from several drugs of the same class relies on putative differences in cost, ease of administration, or patient tolerability. Assuming class effects, however, assumes there are rarely differences between drugs of the same

Table 5 Results for each arm of PROGRESS⁶²

Outcome	Perindopril + Indapamide (%)	Double placebo (%)	RRR (95% CI)	NNT (95% CI)
Stroke	8.5	14	43% (30–54)	17 (13–27)
Major vascular event	13	21	40% (29–49)	14 (10–20)
	Perindopril alone (%)	Single placebo (%)	RRR (95% CI)	NNT
Stroke	12.3	12.9	5% (–19–23)	Not significant
Major vascular event	17.7	18.5	4% (–15–20)	Not significant

CI, confidence interval; NNT, number needed to treat; RRR, relative risk reduction

class in clinically important physiological or pharmacokinetic effects.

Example 1. The glycoprotein IIb/IIIa inhibitors tirofiban and eptifibatid were shown in separate placebo-controlled trials to be more effective than usual care alone in reducing rates of death and infarction in patients with acute coronary syndrome.^{66,67} In contrast, another inhibitor, abciximab, proved ineffective in the same clinical settings,⁶⁸ possibly explained by dose-related variation in platelet inhibition.⁶⁹

Example 2. A class effect of ACE inhibitors in patients with stable cardiovascular disease and preserved left ventricular function is challenged by the discordant results of the HOPE,⁷⁰ EUROPA,⁷¹ and PEACE⁷² studies at 4–5 years follow up. In HOPE, ramipril (target dose 10 mg/day) compared to placebo reduced cardiovascular events by 18% (95% CI 14–30%; $P < 0.001$) in 9297 patients with known vascular disease or diabetes plus another cardiovascular risk factor.⁷⁰ Similarly, in EUROPA, events were reduced by 20% (95% CI 9–29%; $P = 0.003$) by perindopril (target dose 8 mg/day) in 13 655 patients with stable coronary disease.⁷¹ In contrast, in PEACE, trandolopril (target dose 4 mg/day) exerted no effect on events in 8290 patients with stable coronary disease.⁷²

These discordant results might be explained by different levels of tissue penetration⁷³ or, alternatively, between-trial differences in end-point definitions, trial duration, baseline risk of adverse events, and the type and intensity of cointerventions.⁷⁴ Proving superiority or equivalence of members of a drug class requires head-to-head trials using identical study designs and bioequivalent drug dosing; that is, direct within-study comparisons.

As such trials are difficult to fund because of large samples required to show significant differences, and because results might threaten commercial viability of competing products, we are frequently forced to rely on indirect comparisons across trials which compare different drugs with placebo, which are of similar design and quality.

Drugs A and B can be compared indirectly by comparing RR estimates from studies of drug A versus placebo (**P**) and those of drug B versus p, such that $RR_{A \text{ versus } B} = RR_{A \text{ versus } p} / RR_{B \text{ versus } p}$.⁷⁵ Such comparisons assume absence of potential confounders, and consistency of drug effect (or RR) across different patient subgroups. Although indirect methods have validity when comparing different interventions for the same clinical condition,⁷⁶ they tend to overestimate differences in effect when comparing different drugs of the same class.⁷⁵ We propose criteria listed in Table 6 as guides for deciding the existence of class effects.⁷⁷

Beware equivalence trials

Most therapeutic RCTs are designed as superiority trials, attempting to show one agent is better than another (**I**). In contrast, increasing numbers of studies, including those of same class drugs, aim to demonstrate equivalence (or non-inferiority); that is, one is not worse than another by a prespecified amount of clinically important effect (**O**). A treatment that is judged not to be inferior to, or indeed better than, another might be preferred because of less cost, toxicity or invasiveness.

However, of 88 equivalence trials, only 45 (51%) specifically stated equivalence as the study objective, and only 19 (22%) satisfied all validity criteria listed in Table 7.⁷⁸ The most common flaw was poor definition, or marked inconsistency, of quantitative boundaries used to determine, a priori, the magnitude of difference deemed to be clinically important. Clinical judgement should be applied in deciding the appropriateness of arbitrarily defined equivalence boundaries.

Example. In the INJECT trial,⁷⁹ two thrombolytic treatments, reteplase and streptokinase, were to be considered equivalent if the absolute mortality difference between groups was $<1\%$, this boundary being derived from the GUSTO trial where a 1% mortality difference was considered clinically significant.⁸⁰ However, in another thrombolytic trial, COBALT,⁸¹ comparing continuous with double-bolus reteplase, equivalence was defined as a mortality difference $<0.4\%$, this being the lower confidence limit of the 1% mortality difference

Table 6 Criteria for determining likelihood of class effect⁷⁴

Essential

- Clearly defined, common chemical structure, mechanism of action and biological target (or pathway)
- Comparable efficacy demonstrated for multiple agents within the class (either by head-to-head RCT or multiple RCT involving different agents being compared with a common comparator (placebo or 'conventional care')
- Absence of convincing evidence that there is a member of the class that does not have comparable clinical benefit to that of other agents in the class

Supportive

- Clinical circumstances (patient selection criteria, concomitant treatments, clinical setting etc.) under which benefit (and harm) have been determined is the same for all agents
- Particular patient subgroups that show benefit (or lack of benefit or harm) are similar for all agents, and all agents show similar broad (or narrow) clinical benefit
- Absolute and relative levels of benefit are comparable among agents, and the extent and depth of evidence of benefit are similar among agents
- Safety profile and tolerability of individual agents have been evaluated and seen to be similar

RCT, randomised clinical trials.

Table 7 Validity criteria for equivalence (non-inferiority) trials⁷⁸

-
- Equivalence is stated as the study objective
 - Method and units of measurement for detecting differences between different agents are clearly specified:
 - Absolute differences in means, medians or rates
 - Proportionate (or relative) differences
 - Standardised effect sizes
 - Maximum value of clinically important difference in effect beyond which a difference would deem drugs being compared as nonequivalent (also called the boundary of equivalence) is prespecified

Although it is not possible to specify a single value of clinically important difference in all cases, some acceptable boundaries of equivalence might include:

 - Absolute differences of 3%
 - Proportionate differences of 10%
 - Standardised effect size of 0.25
 - Method of statistical testing

Testing for equivalence should not rely on a failed test for superiority, but instead rely on comparing observed differences against a specific equivalence boundary. A direct test of equivalency is intended to reject the 'alternative hypothesis' that the true difference is larger than the boundary limit, whereas statistical tests for superiority are aimed at rejecting a null hypothesis of no difference.
 - Calculation of sample size

Study must be adequately powered, with advance calculation of minimum sample size, to prevent observed differences, which would suggest inferiority, failing to reach statistical significance
-

used in GUSTO.⁸⁰ The INJECT mortality results of 9.53 versus 9.02% led to a declaration that both drugs were equivalent (absolute difference = 0.51% and <1.0%, the chosen equivalence boundary), whereas in COBALT, they were declared not equivalent despite an absolute mortality difference of 0.44% (7.98 *vs* 7.54%), as the equivalence boundary here was 0.4%.

Beware sponsor bias

Funding of trials by pharmaceutical company sponsors raises concerns about undue influence of commercial interests on design and conduct of trials and analysis and reporting of their results.⁸² Although methodologic quality of pharmaceutically sponsored trials seems no worse than that of trials funded from other sources, the former, for the same intervention, more commonly produce outcomes favouring the sponsor.⁸³ Also, pharmaceutical sponsorship increases the odds of strong recommendations favouring the intervention fivefold, after taking into account large effect sizes (measure of benefit) and use of blinding (measure of trial quality).⁸⁴

We suggest clinicians exercise caution in accepting findings of trials that do not contain a detailed statement on the role of the sponsor or funding agency, or where the sponsor, or its contracted agencies, have been directly involved in data gathering and analysis, or have disproportionate representation in trial governance.

Example. The ESTEEM trial, as its primary end-point, evaluated over 6 months the dose–response of the oral direct thrombin inhibitor, ximelagatran (XN) plus aspirin (*vs* aspirin and placebo) on all-cause death, non-fatal myocardial infarction and severe recurrent ischaemia in 1883 patients following acute myocardial infarction.⁸⁵ The abstract stated: 'oral XN significantly reduced the risk for the primary end-point by 14% (95% CI 2–41; $P = 0.036$) for the *combined XN groups* versus placebo'. (This was not the prespecified primary end-

point for which no significant difference existed between any of the four separate XN dosage groups and placebo. We contend data for all patients receiving XN were aggregated post-hoc to derive a result that just reached statistical significance.) 'There was no indication of a dose–response between the XN groups'. (In the context of the previous sentence, this falsely implies all dosages were effective.)

In the discussion section, 'XN is effective at a wide range of doses with a low risk of bleeding'. (This refers to dosage effects on a predefined secondary, composite end-point for which the results for two doses were non-significant.) Finally, 'the lowest dose of 24 mg XN twice daily achieved maximum efficacy at an acceptable safety profile'. (True only for the secondary end-point, but major bleeding risk was clinically comparable (1–3%) among all other dosage groups and, overall, 15% of XN patients discontinued the drug because of adverse events *vs* 6% for placebo, challenging the repeated inference of relative safety.)

The inconsistencies in data analysis and reporting suggests to us a biased attempt to present ESSENCE in a positive light. Four of 7 authors and 4 of 7 members of the trial executive committee were, or had previously been, drug company employees; the trial executive chairman and the lead author both received company research grants; and the company's research and development centre undertook data co-ordination.

Beware publication bias

Historically, trialists and perhaps medical journal editors have shown a bias against publicising negative trials that show no effect or even inferiority of new therapies. This can lead to biased perceptions of their overall efficacy.

Example 1. In a prospective follow up of 126 trials submitted to the ethics committee of a major Sydney tertiary hospital, those with significantly positive results were more likely to be published (85 *vs* 65% over

10 years), and be published earlier (median time to publication 4.8 years *vs* 8.0 years) than trials showing nil effect.⁸⁶ Similar observations are seen for research conducted within various subspecialties.^{87,88}

Example 2. A meta-analysis of published trials involving 1908 patients with ovarian cancer indicated improved survival with combination chemotherapy compared to alkylating agent monotherapy (RRR = 16%; 95% CI 6–27; $P = 0.004$).⁸⁹ However, an analysis of similar trials in an international registry (8 published and 5 unpublished trials involving 2491 patients) showed no significant difference in survival (RRR = 6%; 95% CI –3 to 15%; $P = 0.17$).⁸⁹

Hopefully, the recent decision of the International Committee of Medical Journal Editors to publish only those trials that, if commenced on or after 1 July 2005, have been registered at their inception with a publicly accessible trial registry, will reduce publication bias.⁹⁰ Pharmaceutical companies have also agreed to make publicly available all on-file information from sponsored trials of new prescription medicine,⁹¹ although this agreement is voluntary among larger companies and ideally should be mandatory for all companies.⁹²

HOW TO AVOID BEING MISLED

In minimising the chances of inappropriately changing practice on the basis of biased interpretations and conclusions, we recommend the following when reviewing the results of the latest trial.

1 Resist the temptation to only read the Introduction and Conclusion sections, and carefully read the Methods and Results sections as well.

2 Look for situations that, on the basis of the preceding discussion, increase the risk of being misled:

- Use of surrogate outcomes
- Reporting of clinically borderline benefits
- Use of composite end-points, especially clinician-driven end-points
- Presentation of results that are inconsistent with stated study objectives and methods
- Use of post-hoc analyses or multiple subgroup analyses
- Absence of reporting of safety and toxicity data
- Trials dealing with class effects or equivalence of effect
- Role of commercial sponsors not stated or suggestion of undue influence on study design, analysis and reporting

3 Look for (or calculate from the data provided) absolute measures of benefit and harm.

4 Ensure interpretations and conclusions are substantiated by factual data; beware subliminal messages contained within selective or obfuscated reporting of results.

5 Compare the way results have been reported and interpreted in original articles with that found in independent, pre-appraised sources such as Evidence-based Medicine,⁹³ ACP Journal Club,⁹⁴ InfoPOEMS,⁹⁵ Cochrane Library,⁹⁶ National Prescribing Service Rational Assessment of Drugs and Adverse Reports (RADAR)⁹⁷ and the Critically Appraised Topic articles in this journal and in the IMSANZ–CATs Library.⁹⁸

REFERENCES

- 1 Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine: How to Practice and Teach EBM, 2nd edn. Edinburgh: Churchill Livingstone; 2000.
- 2 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; 357: 1191–4.
- 3 Internal Medicine Society of Australia and New Zealand Evidence-based Medicine Working Group. Goals and Objectives [online: cited 12 January 2005]. Sydney: Internal Medicine Society of Australia and New Zealand. Available from URL: <http://www.racp.edu.au/imsanz/members>
- 4 Montori VM, Jaeschke R, Schunemann HJ, Bhandari M, Brozek JL, Devereaux PJ *et al.* Users' guide to detecting misleading claims in clinical research reports. *BMJ* 2004; 329: 1093–6.
- 5 Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005; 365: 1159–62.
- 6 Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomised trials: comparison of protocols to published articles. *JAMA* 2004; 291: 2457–65.
- 7 Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. *BMJ* 2002; 325: 249.
- 8 Richardson WS, Wilson MC, Nishikawa J, Hayward RSA. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995; 123: A12–13.
- 9 Julian DG, Pocock SJ. Interpreting a trial report. In: Pitt B, Julian D, Pocock S, eds. *Clinical Trials in Cardiology*. London: Saunders; 1997; 33–42.
- 10 Johnstone D, Limacher M, Rousseau M, Liang CS, Ekelund L, Herman M *et al.* Clinical characteristics of patients in studies of left ventricular dysfunction (SOLVD). *Am J Cardiol* 1992; 70: 894–900.
- 11 Evans A, Kalra L. Are the results of randomised controlled trials on anticoagulation in patients with atrial fibrillation generalisable to clinical practice? *Arch Intern Med* 2001; 161: 1443–7.
- 12 Rothwell PM. Treating individuals 1. External validity of randomised controlled trials: 'to whom do the results of this trial apply?'. *Lancet* 2005; 365: 82–93.
- 13 Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A *et al.* The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomised Aldactone Evaluation Study Investigators. *N Engl J Med* 1999; 341: 709–17.
- 14 Bozkurt B, Agoston I, Knowlton AA. Complications of inappropriate use of spironolactone in heart failure: when an old medicine spirals out of new guidelines. *J Am Coll Cardiol* 2003; 41: 211–14.
- 15 Juurlink DN, Mamdani MM, Lee DS, Kopp A, Austin PC, Laupacis A *et al.* Rates of hyperkalemia after publication of the Randomised Aldactone Evaluation Study. *N Engl J Med* 2004; 351: 543–51.
- 16 Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. Evidence-Based Medicine Working Group. *JAMA* 1998; 279: 545–9.
- 17 Frishman WH, Ruggio J, Furberg C. Use of beta-adrenergic blocking agents after myocardial infarction. *Postgrad Med* 1985; 78: 49–53.
- 18 Salpeter S, Ormiston T, Salpeter E. Cardioselective beta-blockers for reversible airway disease. *Cochrane Database Syst Rev* 2002; (4): CD002992.
- 19 Shorr RI, Ray WA, Daugherty JR, Griffin MR. Antihypertensives and the risk of serious hypoglycaemia in older persons using insulin and sulphonylureas. *JAMA* 1997; 278: 40–3.

- 20 Gerstman BB, Jolson HM, Bayer M, Cho P, Livingston JM, Platt R. The incidence of depression in new users of beta-blockers and selected antihypertensives. *J Clin Epidemiol* 1996; 49: 809–15.
- 21 Gottlieb SS, Mccarter RJ, Vogel RA. Effect of beta-blockade on mortality among high-risk and low-risk patients after myocardial infarction. *N Engl J Med* 1998; 339: 489–97.
- 22 Cannon CP, Braunwald E, McCabe CH, Rader DJ, Rouleau JL, Belder R *et al*. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. Pravastatin or Atorvastatin Evaluation and Infection Therapy-Thrombolysis in Myocardial Infarction 22 Investigators. *N Engl J Med* 2004; 350: 1495–1504.
- 23 La Rosa JC, Grundy SM, Waters DD, Shear C, Barter P, Fruchart JC *et al*. Intensive lipid lowering with atorvastatin in patients with stable coronary disease. Treating to New Target (TNT) Investigators. *N Engl J Med* 2005; 352: 1425–1435.
- 24 Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature. XIX. Applying clinical trial results A. How to use an article measuring the effect of an intervention on surrogate end-points. *JAMA* 1999; 282: 771–8.
- 25 The Cardiac Arrhythmia Pilot Study Investigators. Effects of encainide, flecainide, imipramine and moricizine on ventricular arrhythmias during the year after acute myocardial infarction: the CAPS. *Am J Cardiol* 1988; 61: 501–9.
- 26 Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH *et al*. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med* 1991; 324: 781–8.
- 27 Birks JS, Harvey R. Donepezil for dementia due to Alzheimer's disease. *Cochrane Database Syst Rev* 2003; 3. doi: 10.1002/14651858. CD001190.
- 28 Folstein MF, Folstein SE, Mchugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12: 189–98.
- 29 Weyer G, Erzigkeit H, Kanowski S, Ihl R, Hadler D. Alzheimer's disease assessment scale: reliability and validity in a multicenter clinical trial. *Int Psychogeriatr* 1997; 9: 123–38.
- 30 Courtney C, Farrell D, Gray R, Hills R, Lynch L, Sellwood E *et al*. Long-term donepezil treatment in 565 patients with Alzheimer's disease (AD2000): a randomised double-blind trial. *Lancet* 2004; 363: 2105–15.
- 31 Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med* 1992; 92: 121–4.
- 32 Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness. *Ann Intern Med* 1992; 117: 916–21.
- 33 Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on clinicians' willingness to prescribe. *Lancet* 1994; 343: 1209–11.
- 34 MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J *et al*. Blood pressure, stroke, and coronary heart disease. Part 1. Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990; 335: 765–74.
- 35 Loke TW, Koh FC, Ward JE. Pharmaceutical advertisement claims in Australian medical publications. Is evidence accessible, compelling and communicated comprehensively? *Med J Aust* 2002; 177: 291–3.
- 36 Cooper RJ, Schriger DL, Wallace RC, Mikulich VJ, Wilkes MS. The quantity and quality of scientific graphs in pharmaceutical advertisements. *J Gen Intern Med* 2003; 18: 294–7.
- 37 Scott T, Stanford N, Thompson DR. Killing me softly: myth in pharmaceutical advertising. *BMJ* 2004; 329: 1484–8.
- 38 Wing LMH, Reid CM, Ryan P, Beilin LJ, Brown MA, Jennings GLR *et al*. A comparison of outcomes with angiotensin-converting-enzyme inhibitors and diuretics for hypertension in the elderly. *N Engl J Med* 2003; 348: 583–92.
- 39 ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomised to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA* 2002; 288: 2981–97.
- 40 Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH *et al*. Health outcomes associated with antihypertensive therapies used as first-line agents. A network meta-analysis. *JAMA* 2003; 289: 2534–44.
- 41 Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992; 116: 78–84.
- 42 Packer M, O'Connor CM, Ghali JK, Pressler ML, Carson PE, Belkin RN *et al*. Effect of amlodipine on morbidity and mortality in severe chronic heart failure. Prospective Randomised Amlodipine Survival Evaluation Study Group. *N Engl J Med* 1996; 335: 1107–14.
- 43 Thackray S, Witte K, Clark AL, Cleland JGF. Clinical trials update: OPTIME-CHF, PRAISE-2, ALL-HAT. *Eur J Heart Fail* 2000; 2: 209–12.
- 44 Bresalier RS, Sandler RS, Quan H, Bolognese JA, Oxenius B, Horgan K *et al*. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. Adenomatous Polyp Prevention on Vioxx (Approve) Trial Investigators. *N Engl J Med* 2005; 352: 1092–102.
- 45 Solomon SD, McMurray JJ, Pfeffer MA, Wittes J, Fowler R, Finn P *et al*. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. Adenoma Prevention with Celecoxib (APC) Study Investigators. *N Engl J Med* 2005; 352: 1071–80.
- 46 Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B *et al*. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. The Vigor Study Group. *N Engl J Med* 2000; 343: 1520–8.
- 47 Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A *et al*. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: a randomised controlled trial. *JAMA* 2000; 284: 1247–55.
- 48 Wright J. The double-edged sword of COX-2 selective NSAIDs. *CMAJ* 2002; 167: 1131–7.
- 49 Mukherjee D, Nissen SE, Topol EJ. Risk of cardiovascular events associated with selective COX-2 inhibitors. *JAMA* 2001; 286: 954–9.
- 50 Juni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet* 2004; 364: 2021–9.
- 51 Zhao SZ, Burke TA, Whelton A, von Allmen H, Henderson SC. Comparison of the baseline cardiovascular risk profile among hypertensive patients prescribed COX-2-specific inhibitors or non-specific NSAIDs: data from real practice. *Am J Manag Care* 2002; 8: S392–S400.
- 52 Mamdani M, Juurlink DN, Kopp A, Naglie G, Austin PC, Laupacis A. Gastrointestinal bleeding after the introduction of COX 2 inhibitors: ecological study. *BMJ* 2004; 328: 1415–16.
- 53 Sackett DL, Haynes RB, Gent M. Compliance. In: Inman WH, ed. *Monitoring for Drug Safety*. Philadelphia, PA: Lippincott; 1980; 427–38.
- 54 Brewer T. Postmarketing surveillance and adverse drug reactions. Current perspectives and future needs. *JAMA* 1999; 281: 824–9.
- 55 Temple R. Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. *JAMA* 1999; 281: 841–4.
- 56 Lauer MS, Topol EJ. Clinical trials – multiple treatments, multiple end-points, and multiple lessons. *JAMA* 2003; 289: 2575–7.
- 57 Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomised trials. Greater precision but with greater uncertainty? *JAMA* 2003; 289: 2554–9.

- 58 UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998; 352: 837–53.
- 59 Shaughnessy AF, Slawson DC. What happened to the valid POEMs? A survey of review articles on the treatment of type 2 diabetes. *BMJ* 2003; 327: 266.
- 60 Pitt B, Segal R, Martinez FA, Meurers G, Cowley AJ, Thomas I *et al*. Randomised trial of losartan versus captopril in patients over 65 with heart failure (Evaluation of Losartan in the Elderly Study, ELITE). *Lancet* 1997; 349: 747–52.
- 61 Pitt B, Poole-Wilson PA, Segal R, Martinez FA, Dickstein K, Camm AJ *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.
- 62 Progress Collaborative Group. Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack. *Lancet* 2001; 358: 1033–41.
- 63 Jackson G. Making PROGRESS in stable patients post-stroke or transient ischaemic attack: implications for general practice. *Int J Clin Pract* 2003; 57: 385–7.
- 64 Tonkin AM. Does blood pressure prevent recurrent stroke? *Med J Aust* 2002; 176: 283–4.
- 65 PATS Collaborating Group. Post-stroke antihypertensive treatment study. A preliminary result. *Chin Med J* 1995; 108: 710–17.
- 66 Platelet Receptor Inhibition in Ischemic Syndrome Management (PRISM) Investigators. A comparison of aspirin plus tirofiban with aspirin plus heparin for unstable angina. *N Engl J Med* 1998; 21: 1498–505.
- 67 The Pursuit Trial Investigators. Inhibition of platelet glycoprotein IIb/IIIa with eptifibatid in patients with acute coronary syndromes. Platelet glycoprotein IIb/IIIa in unstable angina: receptor suppression using integrilin therapy. *N Engl J Med* 1998; 339: 436–43.
- 68 Simoons ML. 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. GUSTO IV-ACS Investigators. *Lancet* 2001; 357: 1915–24.
- 69 Batchelor WB, Tolleson TR, Huang Y, Larsen RL, Mantell RM, Dillard P *et al*. Randomised Comparison of the early and late Platelet inhibition with abciximab, tirofiban and eptifibatid during percutaneous coronary intervention: the compare study. *Circulation* 2002; 106: 1470–6.
- 70 Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N Engl J Med* 2000; 342: 145–53.
- 71 Fox KM. Efficacy of perindopril in reduction of cardiovascular events among patients with stable coronary artery disease: randomised, double-blind, placebo-controlled, multicentre trial (the EUROPA study). European Trial On Reduction of Cardiac Events with Perindopril in Stable Coronary Artery Disease Investigators. *Lancet* 2003; 362: 782–8.
- 72 Braunwald E, Domanski MJ, Fowler SE, Geller NL, Gersh BJ, Hsia J *et al*. Angiotensin-converting-enzyme inhibition in stable coronary artery disease. Peace Trial Investigators. *N Engl J Med* 2004; 351: 2058–68.
- 73 Tsikouris JP, Suarez JA, Meyerrose GE, Ziska M, Fike D, Smith J. Questioning a class effect: does ACE inhibitor tissue penetration influence the degree of fibrinolytic balance alteration following an acute myocardial infarction? *J Clin Pharmacol* 2004; 44: 150–7.
- 74 McAlister FA, Laupacis A, Wells GA, Sackett DL. 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. Evidence-Based Medicine Working Group. *JAMA* 1999; 282: 1371–7.
- 75 Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomised controlled trials. *J Clin Epidemiol* 1997; 50: 683–91.
- 76 Song F, Altmann DG, Glenny A-M, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003; 326: 472.
- 77 Antmann EM, Ferguson JJ. Should evidence-based proof of efficacy as defined for a specific therapeutic agent be extrapolated to encompass a therapeutic class of agents? *Circulation* 2003; 108: 2604–7.
- 78 Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: are they supported by the evidence? *Ann Intern Med* 2000; 132: 715–22.
- 79 International Joint Efficacy Comparison Thrombolytics. Randomised, double-blind comparison of reteplase double-bolus administration with streptokinase in acute myocardial infarction: trial to investigate equivalence. *Lancet* 1995; 346: 329–36.
- 80 The GUSTO investigators. An international randomised trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med* 1993; 329: 673–82.
- 81 The Continuous Infusion versus Double-bolus Administration of Alteplase (COBALT) Investigators. A comparison of continuous infusion of alteplase with double-bolus administration for acute myocardial infarction. *N Engl J Med* 1997; 337: 1124–30.
- 82 Angell M. *The Truth About the Drug Companies: How They Deceive Us and What to Do About It*. Boston: Random House; 2004.
- 83 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003; 326: 1167–76.
- 84 Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomised drug trials. A reflection of treatment effect or adverse events. *JAMA* 2003; 290: 921–8.
- 85 Wallentin L, Wilcox RG, Weaver WD, Emanuelsson H, Goodvin A, Nystrom P *et al*. Oral ximelagatran for secondary prophylaxis after myocardial infarction: the ESTEEM randomised controlled trial. *Lancet* 2003; 362: 789–97.
- 86 Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of research subjects. *BMJ* 1997; 315: 640–5.
- 87 Krzyzanowska MK, Pintilie M, Tannock IF. Factors associated with failure to publish large randomised trials presented at an oncology meeting. *JAMA* 2003; 290: 495–501.
- 88 Timmer A, Hilsden RJ, Cole J, Hailey D, Sutherland LR. Publication bias in gastroenterological research – a retrospective cohort study based on abstracts, submitted to a scientific meeting. *BMC Med Res Methodol* 2002; 2: 7.
- 89 Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987; 6: 11–29.
- 90 Deangelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R *et al*. Clinical trial registration. A Statement from the International Committee of Medical Journal Editors. *Med J Aust* 2004; 181: 293–4.
- 91 Mayor S. Drug companies agree to make clinical trial results public. *BMJ* 2005; 330: 109.
- 92 Joint Position on the Disclosure of Clinical Trial Information via Clinical Trials Registries and Databases [online; cited 16 January 2005]. Available from URL: http://www.efpia.org/4_pos/sci_regu/Clinicaltrials2005.pdf
- 93 Evidence-based Medicine On-line [online; cited 12 January 2005]. Available from URL: <http://www.evidence-basedmedicine.com>

- 94 ACP Journal Club On-line [online; cited 12 January 2005]. Available from URL: <http://www.acponline.org>
- 95 InfoPOEMS (Patient Oriented Evidence that Matters) [online; cited 12 January 2005]. Available from URL: <http://www.infopoems.com>
- 96 Cochrane Library [online; cited 12 January 2005]. Available from URL: <http://www.nicsl.com.au/cochrane>
- 97 National Prescribing Service Rational Assessment of Drugs Research (RADAR) [online; cited 12 January 2005]. Available from URL: <http://www.npsradar.org.au>
- 98 Internal Medicine Society of Australia and New Zealand. Critically Appraised Topics (CATs) Library [online; cited 20 August 2005]. Available from URL: <http://www.imsanz.org.au/members/resources/CATs/library.cfm>