STATISTICS IN MEDICINE
Statist. Med. 2002; **21**:2939–2948 (DOI: 10.1002/sim.1298)

Policy developments in regulatory approval[‡]

Robert Temple[*,†]

Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, MD, U.S.A.

SUMMARY

Although radical changes in drug regulation are rare (e.g., the Federal Food, Drug and Cosmetic Act of 1938 and the 1962 amendment to the Act creating an effectiveness requirement), regulations and guidance do evolve significantly in the face of new problems and accumulating experience. Recent changes have been driven by the Food and Drug Administration Modernization Act (FDAMA), user fee legislation, the International Conference on Harmonization, recent safety related drug withdrawals, and concerns about trial ethics and investigator conflict of interest.

FDAMA and guidance developed in response to it has helped circumstances in which FDA would rely on a single study to support effectiveness and the circumstances in which surrogate endpoints could support approval. An ICH Document 'Choice of control group and related design issues in clinical trials' focussed attention on the ethics of placebo controls (acceptable, even if there is existing therapy, when the placebo-treated patient will suffer no irreversible injury) and the design of 'equivalence' or 'non-inferiority' trials.

There has been greatly increased attention to obtaining good dose-response information and to assessing need for modifying treatment in demographic (age, gender, race) and concomitant disease (renal or hepatic function abnormalities) subgroups, and in assessing drug–drug interactions.

Other important trends are increasing reliance on non-U.S. data, increasing numbers of FDA-industry meetings during drug development, and new focus on risk assessment and risk management. Published in 2002 by John Wiley & Sons, Ltd.

KEY WORDS: placebo controls; surrogate endpoints; single study; non-inferiority studies

INTRODUCTION

Real revolutions in drug regulation are rare (The Food, Drug and Cosmetic Act of 1938 and the Kefauver-Harris amendments to the Act in 1962) but regulation does evolve. I would identify six major 'change engines' that have had an effect in recent years:

1. FDAMA (Food and Drug Administration Modernization Act, 1997);
2. PDUFA (Prescription Drug User Fee Act, 1992);
3. conspicuous drug withdrawals;
4. ICH (International Conference on Harmonization) and its guidelines;

[*] Correspondence to: Robert Temple, Center for Drug Evaluation and Research, 5600 Fishers Lane, HFD-101/WOC II-6014, Rockville, MD 20857, U.S.A.
[†] E-mail: temple@cder.fda.gov

Published in 2002 by John Wiley & Sons, Ltd.

Table I. Critical considerations in current standards evidence of treatment efficacy.

Number of studies
Surrogate endpoints
Equivalence/non-inferiority
Individualization of therapy
Monitoring/data collection – drug industry standard versus 'other'
Reliance on non-U.S. data

5. attention to trial ethics and growth of investigator entrepreneurs and contract research organizations (CROs);
6. The conduct of large outcome trials by commercial sponsors.

The most important changes have been in the standard of evidence (Table I).

Other important changes include Food and Drug Administration (FDA)/Industry meetings and agreements and completion of phase 4 agreements, focus on safety/risk management, and rising public ethical concerns.

STANDARD OF EVIDENCE

The modern standard for evidence of effectiveness was set forth in the 1962 amendments to the Food, Drug and Cosmetic (FD and C) Act. They required that for approval there must be 'substantial evidence' that the drug would have its claimed effect. 'Substantial evidence' meant evidence derived from adequate and well-controlled clinical studies. In general, that phrase (studies, plural) has been interpreted as meaning that at least two such studies were needed. Over the years, however, exceptions to this rule were sometimes made for good reasons. In addition, the courts have established that the claimed effect needs to be clinically meaningful. In general, this meant that the FDA was very careful about relying on a drug effect on a surrogate endpoint for approval; an exception could be made, however, for serious and life-threatening diseases. The FDAMA put into law two existing FDA practices:

1. The FDA would rely in some cases on a single persuasive study, particularly where the study showed a survival effect. There was no FDA regulation specifically allowing this, but guidance was being developed to describe when the agency might rely on a single trial.
2. In serious or life-threatening diseases, where a treatment appeared to represent a clear improvement over available therapy, the FDA would rely on an effect on a surrogate endpoint 'reasonably likely' to predict a clinical benefit as a basis for 'accelerated approval'. Accelerated approval is really a conditional approval, with approval conditioned on the post-approval conduct of well-controlled studies to demonstrate the expected clinical benefit, and with provision for prompt withdrawal of the drug if the manufacturer fails to do so. Accelerated approval was incorporated into regulations in 1992.

The FDAMA incorporates accelerated approval (and ability to rely on certain surrogate endpoints) into a 'fast-track' provision and explicitly allows reliance on a single study.

RELIANCE ON A SINGLE STUDY

The FD and C Act (1962) refers to results of 'adequate and well-controlled studies' as the only acceptable basis for concluding that there is 'substantial' evidence of effectiveness, the legal standard for approval. Legislative history says the plural was intended to refer to the 'quantity' of evidence needed and the FDA's long-standing policy and legal position were that replication (not of the exact study, but of the finding) was expected. In fact, however, unless two studies happened to be ongoing together or the result of the first study was in doubt, successful mortality studies have rarely been repeated (post-infarction beta blockers, CHF outcome trials of ACE inhibitors at any given severity, statin outcome trials in a given population), for fairly obvious ethical reasons.

The FDAMA made explicit the FDA's legal ability to rely on a single adequate and well controlled clinical investigation 'with confirmatory evidence' (not further defined) to conclude there is substantial evidence of effectiveness, that is, to conclude that the legal standard for approval had been met. The FDA then wrote a guidance document in 1998 ('Providing clinical evidence of effectiveness for human drug and biological products') that explained the circumstances in which we were most likely to rely on a single study (no one of these is determinative; each can contribute support):

1. situations in which a trial has demonstrated an effect on mortality or irreversible morbidity, and replication would be ethically impossible;
2. large multi-centre study, with no single investigator or site providing a disproportionate number of patients;
3. consistency across study subsets (demographic, severity);
4. the presence of multiple studies within a study, for example, monotherapy plus add-on in the same study (ISIS II), multiple doses, each showing an effect versus placebo;
5. multiple endpoints involving different events, such as both death and AMI (This would not be reassuring with respect to possible bias in a single study but would be reassuring as to the possibility that random error led to the observed effect. The events need to be clearly different. Two different depression scales, for example, would not be different events but two measures of the same event);
6. statistically very persuasive finding;
7. no inconsistencies or problems with the study;
8. (one not in the guidance) similar findings with a closely related drug or in a closely related population (demonstration of ACEI benefit in CHF in progressively less ill populations).

RELIANCE ON A SURROGATE ENDPOINT

Some (but not many) surrogate endpoints are used for 'ordinary' approval, such as blood pressure lowering for antihypertensives. 'Accelerated approval' allows reliance on less well-established surrogates (that is, not as well-established as BP, cholesterol, blood sugar and HgAIC) when a drug is intended for use in a serious or life-threatening disease without adequate available therapy. Since the rule was promulgated in 1992, it has primarily been used for drugs for HIV infection (with effects on T4 lymphocytes counts and various measures of

viral load serving as surrogate endpoints) and cancer, where tumour response rates in refractory illness have led to approval. There are only a few other examples.

Midodrine (1996) was approved for orthostatic hypotension based on clear improvement in blood pressure fall on standing but without clear effect on symptoms. Approval was conditioned on the conduct of controlled trials to document clinical benefit.

Celecoxib (1999) was approved for use in reducing the number of adenomatous polyps in familial adenomatous polyposis (FAP), as an adjunct to usual care (for example, endoscopic surveillance, surgery). As the clinical consequences of such a decrease were not certain, approval was conditioned on agreement to conduct a randomized trial in FAP genotypically positive adolescents without polyps, a group in which delayed appearance of polyps would be considered a clear benefit.

Beta-interferon was approved for MS partly based on MRI findings; approval was conditioned on studies to explore the effects of beta-interferon on progression and disability.

The FDAMA also expresses a generally favourable attitude toward reliance on surrogates and requires the FDA to 'establish a program to encourage the development of surrogate endpoints that are reasonably likely to predict clinical benefit for serious or life-threatening conditions for which there exists significant unmet medical needs'. Conferences with the NIH have to date not produced agreement on other areas where surrogates should be relied on, although their use during drug development is strongly urged. An area for future discussion will surely be whether and when to rely on endpoints such as prevention of hyperplasia and other pre-neoplastic states, or of intraepithelial neoplasia, as a basis for cancer prevention claims.

EQUIVALENCE/NON-INFERIORITY STUDIES

Use of placebos when there is an existing effective therapy has been widely discussed in recent years as an ethical issue and as a study design issue. We believe, and a recent ICH document (ICH-E10 'Choice of control group and related design issues in clinical trials') reaches a similar conclusion, that symptomatic therapies can almost always be studied in placebo-controlled trials in consenting informed subjects, even when there is existing treatment, because the patient will suffer no permanent damage and can decide on whether he or she is willing to accept the possible discomfort of delayed treatment. There are some specific situations in which there may be controversy about this. The issue is important because non-inferiority/equivalence studies cannot provide evidence of effectiveness unless one can be assured that they have assay sensitivity [1, 2], that is, the ability to distinguish active from inactive therapies. Fortunately, studies of symptoms, where placebos generally can be used, are also the situation in which non-inferiority trials are least likely to be persuasive because assay sensitivity cannot be supported.

Where available treatments affect survival or irreversible morbidity, however, placebo-controlled trials cannot be ethically conducted. In that case, it becomes particularly important to see if it is possible to identify a placebo-controlled study design that is ethically acceptable or to identify a non-inferiority margin [1, 2] that will allow a credible non-inferiority study to be conducted.

We now have a variety of treatments of cardiovascular disease that affect survival or major morbidity, for example, that decrease congestive heart failure (CHF) mortality, decrease post-

infarction mortality, and prevent morbid events after acute coronary syndrome or angioplasty. We cannot leave patients without these treatments in a placebo-controlled study. We can, however, still do add-on placebo-controlled studies of pharmacologically distinct drugs, adding the new treatment or placebo to existing effective treatments, probably a more interesting study anyway than studying a new agent that does no more than previous drugs. Note that this approach does not work for assessment of pharmacologically similar interventions as they could not be expected to have an added effect. These can be studied, if at all, in non-inferiority studies or, in rare cases, by proving more effective than the pharmacologically similar drug.

ICH-E10 ('Choice of control group and related design issues in clinical trials') represents a clear international understanding of the theory and use of non-inferiority studies, and the cautions needed when using them, but leaves to regulatory authorities the decision as to whether a non-inferiority study is credible in a given case and, if so, exactly how to determine the non-inferiority margin, the degree of inferiority of the new treatment to the control that will be ruled out statistically.

Except when it is relatively easy to determine the actual effect size of the control treatment (for example, antibiotics in many situations), the FDA has so far been conservative in its approach to choosing the non-inferiority margin. For thrombolytics, we used the 95 per cent lower bound of the confidence interval from a meta-analysis of thrombolytic placebo-controlled trials to determine the difference between treatment and placebo that we would expect to be present in a new, active-control thrombolytic trial; we then asked that a loss of more than half that effect with the new thrombolytic be ruled out (the non-inferiority margin). We accepted this non-inferiority design because thrombolytic treatment was consistently superior to placebo. If a treatment is not consistently superior to placebo in previous trials, we would consider the non-inferiority design inappropriate, because we could not presume that the new trial would have assay sensitivity. Once the non-inferiority margin is chosen, inferiority to that margin is ruled out at the 5 per cent level.

It has been argued that using such a 'worst case' (95 per cent lower bound of the historical experience to choose the margin and 95 per cent upper bound for the difference between the new drug and the control in the new trial) is over-conservative, but the properties of alternative approaches are not yet well worked out. This is an area that deserves significant attention, probably including simulation efforts.

Even where placebo controls have been regularly inferior to experimental treatments in the past, changes in practice may decrease the effect a control treatment would be expected to have in a new study. For example, thrombolytics are often now used in the United States with angioplasty; angioplasty might markedly decrease the size of the effect of the thrombolytic, making past experience of uncertain relevance to current studies when choosing the non-inferiority margin.

INDIVIDUALIZATION OF THERAPY

If one thinks of drug development in the U.S. in three great 'Ages', after the Age of Safety (beginning in 1938) and the Age of Efficacy (beginning in 1962), we find ourselves in the Age of Individualization of Therapy, the third Age of Drug Development, reflecting recognition that safety and effectiveness can vary with dose, patient characteristics and other therapy.

Developments in the Age of Individualization include:

1. Beginning in 1970s, adequate methods of dose–response (D/R) assessment were developed. ICH E-4 ('Dose–response information to support drug registration', 1994) recognized the importance of dose–response data and the best methods for obtaining it.
2. Interest in possible differences among demographic subgroups has grown and is reflected in a number of FDA guidances and a regulation:
   Elderly (1989 FDA guideline and ICH E-7 in 1994);
   Gender (1993 FDA guideline);
   1988 FDA Guideline on the Content and Format of the Clinical and Statistical Sections of New Drug Applications; the guideline called for analysis of demographic subgroups for safety, effectiveness, dose–response;
   1998 regulation modifying 1985 regulation requires analysis of effectiveness and safety results by demographic subgroups (specifically, age, sex and race).
3. Pharmacokinetics in recent years are routinely assessed by age, gender, renal functional status, and often liver functional status, as recommended in the elderly and gender guidelines.
4. Assessment of drug–drug interactions is a major growth area, supported by recent FDA *in vitro* and *in vivo* guidances.

STUDY MONITORING/DATA COLLECTION

The industry monitoring model (visit all sites every month and examine most or all data collected) is attractive, albeit costly, and was very nearly ensconced as a universally expected standard in ICH E-6 (Good Clinical Practice – GCP – guidance). Fortunately, the ICH working group, pressed by U.S. regulators, ultimately recognized that few, if any, of the outcome studies we so value have used such methods, and, indeed, that many NIH-sponsored trials in the U.S., MRC-sponsored trials in the U.K., ISIS, GISSI, etc. had had no regular on-site monitoring at all. There are other monitoring models that are reasonable. NCI Cooperative groups, for example, carefully monitor standing site quality periodically for performance, covering many projects, not usually with a focus on individual studies, a very different, but apparently effective, model. Despite the great differences between these approaches and the industry model, these alternative methods have been accepted, perhaps because almost all of the alternative-model monitored trials were government sponsored or independent, not conducted by commercial sponsors. Data from these sources have certainly been relied on by regulators and practitioners. Application of the industry model to these designs would probably make the large trials impossible to conduct or at least would reduce the number of such trials.

Instead of endorsing a single approach, ICH E-6 says that every study needs an adequate monitoring/auditing plan, appropriate to the study; onsite monitoring can range from the industry model (frequent, all sites) to no onsite monitoring at all when 'central monitoring in conjunction with procedures such as investigators' training and meetings and extensive written guidance can assure appropriate conduct of the trial in accordance with GCP. Statistically controlled sampling may be an acceptable method for selecting the data to be verified'.

How much data must be collected in a study is also flexible, depending on the stage of development of the drug (much more earlier and pre-marketing) and the nature of the trial. In large cardiovascular outcome trials, usually involving marketed drugs whose side-effects are already familiar, we have at various times agreed to: recording only cardiovascular-active concomitant therapy, instead of all concomitant therapy; recording only adverse drug reactions (ADRs) that led to change in treatment (discontinuation, lower dose); collecting less baseline information; obtaining infrequent laboratory data; and recording and submitting only abnormal laboratory values.

If the drug is a new (not yet marketed) entity, it may be possible to follow a subset of a large study population more intensely to gain needed safety information while keeping the rest of the (large) trial simple.

As a general matter, late phase 3 and outcome studies should move closer to 'real life', with less frequent monitoring (unless known to be needed), and fewer exclusions because of age, concomitant illness or concomitant therapy.

RELIANCE ON NON-U.S. DATA

ICH E-5 ('Guidance on ethnic factors in the acceptability of foreign clinical data') indicates that if a regulatory agency is supplied with all the information necessary for approval, but the data are wholly from outside the region it should be able to ask for two additional studies, if they are considered necessary. First, a study in the *new* region to 'bridge' the foreign findings to the new region may be required. This could be a new clinical trial (for example, a randomized fixed dose, dose–response study), but for a familiar drug class, a pharmacodynamic dose–response (PK/PD) study might be sufficient. Second, broader safety experience may be required.

Whether a bridging clinical trial is needed, and what kind of trial it should be, depends, among other things, on how well the dose is defined, the amount of experience the new region has had with the region that is the source of the data, the similarity of critical aspects of medical practice in the two regions, and familiarity with the drug class in the new region.

The FDA *can* rely on foreign data [21 CFR 314.106] if the studies are relevant, if inspection can be carried out or is considered unnecessary, if the investigators are of recognized competence, and [21 CFR 312.120] if studies are conducted in accordance with accepted ethical principles.

It is now routine for some critical studies, or parts of those studies, in an application to be conducted in Western Europe, and there are instances in which all of the studies are conducted outside of the U.S. We have less experience with studies from the Far East, including Japan, and South America, Eastern Europe, or Africa, than from Western Europe but clearly studies are increasingly being carried out in those locations.

There are occasional examples of what seem to be 'trans-Atlantic differences', but it is hard to tell if they are real or random differences. It is very common to have multi-site studies using both European and U.S. sites. An interesting issue is what to do if results seem different in the country-defined subsets. Concerns about reliance on apparent subgroup differences are certainly appropriate, yet it can be tempting to pay attention to what seem to be major differences.

It is ironic and interesting that despite the often-voiced concern about possible ethnic and regional differences in response, international outcome studies (4S [3], WOSCOPS [4], ISIS I [5] and II [6], GISSI [7], CONSENSUS [8], timolol post-infarction [9] etc.) are routinely accepted globally, perhaps because repeating them in a new region seems ethically impossible. It is therefore the relatively trivial results we worry about the most.

FDA/INDUSTRY MEETINGS AND AGREEMENTS

FDA/Industry meetings have been strongly encouraged since the 1970s. There has been a steady increase in meetings to agree on the critical study design issues and the FDA's data expectations (end of phase 2 meetings) as well as meetings to agree on format and presentation of NDAs (pre-NDA meetings). It is not unusual to have several 'peri-end of phase 2' meetings. We have recently agreed to aggressive goals for timeliness of minutes' preparation and for scheduling these meetings (most kinds within 60 days).

The FDAMA (section 118) requires that we meet with any sponsor that makes a reasonable written request for a meeting to agree 'on the design and size of clinical trials intended to form the primary basis of an effectiveness claim... . Any agreement regarding... the design and size of clinical trials... shall be reduced to writing [and] shall not be changed [unless the sponsor agrees or the Division Director decides that there is a] substantial scientific issue essential to determining the safety or effectiveness of the drug...'

Except in the most routine situations, we recommend and urge such meetings. They improve study design and avoid needless delay and disappointment. We have been told repeatedly by industry that these meetings are of great value.

COMPLETION OF PHASE 4 AGREEMENTS

There are a great many agreements between the FDA and applicants on the conduct of post-marketing studies. Only those related to accelerated approval are actual conditions of approval, but we also expect the others to be completed. The history of the conduct of those studies is 'chequered', although not as bad as has sometimes been suggested. The FDAMA requires an annual public report on the status of agreements to conduct post-marketing studies. We expect and hope that this will increase the likelihood that the studies will be carried out by focusing attention of the FDA, the public, and Congress on failure to do so. A regulation to describe procedures for submitting status reports is under development.

FOCUS ON SAFETY AND RISK MANAGEMENT

Evaluation of safety has always been at the heart of the FDA's review and of the drug development process. Nonetheless, such events as fenfluramine valulopathy, terfenidine, astemizole and cisapride torsade de pointes arrhythmias, bromfenac, troglitizone and trovafloxacin hepatotoxicity and mibefridil serious and fatal consequences of inhibition of CYP3A4 metabolism of other drugs have reminded the medical community and the public that drugs can have very serious adverse effects. Much attention has focused on post-marketing detection and

'risk management', but there are pre-marketing implications as well. Moreover, reduction in action/approval times resulting from PDUFA and better management inevitably leads to the question of whether sufficient care is being taken. The FDA has responded in a number of ways.

We are developing Good Review Practices (GRP) guidance, a visible set of agreed upon review practices leading to comprehensive reviews and availability of a standard against which performance can be evaluated. A draft of a safety GRP internal guidance is available.

We have taken a new look at how much risk can be accepted, particularly if the drug has no advantage over alternatives. Dilevalol, a beta blocker, was rejected some years ago because of suggestions of potentially serious, albeit rare, hepatotoxicity, that made it less desirable than alternatives. Recently, an anticholinesterase intended for Alzheimer's disease patients was rejected because it was severely emetogenic. Terfenidine, astemizole and cisapride are all no longer marketed because rare, avoidable, potentially fatal induced torsade de pointes arrhythmias continued to occur despite aggressive labelling and educational efforts.

We have considered ways, other than labelling, to assure proper use of drugs. Clozapine, for example, can be given only if patients obtain white blood cell counts every 1–2 weeks (depending on duration of use). Dofetilide, for delaying recurrence of atrial fibrillation, must be started in a monitored facility so that dose can be chosen and effects on QT interval can be assessed.

We have also reconsidered the sufficiency of labelling warning of severe risks when the risk is not accompanied by a benefit (for example, dilevalol, the emetogenic Alzheimer's drug) and of labelling's ability to control risk (for example, it did not do well with bromfenac, cisapride, terfenidine, mibefridil), and have introduced alternatives (for example, clozapine, dofetilide) that go beyond labelling.

We have given further consideration to assessing risk prior to approval in special cases, including enhancing the FDA and industry awareness of signals of hepatic injury and proper evaluation of QT interval prolongation.

Finally, there has been a revolution in assessment of drug–drug interactions (terfenidine) and activity of metabolites.

RISING PUBLIC ETHICAL CONCERNS

In a series of U.S. Department of Health and Human Services (HHS) Inspector General reports and newspaper series, growing concerns have been raised about Institutional Review Boards (IRB) function, recruiting practices, investigator conflicts of interest and trial ethics and adequacy of informed consent. Apart from the findings themselves, these reports should concern all who value clinical trials as the fundamental basis of medical advances, because the general concern casts a shadow on all trials. Many of these issues will become the subject of workshops and guidance. We need to listen carefully, improve what is really broken, and be prepared to explain the value of evidence-based medicine and clinical trials.

There are a few areas of special concern:

1. Conflict of interest of investigators is of concern when the investigators have a financial interest in the product under study (a problem with gene therapy and devices). A trend

to watch closely is 'shared risk' arrangements of contract research organizations (CROs) and others, making the CRO a 'non-neutral' investigator.

2. The effectiveness and adequacy of informed consent has been challenged.
3. The overworked state of IRBs has been noted, and their ability to provide adequate review of protocols and also conduct 'continuous monitoring' has been questioned. Assessment of IRBs' real role and responsibilities is in order.
4. The increasing number of financial relationships between investigators and sponsors has also been of concern. The FDA's financial disclosure experience suggests that various substantial payments (lecture fees etc.) and 'perks' (trips, concerts, vacations etc.) are increasing in number and value. Does this threaten the very idea of the 'independent investigator'?

CONCLUSIONS

The basic regulatory structure is substantially intact but scientific, public, legislative and managerial concerns and initiatives have led to substantial changes, largely for the better, in regulation clinical trials and drug development.

REFERENCES

1. Temple R, Ellenberg S. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: Ethical and scientific issues. *Annals of Internal Medicine* 2000; **133**:455–463.
2. Ellenberg S, Temple R. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 2: Practical issues and specific cases. *Annals of Internal Medicine* 2000; **133**:464–470.
3. Scandinavian Simvastatin Survival Study Group. Randomized trail of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet* 1994; **344**:1383–1389.
4. Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, Macfarlane PW, McKillop JH, Packard CJ. Prevention of coronary heart disease with provastatin in men with hypercholesterolemia. *New England Journal of Medicine* 1995; **333**:1301–1307.
5. ISIS-1 Collaborative Group. Randomized trial of intravenous atenolol among 16,027 cases of suspected acute myocardial infarction: ISIS-1. *Lancet* 1986; **2**:57–66.
6. ISIS-2 Collaborative Study Group. Randomized trial of intravenous streptokinase, oral aspirin, or both or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988; **2**:349–360.
7. Gruppo Italiano Perlo Studio Della Streptochinasi Nell'infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986; **1**:397–402.
8. The CONSENSUS Trial Study Group. Effects of enalapril on mortality in severe congestive heart failure: results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). *New England Journal of Medicine* 1987; **316**:1429–1435.
9. The Norwegian Multicenter Study Group. Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction. *New England Journal of Medicine* 1981; **304**:801–807.
10. Temple R. Problems in interpreting active control equivalence trials. *Accountability in Research* 1996; **4**: 267–275.