Treating Individuals 1

External validity of randomised controlled trials: "To whom do the results of this trial apply?"

Lancet 2005; 365: 82–93

See Comment page 13

Stroke Prevention Research Unit, University Department of Clinical Neurology, Radcliffe Infirmary, Oxford OX2 6HE, UK
(P M Rothwell FRCP)
peter.rothwell@cneuro.ox.ac.uk

Peter M Rothwell

In making treatment decisions, doctors and patients must take into account relevant randomised controlled trials (RCTs) and systematic reviews. Relevance depends on external validity (or generalisability)—ie, whether the results can be reasonably applied to a definable group of patients in a particular clinical setting in routine practice. There is concern among clinicians that external validity is often poor, particularly for some pharmaceutical industry trials, a perception that has led to underuse of treatments that are effective. Yet researchers, funding agencies, ethics committees, the pharmaceutical industry, medical journals, and governmental regulators alike all neglect external validity, leaving clinicians to make judgments. However, reporting of the determinants of external validity in trial publications and systematic reviews is usually inadequate. This review discusses those determinants, presents a checklist for clinicians, and makes recommendations for greater consideration of external validity in the design and reporting of RCTs.

> "Between measurements based on RCTs and benefit . . . in the community there is a gulf which has been much under-estimated"
>
> A L Cochrane, 1971[1]

> "At its best a trial shows what can be accomplished with a medicine under careful observation and certain restricted conditions. The same results will not invariably or necessarily be observed when the medicine passes into general use."
>
> Austin Bradford Hill, 1984[2]

Randomised controlled trials (RCTs) and systematic reviews are the most reliable methods of determining the effects of treatment. They must be internally valid (ie, design and conduct must keep to a minimum the possibility of bias),[3,4] but to be clinically useful the result must also be relevant to a definable group of patients in a particular clinical setting; this is generally termed external validity, applicability, or generalisability. The beneficial effects of some interventions, such as blood pressure lowering in chronic uncontrolled hypertension, are generalisable to most patients and settings, but the effects of other interventions can be very dependent on factors such as the characteristics of the patient, the method of application of the intervention, and the setting of treatment. How these factors are taken into account in the design and performance of an RCT and in the reporting of the results can have a major effect on external validity.

Lack of consideration of external validity is the most frequent criticism by clinicians of RCTs, systematic reviews, and guidelines,[5–13] and is one explanation for the widespread underuse in routine practice of treatments that were beneficial in trials and that are recommended in guidelines.[14–26] Neither Cochrane nor Bradford Hill were practising clinicians, but they understood the limitations of the methodology that they had pioneered. Although what little systematic evidence we now have confirms that RCTs do often lack external validity,[27–41] this issue is neglected by current researchers, medical journals, funding agencies, ethics committees, the pharmaceutical industry, and governmental regulators alike (panel 1).[42–50] Admittedly, assessment of external validity is complex

Panel 1: **Evidence of the neglect of consideration of external validity of RCTs and systematic reviews**

- Research into internal validity of RCTs and systematic reviews far outweighs research into how results should best be used in practice.[42,43]
- Rules governing the performance of trials, such as good clinical practice,[44] do not cover issues of external validity.
- Drug licensing bodies, such as the US Food and Drug Administration, do not require evidence that a drug has a clinically useful treatment effect, or a trial population that is representative of routine clinical practice.[45]
- Guidance on the design and performance of RCTs from funding agencies, such as that from the UK Medical Research Council,[46,47] makes virtually no mention of issues related to external validity.
- Guidance from ethics committees, such as that from the UK Department of Health,[48] indicates that clinical research should be internally valid, and raises some issues that relate to external validity, but makes no explicit recommendations about the need for results to be generalisable.
- Guidelines on the reporting of RCTs and systematic reviews focus mainly on internal validity and give very little space to external validity.[49,50]
- None of the many scores for judging the quality of RCTs address external validity adequately.[31]
- There are no accepted guidelines on how external validity of RCTs should be assessed.

***Panel 2:* Issues that potentially affect external validity**

Setting of the trial
- Healthcare system
- Country
- Recruitment from primary, secondary, or tertiary care
- Selection of participating centres
- Selection of participating clinicians

Selection of patients
- Methods of prerandomisation diagnosis and investigation
- Eligibility criteria
- Exclusion criteria
- Placebo run-in period
- Treatment run-in period
- Enrichment strategies
- Ratio of randomised patients to eligible non-randomised patients in participating centres
- Proportion of patients who declined randomisation

Characteristics of randomised patients
- Baseline clinical characteristics
- Racial group
- Uniformity of underlying pathology
- Stage in the natural history of their disease
- Severity of disease
- Comorbidity
- Absolute risks of a poor outcome in the control group

Differences between the trial protocol and routine practice
- Trial intervention
- Timing of treatment
- Appropriateness/relevance of control intervention
- Adequacy of non-trial treatment—both intended and actual
- Prohibition of certain non-trial treatments
- Therapeutic or diagnostic advances since trial was done

Outcome measures and follow-up
- Clinical relevance of surrogate outcomes
- Clinical relevance, validity, and reproducibility of complex scales
- Effect of intervention on most relevant components of composite outcomes
- Who measured outcome
- Use of patient-centred outcomes
- Frequency of follow-up
- Adequacy of the length of follow-up

Adverse effects of treatment
- Completeness of reporting of relevant adverse effects
- Rates of discontinuation of treatment
- Selection of trial centres and/or clinicians on the basis of skill or experience
- Exclusion of patients at risk of complications
- Exclusion of patients who experienced adverse effects during a run-in period
- Intensity of trial safety procedures

and needs clinical rather than statistical expertise, but it is vital if treatments are to be used correctly in as many patients as possible in routine clinical practice. We cannot expect the results of RCTs and systematic reviews to be relevant to all patients and all settings (that is not what is meant by external validity) but they should at least be designed and reported in a way that allows clinicians to judge to whom they can reasonably be applied. This article considers how external validity should be assessed (panel 2). Illustrative examples are drawn mainly from treatments for cerebrovascular or cardiovascular disease but the general principles are relevant to all areas of medicine and surgery.

Limits on external validity

RCTs and systematic reviews are the most reliable methods of determining moderate treatment effects, but external validity is inevitably less than perfect, at least in theory, because the aim is not to measure the benefit that will be derived from treatment in clinical practice. The response to and/or compliance with a treatment can be influenced strongly by the doctor-patient relationship,[51–53] placebo effects,[54,55] and patient preference.[56–58] Yet trialists rightly try where possible to eliminate any effect of these factors by using blinded treatment allocation, placebo control, and exclusion of patients or clinicians who have strong treatment preferences. These procedures increase the internal validity of an RCT, but will often lead to underestimation of the benefits of treatment in clinical practice, especially for patient-centred outcomes.

Patient preference can cause particular problems for external validity. For example, some women with early breast cancer have a strong preference for lumpectomy, whereas others are far happier in the belief that all the cancer has been removed by a mastectomy. However, only women who did not have a strong preference for a particular treatment could be recruited into the relevant RCTs, and as few as 10% agreed to have their treatment chosen at random.[59] If RCTs show a major advantage for one treatment, then external validity is not a problem. Difficulties arise when one treatment is only moderately more effective but the patient has a strong personal preference for the less effective option. Would the results of the breast surgery RCTs, particularly in relation to psychological wellbeing, have been the same if such patients had been randomised?

These inevitable limitations do not invalidate the results of RCTs and systematic reviews, and they are mentioned here partly for the sake of completeness, but the importance of patient preference, placebo effects, and the doctor-patient relationship outside trials should not be underestimated. This fact is perhaps best illustrated by the popularity of alternative therapies, such as homoeopathy, in which such factors are the only active ingredients. The remainder of this review will concentrate on those factors in the design and reporting

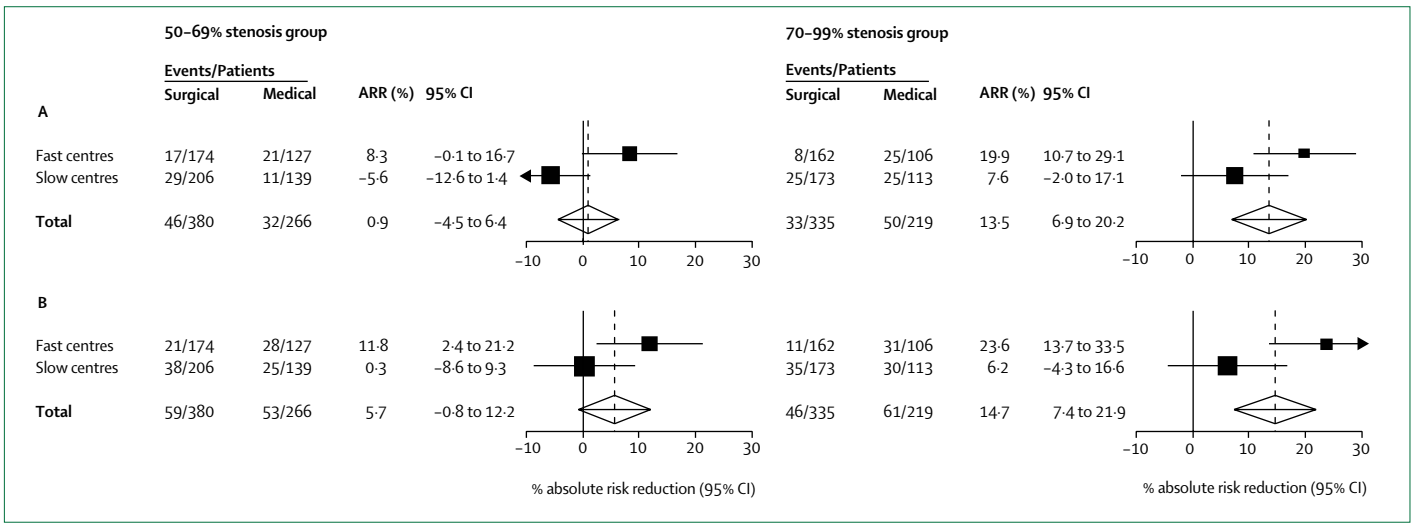

Figure 1: **Absolute reductions in 5-year risks of ipsilateral ischaemic stroke (upper) and any stroke or death (lower) with surgery in ECST[60] centres**
Median delay from last symptomatic event to randomisation in these centres was ⩽50 days (fast centres) compared with centres with a longer delay (slow centres). Data are shown separately for patients with moderate (50–69% [left of figure]) and severe (70–99% [(right of figure]) carotid stenosis. A: ipsilateral ischaemic stroke or operative stroke/death. B: any stroke or operative death.

of RCTs and systematic reviews that limit external validity but which are not unavoidable.

Trial setting

There is often concern about the generalisability of trials done in secondary or tertiary care to practice in primary care,[26–29] but there are several other ways in which the setting of an RCT can affect external validity.

The health-care system

Differences between health-care systems can affect external validity. For example, in the European Carotid Surgery Trial (ECST),[60] an RCT of endarterectomy for recently symptomatic carotid stenosis, there were national differences in the speed with which patients were investigated, with a median delay from last symptoms to randomisation of greater than 2 months in the UK, for example, compared with 3 weeks in Belgium and Holland. Figure 1 shows that separate trials in these systems would have produced very different results, because of the narrow time window for prevention of stroke.[61] These differences were not mentioned in any of the ECST publications or in any subsequent guidelines. Similar differences in performance between health-care systems will exist for other disorders, and there is, of course, the broader issue of how trials done in the developed world apply to the developing world.

The country

Even if the health-care systems are similar, other national differences can still affect generalisability. Continuing with the example of cerebrovascular disease, there are many differences between countries in methods of diagnosis and management,[62] as well as important racial differences in susceptibility to disease and natural history of the disease,[63] all of which could affect the external validity of trial results. Examples in other areas of medicine include the substantial heterogeneity between trials of BCG in prevention of tuberculosis done in different populations, with great efficacy in countries at northern latitudes but a gradual loss of efficacy with decreasing latitude ($p<0·00001$).[64] There are also often striking national differences in the use of ancillary so-called non-trial treatments. In one international RCT of aspirin and heparin in almost 20 000 patients with acute ischaemic stroke, glycerol was used in 50% of the 1473 patients in Italy versus 3% elsewhere, steroids in 32% of the 225 patients in Turkey versus 4% elsewhere, and haemodilution in 44% of the 597 patients in Austria and the Czech Republic versus 3% elsewhere.[65] More extreme differences between countries were recorded in the use of two important non-trial surgical techniques in the ECST (figure 2). There is evidence that the use of both techniques does affect operative risk, but irrespective of this fact, the data illustrate the extent to which clinical practice varies between countries. RCTs done in one country will usually be generalisable to others, but this generalisability should not be taken for granted.

Selection of participating centres and clinicians

Selection of participating centres from secondary care as opposed to primary care has obvious implications for external validity, but RCTs of interventions that are confined to secondary care may also be undermined if they are restricted to specialist units.[66–68] In one systematic review of laparoscopic cholecystectomy, for example, all 15 RCTs were based solely in university hospitals.[67] Problems also arise if participating

clinicians are chosen because of their track record. For example, results from the Asymptomatic Carotid Artery Surgery (ACAS) trial showed that endarterectomy for asymptomatic carotid stenosis reduced the 5-year absolute risk of stroke by about 5%.[69] However, ACAS only accepted surgeons with a good safety record, initially rejecting 40% of applicants,[70] and subsequently barring those who had adverse operative outcomes in the trial from further participation. The benefit from surgery in ACAS was due largely to the consequently low operative risk.[69] Figure 3 compares the ACAS risks with the results of a meta-analysis of the 46 surgical case series that published operative risks during the 5 years after ACAS.[71] Operative mortality was eight-fold higher than in ACAS (1·11% *vs* 0·14%, p=0·01) and the risk of stroke and death was about three-fold higher in comparable studies in which outcome was also assessed by a neurologist (4·30% versus 1·50%, p<0·001). Trials should not include centres that do not have the competence to treat patients safely, but selection should not be so exclusive that the results cannot be generalised to clinical practice. For example, surgeons rejected by ACAS are unlikely to have stopped operating outside the trial.

Selection of patients

Only a small proportion of patients with a specific disorder participate in a particular trial. For example, only about one in every 200–300 patients undergoing carotid endarterectomy in North America at the time got into the large, multicentre RCTs,[72] and similar proportions have been reported in breast cancer trials.[73] These low rates mean that trials take many years to recruit but are not a problem for external validity as long as the patients randomised in the participating centres are representative of the whole. As outlined below this is not always the case.

Selection before consideration of eligibility

Concern is often expressed about highly selective trial eligibility criteria, but there are several earlier stages of selection that can be more problematic. Figure 4 shows that the proportion of patients with a particular disorder in the local community served by a participating centre who are considered for recruitment into a trial will often be well below 1%. For example, consider a trial of a new blood pressure-lowering drug, which like most similar trials is done in a hospital clinic. Fewer than 10% of patients with hypertension are managed in hospital clinics and this group will differ from those managed in primary care. Moreover, only one of the ten doctors who see hypertensive patients in this hospital is taking part in the trial, and this particular doctor mainly sees cross-referrals of young patients with resistant hypertension. Thus even before any consideration of eligibility or exclusion criteria, potential recruits are already very unrepresentative of patients in the local community.

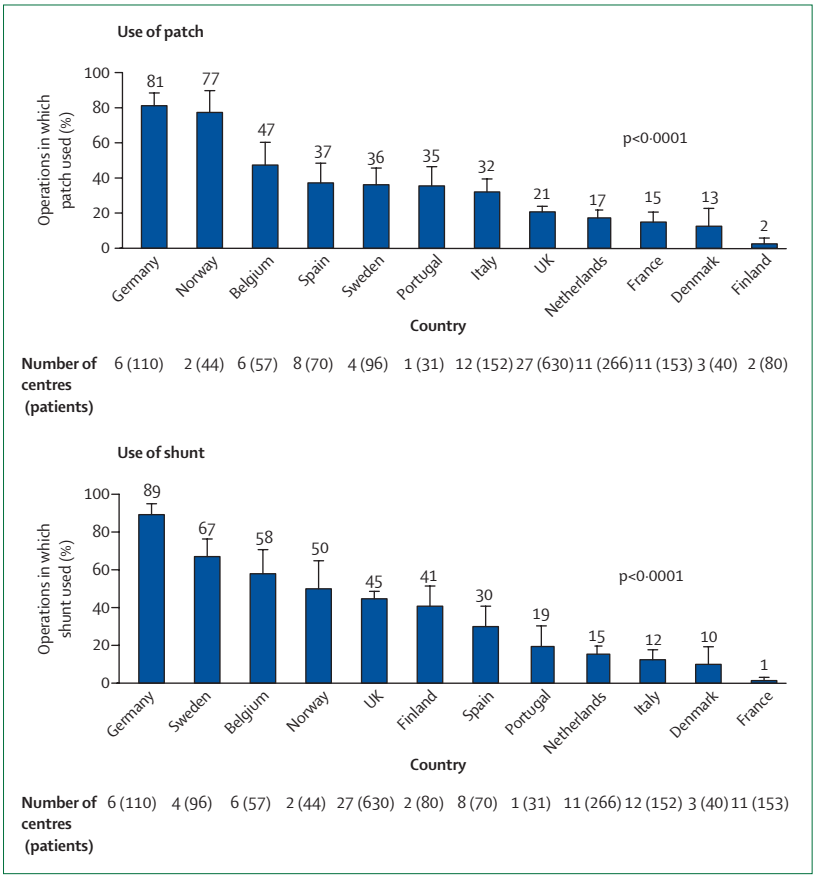

Figure 2: **Variation in rate of use of two ancillary surgical techniques used during carotid endarterectomy (patch angioplasty [upper] and use of an intraoperative shunt [lower]) by country in the ECST[60]**

Selection by eligibility criteria

Patients are then further selected according to trial eligibility criteria (figure 4). Some RCTs exclude women and many exclude the elderly.[74,75] One review of

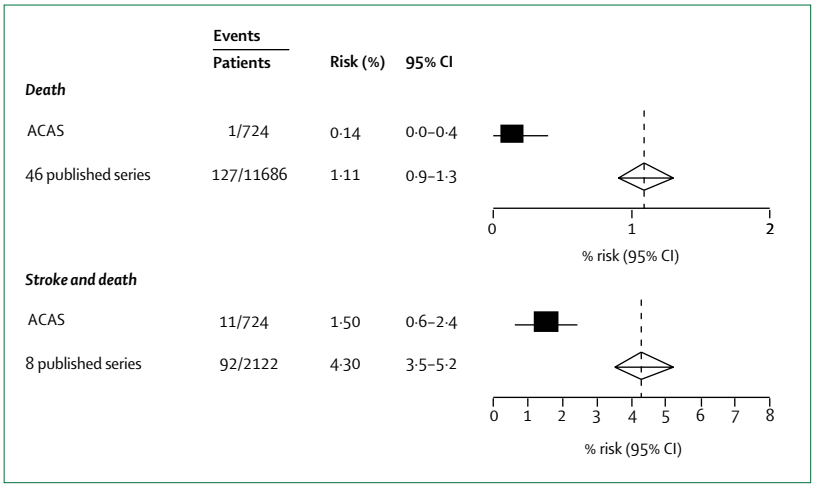

Figure 3: **Overall results of a meta-analysis of the operative risk of death (upper) and stroke and death (lower) from all studies published between 1990 and 2000 inclusive that reported risks due to carotid endarterectomy for asymptomatic stenosis[71] compared with the same risks in ACAS[69]**

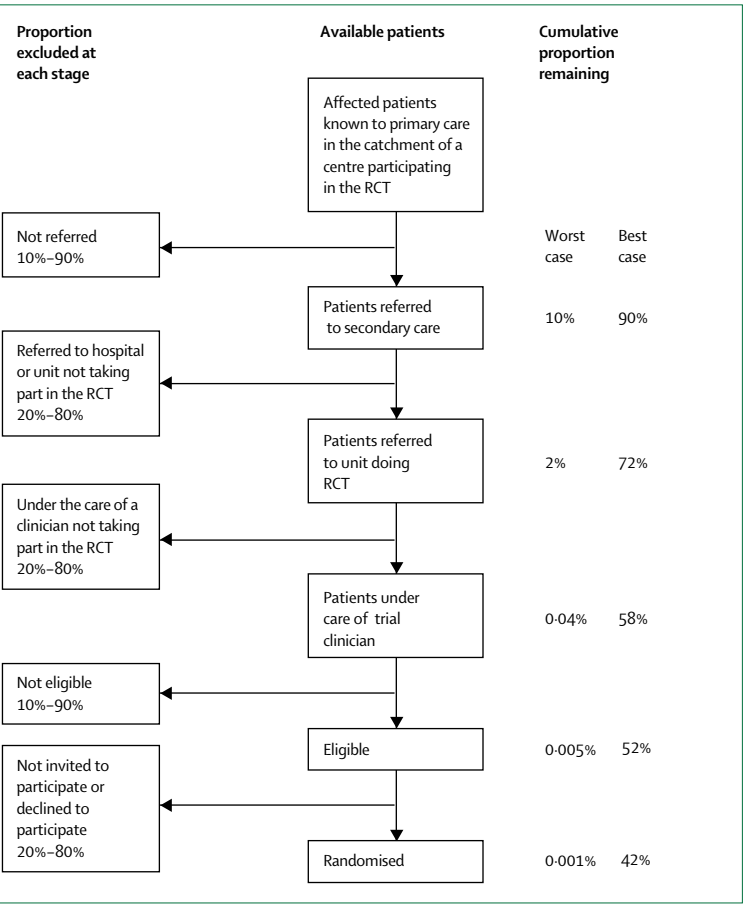

Figure 4: **Schematic diagram illustrating effect of multiple stages of selection inherent in clinical practice on proportion of patients in catchment of participating centre entered into an RCT done in secondary care**
Worst case assumes proportion of patients excluded at each stage is at top of range and best case is based on lowest proportion of patients excluded.

214 drug trials in acute myocardial infarction found that over 60% excluded patients aged over 75 years.[74] Many RCTs also exclude patients with highly prevalent comorbidity. For example, trials of antiplatelet or non-steroidal anti-inflammatory drugs (NSAIDS) often bar recruitment of patients with any history of dyspepsia, even though this criterion excludes over 50% of elderly patients in clinical practice.[76,77] Exclusion rates can be very high. In acute stroke, one study showed that of the small proportion of patients admitted sufficiently quickly to be suitable for thrombolysis,[78] 96% were ineligible based on the various other criteria of the relevant RCT.[79] One centre in another acute stroke trial had to screen 192 patients over 2 years to find an eligible patient.[80] A review of 41 US National Institutes of Health RCTs found an average exclusion rate of 73%.[39]

Selection beyond the eligibility criteria

Strict eligibility criteria can limit the external validity of RCTs and result in lower rates of treatment in clinical practice,[81] but the criteria are at least available for scrutiny, or should be (discussed below). More difficult is the extent to which trials with seemingly reasonable eligibility criteria end up with highly selected populations. Recruitment of less than 10% of patients with the relevant disorder in participating centres is common in pragmatic RCTs in all areas of medicine and surgery in which the data have been gathered.[24,33,82–87] These low rates of recruitment are partly due to additional selection by participating clinicians beyond that required by the eligibility criteria. Patients recruited into RCTs differ from those who are eligible but not recruited in terms of age, sex, race, severity of disease, educational status, social class, and place of residence.[30,34,35,37,59,74,75,88–90] The outlook for patients included in RCTs is also usually better than those who do not participate in trials,[91] often greatly so.[6,92] Yet highly selective recruitment is not inevitable. The GISSI-1 trial of thrombolysis for acute myocardial infarction, for example, recruited 90% of patients admitted within 12 h of the event with a definite diagnosis and no contraindications.[93] As a consequence it has excellent external validity and is one of only a very few RCTs in acute myocardial infarction that had a control group mortality rate (13%) that was remotely consistent with routine clinical practice at the time.

Run-in periods

Pre-randomisation run-in periods are also often used to select or exclude patients.[94] In a placebo run-in, all eligible patients receive placebo and those who are poorly compliant are excluded.[95] There can be good reasons for doing this, but high rates of exclusion will reduce external validity. For example, one trial of the effect of salts on blood pressure excluded 93% of patients in a placebo run-in period.[96] Active treatment run-in periods in which patients are excluded if they have adverse events or show signs that treatment may be ineffective are excluded are more likely to undermine external validity. For example, two RCTs of carvedilol, a vasodilatory β blocker, in chronic heart failure excluded 6% and 9% of eligible patients in treatment run–in periods[97,98] because of worsening heart failure and other adverse events, some of which were fatal. In both trials, the complication rates in the subsequent randomised phase were much lower than in the run-in phase.[97,98]

Enrichment strategies

Patients who are likely to respond well to treatment are sometimes actively recruited.[99–101] For example, some trials of antipsychotic drugs have selectively recruited patients who had previously had a good response to such drugs.[102] Other trials have excluded non-responders in a run-in phase. One RCT of the acetylcholinesterase inhibitor tacrine in Alzheimer's disease recruited 632 patients to a 6-week enrichment phase in which patients were given different doses of tacrine or placebo.[103] After a

washout period, only the 215 (34%) patients who had a measured improvement on tacrine in the enrichment phase were randomised to tacrine (at their most effective dose) versus placebo in the main phase of the trial. External validity is clearly undermined here.

Reporting of patient selection

The number of eligible non–randomised patients can be recorded, but is difficult to determine reliably, and underestimates selection because logs usually only cover patients referred to the participating clinician. Another useful index is the number of patients who are invited to participate and decline, but neither statistic is usually reported. However, inadequate reporting of trial eligibility criteria is a far greater barrier to the assessment of external validity.[104] The CONSORT guidelines[49] and the Uniform Requirements for Manuscripts Submitted to Biomedical Journals (http://www.ICMJE.org) require that all eligibility criteria should be reported, but a review of trials leading to clinical alerts by the US National Institutes of Health revealed that of an average of 31 eligibility criteria, only 63% were published in the main trial report and only 19% in the clinical alert.[105] Inadequate reporting is a major problem in secondary publications, such as systematic reviews and clinical guidelines, in which space limitations and the need for a succinct message do not usually allow detailed consideration of the eligibility and exclusion criteria of trials or other determinants of external validity. The same also applies to pharmaceutical marketing, although the motivation for concealing poor external validity may be different here.

Characteristics of randomised patients

Trial reports usually include the baseline clinical characteristics of randomised patients and so it is argued that clinicians can assess external validity by comparison with their patient(s).[49,50] This theory is clearly sensible, but baseline clinical characteristics can be misleading. The difficulty in extrapolating from baseline clinical characteristics is illustrated by the patients who were randomised to endarterectomy in the ECST[60] but who did not have surgery because their surgeon and/or anaesthetist judged them to be too frail. Although this clinical impression was confirmed by a much worse outcome during follow-up compared with patients who were randomised to medical treatment (5-year risks: stroke=36% *vs* 18%, p<0·001; stroke or death=52% *vs* 27%, p<0·0001), their baseline clinical characteristics did not differ (table 1).

A patient may also differ from the trial population in a way that seems unimportant but which may have a major effect on external validity. For example, table 2 shows the baseline clinical characteristics of the patients randomised to warfarin in two RCTs of secondary prevention of stroke.[106–108] In one trial, patients were in atrial fibrillation and in the other they were in sinus rhythm. This difference might be expected to affect the risk of further ischaemic stroke, but would not be expected to affect the safety of warfarin. In fact, the risk of intracranial haemorrhage on warfarin was 19-fold higher (p<0·0001) in SPIRIT (Stroke Prevention in Reversible Ischaemia Trial) than in EAFT (European Atrial Fibrillation Trial) after adjustment for differences in baseline clinical characteristics and the intensity of anticoagulation (table 2).[108] Seemingly irrelevant differences between patients can have major effects on

	Surgery (n=1807)		No surgery (n=1211)
	Not operated (n=62)	Operated (n=745)	
Demography			
Male sex	36 (58%)	1263 (72%)	869 (72%)
Age in years	64·1 (8·7)	62·5 (8·1)	62·3 (8·0)
Cerebrovascular events within past 6 months			
Hemispheric TIA or stroke	56 (90%)	1495 (85%)	1038 (86%)
Ocular event only	6 (10%)	250 (15%)	173 (14%)
Residual neurological signs	2 (3%)	106 (6%)	78 (7%)
Days since last symptoms	74 (56)	62 (53)	62 (52)
Other clinical data			
Previous stroke	2(3%)	101 (6%)	78 (7%)
Systolic BP (mm Hg)	154 (27·2)	151 (22·3)	150·2 (21·3)
Diastolic BP (mm Hg)	89·0 (13·0)	86·2 (11·4)	86·3 (10·8)
Angina	15 (24%)	305 (18%)	190 (16%)
Previous myocardial infarction	7 (11%)	219 (13%)	136 (11%)
Previous coronary artery surgery	2 (3%)	47 (3%)	23 (2%)
Peripheral vascular disease	7 (11%)	292 (16%)	203 (17%)
Diabetes	8 (13%)	208 (12%)	145 (12%)
Current cigarette smoking	25 (40%)	844 (48%)	557 (46%)
Blood cholesterol (mmol/L)	6·4 (1·6)	6·4 (1·4)	6·4 (1·4)
Mean symptomatic carotid stenosis	60% (25)	62% (21)	59% (22)
Mean contralateral carotid stenosis	37% (26)	42% (26)	37% (27)

Patients who were randomised to surgery but did not have the operation are compared with those who did and with patients who were randomised to medical treatment only. TIA=transient ischaemic attack. BP=blood pressure. Data are mean (SD) or number (%).

Table 1: **Baseline clinical characteristics and outcomes of patients randomised in ECST[60]**

	SPIRIT (n=651)	EAFT (n=225)
Baseline clinical characteristics		
Male sex	66%	55%
Age >65 years	47%	81%
Hypertension	39%	48%
Angina	9%	11%
Myocardial infarction	9%	7%
Diabetes	11%	12%
Leukoaraiosis on CT brain scan	7%	14%
Outcomes during trial		
INR during trial (mean, SD)	3·3 (1·1)	2·9 (0·7)
Patient-years of follow-up	735	507
Intracranial haemorrhage	27	0 *
Extracranial haemorrhage	26	13
Adjusted hazard ratio (95% CI)*		
Intracranial haemorrhage	19·0 (2·4–250)	p<0·0001
Extracranial haemorrhage	1·9 (0·8–4·7)	p=0·15

INR=international normalised ratio. CT=computed tomography. *There were no proven intracranial haemorrhages, but no CT scan was done in two strokes. For calculation of adjusted hazard ratio for haemorrhage these two strokes were categorised as having been due to intracranial haemorrhage.

Table 2: **Baseline clinical characteristics and outcomes of patients randomised to anticoagulation with warfarin in EAFT[106] and SPIRIT[107]**

	Disorder	Surrogate outcome	Clinical outcome
Treatment			
Fluoride	Osteoporosis	Increase in bone density[112]	Major increase in fractures[112]
Antiarrhythmic drugs	Post–myocardial infarction	Reduction in ECG abnormalities[113]	Increased mortality[114]
Interferon β	Multiple sclerosis	70% reduction in new brain lesions using MRI [115–118]	No convincing effect on disability[115–118]
Milrinone and Epoprostanol	Heart failure	Improved exercise tolerance [119,120]	Increased mortality [121,122]
Ibopamine	Heart failure	Improved ejection fraction, and heart rate variability[123]	Increased mortality[124]

ECG=electrocardiogram.

Table 3: **Examples of trials with misleading surrogate outcomes**

risks and benefits of treatment. This fact was further highlighted in SPIRIT by patients with leucoaraiosis on baseline brain imaging who had a nine-fold greater risk of intracranial haemorrhage on warfarin than patients who did not.[106,108]

There are many other factors related to patient characteristics that can determine the relevance of a trial to a particular patient, including the underlying pathology, the severity of disease, the stage in the natural history of the disease, comorbidity, and the probable absolute risk of a poor outcome without treatment. These issues are considered in detail in subsequent articles in this series.

The intervention, control treatment, and pre-trial or non-trial management

External validity can also be affected if trials have protocols that differ from usual clinical practice. For example, before randomisation in the RCTs of endarterectomy for symptomatic carotid stenosis patients had to be diagnosed by a neurologist and have conventional arterial angiography,[109] neither of which are routine in many centres. The trial intervention itself may also differ from that used in current practice, such as in the formulation and bioavailability of a drug, or the type of anaesthetic used for an operation. The same can be true of the treatment in the control group in a trial, which may use an especially low dose of the comparator drug, or fall short of best current practice in another way. External validity can also be undermined by too stringent limitations on the use of non-trial treatments. For example, antihypertensive drugs or drugs for the treatment of cardiac failure will be less effective in elderly patients outside trials who are unable to stop taking their NSAIDs.[110] Any prohibition of non-trial treatments should be reported in the main trial publications along with details of relevant non-trial treatments that were used. The timing of many interventions is also critical, as is illustrated in figure 1, for endarterectomy for recently symptomatic carotid stenosis, and should be reported when relevant.

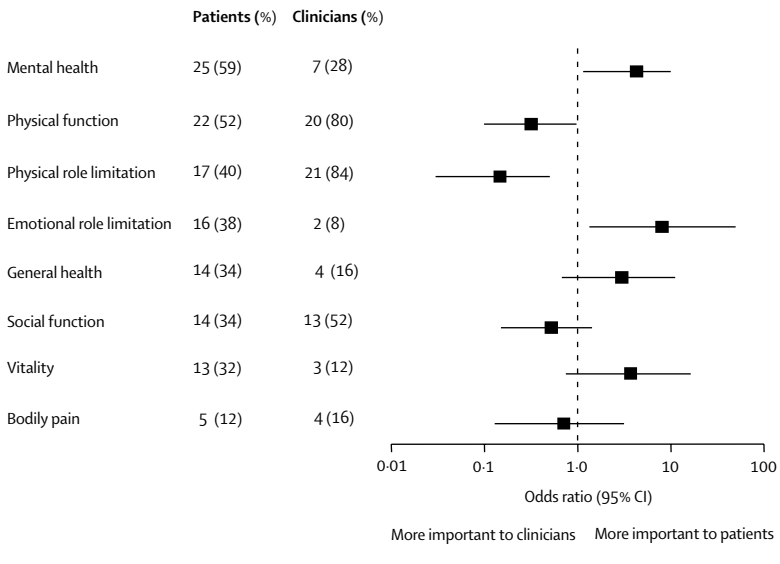

Figure 5: **Comparison of the elements of quality of life of greatest concern to patients with multiple sclerosis with those elements that their clinicians believed most important**
Based on selection of three of eight aspects of health-related quality of life assessed in short form-36 measures.[130]

Outcome measures and follow-up

The external validity of an RCT also depends on whether the outcomes were clinically relevant. This can depend on subtle considerations such as who actually measured the outcome, as is illustrated by the lower operative risks of endarterectomy in studies in which patients were assessed by surgeons rather than by neurologists,[111] but is more often dependent on what was measured and when.

Surrogate outcomes

Many trials use surrogate outcomes, usually biological or imaging markers that are believed to be indirect measures of the effect of treatment on clinical outcomes. However, as well as being of questionable clinical relevance, surrogate outcomes are often misleading. Each of the treatments in table 3 had a major beneficial effect on a surrogate outcome, but even though each surrogate outcome was correlated with a relevant clinical outcome in observational studies, the treatments proved ineffective or harmful in subsequent large RCTs that used these clinical outcomes.[112–124]

Scales

RCTs sometimes use complex scales, often made up of arbitrary combinations of symptoms and clinical signs.[125,126] For example, a review of 196 RCTs of NSAIDs in rheumatoid arthritis identified more than 70 different outcome scales,[127] and a review of 2000 RCTs in schizophrenia identified 640 scales, many of which were devised for the particular RCT and had no supporting data for validity or reliability.[128] These unvalidated scales were more likely to show significant treatment effects than established scales.[129] Moreover, the clinical meaning of apparent treatment effects (eg, a 2·7 point mean reduction in a 100 point outcome scale made up of various symptoms and signs) is usually impossible to discern.

Patient-centred outcomes

Simple clinical outcomes usually have most external validity but only if they indicate the priorities of patients. For example, figure 5 shows the results of a study in which patients with multiple sclerosis and their clinicians were asked independently to select the three aspects of the disease that had the greatest effect on quality of life.[130] Clinicians focused mainly on the physical effects of the disease, whereas patients were more concerned about mental health, emotional wellbeing, general health, and vitality, which are often not measured in RCTs. It is also important that clinical outcomes should be expressed in a way that is most relevant to patients, even if such estimation reduces the statistical power of the trial. For example, patients with epilepsy are much more interested in the proportion of patients rendered free of seizures in RCTs of anticonvulsants than they are in changes in mean seizure frequency.[131,132]

Composite outcome measures

Many trials combine events in their primary outcome measure. Such combination can produce a useful measure of the overall effect of treatment on all the relevant outcomes, and it usually affords greater statistical power, but there are difficulties. For example, the outcome that is most important to a particular patient can be affected differently by treatment than would be the combined outcome. Although the antiplatelet agent dipyridamole reduces the risk of the combined outcome of stroke, myocardial infarction, or vascular death, it seems to have no effect on the risk of myocardial infarction alone[133] and would not be the optimum agent in a patient with unstable coronary artery disease. Composite outcomes also sometimes combine events of very different severity, and treatment effects can be driven by the least important outcome, which is often the most frequent. This is sometimes the case, for example, in trials of prevention of stroke that include transient ischaemic attacks in a composite outcome. An equally problematic composite outcome is the mixture of definite clinical events and episodes of hospitalisation. The fact that a patient is in an RCT will probably affect the likelihood of hospitalisation, and this likelihood will certainly vary between different health-care systems.

Length of treatment and follow-up

Another common difficulty for the external validity of RCTs is an inadequate duration of treatment and/or follow-up. For example, although patients with refractory epilepsy require treatment for many years, most RCTs of new drugs follow up the effects of treatment for only a few weeks.[131,132] Whether initial response is a good predictor of long-term benefit is unknown. The same problem has been identified in RCTs in schizophrenia, with fewer than 50% of trials having greater than 6 weeks follow-up and only 20% following up patients for longer than 6 months.[33,128] The contrast between beneficial effects of treatments in short-term RCTs and the less encouraging experience of long-term treatment in clinical practice has also been emphasised by clinicians treating patients with rheumatoid arthritis.[134]

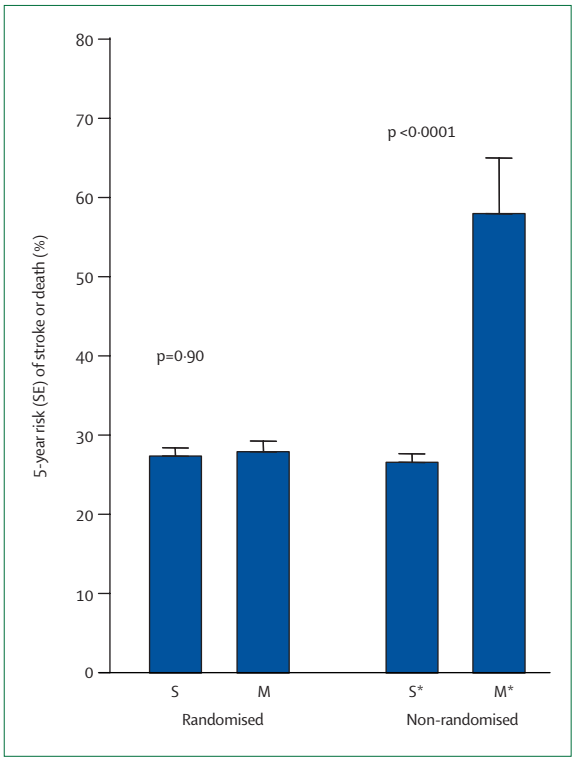

Figure 6: 5-year risks of stroke or death in all patients randomised to surgery (S) versus medical treatment only (M) in the ECST,[60] and in equivalent non-randomised comparison within group of patients randomised to surgery—ie, patients who had surgery (S*) versus those who did not (M*)
See table 1 for baseline clinical characteristics.

Adverse effects of treatment

Reporting of adverse effects of treatment in RCTs and systematic reviews is often poor. In a review of 192 pharmaceutical trials, less then a third had adequate reporting of adverse clinical events or laboratory toxicological findings.[135] Treatment discontinuation rates provide some guide to tolerability but pharmaceutical trials often use eligibility criteria and run–in periods to exclude patients who might be prone to adverse effects. Rates of discontinuation of treatment are therefore greater in clinical practice.[136,137] Publication bias and inadequate reporting of adverse events in RCTs supported by the pharmaceutical industry is a long-standing and unresolved difficulty.[138,139]

Clinicians are usually most concerned about the external validity of RCTs of potentially dangerous treatments. Iatrogenic complications are a leading cause of death in developed countries.[140] Risks can be overestimated in RCTs, particularly during the intro-

duction of new treatments when trials often enrol patients with very severe disease, but stringent selection of patients, confinement to specialist centres, and intensive safety monitoring usually lead to lower risks than in routine clinical practice. RCTs of warfarin in non-rheumatic atrial fibrillation are a good example. All trials reported a benefit with warfarin but complication rates were much lower than in routine practice.[24,141] Consequent doubts about external validity are partly to blame for major under-prescribing of warfarin, particularly in patients aged over 75,[24,142,143] who account for over 70% of non-rheumatic atrial fibrillation cases,[24] and who are at highest risk without treatment.[24,143]

Validity of routinely collected data

Non-randomised treatment comparisons based on routinely collected data are often suggested to be more externally valid than RCTs because they include all patients, are done in the real world, and include the effects of the doctor-patient relationship and patient preference.[144,145] Although with appropriate adjustment for differences in case-mix the results are sometimes similar to those from RCTs,[146,147] it is impossible to be certain that bias will be evident or amenable to correction. The analysis in table 1 of patients who were randomised to endarterectomy in the ECST showed that apparently similar clinical characteristics can hide major differences in prognosis. It also provides an interesting non-randomised comparison—ie, outcome can be compared within the group randomised to surgery between patients who underwent surgery (treatment group) and those who did not (controls). The non-randomised comparison suggests that surgery reduces the odds of stroke or death at 5 years by more than half (odds ratio=0·46, 95% CI=0·28–0·77, p=0·003, figure 6) and benefit is even greater after adjustment for case-mix (0·32, 0·15–0·57, p<0·001). However, the equivalent intention-to-treat analysis in ECST based on randomised treatment allocation shows that endarterectomy is, in fact, of no overall benefit across all degrees of stenosis (1·02, 0·88–1·18, p=0·90, figure 6). The non-randomised comparison is, of course, contrived and unreasonable because, as discussed earlier, there were specific reasons why patients who were randomised to surgery did not have the operation, but it illustrates the fact that such bias cannot be reliably corrected by adjustment for case-mix.[148] Routinely collected data are useful where RCTs are impractical, such as in evaluating rare adverse events, but are an adjunct rather than an alternative.

Panel 3: Examples of interventions believed beneficial (or harmful) but subsequently shown to be harmful (or beneficial) in RCTs

Judged beneficial, shown to be harmful

- High-dose oxygen therapy in neonates
- Antiarrhythmic drugs after myocardial infarction
- Fluoride treatment for osteoporosis
- Bed rest in twin pregnancy
- Hormone replacement therapy in vascular prevention
- Extracranial to intracranial arterial bypass surgery in stroke prevention
- High-dose aspirin for carotid endarterectomy

Judged harmful, shown to be beneficial

- β blockers in heart failure
- Digoxin after myocardial infarction

Summary and recommendations

Randomised trials and systematic reviews provide the most reliable data on the effects of treatment and many serious errors have resulted from relying on other types of evidence (panel 3). However, although dogmatic refusal by clinicians to accept the results of RCTs is unacceptable, there is justifiable concern that external validity is often poor. This perception is leading to the under-use in routine practice of treatments that have been shown to be effective in trials. Some trials have very good external validity,[93,149] but, as outlined above, many do not, especially some of those done by the pharmaceutical industry. Yet researchers, funding agencies, ethics committees, medical journals, the pharmaceutical industry, and their governmental regulators all neglect proper consideration of external validity (panel 1). Judgment is left to clinicians, but reporting of the determinants of external validity in trial publications, and particularly in secondary reports and clinical guidelines, is rarely adequate. Some information is sometimes published in a preceding methods paper, but this is often buried in an obscure journal not readily accessible to busy clinicians, and much relevant information is never published. RCTs and systematic reviews cannot be expected to produce results that are directly relevant to all patients and all settings, but to be externally valid they should at least be designed and reported in a way that allows patients and clinicians to judge to whom they can reasonably be applied. While cognisant of the risks of over–regulation, some of the following recommendations might be worthwhile:

- Further research into the external validity of RCTs, particularly in relation to the measured treatment effect.
- Stricter requirements than previously for the external validity of RCTs submitted to pharmaceutical licensing authorities.
- Increased consideration of external validity in the CONSORT guidelines on the reporting of RCTs,[49] and the Cochrane Collaboration guidelines on systematic reviews,[50] and agreement on a checklist similar to panel 2.

- The International Committee of Medical Journal Editors should require that all primary reports of RCTs or systematic reviews should contain a section entitled "To whom do these results apply?"

References

1 Cochrane AL. Effectiveness and Efficiency: random reflections on Health Services. London: Nuffield Provincial Hospitals Trust, 1972.

2 Horton R. Common sense and figures: the rhetoric of validity in medicine (Bradford Hill Memorial Lecture 1999). *Stat Med* 2000; **19:** 3149–64.

3 Pocock SJ. Clinical trials: a practical approach. Chichester: John Wiley, 1983.

4 Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials. 3rd edn. New York: Springer 1998.

5 Black D. The limitations of evidence. *J R Coll Physicians Lond* 1998; **32:** 23–26.

6 Hampton JR. Size isn't everything. *Stat Med* 2002; **21:** 2807–14.

7 Caplan LR. Evidence based medicine: concerns of a clinical neurologist. *J Neurol Neurosurg Psychiatry* 2001; **71:** 569–74.

8 Evans JG. Evidence-based and evidence-biased medicine. *Age Ageing* 1995; **24:** 461–63.

9 Charlton BG, Miles A. The rise and fall of EBM. *Q J Med* 1998; **91:** 371–74.

10 Naylor C. Grey zones in clinical practice: some limits to evidence-based medicine. *Lancet* 1995; **345:** 840–42.

11 Swales JD. Evidence-based medicine and hypertension. *J Hypertens* 1999; **17:** 1511–16.

12 Morgan WKC. On evidence, embellishment and efficacy. *J Eval Clin Pract* 1997; **3:** 117–22.

13 Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine". *Am J Med* 1997; **103:** 529–35.

14 Pashos CL, Normand ST, Garfinkle JB, Newhouse JP, Epstein AM, McNeil B. Trends in use of drug therapies in patients with acute myocardial infarction: 1988–1992. *J Am Coll Cardiol* 1994; **23:** 1023–30.

15 Garfield FB, Garfield JM. Clinical judgement and clinical practice guidelines. *Int J Technol Assess Health Care* 2000; **16:** 1050–60.

16 Cabana MB, Rand CS, Powe NR, et al. Why don't clinicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999; **282:** 1458–65.

17 Davis DA, Taylor-Vaisey A. Translating guidelines into practice, a systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *Can Med Assoc J* 1997; **157:** 408–16.

18 Ford LG, Hunter CP, Diehr P, Frelick RW, Yates J. Effect of patient management guidelines on physician practice patterns: the community hospital oncology program experience. *J Clin Oncol* 1987; **5:** 504–11.

19 Grol R, Dalhuijsen J, Thomas S, et al. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998; **317:** 858–61.

20 Messerli FH. Antihypertensive therapy: β-blockers and diuretics - why do physicians not always follow guidelines? *Proceedings/Baylor University Medical Center* 2000; **13:** 128–31.

21 Sonis J, Doukas D, Klinkman M, Reed B, Ruffin MT. Applicability of clinical trial results to primary care. *J Am Med Soc* 1998; **280:** 1746.

22 Wilson S, Delaney BC, Roalfe A, et al. Randomized controlled trials in primary care: case study. *BMJ* 2000; **321:** 24–27.

23 Sellors J, Cosby R, Trim K, et al, for the Seniors Medication Assessment Research Trial (SMART) Group. Recruiting family physicians and patients for a clinical trial: lessons learned. *Fam Pract* 2002; **19:** 99–104.

24 Oswald N, Bateman H. Applying research evidence to individuals in primary care: a study using non-rheumatic atrial fibrillation. *Fam Pract* 1999; **16:** 414–19.

25 Fahey T. Applying the results of clinical trials to patients in general practice: perceived problems, strengths, assumptions, and challenges for the future. *Br J Gen Pract* 1998; **48:** 1173–78.

26 Mant D. Can randomised trials inform clinical decisions about individual patients? *Lancet* 1999; **353:** 743–46

27 Jacobson LD, Edwards AGK, Granier SK, Butler CC. Evidence-based medicine and general practice. *Br J Gen Pract* 1997; **47:** 449–52.

28 McCormick JS. The place of judgement in medicine. *Br J Gen Pract* 1994; **44:** 50–51.

29 Benech I, Wilson AE, Dowell AC. Evidence-based practice in primary care: past, present and future. *J Eval Clin Prac* 1996; **2:** 249–63.

30 Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials: exclusions and selective participation. *J Health Serv Res Policy* 1999; **4:** 112–21.

31 Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non–randomised studies of health care interventions. *J Epidemiol Community Healt*h 1998; **52:** 377–84.

32 Egglin TKP, Horwitz RI. The case for better research standards in peripheral thrombolysis: poor quality of randomised trials during the past decade. *Acad Radiol* 1996; **3:** 1–9.

33 Gilbody S, Wahlbeck K, Adams C. Randomized controlled trials in schizophrenia: a critical perspective on the literature. *Acta Psychiatr Scand* 2002; **105:** 243–51.

34 Licht RW, Gouliaev G, Vestergaard, Frydenberg M. Generalisability of results of randomised drug trials. *Br J Psychiatry* 1997; **170:** 264–67.

35 Camm AJ. Clinical trials of arrhythmia management: methods or madness. *Control Clin Trials* 1996; **17:** 4s–16s.

36 Norris SL, Engelgau MM, Narayan KMV. Effectiveness of self-management training in type 2 diabetes: a systematic review of randomised controlled trials. *Diabetes Care* 2001; **24:** 561–87.

37 Moore DAJ, Goodall RL, Ives NJ, Hooker M, Gazzard BG, Easterbrook PJ. How generalisable are the results of large randomised controlled trials of antiretroviral therapy. *HIV Med* 2000; **1:** 149–54.

38 Brown N, Melville M, Gray D, et al. Relevance of clinical trial results in myocardial infarction to medical practice: comparison of four year outcome in participants of a thrombolytic trial, patients receiving routine thrombolysis, and those deemed ineligible for thrombolysis. *Heart* 1999; **81:** 598–602.

39 Charleson ME, Horwitz RI. Applying results of randomised trials to clinical practice: impact of losses before randomisation. *BMJ* 1984; **289:** 1281–84.

40 Simon GE, Vonkorff M, Heilingenstein JH, et al. Initial antidepressant choice in primary care: effectiveness of fluoxetine vs. tricyclic antidepressants. *JAMA* 1996; **275:** 1897–902.

41 Davey Smith G, Egger M. Incommunicable knowledge? Interpreting and applying the results of clinical trials and meta-analyses. *J Clin Epidemiol* 1998; **51:** 289–95.

42 Hennen BK. Measuring complexity of clinical problems. *J Med Educ* 1984; **59:** 487–93.

43 Julian DG, Pocock SJ. Interpreting a trials report. In Pitt B, Julian D, Pocock S, eds. *Clinical Trials in Cardiology* London: WB Saunders, 1997: 33–42.

44 Idanpaan-Heikkila JE. WHO guidelines for good clinical practice (GCP) for trials on pharmaceutical products: responsibilities of the investigator. *Ann Med* 1994; **26:** 89–94.

45 Wermeling DP. Clinical research: regulatory issues. *Am J Health Syst Pharm* 1999; **56:** 252–56.

46 Medical Research Council. MRC Guidelines for Good Clinical Practice in Clinical Trials. London: Medical Research Council, 1998.

47 Medical Research Council. Clinical Trials for Tomorrow. London: Medical Research Council, 2003.

48 Governance arrangements for NHS Research Ethics Committees. www.dh.gov.uk/assetRoot/04/05/86/09/04058609.pdf (accessed July 21, 2004).

49 Altman DG, Schulz KF, Moher D, et al, for the Consort Group. The revised CONSORT statement for reporting randomised trials: explanation and elaboration. *Ann Intern Med* 2001; **134:** 663–94.

50 Alderson P, Green S, Higgins JPT, eds. Cochrane Reviewers' Handbook 4.2.1 [updated December 2003]. In: *The Cochrane Library*, Issue 1, 2004. Chichester: John Wiley & Sons Ltd.

51 LeBaron S, Reyher J, Stack JM. Paternalistic vs egalitarian physician styles: the treatment of patients in crisis. *J Fam Pract* 1985; **21:** 56–62.

52 Thomas KB. General practice consultations: is there any point in being positive? *BMJ* 1987; **294:** 1200–02.

53 Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. *Lancet* 2001; **357:** 757–62.

54 Kleijnen J, de Craen AJM, van Everdingen J, Krol L. Placebo effect in double-blind clinical trials: a review of interactions with medications. *Lancet* 1994; **344:** 1347–49.

55 Kaptchuk TJ. Powerful placebo: the dark side of the randomised controlled trial. *Lancet* 1998; **351:** 1722–25.

56 Onel E, Hammond C, Wasson JH, et al. An assessment of the feasibility and impact of shared decision making in prostate cancer. *Urology* 1998; **51:** 63–66.

57 Benson J, Britten N. Patients decisions about whether or not to take antihypertensive drugs: qualitative study. *BMJ* 2002; **325:** 873–76.

58 Redelmeier DA, Rozin P, Kahneman D. Understanding patients' decisions. *JAMA* 1993; **270:** 72–76.

59 Olschewski M, Schumacher M, Davis KB. Analysis of randomised and non-randomized patients in clinical trials using the comprehensive cohort follow-up study design. *Control Clin Trials* 1992; **13:** 226–39.

60 European Carotid Surgery Trialists' Collaborative Group. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). *Lancet* 1998; **351:** 1379–87

61 Lovett JK, Coull A, Rothwell PM, on behalf of the Oxford Vascular Study. Early risk of recurrent stroke by aetiological subtype: implications for stroke prevention. *Neurology* 2004; **62:** 569–74.

62 Masuhr F, Busch M, Einhaupl KM. Differences in medical and surgical therapy for stroke prevention between leading experts in North America and Western Europe. *Stroke* 1998; **29:** 339–45.

63 Sacco RL, Kargman DE, Gu Q, Zamanillo MC. Race-ethnicity and determinants of intracranial atherosclerotic cerebral infarction. The Northern Manhattan Stroke Study. *Stroke* 1995; **26:** 14–20.

64 Fine PEM. Variation in protection by BCG: implications of and for heterologous immunity. *Lancet* 1995; **346:** 1339–45.

65 Ricci S, Celani MG, Righetti E, Cantisani AT, for the International Stroke Trial Collaborative Group. Between country variations in the use of medical treatments for acute stroke: An update. *Cerebrovasc Dis* 1996; **6** (suppl 2)**:** 133

66 Roberts C. The implications of variation in outcome between health professionals for the design and analysis of randomized controlled trials. *Stat Med* 1999; **18:** 2605–15.

67 Downs SH, Black NA, Devlin HB, Royston CMS, Rusell RCG. Systematic review of the effectiveness and safety of laparoscopic cholecystectomy. *Ann R Coll Surg Engl* 1996; **78:** 241–323.

68 Black NA, Downs SH. The effectiveness of surgery for stress incontinence in women: a systematic review. *Br J Urol* 1996; **78:** 497–510.

69 Asymptomatic Carotid Atherosclerosis Study Group. Carotid endarterectomy for patients with asymptomatic internal carotid artery stenosis. *JAMA* 1995; **273:** 1421–28.

70 Moore WS, Young B, Baker WH, et al. Surgical results: a justification of the surgeon selection process for the ACAS trial. *J Vasc Surg* 1996; **23:** 323–38.

71 Bond R, Rerkasem K, Rothwell PM. High morbidity due to endarterectomy for asymptomatic carotid stenosis. *Cerebrovasc Dis* 2003: **16** (suppl 4)**:** 65.

72 Barnett HJM, Barnes RW, Clagett GP, Ferguson GG, Robertson JT, Walker PM. Symptomatic carotid artery stenosis: a solvable problem. The NASCET trial. *Stroke* 1992; **23:** 1048–53

73 De Vita VT. Breast cancer therapy; exercising all our options. *N Engl J Med* 1989; **320:** 527–29.

74 Gurwitz JH, Col NF, Avorn J. The exclusion of elderly and women from clinical trials in acute myocardial infarction. *JAMA* 1992; **268:** 1417–22.

75 Bungeja G, Kumar A, Banerjee AK. Exclusion of elderly people from clinical research: a descriptive study of published reports. *BMJ* 1997; **315:** 1059.

76 Jones R, Lydeard S. Prevalence of symptoms of dyspepsia in the community. *BMJ* 1989; **298:** 30–32.

77 Kay L. Prevalence, incidence and prognosis of gastrointestinal symptoms in a random sample of an elderly population. *Age Ageing* 1994; **23:** 146–49.

78 Jorgensen HS, Nakayama H, Kammersgaard LP, et al. Predicted impact of intravenous thrombolysis on prognosis of general population of stroke patients: simulation model. *BMJ* 1999; **319:** 288–89.

79 National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischaemic stroke. *N Engl J Med* 1995; **333:** 1581–87.

80 LaRue LJ. Alter M, Traven ND, et al. Acute stroke therapy trials: problems in patient accrual. *Stroke* 1988; **19:** 950–54.

81 Maynard C, Althouse R, Cerqueira M, Olsufka M, Kennedy JW. Underutilisation of thrombolytic therapy in eligible women with acute myocardial infarction. *Am J Cardiol* 1991; **68:** 529–30.

82 Steiner TJ, Clifford Rose F. Towards a model stroke trial. The single centre naftidrofuryl study. *Neuroepidemiology* 1986; **5:** 121–47.

83 Rovers MM, Zielhuis GA, Bennett K, Haggard M. Generalisability of clinical trials in otitis media with effusion. *Int J Pediatr Otorhinolaryngol* 2001; **60:** 29–40.

84 Califf RM, Pryor DB, Greenfield JC. Beyond randomised clinical trials: applying clinical experience in the treatment of patients with coronary artery disease. *Circulation* 1986; **74:** 1191–94.

85 Stroke Prevention in Atrial Fibrillation Investigators. Stroke Prevention in Atrial Fibrillation Study. *Circulation* 1991; **84:** 527–39.

86 Muller DWM. Selection of patients with acute myocardial infarction for thrombolytic therapy. *Ann Intern Med* 1990; **113:** 949–60.

87 Henderson RA, Raskino CL, Hampton JR. Variations in the use of coronary arteriography in the UK: the RITA trial coronary arteriogram register. *Q J Med* 1995; **88:** 167–73.

88 Bjorn M, Brendstrup C, Karlsen S, Carlsen JE. Consecutive screening and enrolment in clinical trials: the way to representative patient samples? *J Card Fail* 1998; **4:** 225–30.

89 Bowen JT, Barnes TRE. The clinical characteristics of schizophrenic patients consenting or not consenting to a placebo controlled trial of antipsychotic medication. *Hum Psychopharmacol* 1994; **9:** 432–33.

90 Schmoor C, Olschewski M, Schumacher M. Randomized and non-randomized patients in clinical trials: experiences with comprehensive cohort studies. *Stat Med* 1996; **15:** 263–71.

91 Stiller CA. Centralised treatment, entry to trials and survival. *Br J Cancer* 1994; **70:** 352–62.

92 Woods Kl, Ketley D. Intravenous β blockade in acute myocardial infarction. Doubt exists about external validity of trials of intravenous β blockade. *BMJ* 1999; **318:** 328–29.

93 Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986; **1:** 397–402

94 Pablos-Mendez A, Barr RG, Shea S. Run-in periods in randomised trials: implications for the application of results in clinical practice. *JAMA* 1998; **279:** 222–25.

95 Haynes RB, Dantes R. Patient compliance and the conduct and interpretation of therapeutic trials. *Control Clin Trials* 1987; **8:** 12–19.

96 Gomez-Marin O, Prineas RJ, Sinaiko AR. The sodium-potassium blood pressure trial in children: design, recruitment and randomisation. *Control Clin Trials* 1991; **12:** 408–23.

97 Australia–New Zealand Heart Failure Research Collaborative Group. Effects of carvedilol, a vasodilatory β-blocker, in patients with congestive heart failure due to ischaemic heart disease. *Circulation* 1995; **92:** 212–18.

98 Packer M, Bristow MR, Cohn JN et al, for the US Carvedilol Heart Failure Study Group. The effects of carvedilol on morbidity and mortality in patients with chronic heart failure. *N Engl J Med* 1996; **334:** 1349–55.

99 Amery W, Dony J. A clinical trial design avoiding undue placebo treatment. *J Clin Pharmacol* 1975; **15:** 674–79.

100 Quitkin FM, Rabkin JG. Methodological problems in studies of depressive disorder: utility of the discontinuation design. *J Clin Psychopharmacol* 1981; **1:** 283–88.

101 Hallstrom A, Verter J, Friedman L. Randomising responders. *Control Clin Trials* 1991; **12:** 486–503.

102 Leber PD, Davis CS. Threats to the validity of clinical trials employing enrichment strategies for sample selection. *Control Clin Trials* 1998; **19:** 178–87.

103 Davis KL, Thal LJ, Gamzu ER, et al, and the Tacrine Collaborative Study Group. A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease. *N Engl J Med* 1992; **327:** 1253–59.

104 Hall JC, Mills B, Nguyen H, Hall JL. Methodologic standards in surgical trials. *Surgery* 1996; **119:** 466–72.

105 Shapiro SH, Weijer C, Freedman B. Reporting the study populations of clinical trials. Clear transmission or static on the line? *J Clin Epidemiol* 2000; **53:** 973–79.

106 European Atrial Fibrillation Trial (EAFT) Study Group. Secondary prevention in non-rheumatic atrial fibrillation after transient ischaemic attack or minor stroke. *Lancet* 1993; **342:** 1255–62.

107 Algra A, Francke CL, Koehler PJJ, for the Stroke Prevention in Reversible Ischaemia Trial (SPIRIT) group. A randomized trial of anticoagulants versus aspirin after cerebral ischaemia of presumed arterial origin. *Ann Neurol* 1997; **42:** 857–65.

108 Gorter JW, for the Stroke Prevention in Reversible Ischaemia Trial (SPIRIT) and European Atrial Fibrillation Trial (EAFT) groups. Major bleeding during anticoagulation after cerebral ischaemia: patterns and risk factors. *Neurology* 1999; **53:** 1319–27.

109 Rothwell PM, Eliasziw M, Gutnikov SA, et al, for the Carotid Endarterectomy Trialists' Collaboration. Pooled analysis of individual patient data from randomised controlled trials of endarterectomy for symptomatic carotid stenosis. *Lancet* 2003; **361:** 107–16

110 Merlo J, Broms K, Lindblad U, et al. Association of outpatient utilisation of non-steroidal anti-inflammatory drugs and hospitalised heart failure in the entire Swedish population. *Eur J Clin Pharmacol* 2001; **57:** 71–75.

111 Rothwell PM, Slattery J, Warlow CP. A systematic review of the risks of stroke and death due to carotid endarterectomy for symptomatic stenosis. *Stroke* 1996; **27:** 260–65.

112 Riggs BL, Hodgson SF, O'Fallon WM, et al. Effect of fluoride treatment on fracture rate in postmenopausal women with osteoporosis. *N Engl J Med* 1990; 322: 802–09.

113 McAlistair FA, Teo KK. Antiarrhythmic therapies for the prevention of sudden cardiac death. *Drugs* 1997; **54:** 235–52.

114 The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989; **321:** 406–12.

115 The IFNB Multiple Sclerosis Study Group. Interferon β-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology* 1993; **43:** 655–61.

116 The IFNB Multiple Sclerosis Study Group and the University of British Columbia MS/MRI Analysis Group. Interferon β-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. *Neurology* 1995; **45:** 1277–85.

117 Jacobs LD, Cookfair DL, Rudick AR, et al. Intramuscular interferon β-1a for disease progression in relapsing multiple sclerosis. *Ann Neurol* 1996; **39:** 285–94.

118 Ebers GC for the PRISMS Study Group. Randomised double-blind placebo-controlled study of interferon β-1a in relapsing-remitting multiple sclerosis. *Lancet* 1998; **352:** 1498–504.

119 Di Bianco R, Shabetai R, Kostuk W, et al. A comparison of oral milrinone, digoxin and their combination in the treatment of patients with chronic heart failure. *N Engl J Med* 1989; **320:** 677–83.

120 Sueta CA, Gheorghiade M, Adams KF, et al, and the Epoprostenol Multicentre Research Group. Safety and efficacy of epoprostenol in patients with severe congestive heart failure. *Am J Cardiol* 1995; **75:** 34A–43A.

121 Packer M, Carver JR, Rodeheffer RJ, et al, for the Promise Study Research Group. Effect of oral milrinone on mortality in severe chronic heart failure. *N Engl J Med* 1991; **325:** 1468–75.

122 Califf RM, Adams KF, McKenna WJ, et al. A randomized controlled trial of epoprostenol therapy for severe congestive heart failure. *Am Heart J* 1997; **134:** 44–54.

123 Yee KM, Struthers AD. Can drug effects on mortality in heart failure be predicted by any surrogate outcome measure? *Eur Heart J* 1997; **18:** 1860–64.

124 Hampton JR, van Veldhuisen DJ, Kleber FX, et al. Randomised study of ibopamine on survival in patients with advanced severe heart failure. *Lancet* 1997; **349:** 971–77.

125 Coste J, Fermanian J, Venot A. Methodological and statistical problems in the construction of composite measurement scales: a survery of six medical and epidemiological journals. *Stat Med* 1995; **14:** 331–45.

126 van Gijn J, Warlow C. Down with stroke scales. *Cerebrovasc Dis* 1992; **2:** 244–46.

127 Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; **10:** 31–56.

128 Thornley B, Adams CE. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998; **317:** 1181–84.

129 Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales—a major source of bias in randomised controlled trials of treatments for schizophrenia? *Br J Psychiatry* 2000; **176:** 249–52.

130 Rothwell PM, McDowell Z, Wong CK, P Dorman. Doctors and patients don't agree: cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. *BMJ* 1997; **314:** 1580–83.

131 Binnie CD. Design of clinical antiepileptic drug trials. *Seizure* 1995; **4:** 187–92.

132 Walker MC, Sander JW. Difficulties in extrapolating from clinical trial data to clinical practice: the case of antiepileptic drugs. *Neurology* 1997; **49:** 333–37.

133 Diener HC, Cunha L, Forbes C, Sivenius J, Smets P, Lowenthal A, for the European Stroke Prevention Study. 2. Dipyridamole and acetylsalicylic acid in the secondary prevention of stroke. *J Neurol Sci* 1996; **143:** 1–13.

134 Pincus T. Rheumatoid arthritis: disappointing long-term outcomes despite successful short–term clinical trials. *J Clin Epidemiol* 1998; **41:** 1037–41.

135 Ioannidis JP, Contopoulos-Ioannidis DG. Reporting of safety data from randomised trials. *Lancet* 1998; **352:** 1752–53.

136 Jones J, Gorkin L, Lian J, Staffa JA, Fletcher AP. Discontinuation of and changes in treatment after start of new courses of antihypertensive drugs: a study of a United Kingdom population. *BMJ* 1995; **311:** 293–95.

137 Andrade SE, Walker AM, Gottlieb LK, et al. Discontinuation of antihyperlipidaemic drugs—do rates reported in clinical trials reflect rates in primary care settings. *N Engl J Med* 1995; **332:** 1125–31.

138 Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ* 1980; **280:** 833–36.

139 McPherson K, Hemminki E. Synthesising licensing data to assess drug safety. *BMJ* 2004; **328:** 518–20.

140 Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalised patients. *JAMA* 1998; **279:** 1200–05.

141 Landefeld CS, Beyth RJ. Anticoagulant-related bleeding: clinical epidemiology, prediction, and prevention. *Am J Med* 1993; **95:** 315–28.

142 Lip GH, Golding DJ, Masood N, Beevers DG, Child DL, Fletcher RI. A survey of atrial fibrillation in general practice. *Br J Gen Pract* 1997; **47:** 285–89.

143 Antani MR, Beyth RJ, Covinsky KE, et al. Failure to prescribe warfarin to patients with non-rheumatic atrial fibrillation. *J Gen Intern Med* 1996; **11:** 713–20.

144 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; **312:** 1215–18.

145 McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Methods in health service research. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999; **319:** 312–15.

146 Concato J, Shah N, Horwitz RI. Randomised, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; **342:** 1887–92.

147 Benson K, Hartz AJ. A comparison of observational studies and randomized controlled trials. *N Engl J Med* 2000; **342:** 1878–86.

148 Coronary Drug Project Research Group. Influence of adherence to treatment and response to cholesterol on mortality in the Coronary Drug Project. *N Engl J Med* 1980; **303:** 1038–41.

149 Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002; **360:** 7–22.