

TO WHOM DO THE RESEARCH FINDINGS APPLY?

Curt D Furberg

570

Heart 2002;**87**:570–574

When a new intervention (drug, procedure or device) becomes mainstream care, one hopes that all groups of patients for whom this intervention is intended have been properly studied and, thus, are well defined. This ideal situation rarely applies. The clinical trials conducted to determine efficacy and safety of new interventions are typically designed to be feasible and time and cost efficient. As a consequence, trial populations are typically highly selected and may represent only a subset of the patients for whom the intervention is targeted. Thus, the applicability of the trial findings to other subpopulations has to be based on extrapolations. Some of these extrapolations are reasonable, while others are debatable.

Five considerations often influence trial design¹: the desire for a study population that (1) is aetiologically homogeneous, (2) is most likely to respond favourably to the intervention, (3) is least likely to suffer adverse events, (4) has no or limited co-morbidity, and (5) most likely will consist of good compliers. The inclusion and exclusion criteria in the trial protocol define those patients with a given condition who are eligible for trial participation or the so-called study population. In addition, all trial participants, by definition, must consent to participate in a research project. Those enrolled constitute the study sample.

This article highlights the conflict between the needs of an optimal research design and a desire from the clinical perspective to determine if all patient groups stand to benefit from a new intervention. The outcome chosen for a clinical trial often influences the interpretation of results. The problem of application of research findings will be illustrated by examples from the literature.

▶ HOW ELIGIBILITY CRITERIA LIMIT THE ABILITY TO GENERALISE FINDINGS

From the point of view of generalisability, the ideal trial would have no exclusion criteria, other than exclusions that reflect known contraindications to the study intervention. All other patients with a given condition would be eligible for enrolment. In addition, the sample size chosen would allow enrolment of sufficient numbers of participants in defined subgroups of interest, so that adequately powered subgroup analyses could be conducted. Unfortunately, such trials are not feasible. Rarely do we have enough statistical power to determine the efficacy and safety of an intervention in even major subgroups that are defined by co-variables such as age, sex, ethnicity, disease severity and stage, co-morbidity, use of other major interventions (interactions), and presence of specific genetic polymorphism that may influence treatment response. Readers of scientific articles should be aware of the “leaps of faith” that are inherent in interpreting research findings.

Homogeneity

Patients who could potentially benefit the most from a new intervention represent the preferred candidates for enrolment into a trial. Decisions regarding eligibility are often based on knowing the mechanism(s) of action of an intervention, thus enabling investigators to identify those most likely to respond favourably. Knowledge of the microorganism causing a specific infection is an important consideration when designing a trial of a new antibiotic agent. Those with the same clinical diagnosis caused by other types and strains of bacteria may be excluded. Exclusion of otherwise eligible patients based on age, impaired renal or liver function, and other co-morbidity creates a more homogeneous group that is more likely to benefit maximally. The desire to create a well defined, homogeneous study population that optimises the likelihood of a favourable trial outcome, however, may limit the ability to generalise the findings.

Likelihood of benefit

Behind the careful selection of study participants is also the desire to obtain results within a reasonable time and with a finite amount of funding. For a new anti-anginal drug, one would probably exclude those with mild angina as well as those with the most severe pain, thus focusing on patients who fall between these extremes. It could be difficult to demonstrate benefit in a patient who only has chest pain once a month. Patients at the other end of the disease spectrum—those with very severe or intractable chest pain—may be too incapacitated to respond to a typical new

Correspondence to:
Curt D Furberg, MD, PhD,
Department of Public Health
Sciences, Wake Forest
University School of Medicine,
Medical Center Boulevard,
Winston-Salem, NC
27157-1063, USA;
cfurberg@wfubmc.edu



- RESEARCH SOMETIMES
ADVANCES THE SCIENTIST
MORE THAN SCIENCE.

anti-anginal agent. The aetiology behind their pain may be different from that of ambulatory patients with modest angina pectoris. This selection of a study population most likely to respond favourably may come at the expense of not knowing whether and to what extent the drug works in the mildest and most severe cases. Once again, the desire to optimise the outcome of a research study could limit the ability to generalise study findings.

Avoiding adverse effects

Since most (all?) interventions have adverse effects, investigators who design trials prefer excluding patients who are likely to experience these. This consideration is in accordance with the ethical guidelines defined in the Declaration of Helsinki. Many exclusion criteria in a randomised clinical trial indeed reflect potential safety problems. Because such exclusions include various types and severities of potential adverse effects of the intervention, these constitute relative and absolute contraindications. Teratogenicity is a common concern, and pregnant women are typically excluded from trial participation. Excluding patients who are at increased risk for developing adverse events makes sense. Patients with a history of gastric bleeding are typically excluded from trials testing agents that may cause gastric bleeding, such as anti-inflammatory drugs. Thus, trials are designed to enroll uncomplicated cases, in which the risk of adverse effects is small. Low rates also help in the regulatory approval process and in the subsequent marketing of the new product. Co-morbidity is avoided, which often means an underrepresentation of older patients in the study population. In real life, the most likely candidates for prescription of a newly marketed drug are those with some form of co-morbidity or more advanced disease. They may have failed to respond to existing drugs or developed adverse effects. Thus, the desire for a well defined study population with no or limited co-morbidity comes with a cost, in terms of general applicability and an underestimation of adverse effects.

Avoidance of competing risk

A related issue is that of so-called competing risk. A general principle in trial design is to exclude certain patients who are

at increased risk of developing the clinical outcome that investigators are trying to prevent. For example, in a lipid lowering trial with all cause mortality as the primary outcome, patients with an increased risk of dying from reasons unrelated to lipids/lipoproteins are excluded. This would, for example, apply to those with cancer or serious kidney or liver damage who can be expected to have shortened life expectancy. Inclusion of patients who are dying from other conditions during a trial will add background "noise" to the trial findings by diluting any mortality effect of the new lipid lowering agent. Thus, the ability to ascertain the true effect of an intervention is lessened in the presence of competing risk.

Avoiding potential non-compliers

Every investigator's nightmare is the patient who stops taking the study medication, especially shortly after he or she has been enrolled. The impact of non-compliers as well as poor compliers on sample size can be substantial. These patients also require major staff commitment during the trial. For analytic purposes, they have to be contacted and monitored for the occurrence of trial outcomes. For proper reporting of trial findings, events in all randomised patients are expected to be collected and reported. Therefore, investigators endeavour to exclude from trial participation anticipated non-compliers or poor compliers. This would include those with a history of adherence problems, alcohol and drug abusers, and those with mental problems. It makes sense from a design efficiency perspective to enrich the study population with potentially good compliers. However, it should be noted that poor and good compliers might differ in other respects. Canner and colleagues² reported that the risk of major coronary events differed among compliers and non-compliers in the placebo group of the coronary drug project. The non-compliers were at a significantly higher risk. It is not known why non-compliers on placebo have more coronary events. Thus, the focus in clinical trials on good compliers can overestimate the favourable findings of a trial.

Volunteers

Finally, clinical trial participants all volunteer to enroll by signing an informed consent. It has been argued that volunteers and non-volunteers (those who qualify but decline an invitation to participate) differ. There is scientific evidence to support either side of that argument. Efforts were made to address this question in the coronary artery surgery study.³ The event rate in the non-surgical (medically treated) control group of the trial was comparable to that of patients who met the inclusion criteria, but declined randomisation. In contrast, Smith and Arnesen⁴ found that non-consenters had a higher mortality than consenters in a postinfarction trial.

In summary, clinical trials are typically designed to test an intervention in patients: (1) who are carefully chosen to respond optimally based on the presumed mechanism(s) of action of the intervention and disease severity, (2) who are at low risk of adverse effects and free of co-morbid conditions, and (3) who are likely to be compliant. Compared to an unselected population with the same condition, one could expect trials to provide results in terms of both efficacy and safety that are more favourable to the new intervention. Extrapolation of the research findings to patients with characteristics that disqualified them from trial participation may present a challenge. Readers of scientific reports need to consider carefully the eligibility criteria and accept that the benefit versus risk balance may differ for patients not meeting these criteria. Clinical trials with few exclusion criteria (other than major contraindications) are more applicable to clinical practice.



— THE TRIAL WAS SO EXCLUSIVE THAT
NO ONE WAS EVER RANDOMIZED.

HOW THE TYPE OF INTERVENTION OUTCOME INFLUENCES APPLICABILITY

Most medical interventions are aimed at alleviating an existing symptom or sign, such as pain. Others directed at acute conditions such as an infection may accelerate cure or recovery. A third type of intervention is directed at altering the future course of a disease by preventing its complications, including premature death. Antihypertensive treatment is prescribed to prevent or reduce the risks of developing the devastating cardiovascular complications of hypertension.

Intervention trials assume varying designs, depending, in part, on whether they address existing conditions or endeavour to prevent complications that may occur. Of paramount importance are sample size requirements, which can differ enormously. It takes fewer patients to document a symptomatic benefit of a new agent. Whether such a treatment is beneficial in individual patients is easy to determine clinically. The patient can serve as his or her own control and an improvement may be “credited” to the intervention. This concept is behind the “trial of $n = 1$ ” approach.⁵

Preventing a future stroke in a hypertensive subject is a different story. If the risk of stroke is 2% per annum and the risk is reduced by half, of 100 hypertensive subjects treated, on average, one stroke will be prevented, one subject will suffer a stroke in spite of effective treatment, and the other 98 subjects will experience no strokes during the year of treatment. The problem with prevention is that no one can project who will suffer a complication that is preventable, who will suffer a complication in spite of treatment, and who will be treated unnecessarily and only be at risk of possible adverse events. Until we learn how to predict the course of a disease in individual patients better, prevention will always involve playing the odds.

Applying research findings to individual patients is more straightforward for interventions that alleviate symptoms or accelerate recovery from an acute condition. The individual patient’s response after exposure to the intervention will tell whether it “works”. There is no such direct feedback in prevention. Typically a large number of patients have to be treated for extended periods in order to help a few.

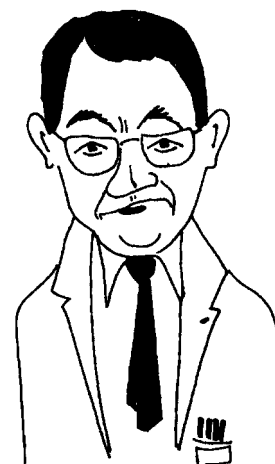
HOW CHANGES IN SURROGATE MARKERS PREDICT CLINICAL OUTCOMES

To avoid large and lengthy clinical trials, investigators and trial sponsors often resort to surrogate markers in the testing of an intervention. The blood pressure lowering effect of a new antihypertensive agent can be documented in a placebo controlled trial of 50–100 hypertensive subjects treated for 8–12 weeks. A stroke prevention trial of the same agent would require 4–5000 subjects treated for 4–5 years. Thus, small, short term trials with surrogate markers offer obvious advantages. Other examples of common surrogates in the cardiovascular field include low density and high density lipoprotein (LDL and HDL) cholesterol, Hb_{A1c} , premature ventricular depolarisations, ejection fraction, other haemodynamic measures, and angiographic changes.

A valid surrogate marker is one whose response to an intervention *closely* mimics that of the real (clinical) outcome it is supposed to represent. Unfortunately, this requirement is seldom met. The Veterans Affairs high density lipoprotein intervention trial⁶ reported that gemfibrozil reduced the risk of major coronary events in coronary patients with normal LDL cholesterol, but low HDL cholesterol. The assumption was that benefit was mediated through gemfibrozil induced increases in HDL cholesterol. When the investigators analysed the trial data to determine how much of the health benefit could be explained by individual changes in the surrogate marker (HDL cholesterol), they came up with the surprising finding that only 22% of the benefit could be attributed to gemfibrozil induced increases in HDL cholesterol. Similar observations have been reported for raised blood pressure (CD Furberg, unpublished data).

By contrast, sometimes drugs have favourable effects on surrogates, but actually cause harm. The cardiac arrhythmia suppression trial⁷ reported that even though encainide and flecainide notably reduced the number of premature ventricular depolarisations (a surrogate for sudden death), these drugs increased the risk of sudden death. A handful of inotropic agents have been shown to improve haemodynamic parameters in patients with congestive heart failure, but they were later shown to increase mortality.

The magnitude of the “improvement” of a surrogate marker cannot be assumed to predict, with high precision, the magnitude of a health benefit in individual patients. The expectation



— DARN, THIS TRIAL WOULD BE SIGNIFKANT
IF I COULD JUST EXCLUDE ONE MORE
EVENT FROM THE TREATMENT GROUP.

that common surrogates are clinically useful and predictive rests on the assumptions that drugs have only one mechanism of action (that of the surrogate) and that the development of clinical complications evolves through a single mechanism (mediated through the surrogate). All antihypertensive drugs lower raised blood pressure, but they differ greatly in their blood pressure independent actions. Hypertension is not just high blood pressure. Thus, there are good scientific reasons to expect that different classes of antihypertensive agents differ in how they reduce risk.⁸

It is important to remember that clinical trials investigate and report results for groups of subjects, not individual subjects. When we interpret trials, we assume that the group data apply equally to all individuals. Two recent articles^{9, 10} highlight the issues of interpreting and applying research findings to individuals. Caution is advised in inferring that a large change in a surrogate marker in an individual automatically translates to a greater clinical benefit than a small marker change. Subjects with small changes may also stand to benefit clinically.

ILLUSTRATIONS FROM CLINICAL TRIALS

To illustrate how highly selected the cohort of eligible trial patients are, Kääriäinen and colleagues¹¹ analysed 397 consecutively hospitalised cases of gastric ulcer to determine what proportion would be eligible for participation in drug trials and how the eligibility criteria affected generalisability. When the commonly used exclusion criteria were applied, 282 patients (71%) met at least one of them. Several patients had two or more reasons for exclusion. The most troubling findings came from an extended follow up of all 397 patients. Major complications of gastric ulcer—bleeding, perforation, gastric retention, and deaths—occurred in 71 patients, and only two of those were observed in the 115 patients who met the typical eligibility criteria for trials of gastric ulcer. Patients with the worst prognosis would have been excluded. The authors concluded: “when many patients are excluded, the applicability of the results to the whole material is questionable.”

Under representation of certain subgroups of patients in randomised clinical trials is another problem. Women and minorities are often under represented.^{12, 13} So are patients aged 65 years or older,¹⁴ who are the most likely to develop adverse effects. This failure to enroll certain groups of patients has led to a change in federal policies in the USA. It is important that patients enrolled in a trial represent the entire spectrum of patients with a given condition, to enhance the clinical applicability of the results.

ILLUSTRATIONS FROM OTHER TYPES OF RESEARCH STUDIES

Many of the methodological issues of randomised clinical trials also apply to other types of research studies. The latter studies are susceptible to additional problems/biases caused by lack of randomisation, comparable control groups, and blinding. This is illustrated by the following example.

In early July 1997, the US Food and Drug Administration (FDA) reported that it had received 33 reports of unusual valvar morphology and regurgitation among users of combined fenfluramine and phentermine, “fen-phen”.^{15, 16} Half of the cases, all women, who had used the drug combination from one month to more than 16 months (mean 10 months) also had pulmonary hypertension.



- SOMETIMES I HAVE TO GO THROUGH
MANY DIFFERENT STATISTICIANS TO
GET THE RIGHT RESULTS.

To determine the magnitude of the problem nationwide, the FDA strongly encouraged all healthcare professionals to report suspected cases of cardiac valvar disease associated with fen-phen use. It was known that between 1.2–4.7 million persons had been “exposed” (14 million prescriptions). Obesity clinics from five states reported echocardiographic findings from 284 subjects. The prevalence of valvulopathy was a staggering 32.8%; 22% in those with exposures < 6 months and 35% in those with longer exposures. Multiplying the number of persons exposed with the risk of valvulopathy gives a number of persons affected ranging from 130 000–500 000. These estimates, of epidemic proportions, raised several questions regarding their reliability.

A closer look at the data revealed a sampling bias. The cases in the Mayo Clinic report¹⁵ and the FDA sample had a much longer exposure than the 1.2–4.7 million users. Expectation bias created by all the publicity was another factor. The sonographers and the readers were not blinded and the readings were subjective (non-standardised). No consideration was given to the fact that valvulopathy is not uncommon in obese, middle aged persons.

Interestingly, the *Wall Street Journal*¹⁷ subsequently conducted its own survey, which among 746 persons found 57 leaky valves (8%). Subsequent scientific studies confirmed an even lower prevalence and also concluded that most cases were mild, with a large majority of confirmed cases having an exposure duration > 3 months.

Several methodologic lessons were learned: (1) defined cohorts, including unexposed persons, are more reliable sources of data than case series, (2) random sampling is preferable to self selection, (3) standardisation (explicit diagnostic criteria) trumps non-standardisation, and (4) blinded readings are superior to unblinded readings. Adjustment for “background noise” is another important consideration. Routine clinical echocardiograms are rarely of the highest scientific quality and should not be relied on for estimation of prevalence rates.

To whom do the research findings apply? Key points

- ▶ Design considerations tend to limit the broad applicability of findings from randomised clinical trials
- ▶ Trials of new interventions are typically designed to optimise the benefit-versus-harm balance
- ▶ The application of research findings to individual patients in clinical practice often requires leaps of faith, some being reasonable, others less so
- ▶ Reliance on surrogate markers in lieu of health outcomes can be misleading
- ▶ Poorly standardised clinical data from selected case series may be highly biased

574

ACKNOWLEDGEMENTS

The cartoons appearing in this article are reproduced from *All that glitters is not gold: what clinicians need to know about clinical trials*, by B Furberg and C Furberg, with permission.

REFERENCES

- 1 **Friedman LM, Furberg CD, DeMets DL.** *Fundamentals of clinical trials*, 3rd ed. St. Louis: Mosby, 1996; New York: Springer-Verlag, 1998.
 - ▶ **One of the leading texts on clinical trials methodology.**
- 2 **The Coronary Drug Project Research Group.** Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project. *N Engl J Med* 1980;**303**:1038–41.
 - ▶ **Classic paper showing that poor compliers to placebo had a worse survival experience than good compliers.**
- 3 **CASS Principal Investigators.** Coronary artery surgery study (CASS): a randomized trial of coronary artery bypass surgery. Comparability of entry characteristics and survival in randomized patients and non-randomized patients meeting randomization criteria. *J Am Coll Cardiol* 1984;**3**:114–28.
- 4 **Smith P, Arnesen H.** Mortality in non-consenters in a post-myocardial infarction trial. *J Intern Med* 1990;**228**:253–6.
- 5 **Guyatt GH, Keller JL, Jaeschke R, et al.** The n-of-1 randomized controlled trial: clinical usefulness. Our three-year experience. *Ann Intern Med* 1990;**112**:293–9.
- 6 **Robins SJ, Collins D, Wittes J, et al.** Relation of gemfibrozil treatment and lipid levels with major coronary events. VA-HIT: a randomized controlled trial. *JAMA* 2001;**285**:1585–91.
 - ▶ **Examined how much of the mortality/morbidity benefit of gemfibrozil could be explained by individual changes in HDL cholesterol. Only one quarter of the benefit was explained, thus concluding that gemfibrozil has important HDL cholesterol independent mechanism(s) of action.**
- 7 **Echt DS, Liebson PR, Mitchell LB, et al.** Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The cardiac arrhythmia suppression trial. *N Engl J Med* 1991;**324**:781–8.
 - ▶ **CAST surprised the cardiology community by showing that drug induced suppression of premature ventricular depolarisations increased rather than decreased the risk of sudden death. The findings were a setback for believers in surrogate markers.**
- 8 **Furberg CD, Psaty BM, Pahor M, et al.** Clinical implications of recent findings from the antihypertensive and lipid-lowering treatment to prevent hearth attack trial (ALLHAT) and other studies of hypertension. *Ann Intern Med* 2001;**135**:1074–8.
 - ▶ **A recent review of comparative trials strongly suggesting that it matters how raised blood pressure is lowered.**
- 9 **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.
 - ▶ **A thoughtful review of the central role clinicians have in applying trial results to individual patients.**
- 10 **Chalmers I.** A patient's attitude to the use of research evidence for guiding individual choices and decisions in health care. *Clinical Risk* 2000;**6**:227–30.
 - ▶ **An interesting commentary from the perspective of the patient.**
- 11 **Kääriäinen I, Sipponen P, Siurala M.** What fraction of hospital ulcer patients is eligible for prospective drug trials? *Scand J Gastroenterol* 1991;**26**:73–6.
 - ▶ **One of the few studies that documented how trial eligibility criteria can undermine generalisability.**
- 12 **Schmucker DL, Vesell ES.** Underrepresentation of women in clinical drug trials. *Clin Pharmacol Ther* 1993;**54**:11–15.
- 13 **El-Sadr W, Capps L.** The challenge of minority recruitment in clinical trials for AIDS. *JAMA* 1992;**267**:954–7.
- 14 **Hutchins LF, Unger JM, Crowley JJ, et al.** Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med* 1999;**341**:2061–7.
- 15 **Connolly HM, Crary JL, McGoan MD, et al.** Valvular heart disease associated with fenfluramine-phentermine. *N Engl J Med* 1997;**337**:581–8.
- 16 **US Food and Drug Administration.** FDA Public Health Advisory. Reports of valvular heart disease in patients receiving concomitant fenfluramine and phentermine. *FDA Medical Bulletin* 8 July 1997.
 - ▶ **A premature alarm overstating the magnitude of the fen-phen epidemic.**
- 17 **Anon.** WSJ Survey, *The Wall Street Journal*. 31 October 1997.
 - ▶ **An example of balanced investigative reporting.**