

Circulation

JOURNAL OF THE AMERICAN HEART ASSOCIATION



Lessons Learned From Recent Cardiovascular Clinical Trials: Part I

David L. DeMets and Robert M. Califf

Circulation 2002;106;746-751

DOI: 10.1161/01.CIR.0000023219.51483.66

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214
Copyright © 2002 American Heart Association. All rights reserved. Print ISSN: 0009-7322. Online
ISSN: 1524-4539

The online version of this article, along with updated information and services, is
located on the World Wide Web at:

<http://circ.ahajournals.org/cgi/content/full/106/6/746>

Subscriptions: Information about subscribing to *Circulation* is online at
<http://circ.ahajournals.org/subscriptions/>

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters
Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax:
410-528-8550. E-mail:
journalpermissions@lww.com

Reprints: Information about reprints can be found online at
<http://www.lww.com/reprints>

Lessons Learned From Recent Cardiovascular Clinical Trials: Part I

David L. DeMets, PhD; Robert M. Califf, MD

This is the first installment in a 4-part series of articles about our experiences with cardiovascular clinical research during the past 2 decades. From our vantage points as statistical and clinical researchers, we derive a series of lessons from recent clinical trials and incorporate them into principles designed to guide the practicing clinician who cares for patients with heart diseases.

We are entering an era in which the imperative to understand the rational basis for diagnostic and therapeutic options has become a major force in medical care. Medical products (drugs, devices, and biologics) are proliferating simultaneously with a substantial restructuring of the delivery of health care, with a focus on evidence to support medical interventions. As the texture of evidence to support clinical practice becomes clearer, the imperative to investigate the effectiveness of behavioral interventions becomes more important. When coupled with concern about the rising cost of medical care, this proliferation of technological and behavioral intervention reinforces the view that we cannot afford to offer all possible diagnostic and therapeutic options to all patients. Recent publicity about therapies that were in widespread use and were later found to be detrimental¹⁻³ has fueled this concern to extend well beyond the cost implications. The advent of genomics, proteomics, combinatorial chemistry, and advanced biomedical engineering will increase the importance of developing rational evidence about therapeutic risks and benefits. This effort to develop a rational basis for diagnostic and therapeutic decisions on the basis of quantitative assessments is a major component of evidence-based medicine.

The randomized clinical trial (RCT) has emerged as the principal research tool for developing evidence to inform and influence clinical practice, particularly in the cardiovascular field. The practice of clinical trials is built on an understanding of the biology of disease, the structure of medical care delivery, statistical design and analysis, data management, clinical research quality, and scientific and medical integrity. As clinical trials continue to evolve, investigators will be challenged to find the optimum balance among these factors that best evaluates new interventions and places older ap-

proaches in the proper context as new technology and practice patterns emerge.

Because the RCT is relatively new as a scientific technique, little empirical experience could be used to test this balance in the past. However, the recent explosive growth of RCTs with their widely discussed methodological approaches has been valuable in guiding current RCT design and deployment.^{4,5} In the first 2 parts of this series of articles, we will review some of the lessons we have learned from recent cardiovascular trials about the conduct of trials. In the second two parts, we will examine proposed implications of these lessons for medical practice and focus on applying the concepts in clinical trials to patient care decisions. These concepts represent accumulated wisdom from many collaborators and colleagues over the past several decades.

Surrogate Outcome Measures

More diagnostic and therapeutic approaches are under consideration than we can afford to evaluate, given the current time consumption and expensive nature of clinical trials. Any efficiency can be of tremendous value to patients, investigators, and sponsors. For example, in the ideal world, the choice of one approach for diagnosis or treatment over another would be based on anticipated benefits in longevity, prevention of events, and improvement in quality of life at a reasonable cost. Given the large number of issues to be addressed, however, the use of a substitute or "surrogate" outcome measure in place of these clinical outcomes could allow shorter, less costly trials with fewer subjects.

Surrogates are required for early-phase research because deductive, pathophysiological reasoning must be applied when developing new therapies. Surrogates also provide a potential means of monitoring the effect of therapy in an individual patient. Lowering blood pressure in a population of hypertensive patients is an example of a potential surrogate for the clinical outcome of death, myocardial infarction (MI), or stroke.

Composite outcomes are also used to evaluate treatments. A composite outcome may be composed of clinical events or measures such as survival or disease-free survival, and nonclinical outcomes such as change in blood pressure or CD4 cell counts by a specified amount in the treatment of

From the Duke Clinical Research Institute (R.M.C.), Duke University Medical Center, Durham, NC, and the Department of Biostatistics and Medical Informatics (D.L.D.), University of Wisconsin, Madison, Wis.

This is Part I of a 4-part article. Parts II, III, and IV will appear in the August 13, August 20, and August 27, 2002, issues of *Circulation*, respectively. Correspondence to Robert M. Califf, MD, Duke Clinical Research Institute, PO Box 17969, Durham, NC 27715.

(*Circulation*. 2002;106:746-751.)

© 2002 American Heart Association, Inc.

Circulation is available at <http://www.circulationaha.org>

DOI: 10.1161/01.CIR.0000023219.51483.66

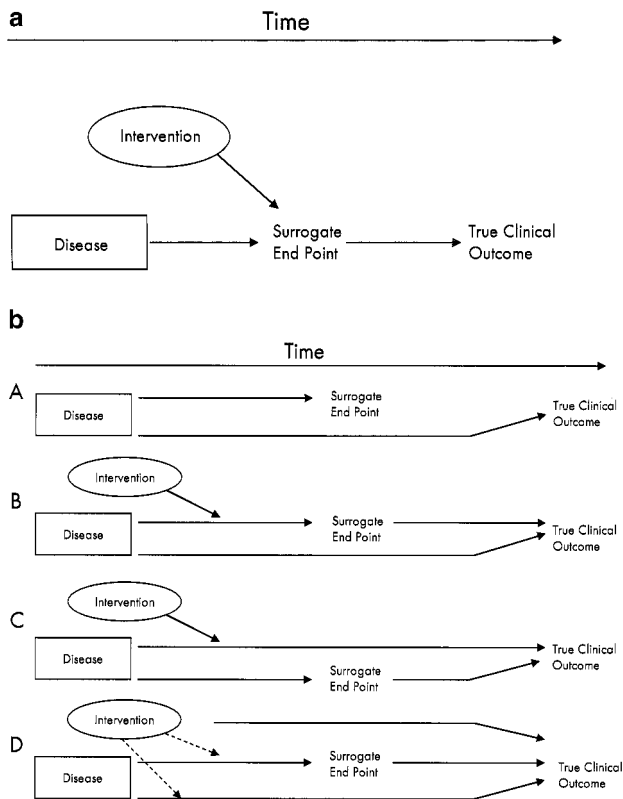


Figure 1. a, The setting that provides the greatest potential for the surrogate end point to be valid. b, Reasons for failure of surrogate end points. The surrogate is not in the causal pathway of the disease process (A). Of several causal pathways of disease, the intervention affects only the pathway mediated through the surrogate (B). The surrogate is not in the pathway of the intervention's effect or is insensitive to its effect (C). The intervention has mechanisms for action independent of the disease process (D). Dotted lines indicate mechanisms of action that might exist. Reprinted with permission from Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996;125:605–613.⁷

HIV-positive patients. The validity of a composite depends on the validity of its individual components.

Unfortunately, potential surrogates have been inadequately validated in most cases. A valid surrogate outcome must fulfill at least 2 statistical and logical criteria.⁶ First, changes in the surrogate must be predictive of the relevant clinical outcome. The second, and more critical requirement, is that a potential surrogate must fully (or nearly so) capture the effect of the intervention on the clinical outcome (Figure 1a).

This second requirement can be difficult to fully appreciate. In considering surrogate outcomes, investigators might think of a simple cause-and-effect pathway for the treatment. If this were the true biological process, then measuring the surrogate would measure all of the treatment effect and would thus be adequate. Many, if not most, treatments have several effect pathways, however, as depicted in Figure 1b. The treatment may alter the clinical outcome in 2 pathways, although the surrogate measures only 1 of them; drugs, for example, may have multiple binding sites. Alternatively, a treatment may affect the clinical outcome on a pathway that is not related to the pathway the surrogate is measuring. In

these cases, measuring only the proposed surrogate does not reflect the ultimate clinical impact of the therapy. In addition, the concept behind the proposed surrogate may be correct, but the measure of the concept may be flawed. However, the proposed surrogate is a specific measure of the concept and if the measure is flawed, the proposed surrogate may be invalid. In the past, these criteria, especially the second, have not been easily met,^{7,8} and proposed surrogate outcome measures in cardiovascular trials have been extremely disappointing.

Perhaps the most dramatic failure of a proposed surrogate outcome was demonstrated in the Cardiac Arrhythmia Suppression Trial (CAST).⁹ This trial tested whether 3 drugs in a class of antiarrhythmic drugs prevented sudden deaths in patients at risk because of premature ventricular contractions. Ventricular arrhythmias have been repeatedly found to be a risk factor for sudden death and total cardiovascular death. As a result, a class of drugs was developed that suppressed arrhythmias, and these drugs were approved for this indication by regulatory agencies. The cardiovascular community believed that because ventricular arrhythmias were predictive of sudden death and because these drugs could suppress ventricular arrhythmias, the drugs would reduce sudden death and total cardiovascular death.

CAST tested encainide, flecainide, and moricizine in patients with documented arrhythmias. Patients first had to go through a prerandomization “run-in” period that determined whether they had suppressible arrhythmias. If so, they were then randomly assigned, either to one of the three drugs or to a matching placebo. The primary outcome was sudden death, and the secondary outcome was death from any cause. The results were surprising.^{9,10} Very early in CAST, 2 arms (encainide and flecainide) were terminated because of a highly significant increase in both sudden death and all-cause mortality for patients on active treatment. The third arm was later stopped because of an increase in mortality.¹⁰

In another example, improved cardiac function (as estimated by cardiac output) was named as a surrogate in patients with chronic heart failure. Several inotropic drugs were developed that improved various heart function measures, such as cardiac output, and these were later tested in a series of trials.^{11–14} Despite the demonstration that these drugs would improve cardiac output in the short term, many of them increased mortality. Three trials were terminated early because of the unethical nature of continuing an experimental treatment that increased the risk of mortality. Thus, despite being predictive of patient survival, improved cardiac output was not a valid surrogate for evaluating this new class of drugs.

Even the previously accepted surrogate of coronary perfusion status for coronary artery reperfusion strategies has recently been debated.⁷ Many agents with superior efficacy for achieving TIMI grade 3 flow have not proven to be superior in reducing mortality.¹⁵ Although more complex measures such as echocardiographic contrast¹⁶ or myocardial blush score,¹⁷ or simple measures such as the ST-segment resolution on the ECG may correlate even more closely with outcome, they have not been validated as surrogates, and in the recently completed Global Use of Strategies To Open occluded coronary arteries V in Acute Myocardial Infarction

(GUSTO V AMI) trial, they predicted an effect on mortality that did not occur in the trial.¹⁸ These examples offer sound evidence that clinical trials intended to inform clinical practice must not rely on unproven surrogates and that strict criteria should be met before a nonclinical outcome can be accepted as a valid surrogate.

Until recently, blood pressure and low-density lipoprotein (LDL) cholesterol levels were considered valid surrogates for predicting the effect of treatment on outcome, but the results of the Antihypertensive and Lipid Lowering treatment to Prevent Heart Attack Trial (ALLHAT) point out the problems with this assumption.¹⁹ Although the trial is not complete, the Data and Safety Monitoring Board recommended discontinuing the use of doxazosin, an α -adrenergic blocking agent, because of a sharp increase in heart failure in the treated patients relative to the reference arm of the study (chlorthalidone). According to the investigators, this happened despite the similar efficacy of doxazosin and chlorthalidone on blood pressure and well-documented improvements in lipids and glucose control with doxazosin relative to chlorthalidone. Similar data have recently been published about stroke in a comparison of the angiotensin II receptor–blocker, losartan, with the β -blocker, atenolol.²⁰ Recent indirect comparisons²¹ have called into question whether different agents with the same effect on LDL cholesterol would have an equal effect on reducing clinical events. Furthermore, the recent product recall of cerivastatin resulted from a known toxicity (rhabdomyolysis) in the absence of quantification of benefit. The drug had been marketed on the basis of the demonstration of a reduction in LDL cholesterol superior to other “statins” without any evidence of superior prevention of vascular events. Most recently, the Heart Protection Study²² found no relationship between baseline cholesterol values and benefit of simvastatin in preventing cardiac events, and the Air Force/Texas Coronary Atherosclerosis Prevention Study²³ study found a benefit of pravastatin in patients with LDL cholesterol <100 mg/dL and elevated C-reactive protein level, providing further evidence that the benefits of statins may not be mediated solely by lipid levels. Thus, it seems highly unlikely that biomarkers will ever be able to predict all possible toxicities of a systemically administered drug.

Because of these experiences, it is preferable to use the term *physiological measure* to describe biological measures that give insight into disease progression.²⁴ Biomarkers proposed as substitutes for tangible human outcomes should be evaluated rigorously, and this evaluation should employ modern statistical assessment in the context of current pathophysiological thinking. Trialists must distinguish between the correlation of a biomarker with outcome and the causation of a relevant clinical outcome difference. Furthermore, even if a proposed surrogate was determined to be valid for one class of treatments, it may be invalid for another class of treatments, and a surrogate may not accurately predict outcome for another drug or biologic in the same general class. Relying on nonvalidated surrogates only encourages the use of ineffective therapies and may even promote the use of harmful treatments.

Composite Clinical End Points

For many therapies, the goal is not to reduce mortality but to prevent nonfatal events or improve symptoms. In many of these cases, a variety of negative outcomes are important, such as nonfatal MI, stroke, and unstable angina. Furthermore, it is important with a new therapy to be relatively certain that excess mortality is not being caused in the pursuit of fewer symptoms. An approach to account for these goals is to use a composite end point that includes all relevant events. If all major relevant end points are included, a valid conclusion from a positive trial with a composite of death and nonfatal MI or death and admission for heart failure would be that the superior treatment gave the patients a greater chance of being free of an event over the course of follow-up. This type of end point is highly valuable and has resulted in widespread acceptance of therapies for heart failure and acute coronary syndromes. The interpretation becomes problematic when elements of the composite go in opposite directions (the treatment reduces death, whereas nonfatal MI is increased). Furthermore, the common approach to composite end points counts each end point with equal weight, which is counterintuitive. Although the development of weighting schemes for end points would seem to be a good approach to this problem,²⁵ clinicians typically have a difficult time agreeing on the appropriate weights.

Subgroups and Treatment Interactions

Clinicians do not treat groups of patients; rather they help individual patients who have complex combinations of characteristics. It seems intuitively correct, therefore, that dividing a trial's results into multiple smaller units of analysis on the basis of patient characteristics would offer clinicians insight into treating their individual patients. Mostly for this reason, it is common in large-scale multicenter trials to evaluate the treatment effect in subgroups with the use of baseline variables. This evaluation can measure the consistency of the treatment effect across different risk groups and demographic groups and identify particular subgroups at excessive risk from treatment.

However, the evaluation of multiple subgroups can lead to nominally significant results by chance alone. To avoid spurious results as a result of multiple testing, clinical trialists prefer for subgroup analyses to be specified in advance and not performed as a result of conducting a large number of divisions of the cohort. However, safety analyses do not readily allow all potential risk groups to be identified in advance. Thus, subgroup analyses require careful interpretation. Despite the attractiveness of these subgroup analyses to many clinicians,²⁶ the best estimate of treatment effect to be expected from a patient treated outside the trial is still the overall estimate.²⁷ This empirical observation about subgroups has profound implications for genetic polymorphism analysis done to identify differential responses to treatment (pharmacogenomics). Such observations should be replicated in independent samples before they are accepted because of the high likelihood of spurious associations when multiple polymorphisms are tested.

When a trial demonstrates either a significant beneficial or a harmful treatment effect, the direction of this effect is

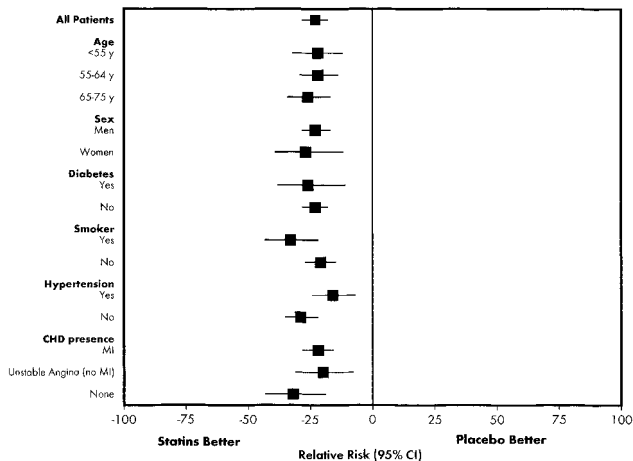


Figure 2. Overview of the treatment effect of one statin (prava-statin) on secondary prevention in patients with coronary disease. The figure demonstrates the absence of qualitative interactions over a large number of subgroups and the frequent occurrence of quantitative interactions. Adapted with permission from Sacks FM, Tonkin AM, Shepherd J, et al. Effect of prava-statin on coronary disease events in subgroups defined by coronary risk factors: the prospective pravastatin pooling project. *Circulation* 2000;102:1893–1900.²¹

usually consistent within subgroups. The size of the effect within any particular subgroup can vary from the overall effect. For example, in recent trials focusing on heart failure,^{28–31} coronary intervention,^{32–35} coronary reperfusion,^{36–39} and secondary prevention with statins (Figure 2),⁴⁰ the treatment effect across known risk factors was highly consistent in terms of direction and relative size. Any variations were well within the expected range, as compared with other trials with the same number of events. This same trend has been reported in trials with a harmful or negative effect, such as those evaluating inotropic drugs for heart failure^{11,13,14} and orally administered glycoprotein IIb/IIIa receptor antagonists.^{41–45}

Trialists have known for some time that when a trial demonstrates an overall effect, failure to have a statistically significant finding in an individual subgroup does not mean that the treatment is ineffective in patients with that particular characteristic. In studies of primary⁴⁶ or secondary⁴⁷ prevention with aspirin, few women were randomized, and the results in women were not statistically significant. Unfortunately, literal interpretation of this subgroup analysis led to the undertreatment of women for years,⁴⁸ although we now know that aspirin is effective in women.

Not all trials provide perfect consistency. When a particular subgroup suggests a different treatment effect from the overall effect, caution is needed in the interpretation. The Prospective Randomized Amlodipine Survival Evaluation (PRAISE I) trial⁴⁹ evaluated amlodipine, a calcium channel blocker, in the treatment of congestive heart failure. The randomization was stratified by ischemic and nonischemic pathogenesis because of concern that the treatment might not be effective in patients with nonischemic heart disease. Mortality plus hospitalization was the primary outcome, and mortality from all causes was the secondary outcome. Three thousand patients were randomly assigned to best available

care plus either drug or placebo. Overall results of the mortality outcome provided a nearly significant log rank test ($P=0.06$) favoring amlodipine. The treatment by pathogenesis subgroup interaction test was highly significant ($P=0.004$), indicating a hazard ratio of 1.0 in the ischemic subgroup and 0.6 in the nonischemic subgroup. Thus, the observed results were contrary to the prior belief that treatment would be more effective in patients with ischemic heart disease. With such a highly significant interaction test, statistical theory and practice suggested that the trial be interpreted separately for each subgroup. Accordingly, the standard interpretation for these results would judge amlodipine beneficial in the nonischemic subgroup.

The trial investigators and the sponsor were encouraged by the positive results in the nonischemic subgroup but were reluctant to endorse the treatment effect without further confirmation. Thus, the Prospective Randomized Amlodipine Survival Evaluation (PRAISE II) trial⁵⁰ was conducted in a nonischemic heart failure population. Although the design of PRAISE II was similar to PRAISE I (except that enrollment was limited to patients without coronary artery disease), the results proved to be different. The overall hazard ratio for PRAISE II was 1.0, showing no apparent treatment benefit for nonischemic patients. No differences in patient baseline characteristics or concomitant therapy between PRAISE I and PRAISE II could explain the discrepant results.

Whether the discrepancies were due to a chance finding or to a change in background therapy in the niche for amlodipine identified in PRAISE I is unknown. If the overall results of PRAISE I had produced a probability value <0.05 , a non-conservative interpretation might have recommended a positive benefit of amlodipine overall, or at least a positive benefit in the nonischemic subgroup. The results of PRAISE II remind us of the need to confirm results, especially when such results are not expected. Even if results are as expected, the most reliable estimate of a subgroup's results are still the overall results, not the estimate of the particular subgroup.

Trials should not be expected to provide absolute consistency, although further analyses can often provide insight into the plausibility of observed variations. For example, 3 trials recently investigated β -blocker therapy in patients with heart failure: the Metoprolol CR/XL Randomized Intervention Trial in congestive heart failure (MERIT),²⁸ the Cardiac Insufficiency Bisoprolol Study II (CIBIS-II),²⁹ and the Carvedilol Prospective Randomized Cumulative Survival (COPERNICUS) trial.³⁰ All 3 trials terminated early because of positive overall results on rates of mortality and mortality plus hospitalization. In keeping with the usual practice, prespecified subgroup analyses were conducted, and the results were extremely consistent.

However, one post hoc subgroup analysis in the MERIT trial suggested that the mortality results in the United States were not as impressive as the overall mortality results.⁵¹ The US results for the combined end point of mortality and hospitalization were consistent with the other countries and with the overall results,⁵² and the US subgroup results were consistent with the overall study results in the other 2 trials. Thus, it is highly unlikely that best available care in the United States diminishes the benefit of β -blockers in heart

failure patients. A similar variation across clinical sites was observed in the β -Blocker Heart Attack Trial (BHAT),⁵³ a trial that used β -blockers in patients surviving a recent MI. Overall results showed a highly significant reduction in total mortality, yet an increase in risk existed in a few sites.

The complexity of this lesson is exemplified by the difficulty of guaranteeing adequate participation of underrepresented minorities in clinical trials. Although a laudable goal, it is clear that in a clinical trial with a sample size designed to detect a clinically meaningful overall effect, there is usually not sufficient minority participation to ensure detection of a heterogeneous treatment effect as a function of race or ethnic background. When differences are found, as in the recent trials of β -blocker and ACE inhibitor treatment for heart failure,^{54,55} they can serve only as hypothesis-generating experiences to stimulate further research.

The lesson for clinical trialists here is that a trial is typically designed to detect an effect in the whole population. Subgroup findings should be regarded with suspicion unless they are independently confirmed or expected on the basis of prior findings. In retrospect, almost any subgroup finding can be justified on the basis of a theory, but failing to recognize the capriciousness of random variation often leads to premature acceptance of the results, risking the adoption of inferior or unnecessarily costly treatments for patients.

The next article in this series will conclude the lessons we have learned about the conduct of clinical trials on the basis of recent cardiovascular clinical research. The final 2 articles will translate those lessons into principles to assist the clinician in applying the results of well-done clinical trials to the care of individual patients.

References

- Smalley W, Shatin D, Wysowski DK, et al. Contraindicated use of cisapride: impact of food and drug administration regulatory action. *JAMA*. 2000;284:3036–3039.
- Mullins M, Horowitz BZ, Linden DHJ, et al. Life-threatening interaction of mibefradil and beta-blockers with dihydropyridine calcium channel blockers. *JAMA*. 1998;280:157–158.
- St Jude Medical, Inc. St Jude Medical recalling heart valve products with silzone coating. St Jude Medical News Room/Press Releases [web site]. January 24, 2000. Press release available at: http://www.corporate-ir.net/ireye/ir_site.zhtml?ticker=STJ&script=460&layout=-6&item_id=228634. Accessed July 12, 2002.
- Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet*. 2001;357:373–380.
- Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med*. 1984;3:409–422.
- Prentice RL. Surrogate end points in clinical trials: definition and operational criteria. *Stat Med*. 1989;8:431–440.
- Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996;125:605–613.
- Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA*. 1999;282:790–795.
- 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;321:406–412.
- The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med*. 1992;327:227–233.
- Packer M, Carver JR, Rodeheffer RJ, et al. Effect of oral milrinone on mortality in severe chronic heart failure. The PROMISE Study Research Group. *N Engl J Med*. 1991;325:1468–1475.
- Moe GW, Rouleau JL, Charbonneau L, et al. Neurohumoral activation in severe heart failure: relations to patient death and the effect of treatment with flosequinan. *Am Heart J*. 2000;139:587–595.
- Feldman AM, Bristow MR, Parmley WW, et al. Effects of vesnarinone on morbidity and mortality in patients with heart failure. Vesnarinone Study Group. *N Engl J Med*. 1993;329:149–155.
- Cohn JN, Goldstein SO, Greenberg BH, et al. A dose-dependent increase in mortality with vesnarinone among patients with severe heart failure. Vesnarinone Trial Investigators. *N Engl J Med*. 1998;339:1810–1816.
- Fleming TR. Surrogate end points in cardiovascular disease trials. *Am Heart J*. 2000;139:S193–S196.
- de Marchi SF, Schwerzmann M, Fleisch M, et al. Quantitative contrast echocardiographic assessment of collateral derived myocardial perfusion during elective coronary angioplasty. *Heart*. 2001;86:324–329.
- Vanhof AWJ, Liem A, Suryapranata H, et al. Angiographic assessment of myocardial reperfusion in patients treated with primary angioplasty for acute myocardial infarction: myocardial blush grade. *Circulation*. 1998;97:2302–2306.
- The GUSTO V Investigators. Reperfusion therapy for acute myocardial infarction with fibrinolytic therapy or combination reduced fibrinolytic therapy and platelet glycoprotein IIb/IIIa inhibition: the GUSTO V randomised trial. *Lancet*. 2001;357:1905–1914.
- Whelton PK, Williamson JD, Louis GT, et al. Experimental approaches to determining the choice of first-step therapy for patients with hypertension. The ALLHAT Research Group. *Clin Exp Hypertens*. 1996;18:569–579.
- Dahlof B, Devereux RB, Kjeldsen SE, et al. Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. *Lancet*. 2002;359:995–1003.
- Sacks FM, Tonkin AM, Shepherd J, et al. Effect of pravastatin on coronary disease events in subgroups defined by coronary risk factors: the prospective pravastatin pooling project. *Circulation*. 2000;102:1893–1900.
- Biggest ever statin study will change practice worldwide. Heartwire [web site]. November 13, 2001. Available at: <http://www.theheart.org>. Accessed July 12, 2002.
- De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate end points in clinical trials: summary of a National Institutes of Health workshop. *Control Clin Trials*. 2001;22:485–502.
- Downs JR, Clearfield M, Weis S, et al. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. Air Force/Texas Coronary Atherosclerosis Prevention Study. *JAMA*. 1998;279:1615–1622.
- Califf RM, Woodlief LH. Pragmatic and mechanistic trials. *Eur Heart J*. 1997;18:367–370.
- Julian DG. Treatment for survivors of acute myocardial infarction: what have we learned from large intervention trials? *Cardiovasc Drugs Ther*. 1995;9:495–502.
- Sleight P. Subgroup analysis in clinical trials: fun to look at—but don't believe them. *Curr Control Trials Cardiovasc Med*. 2001;1:25–27. Available at: <http://cvm.controlled-trials.com/content/1/1/025>. Accessed July 1, 2002.
- The International Steering Committee. Rationale, design, and organization of the Metoprolol CR/XL Randomized Intervention Trial in Heart Failure (MERIT-HF). *Am J Cardiol*. 1997;80:54J–58J.
- The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II). a randomised trial. *Lancet*. 1999;353:9–13.
- Packer M, Coats AJS, Fowler MB, et al. Effect of carvedilol on survival in severe chronic heart failure. *N Engl J Med*. 2001;344:1651–1658.
- Flather MD, Yusuf S, Kober L, et al. Long-term ACE-inhibitor therapy in patients with heart failure or left-ventricular dysfunction: a systematic overview of data from individual patients. *Lancet*. 2000;355:1575–1581.
- The EPISTENT Investigators. Randomised placebo-controlled and balloon-angioplasty-controlled trial to assess safety of coronary stenting with use of platelet glycoprotein-IIb/IIIa blockade. Evaluation of Platelet IIb/IIIa Inhibitor for Stenting. *Lancet*. 1998;352:87–92.
- The ESPRIT Investigators. Novel dosing regimen of eptifibatid in planned coronary stent implantation (ESPRIT): a randomised, placebo-controlled trial. Enhanced Suppression of the Platelet IIb/IIIa Receptor with Integrilin Therapy. *Lancet*. 2000;356:2037–2044.
- Slota PA, Fischman DL, Savage MP, et al. Frequency and outcome of coronary artery aneurysm after intracoronary stent placement and angioplasty. STRESS Trial Investigators. *Am J Cardiol*. 1997;79:1104–1106.

35. Serruys PW, de Jaegere P, Kiemeneij F, et al. A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. BENESTENT Study Group. *N Engl J Med.* 1994;331:489–495.
36. The GUSTO investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med.* 1993;329:673–682.
37. The Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO) IIa Investigators. Randomized trial of intravenous heparin versus recombinant hirudin for acute coronary syndromes. *Circulation.* 1994;90:1631–1637.
38. The Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO III) Investigators. A comparison of reteplase with alteplase for acute myocardial infarction. *N Engl J Med.* 1997;337:1118–1123.
39. Assessment of the Safety, and Efficacy of a New Thrombolytic Investigators. Single-bolus tenecteplase compared with front-loaded alteplase in acute myocardial infarction: the ASSENT-2 double-blind randomised trial. *Lancet.* 1999;354:716–722.
40. Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet.* 1994;344:1382–1389.
41. PRISM Study Investigators. A comparison of aspirin plus tirofiban with aspirin plus heparin for unstable angina. *N Engl J Med.* 1998;338:1498–1505.
42. PRISM-PLUS Study Investigators. Inhibition of the platelet glycoprotein IIb/IIIa receptor with tirofiban in unstable angina and non-Q-wave myocardial infarction. *N Engl J Med.* 1998;338:1488–1497.
43. The PARAGON Investigators. International, randomized, controlled trial of lamifiban (a platelet glycoprotein IIb/IIIa inhibitor), heparin, or both in unstable angina. *Circulation.* 1998;9:2386–2395.
44. The PURSUIT Trial Investigators. Inhibition of platelet glycoprotein IIb/IIIa with eptifibatid in patients with acute coronary syndromes. Platelet glycoprotein IIb/IIIa in Unstable angina: Receptor Suppression Using Integrilin Therapy. *N Engl J Med.* 1998;339:436–443.
45. Newby LK, McGuire DK. Oral platelet glycoprotein IIb/IIIa inhibition. *Curr Cardiol Rep.* 2000;2:372–377.
46. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med.* 1989;321:129–135.
47. Second International Study of Infarct Survival Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both or neither among 17 187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet.* 1988;2:349–360.
48. Shaw LJ, Miller DD, Romeis JC, et al. Gender differences in the non-invasive evaluation and management of patients with suspected coronary artery disease. *Ann Intern Med.* 1994;120:559–566.
49. Packer M, O'Connor CM, Ghali JK, et al. Effect of amlodipine on morbidity and mortality in severe chronic heart failure. Prospective Randomized Amlodipine Survival Evaluation Study Group. *N Engl J Med.* 1996;335:1107–1114.
50. Carson PA, O'Connor CM, Miller AB, et al. Circadian rhythm and sudden death in heart failure: results from Prospective Randomized Amlodipine Survival Trial. *J Am Coll Cardiol.* 2000;36:541–546.
51. AstraZeneca Pharmaceuticals LP. FDA approves AstraZeneca's Toprol-XL (metoprolol succinate) for heart failure. February 6, 2001. AstraZeneca US Gateway News Room [web site]. Available at: <http://www.astrazeneca-us.com/news/archive.asp>. Accessed July 12, 2002.
52. Hjalmarson A, Goldstein S, Fagerberg B, et al. Effects of controlled-release metoprolol on total mortality, hospitalizations, and well-being in patients with heart failure: the Metoprolol CR/XL Randomized Intervention Trial in congestive heart failure (MERIT-HF). *JAMA.* 2000;283:1295–1302.
53. Eichhorn E, Domanski M, Krause-Steinrauf H, et al. A trial of the beta-blocker bucindolol in patients with advanced chronic heart failure. *N Engl J Med.* 2001;344:1659–1667.
54. Eichhorn EJ. Experience with beta-blockers in heart failure mortality trials. *Clin Cardiol.* 1999;22:V21–V29.
55. Yancy CW, Fowler MB, Colucci WS, et al. Race and the response to adrenergic blockade with carvedilol in patients with chronic heart failure. *N Engl J Med.* 2001;344:1358–1365.

KEY WORDS: trials ■ cardiovascular diseases ■ therapy ■ outcome assessment ■ statistics