Hypertension and Outcomes Research
From Clinical Trials to Clinical Epidemiology
Bruce M. Psaty, David S. Siscovick, Noel S. Weiss, Thomas D. Koepsell, Frits R. Rosendaal, Danyu Lin, Susan R. Heckbert, Edward H. Wagner, and Curt D. Furberg

Outcomes research seeks to identify effective evidence-based methods of providing the best medical care. While randomized clinical trials (RCT) usually provide the clearest answers, they are often not done or not practicable. More than a decade after the introduction of calcium channel blockers and angiotensin converting enzyme (ACE) inhibitors, clinical trial data about their effect on major disease endpoints in patients with hypertension are still not available. The primary alternatives are the use of randomized trials that include surrogate endpoints, such as level of blood pressure or extent of carotid atherosclerosis, and the use of observational studies that include major disease endpoints. Both approaches, their strengths and limitations, are discussed in detail. The possibility of residual confounding limits the strength of inferences that can be drawn from observational studies. Similarly, the possibility of important drug effects, other than those involving the surrogate endpoint, limits the inferences that can be drawn from randomized trials that rely solely on surrogate outcomes as guides to therapy. In the absence of evidence from large clinical trials that include major disease endpoints, treatment decisions and guidelines need to synthesize the best available information from a variety of sources. Consistency of findings across various study designs, outcomes, and populations is critical to the practice of evidence-based medicine and the effort to maximize the health benefits of antihypertensive therapies. Am J Hypertens 1996;9:178-183

KEY WORDS: Hypertension, drug therapy, evaluation, randomized trials, observational studies, confounding variables, inference.

High blood pressure causes a variety of clinical events, including stroke, myocardial infarction, and congestive heart failure. The primary goal of antihypertensive therapy is to reduce the incidence of these major clinical cardiovascular complications of high blood pressure. Since all antihypertensive drugs have multiple effects, some of them unpredictable, therapeutic choices should be based ideally on the results of large randomized clinical trials that use the occurrence of these clinical events as their primary outcome. But so many drugs, formulations, and doses require evalu-
tion that it is not practicable to evaluate all of them in large clinical trials.

The two principal alternative strategies are the use of surrogate endpoints in small short-term randomized clinical trials and the use of major disease endpoints in observational studies. We consider both approaches in this article. In the first part, we inquire whether surrogate endpoints, such as level of blood pressure, in randomized clinical trials are valid proxies for major disease endpoints. In the second, we examine the conditions under which observational studies, such as case-control studies, can provide valid estimates of the risk or benefit of a therapy.

The purpose of this article is to provide an overview of the strengths, limitations, and implications of the evidence provided by studies that use various study designs and outcomes. The integration of information from these sources helps to enhance the quality of the clinical decision-making process and to assure the development of practice guidelines that rely on the best available evidence.

**SURROGATE ENDPOINTS IN CLINICAL TRIALS**

The logic of surrogate endpoints seems compelling: high blood pressure causes morbidity and mortality; a drug therapy reduces the level of blood pressure; thus the drug therapy reduces the risk of morbidity and mortality. Surrogate endpoints may include not only physiologic measures, such as blood pressure, but also measures of subclinical disease, such as left ventricular mass, or carotid or coronary atherosclerosis as assessed by echocardiography, ultrasonography, or angiography. The Treatment of Mild Hypertension Study, for example, is a 4-year prospective clinical trial that randomized participants to placebo or one of four antihypertensive agents, and the endpoints included not only measures of compliance and levels of blood pressure, but also electrocardiograms, ambulatory electrocardiograms, and echocardiography. Because these measures of subclinical disease are themselves powerful risk factors for the major clinical events, they are particularly appealing as surrogate endpoints. In general, clinical trials using these continuous surrogate endpoints can be much smaller and shorter than trials using major clinical events as the primary outcome. This economic advantage is worth pursuing, of course, only if surrogate endpoints serve as valid proxies for major disease endpoints in clinical trials of antihypertensive therapies.

In order to make a definitive treatment decision from a randomized clinical trial using a surrogate endpoint, a test of the null hypothesis of no treatment effect based on the surrogate endpoint should also be a valid test of the corresponding null hypothesis based on the true clinical endpoint. This correspondence will occur if the surrogate endpoint is predictive of the true endpoint and the surrogate fully captures the effect of treatment on the true clinical endpoint. Most surrogate endpoints are selected because they have been demonstrated to be strong predictors for the occurrence of the clinical outcome in observational studies—namely, because they satisfy the first condition. All antihypertensive therapies have multiple effects, some of which may not be mediated through a particular surrogate endpoint. Under these circumstances, the second condition is rarely satisfied. As a result, the use of evidence from surrogate endpoints to infer the effects on long-term clinical outcomes can produce highly misleading conclusions.

Though summarized only in conference reports, several recent randomized clinical trials in hypertension raise questions about the validity of surrogate endpoints in hypertension trials. In one clinical trial from the Evaluation Group of Long-Term Antihypertensive Treatment (CLANT) in Japan, hypertensive patients were randomized to a calcium channel blocker, either nifedipine or manidipine (n = 1017), or to the angiotensin-converting enzyme (ACE) inhibitor delapril (n = 1025), and followed for 1 year. Blood pressure reduction was greater in subjects taking the calcium channel blocker. Despite this effect on level of blood pressure, subjects treated with the calcium-channel blocker had a significantly higher incidence of cerebrovascular events than the patients randomized to treatment with an ACE inhibitor (risk ratio [RR] = 3.0, 95% confidence interval [CI] = 1.1 to 8.3, P = .02). These clinical trial data illustrate the point that because drugs have multiple effects, the use of blood pressure may not be adequate as a surrogate for the effect of antihypertensive therapies on major disease endpoints.

In the Multicenter Isradipine Diuretic Atherosclerosis Study (MIDAS), hypertensive patients with early carotid atherosclerosis were randomized to isradipine (2.5 to 5.0 mg twice a day) or hydrochlorothiazide (12.5 to 25.0 mg twice a day) and followed for 3 years with serial carotid ultrasound examinations. While the reduction in systolic blood pressure was larger among those randomized to hydrochlorothiazide, the reduction in diastolic blood pressure was almost identical in the two groups over 3 years. The primary outcome was subclinical atherosclerosis, and long-term progression was small and similar in both groups. Yet major vascular events were more common among those randomized to isradipine (25 [4.9%] of 442) than among those randomized to hydrochlorothiazide (14 [3.3%] of 44; RR = 1.78, 95% CI = 0.94 to 3.38; P = .07). Despite the fact that this calcium channel blocker is supposed to have antianginal properties, hospitalization for angina was almost four times more common among subjects randomized to isradipine.
(RR = 3.66; 95% CI = 1.02 to 13.0; \( P < .05 \)). In this randomized trial, while the effect on the surrogate endpoint was nil, low-dose diuretic therapy appeared to reduce the occurrence of major disease endpoints better than the calcium channel blocker.

The International Nifedipine Trial on Antiatherosclerotic Therapy (INTACT) randomized 425 subjects with mild coronary atherosclerosis to short-acting nifedipine or placebo.\(^7\) While participants in the study did not necessarily have high blood pressure, the primary outcome was a surrogate endpoint—the development of new coronary lesions over a 3 year period. Indeed, the use of nifedipine significantly reduced the average number of new coronary lesions per patient (0.82 for placebo vs 0.59 for nifedipine, \( P = .03 \)). But there were 12 deaths among subjects randomized to nifedipine compared to only two deaths among subjects randomized to placebo. The RR for mortality was 5.92 (95% CI = 1.34 to 26.2; \( P = .008 \)). In INTACT, the findings for subclinical and clinical endpoints were both statistically significant. While the subclinical surrogate endpoint favored nifedipine, one of the most clinically relevant endpoints, total mortality, favored placebo.

The most likely explanations for the disparity between the results for the surrogate and the major disease endpoints are 1) the fact that drugs have multiple effects, some beneficial and some potentially harmful; or 2) for a variety of reasons, either genetic or environmental, some patients may be particularly susceptible to certain drug effects compared to others. Because these studies were well-conducted randomized clinical trials, the findings for both the surrogate endpoint and the clinical endpoint are likely to be valid estimates of the drug effect on the respective outcomes. But since calcium channel blockers are powerful cardiovascular agents that have multiple effects, it is not possible to generalize, for instance, from a blood pressure lowering effect to an effect on the risk of coronary or cerebrovascular events. There are other recent examples of the failure of surrogate endpoints in cardiovascular medicine: 1) the use of suppression of premature ventricular contractions by antiarrhythmic agents as a proxy for the incidence of sudden death among postmyocardial infarction patients;\(^8,9\) 2) the use of cholesterol lowering by clofibrate as a proxy for total mortality;\(^10\) and 3) the use of oral milrinone to improve hemodynamics as a proxy for total mortality.\(^11,12\)

As proxies for major disease endpoints, the level of blood pressure in GLANT and the change in carotid atherosclerosis in MIDAS were clearly inadequate or partially misleading. While randomized controlled clinical trials are appropriately regarded as the gold standard of evidence, the use of clinical trials alone does not assure proper inference. The choice of outcome is critical. The findings from GLANT illustrate the limitations of drawing inferences from results based on the surrogate endpoints: interpretations based on level of blood pressure control and risk of cerebrovascular events would be diametrically opposed in this case. We might revise the syllogism with which we started this section in the following way: all drugs have multiple effects; a single surrogate endpoint is likely to measure only a subset of those effects; therefore, studies using surrogate endpoints may be inadequate or misleading as a basis for the choice of a therapy.

**MAJOR DISEASE ENDPOINTS**

While the choice of the outcome is obviously critical, studies that examine drug effects on major disease endpoints are often lengthy and expensive to conduct well. At the same time, the number of potentially evaluable antihypertensive agents is exceedingly large, especially if there are clinically important differences among the various formulations of specific drugs or among the various doses of those formulations. It is simply not practicable to evaluate all antihypertensive drugs, doses, and formulations in large endpoint trials. The newest agents also tend to be the least well evaluated.

One alternative to large long-term clinical trials is the use of metaanalysis to combine the results from multiple smaller short-term clinical trials. Metaanalysis is a tool for the quantitative review of existing data.\(^13-17\) In metaanalysis, the individual clinical trials serve as the unit of analysis, and the within-trial estimates of effect are preserved and summed across studies. Dose-response analyses are also possible.\(^15,16\) Metaanalysis is, however, subject to some of the same potential biases as observational studies. For instance, lack of information from published studies and incomplete reporting of events in published studies may introduce bias. A major drawback of metaanalysis is the inability to evaluate new therapies. Even combining the endpoint data from several small trials designed to assess the effect of a new drug or formulation on the level of blood pressure would in general lack adequate power to assess major disease endpoints.

The other major alternatives to large long-term randomized trials are observational studies, such as the cohort or case-control study.\(^19-21\) With these designs, it is possible to use major disease endpoints, such as stroke or myocardial infarction, as the outcome of interest. Compared with randomized clinical trials, they are efficient and relatively inexpensive. These study designs are also capable of evaluating the effects of current practice patterns on health.

In almost all respects, the use of observational studies to evaluate the efficacy or safety of therapies is similar to their use in the study of etiology. For case-control studies, the essential design features include, for instance, the complete ascertainment of cases from a de-
fined population, the use of controls who reflect the exposures sustained by the underlying population from which the cases arose, and the comparable assessment of exposures and covariates in cases and controls alike. The fundamental difference is simply that the validity of an observational study of a therapy depends on the ability to control not only for the traditional risk factors that are associated with the outcome of interest, but also for the clinical factors that may lead physicians to use particular therapies for particular patients. In other words, observational studies of the safety or efficacy of a therapy pose this one additional difficulty, which is the possibility of confounding by indication or contraindication.

For example, an antihypertensive agent, such as an ACE inhibitor, may be preferentially used in patients with diabetes mellitus. Since diabetes mellitus is also a risk factor for myocardial infarction, an observational study may find spuriously, due to confounding, that compared with diuretics, ACE inhibitors appear to increase the risk of myocardial infarction. Similarly, β-blockers are indicated for the treatment of angina as well as hypertension. Since angina is a risk factor for myocardial infarction, an observational study may find that compared with diuretics, the use of β-blockers appears to increase the risk of myocardial infarction. Observational studies of therapeutic efficacy that seek to control for confounding by indication often require a detailed knowledge of how the therapies under study are used in clinical practice and of how the use of those therapies may have changed over the time course of the study.

The approaches to handling confounding by indication—restriction, matching, stratification, and multivariate analysis—are the same as the approaches to handling other potential confounding factors in observational studies. In the ACE inhibitor example, it is important initially to stratify the analysis on the presence or absence of diabetes mellitus. If the association between drug use and myocardial infarction is similar in both strata, then it is reasonable to combine the estimates and adjust for diabetes mellitus. Within each stratum, a fairer comparison of the effects of the alternative drugs is possible. If the associations differ according to the presence or absence of diabetes mellitus, then it may be reasonable to present separate estimates for each group. An alternative is restriction—simply to exclude all subjects with any clinical evidence of diabetes mellitus from the analysis. In other words, confounding by indication can be recognized and dealt with by the same methods that we use to deal with confounding from more traditional sources, such as demographic factors.

For some therapies, it may be difficult or impossible to use observational studies to evaluate their efficacy and safety. Pentoxifylline, for example, is the only approved drug therapy for peripheral vascular disease. With surgery as the alternative therapy, pentoxifylline tends to be used in practice either in patients with mild disease or in patients whose surgical risk is too high. Under these circumstances, the choice of medical versus surgical therapy is strongly and complexly confounded by severity of disease, and the area of overlap, where either therapy is equally indicated, may be small. For hypertension, on the other hand, scores of comparable drugs are available, and variation in drug use depends in large part on physician practice style. These circumstances are precisely the ones in which a nonrandomized study of effectiveness can be expected to have the greatest validity.

Of course, nonexperimental studies may give misleading results. In several instances, including the study of hypertensive therapies, however, observational studies have provided results comparable to those of the randomized trials. In a metaanalysis of randomized clinical trials comparing β-blockers with diuretics in middle-aged adults, Collins and colleagues reported a slight advantage of β-blockers in the prevention of coronary heart disease. Similarly, in a case-control study, the use of β-blockers compared with other therapies, principally high-dose diuretics, was associated with a small reduction in the risk of coronary heart disease in patients with high blood pressure. These estimates of effect, 0.94 and 0.87, are similar.

The findings of observational studies can also complement those of randomized controlled clinical trials. In the large hypertension trials in middle-aged adults, high doses of diuretics, the equivalent of 90 to 100 mg of hydrochlorothiazide per day, were generally used, and the reduction in the incidence of coronary disease was less than expected from observational studies of the association with level of blood pressure. The recent trials in older adults often used low doses of diuretics, the equivalent of 12.5 to 25 mg of hydrochlorothiazide, sometimes in combination with potassium-sparing agents. For diuretics, coronary heart disease risk was reduced by 20%,25 which is within the range of the expected reduction of 20% to 25%. From the results of these clinical trials alone, it is not possible to tell whether the difference between the two generations of large clinical trials is the result of patient age or dose of the diuretic. The recent case-control study by Siscovick and colleagues suggests that compared with high-dose diuretic therapy, low-dose diuretic therapy, with or without potassium-sparing agents, was associated with a substantially reduced risk of primary cardiac arrest—a finding that was present and similar in magnitude in both the middle-aged and the older adult. Importantly, observational studies have the
The study of hypertension serves perhaps as a model of studies in which observational studies can provide results comparable to randomized clinical trials. In part because a large number of alternative therapies are available and commonly used, antihypertensive therapy is well suited to an outcome evaluation by observational methods. In several case-control studies, patient characteristics, such as smoking, diabetes mellitus, and cholesterol level, were only weakly associated with the choice of therapy in clinical practice. The high degree of similarity in treatment regimens between controls with and without various clinical characteristics not only minimizes the possibility of important confounding by those characteristics but also provides some assurance of the validity of the adjusted comparisons.

Like clinical trials, observational studies of therapeutic efficacy and safety must be conducted well. Even the best observational studies nonetheless have important limitations. First, there may be unknown or unmeasured confounding factors for which adjustment is not possible. In observational studies of drug therapies, these may include confounding by indication, in which the selection of antihypertensive treatment by physicians and patients may introduce bias. Secondly, measurement error in the assessment of the presence or the severity of exposures or covariates may result in bias or in incomplete adjustment and residual confounding. Large clinical trials, which are not subject to these limitations, are also important because they can assess the overall risk of benefit of a therapy in terms of a variety of important cardiovascular outcomes—not only a single case group, such as patients with myocardial infarction, but also for the other important outcomes of stroke, congestive heart failure, renal disease, and total mortality.

SYNTHESES AND CLINICAL IMPLICATIONS

Ideally, practitioners would like to base clinical decisions about antihypertensive therapy upon the results of randomized clinical trials that include major disease endpoints. Currently, approval by the Food and Drug Administration requires evidence of efficacy measured solely in terms of the effect of antihypertensive agents on a surrogate endpoint—the level of blood pressure. As a result, clinical trial data for the long-term safety and efficacy of the newer agents—calcium-channel blockers, β-blockers, and ACE inhibitors—are lacking in patients with high blood pressure. In the absence of information from major clinical trials, clinicians still need to make treatment decisions based upon the best available scientific evidence. Inferences from randomized trials that evaluate therapies in terms of their effects on blood pressure may be limited. On the other hand, it is difficult to exclude the possibility of confounding in observational studies. These limitations are complementary.

Treatment decisions and guidelines need to synthesize the best available information from a variety of sources. In the end, consistency of findings across various studies, designs, outcomes, and populations is critical to the conduct of evidence-based medicine and the effort to maximize the health benefits of antihypertensive therapies.

REFERENCES


