SPECIAL ARTICLE

Developing Improved Observational Methods for Evaluating Therapeutic Effectiveness

RALPH I. HORWITZ, M.D., CATHERINE M. VISCOLI, M.Phil., *New Haven, Connecticut*, JOHN D. CLEMENS, M.D., *Baltimore, Maryland*, ROBERT T. SADOCK, M.D., *New Haven, Connecticut*

Therapeutic efficacy is often studied with observational surveys of patients whose treatments were selected nonexperimentally. The results of these surveys are distrusted because of the fear that biased results occur in the absence of experimental principles, particularly randomization. The purpose of the current study was to develop and validate improved observational study designs by incorporating many of the design principles and patient assembly procedures of the randomized trial. The specific topic investigated was the prophylactic effectiveness of β-blocker therapy after an acute myocardial infarction.

To accomplish the research objective, three sets of data were compared. First, we developed a restricted cohort based on the eligibility criteria of the randomized clinical trial; second, we assembled an expanded cohort using the same design principles except for not restricting patient eligibility; and third, we used the data from the Beta Blocker Heart Attack Trial (BHAT), whose results served as the gold standard for comparison.

In this research, the treatment difference in death rates for the restricted cohort and the BHAT trial was nearly identical. In contrast, the expanded cohort had a larger treatment difference than was observed in the BHAT trial. We also noted the important and largely neglected role that eligibility criteria may play in ensuring the validity of treatment comparisons and study outcomes. The new methodologic strategies we developed may improve the quality of observational studies and may be useful in assessing the efficacy of the many medical/surgical therapies that cannot be tested with randomized clinical trials.

From the Departments of Medicine and Epidemiology, Yale University School of Medicine, New Haven, Connecticut, and the University of Maryland, Baltimore, Maryland.

Requests for reprints should be addressed to Ralph I. Horwitz, M.D., Yale Univesity School of Medicine, Room IE-61 SHM, P.O. Box 3333, New Haven, Connecticut 06510.

Manuscript submitted March 22, 1990, and accepted in revised form June 14, 1990.

Evidence that prophylactic β-blocker therapy reduces mortality in patients surviving an acute myocardial infarction comes from a research strategy, the randomized clinical trial, that has become the accepted scientific standard for evaluating therapeutic efficacy [1]. The outstanding feature of such experimental trials is the use of randomization to prevent the compared treatments from being assigned preferentially to groups of patients with particularly good or poor prognoses. This important function serves to create groups that are similar on known or suspected prognostic features. In addition to randomization, however, experimental trials are planned with several other essential methodologic strategies.

Two strategies involve the use of eligibility criteria to exclude patients (1) who are moribund or who have co-morbid conditions likely to hinder the ascertainment of endpoints and (2) who have strong indications for or contraindications to any of the compared treatments. A third strategy involves the analysis of data. The results are usually analyzed according to the treatment group assigned by randomization ("intention-to-treat") [2].

Despite these scientific advantages, experimental trials are often difficult or infeasible to conduct [3]. Many physicians or patients are reluctant to accept randomization as a chance mechanism for assigning treatment, particularly if one of the proposed regimens under comparison seems especially desirable or undesirable.

For these and other reasons, therapeutic efficacy is often studied with observational surveys of patients whose treatments were selected nonexperimentally, according to nonstandardized physician judgments. The results of these observational cohort surveys are often distrusted because of the fear that biased results are produced in the absence of experimental principles, particularly in the absence of randomization [4]. The possible biases of a nonrandomized treatment assignment can be reduced if the therapeutic results are compared within prognostically similar groups of patients. However, the demonstration of prognostic similarity may be quite difficult. Clinicians may disagree about the most cogent data to be used for demarcating prognostic strata. However, when they agree, the re-

quired data may not have been collected amid all the other information that competes for attention in a large-scale cohort study.

The combination of all of these problems leads to a dilemma in which the randomized trials that might provide convincing evidence are often not conducted, and the evidence from observational surveys is frequently unconvincing. The current research was undertaken to provide an escape from this dilemma. The goal of the study was to develop and test improved cohort methods. First, we sought to assess whether a restricted cohort design, using many of the methods of a randomized clinical trial, would achieve results similar to those of a randomized trial. Second, we intended to compare the results of the restricted cohort to the results of an expanded cohort designed according to customary epidemiologic methods. The expanded cohort would enable us to estimate the benefit of treatment in a broader group of patients than are enrolled in most randomized trials. The specific topic investigated was the prophylactic effectiveness of β-blocker therapy for reducing mortality after an acute myocardial infarction.

METHODS

Strategy of Research

To accomplish our research goal, we employed a strategy that required assembling three sets of comparative data. First, we developed and assessed an observational cohort method based on the principles of patient assembly and study design inherent in the randomized trial. Although randomization was not used to assign treatment, we retained many of the other cited principles of design and analysis that are used in conducting randomized clinical trials. We designated this cohort design the restricted cohort. Next, we compared the results of the restricted cohort to the results of an expanded cohort designed without adherence to these experimental principles of patient assembly. We designated this design the expanded cohort. Finally, we compared the results of both the expanded and restricted cohorts to the results of a completed and widely accepted randomized clinical trial. For this comparison, we chose the Beta Blocker Heart Attack Trial (BHAT). Before considering the detailed methods of the current study, we shall briefly describe the design of the BHAT trial.

BHAT Study Design

BHAT was a multicenter, randomized, double-blind clinical trial comparing propranolol and placebo in patients who survived an acute myocardial infarction. Nearly 4,000 men and women aged 30 to 69 were recruited at 134 hospitals and enrolled in the study between June 1978 and October 1980. Zero-time treatment was begun at randomization, while the patient was still in the hospital, 5 to 21 days after hospital admission. Patients returned for clinical evaluation every 3 months, with an average follow-up of all study participants of 25 months. A detailed description of the BHAT design and results has been published elsewhere [5].

TABLE I

Comparison of Expanded and Restricted Cohorts for Use of Randomized Trial Principles

Randomized Trial Principle	Expanded	Restricted
MI definition	Yes	Yes (same)
Zero-time	Yes	Yes (different)
Exclusion	No	Yes
Severity adjustment (clinical)	Yes	Yes

MI = myocardial infarction.

Current Study

OVERVIEW: The current study comprises two cohorts, the expanded and restricted. For both cohorts, patients were assembled who had an acute myocardial infarction and were admitted to Yale-New Haven Hospital between January 1, 1978, and December 31, 1982. (None of the patients were part of the original BHAT study). Data describing the patient's hospital course were obtained from medical records by data technicians unaware of the research hypothesis; survival data describing outcome were obtained from mailed questionnaires sent to patients and physicians, state vital statistics records, and the National Death Index. The primary source of vital status data for death was the National Death Index, which has been extensively validated [6–10].

In **Table I,** we compare the expanded and restricted cohort designs for their adherence to four randomized trial principles. The expanded cohort design uses standard and accepted clinical criteria to define a myocardial infarction and selects a zero time to classify a patient's use of β-blocker therapy (the choice of zero time, which will be considered in detail later, is different in the expanded and restricted cohorts). Most expanded cohorts do not follow the randomized trial principle of using clinical exclusion criteria to assemble the eligible cohort. However, many cohort studies, like randomized trials, do employ clinical and sociodemographic features to adjust for baseline differences in infarct severity.

The restricted cohort design also uses standard

myocardial infarction criteria, but unlike the expanded cohort, the zero time is chosen to approximate the point of randomization in a clinical trial. The restricted cohort is designed to use exclusion criteria to assemble the cohort and to employ clinical features to adjust for differences in infarct severity. Each of the methodologic principles included in the restricted cohort was specified explicitly before conducting the research. The specific tactics for incorporating each of these principles in the two cohorts is considered in the next section.

DEFINITION OF MYOCARDIAL INFARCTION: Both the expanded and the restricted cohorts comply with the first experimental design principle by employing the same diagnostic criteria for myocardial infarction that were used in the BHAT randomized clinical trial [4]. Eligible subjects were required to fulfill at least two of the following three criteria for the diagnosis of acute myocardial infarction: (1) central anterior chest pain of more than 15 minutes' duration, acute pulmonary edema, or cardiogenic shock; (2) characteristic electrocardiographic changes with development of pathologic Q waves or ST-segment elevation followed by T-wave inversion in at least two leads; or (3) serum enzyme elevations consistent with acute myocardial infarction (at least two values of serum glutamic oxaloacetic transaminase or creatinine phosphokinase exceeding twice the upper limits of normal).

CHOICE OF ZERO TIME: Each member of a cohort has a chronologic reference point at which the baseline clinical status is identified and characterized, the treatment category is determined, and the subsequent follow-up period is counted and monitored. In previous clinical trials of this topic, zero time was the time when randomization occurred and patients began taking either the β blocker or the placebo. For the current study, zero time is different in the two cohorts. In the expanded cohort, we chose to follow the practice of most investigators and defined zero time as the date of hospital discharge for all patients. In the restricted cohort, for patients not treated with a β blocker, zero time is the date of hospital discharge, as it was in the expanded cohort. To approximate as closely as possible the design of the randomized trial, we defined the zero time for treated patients in the restricted cohort as the date the β blocker was started. This choice of a zero time for treated patients creates a therapeutic "window" during which a β blocker may be given that begins with a patient's stabilization in the coronary care unit and ends with hospital discharge. This zero-time window corresponds best to the strategy employed in the BHAT trial, in which patients were eligible for the study only after their clinical condition had stabilized and between the 5th and 21st day of hospitalization.

PATIENT ELIGIBILITY AND EXCLUSION CRITERIA: Both the expanded and restricted cohorts exclude patients whose age was less than 30 or greater than 75 and who did not meet the diagnostic criteria for a myocardial infarction. However, only the restricted cohort excluded patients, as was also done in the BHAT study, if any of the following clinical features were present: (1) any contraindication to β-adrenergic blockade including congestive heart failure, severe intermittent claudication, or unstable insulin-dependent diabetes mellitus; (2) any strong need for treatment with a β-blocking agent, including idiopathic hypertrophic subaortic stenosis or severe incapacitating migraine headaches; (3) any patient in whom the qualifying myocardial infarction resulted from a nonatherosclerotic cause such as surgery, trauma, or shock; and (4) any condition likely to hinder or confuse follow-up or endpoint evaluation, such as malignant neoplasm or drug addiction.

PROGNOSTIC ADJUSTMENT OF BASELINE DIFFERENCES: A major scientific hazard in this study, as in all observational studies, is the possibility that the compared groups are substantially unequal in their initial susceptibility to the outcome. With this type of inequality, the subsequent difference in outcomes may be caused by "susceptibility bias," rather than by true differences in the compared treatments.

In the BHAT study, susceptibility bias was reduced or avoided by the use of randomization to assign the treatments. Since randomization is not employed in the current observational study, we have applied several alternative strategies.

One strategy is contained in the specification of the eligibility criteria. The eligibility criteria constrain the target group to only those subjects able to receive both treatments. In doing so, the eligibility criteria, which were only applied to the assembly of the restricted cohort, eliminate possible inequalities that occur when a particular treatment is withheld from patients with contraindications or assigned to patients with a particular therapeutic indication. Since these reasons for treatment decisions often carry strong prognostic implications (e.g., congestive heart failure), they can seriously distort any treatment comparisons.

After eligibility criteria are used to create a relatively homogeneous study population by reducing variability in the candidate population, randomization is employed in a clinical trial to prevent any further risk of susceptibility bias. To compensate for the absence of randomization in both the re-

stricted and expanded cohorts, we collected detailed clinical data describing prognostic features of the patients that enabled multivariate adjustments. The prognostic features used included prior myocardial infarction or angina; congestive heart failure on chest roentgenograph; Killip class 3 or 4 infarct severity in hospital; radionuclide ejection fraction of less than 40%; and more than 10 premature ventricular contractions per hour. In the data display, we use these risk groups to compare the baseline clinical characteristics of patients in each of the three cohorts.

These aforementioned prognostic features were also used to create the following risk groups: severe, including patients with a prior myocardial infarction; moderate, including patients with no prior myocardial infarction and presence of Killip class 3 or 4 infarct severity, ejection fraction less than 40%, or congestive heart failure on chest roentgenograph; and mild, including the remaining patients with none of the aforementioned. (These same risk groups were used for prognostic analyses in the BHAT trial.)

STATISTICAL ANALYSIS: The initial steps in data analysis included univariate and bivariate displays of the information that describes the clinical population at baseline and during the subsequent period of follow-up. In each of these analyses, three groups are compared: the BHAT trial cohort, with data assembled from the published reports of the randomized clinical trial [4]; the expanded cohort of the current study; and the restricted cohort of the current study that was assembled according to the principles of the randomized trial. Since complete 24- and 36-month follow-up was obtained for all members of the current study, mortality rates among these three groups are compared both at 24 and 36 months.

The mortality rates displayed for the BHAT trial at 24 and 36 months are estimates from the lifetable analyses included in their report. Since randomization was used to assign treatment category, and since no important differences were noted between treatment groups at baseline, the compared groups of the BHAT trial do not need to be adjusted for prognostic inequalities at baseline. (Adjustments carried out by the BHAT investigators did not change the trial results.) For the expanded and restricted cohorts, crude (or unadjusted) mortality rates are calculated directly from the 24- and 36-month follow-up data. Adjusted mortality rates were obtained by using multiple logistic regression, including as independent variables the prognostic features noted earlier.

Finally, the BHAT trial, like most randomized trials, employed the analytical strategy called "intention-to-treat." With this strategy, all analyses are conducted according to the original, randomly assigned treatment groups. For both the restricted and the expanded cohorts, we have also employed an approximation of the intention-to-treat principle. Thus, patients are retained in the treatment category determined at zero time, even if the treatment was changed (for example, β blocker was discontinued) after its initiation.

TABLE II

Assembly of Randomizable and Restricted Cohorts

Cohort	Number
Available (all patients, 1978 to 1982)	2,497
Excluded	1,314
•Wrong diagnosis (n = 528)	
•Age <30 or >75 (n = 444)	
•Missing records (n = 342)	
Candidate	1,183
•Expanded	1,059
(124 died before hospital discharge)	
•Restricted	622
(561 ineligible due to clinical indications or contraindications)	

RESULTS

Assembly of the Expanded and Restricted Cohorts

Table II indicates the assembly of the expanded and restricted cohorts of the current study. Available patients included 2,497 with a discharge diagnosis of myocardial infarction between 1978 and 1982. Excluded from both cohorts were 1,314 patients for the following reasons: did not fulfill the diagnostic criteria for a myocardial infarction, 528; age less than 30 or greater than or equal to 75, 444; or a medical record that could not be located, 342. Thus, 1,183 subjects remained. The expanded cohort of 1,059 subjects was formed after excluding 124 patients who died before hospital discharge, their zero time. The restricted cohort of 622 was formed after excluding 561 patients ineligible for the BHAT trial because of clinical features that were indications for or contraindications to β-blocker therapy. (See list of features on pages 632–633).

BASELINE COMPARISON OF THREE COHORTS: **Table III** compares all three cohorts on certain selected baseline features. In the BHAT trial, randomization created similar treatment groups for percentage men, average age, infarct location, and severity of the myocardial infarction (using the "risk-class" severity categories employed in the BHAT trial). In the expanded cohort, more patients were treated with β blockers than were untreated, and treated patients were slightly younger and were

TABLE III

Baseline Comparison of Three Cohorts

Feature	BHAT Trial		Current Study			
			Expanded		Restricted	
	BB	No BB	BB	No BB	BB	No BB
Number	1,916	1,921	626	433	417	205
Males (%)	84	85	74	68	75	73
Average age (years)	55	55	58	60	57	60
Infarct location (%)						
Anterior	28	26	29	33	20	27
Inferior	32	32	38	39	40	45
Nontransmural/ uncertain	32	33	33	29	30	28
Non-BHAT*	9	9	NA	NA	NA	NA
Risk class (%)†						
Mild	58	61	64	51	72	68
Moderate	28	26	15	30	18	16
Severe	14	13	21	18	10	16

NA = not applicable; BB = β blocker.
* In BHAT study, 9% of groups that received β-blocker and placebo did not have a qualifying myocardial infarction.
† Severe = prior myocardial infarction, moderate = no prior myocardial infarction, Killip class 3 or 4 in hospital or congestive failure on chest roentgenograph, or ejection fraction 40%; mild = no prior myocardial infarction, none of the above.

more likely to have had mild-risk infarcts. In the restricted cohort, treated patients were also slightly younger compared with untreated patients. Although there was less inequality, baseline differences for the severity of the myocardial infarct were again noted. These differences between treatment groups for age and infarct severity will require statistical adjustment in the analysis for mortality rates.

MORTALITY RATES AT 24 AND 36 MONTHS: The mortality rates for all three cohorts at 24 months of follow-up are displayed in **Table IV.** Included in the table are data for the overall results as well as results adjusted for baseline differences in age and the clinical severity of the myocardial infarction. In the BHAT trial, the overall death rates (Part A) were 7.3% in the treated group and 9.2% in the untreated group. The treatment differences favoring β blockers included an absolute difference of 1.9% and a proportional reduction in mortality of 21%. Since the BHAT trial was a randomized experiment, the compared groups were, on average, equal at baseline for age and clinical severity of the myocardial infarction. Consequently, the same mortality rates for β-blocker–treated (7.3%) and non-β-blocker–treated (9.2%) subjects are repeated for Parts B and C of Table IV estimating adjusted rates.

In the expanded cohort, the overall unadjusted mortality rate was 9.3% for β-blocker–treated patients and 16.4% for untreated patients. These differences favoring β-blocker therapy (7.1% absolute difference and 43% mortality reduction) are substantially larger than both the mortality rates and the treatment differences observed in the BHAT trial. In the restricted cohort, the death rate was 7.2% in the β-blocker–treated group and 10.7% in the untreated group, for an absolute difference of 3.5% and a mortality reduction of 33%. Using only the principles of patient eligibility and choice of a zero time that correspond to the design of a randomized trial, the restricted cohort, compared with the expanded cohort, more closely approximates both the mortality rates and rate differences observed in the BHAT trial.

We determined the separate effects of the eligibility criteria and the specification of the zero-time date on the unadjusted overall mortality rates at 24 and 36 months. To do this, we maintained the hospital discharge as the zero time for the expanded cohort, but applied the eligibility criteria to the assembly of the compared treatment groups. At 24 months, the use of randomized trial eligibility criteria resulted in the new mortality rates of 5.2% for β-blocker–treated patients and 10.4% for non-β-blocker–treated subjects. At 36 months, the new mortality rates were 7.4% for the former group and 13.3% for the latter group. In both instances, the absolute treatment difference in the expanded cohort is smaller after the use of the eligibility criteria and more closely approximates the treatment difference found in the restricted cohort.

As noted previously, for both the restricted and expanded cohorts, data in Table III demonstrate clinically and statistically significant differences between treatment groups for age and the severity of the myocardial infarction. In Parts B and C of Table IV, we display the mortality results after adjusting first for age alone and then for both age and infarct severity. For the expanded cohort, age ad-

TABLE IV

Comparison of Three Cohorts for 24-Month Mortality Rates

	BHAT Trial		Current Study			
			Expanded		Restricted	
	BB	No BB	BB	No BB	BB	No BB
Overall (%)	7.3	9.2	9.3	16.4	7.2	10.7
Absolute difference	1.9		7.1		3.5	
Mortality reduction	21		43		33	
Age-adjusted (%)	7.3	9.2	9.8	15.2	7.6	9.8
Absolute difference	1.9		5.4		2.2	
Mortality reduction	21		35		22	
Age- and severity-adjusted (%)	7.3	9.2	10.2	14.4	7.6	9.7
Absolute difference	1.9		4.2		2.1	
Mortality reduction	21		29		22	

BB = β-blocker.

justment reduces the absolute difference favoring β blocker to 5.4% (from 7.1%) and the mortality reduction to 35% (from 43%). After adjusting for age and infarct severity, the actual mortality rates and the absolute difference in mortality decline further, although they continue to be larger than the results in the BHAT trial.

For the restricted cohort, a different pattern emerges. Age adjustment alone creates actual mortality rates for treated and untreated patients that are very similar to the BHAT rates, and both the absolute treatment difference and the proportional mortality reduction are nearly identical to the results in the BHAT trial. In contrast to the expanded cohort, for the restricted cohort, adjustments for the severity of the myocardial infarction have little additional impact on the results. We think that no further change in the results occurred because the restricted cohort had already created compared groups that were similar for prognostically important clinical severity through the use of the experimental trial's eligibility criteria and choice of a zero time.

The comparisons for 36-month mortality rates and rate differences for all three cohorts are displayed in **Table V,** using the same format as Table IV. The results at 36 months are consistent with those demonstrated at 24 months. Noted again are the substantially larger estimates in the expanded cohort of actual mortality rates and rate differences in the BHAT trial and the similarity to BHAT results achieved with the restricted cohort.

Examined slightly differently, the overall 36-month mortality rate differences in the expanded cohort exceed the BHAT mortality rate treatment differences by 5.9% (9.4% versus 3.5%). Much of this excess estimate in treatment difference is removed in the restricted cohort simply by using the principles of the eligibility criteria and choice of a zero time. With these features, and without any further adjustment for clinical severity or age, the 5.9% overestimate in the expanded cohort is substantially reduced in the restricted cohort (3.5% in BHAT versus 4.7% in the restricted cohort).

In the original BHAT trial, the upper bound of the 95% confidence limit for the absolute difference in mortality rates was 3.8% at 24 months and 6.1% at 36 months. Thus, the age- and severity-adjusted absolute differences in the restricted cohort at 24 and 36 months (2.1% and 3.3%, respectively) fall within the 95% confidence limits of the BHAT trial. In contrast, the 95% confidence limits of the BHAT trial exclude the absolute differences at 24 and 36 months estimated in the expanded cohort (4.2% and 6.2%, respectively).

TABLE V

Comparison of Three Cohorts for 36-Month Mortality Rates

	BHAT Trial		Current Study: Expanded		Current Study: Restricted	
	BB	No BB	BB	No BB	BB	No BB
Overall (%)	9.0	12.5	12.1	21.5	9.4	14.1
Absolute difference	3.5		9.4		4.7	
Mortality reduction	28		43		33	
Age-adjusted (%)	9.0	12.5	12.8	20.1	9.8	13.0
Absolute difference	3.5		7.3		3.2	
Mortality reduction	28		36		25	
Age- and severity-adjusted (%)	9.0	12.5	12.9	19.1	9.8	13.1
Absolute difference	3.5		6.2		3.3	
Mortality reduction	28		32		25	

BB = β-blocker.

COMMENTS

Two important conclusions emerge from this research. First, an observational cohort method based on the design principles and patient assembly procedures of a randomized clinical trial closely approximates the results of the experimental trial. Second, an expanded cohort design achieves estimates of treatment benefit that are different from those both of an experimental trial and a cohort with restricted methods of patient assembly. In the sections that follow, we discuss the importance of the methodologic strategies that we have developed and then consider how we can reconcile our results with previous research.

Restricted Cohort Study

The restricted cohort method developed in this research is based on specifying and applying the principles and patient assembly procedures used for conducting randomized clinical trials. The implementation of this observational method requires fulfilling four essential procedures. First, the investigator needs to specify a zero time that will be used in determining patient eligibility and adjusting for baseline differences in prognostic risk. Second, eligibility for the study is determined according to the same criteria of inclusion and exclusion that would be used in a randomized clinical trial. Third, patients must be classified according to suitable clinical criteria to enable adjustment for any inequalities in susceptibility to the outcome. Finally, the main analysis should be conducted using the same statistical strategies (e.g., intention-to-treat procedures) as those employed in a clinical trial.

When observational designs are used to study

therapeutic efficacy, the usual method is a survey of cohorts of patients whose treatments were elected nonexperimentally, according to ad hoc patient and physician judgments. The absence of randomization leads to concern about the problem of bias due to baseline inequalities and to a general distrust of the study results. In clinical trials, randomization is generally accorded the major role in ensuring the baseline similarity of the compared treatment groups. One of the main findings that has emerged from the analysis of the restricted cohort is the empiric demonstration of the important and largely neglected role that strict eligibility criteria may play in ensuring the validity of treatment comparisons and study outcomes. For example, compared with the mortality rates in the BHAT trial, 36-month crude mortality rates in the expanded cohort substantially overestimated treatment differences (3.5% versus 9.4%, respectively). Using only the strict eligibility criteria and an appropriate choice of zero time, the restricted cohort substantially reduced any mismeasurement in treatment differences. The use of age and prognostic features to adjust further for remaining baseline inequalities then achieved an even closer correspondence between the BHAT trial estimates and the restricted cohort estimates for mortality rate outcomes. Furthermore, when only the eligibility criteria were applied to the expanded cohort, the mortality rates at 24 and 36 months and the rate differences more nearly approximated the rates estimated in the restricted cohort.

The implications of this finding are critical to our understanding of the validity of research study designs. To date, we have assumed that random allocation is the principal protection against bias in an experimental clinical trial. For this reason, we have been reluctant to consider any study design that did not include randomization of treatments. The result of the current research suggests that new observational designs can be developed that may not need to rely on random allocation. The use of restriction criteria such as eligibility features and the choice of cogent clinical features for prognostic adjustment may enable observational studies to approximate more closely the results that would have been found had a randomized clinical trial been conducted. Other principles of the experimental trial, including the choice of a suitable zero time, would need to be maintained to ensure the inherent validity of the observational study.

Expanded Cohort

The restricted cohort achieved results that approximate the results of the BHAT trial at a considerable cost in the scope of the study participants. Since the restricted cohort uses the same restricted eligibility criteria as the randomized trial, the restricted cohort also excludes many patients who would be considered candidates to receive treatment in customary clinical practice. Indeed, the restricted cohort included only 53% (622 of 1,183) of the subjects who survived a myocardial infarction and were available to receive a β blocker.

An important advantage of the expanded cohort is its ability to estimate treatment effects in this broader spectrum of clinical practice. The hazard of this expanded spectrum is that patients are included for whom the decision not to use a β blocker may be related to clinical features that increase the risk of death after a myocardial infarction. Since randomization is not used in observational studies, and since restriction criteria for patient assembly are avoided to ensure a broader study population, suitable prognostic adjustment on crucial clinical features is needed to minimize the problem of susceptibility bias.

In the expanded cohort, the actual mortality rates for both treatment groups exceed the rates for either the BHAT trial or the restricted cohort. These higher mortality rates were anticipated since the expanded cohort includes many patients with relative contraindications to β blockers such as congestive heart failure or strong indications such as angina pectoris. Thus, we were not surprised to find a larger estimated treatment benefit in the expanded cohort, which reflects in part this susceptibility bias. Interestingly, we suspect that much of the bias is eliminated by the prognostic adjustments for age and clinical severity. The decrease in treatment benefit from 43% to 29% after prognostic adjustment seen in the expanded cohort is nearly identical to the proportional decrease observed in the restricted cohort (33% to 22%). Both age and clinical severity are important for adjustment in the expanded cohort, whereas only age appears important in the restricted cohort.

Reconciliation with Previous Research

Observational study designs have long been sought as a substitute for randomized clinical trials. However, no previous research has demonstrated that observational studies may be valid surrogates. Numerous previous investigators have considered the limitations of alternatives to randomized clinical trials. In an important essay commenting on observational studies, the author noted that numerous methodologic problems would make sound, unbiased inferences from nonrandomized studies of therapy difficult [11]. The problems cited included

difficulties with bias in treatment assignment (susceptibility bias), the use of nonstandard definitions, and the testing for multiple comparisons. Our methods incorporate specific methodologic strategies to minimize each of these problems. In the restricted cohort, susceptibility bias was minimized by using eligibility criteria and prognostic adjustment; nonstandard definitions for myocardial infarction, clinical severity, and other features were eliminated by using accepted, detailed criteria that mimicked those employed in the BHAT trial; and the problem of multiple comparisons was avoided by testing a predetermined hypothesis with a single outcome. In the expanded cohort, all of these procedures were used except for restricted eligibility criteria.

Other investigators compared six therapies in which randomized clinical trials and historical control trials were both performed [4]. Overall, 44 of 56 historical control trials (79%) found the experimental or new therapy better than the control regimen, but only 10 of 50 randomized trials (20%) agreed. Adjustment of the outcomes of the historical control trials for prognostic factors did not appreciably change the results. The authors concluded that "... biases in patient selection may irretrievably weigh the outcome of historical control trials in favor of new therapies" [4].

The research reported here provides support for the empirical evidence the authors present, but argues strongly against a general prohibition against observational studies. Observational research conducted using historical controls does overestimate treatment benefits. However, neither the expanded cohort nor the restricted cohort should be considered vulnerable to the biases that distort the results using historical controls. In both cohorts, we use concurrent, not historical, controls, strict patient eligibility criteria, a suitable zero time, and prognostic adjustment. When these principles are included in an observational study of therapeutic outcomes, the results approximate those reported in a widely accepted randomized clinical trial.

Finally, our data are consistent with two reports that also compared observational studies to randomized controlled trials. In the Coronary Artery Surgery Study [12], a comparison of the randomizable (patients who were eligible but refused participation) and randomized groups showed remarkable similarity in the results for patients treated with surgery or medical therapy of coronary artery disease. For example, at 5-year follow-up, survival in the medically randomized and randomizable patient groups was similar in the aggregate (both 92% at 5 years) and also in all subgroups examined based on the number and severity of diseased vessels, ejection fraction, and clinical classification. Survival in the surgically randomized and randomizable patients was also similar in the aggregate and in all patient subgroups [12]. Finally, the findings of three randomized trials of coronary bypass surgery were compared with the predictions of multivariable statistical models derived from observations in a large cardiovascular disease data bank. Overall, model predictions agreed well with randomized trial results for survival rates [13].

The improved cohort methods developed and tested in this report have importance far beyond the implications of β-blocker therapy. Physicians are increasingly pressured to establish the effectiveness and efficiency of medical interventions. Despite its scientific advantages, the randomized clinical trial, because of problems in logistics, costs, and ethics, cannot be applied to many of the problems in clinical therapy and patient results that remain unresolved. A recent editorial noted that "an exclusive reliance on randomized controlled trials to provide definitive information about effectiveness is not the answer" [14]. The author went on to say that "alternative designs for studies of effectiveness are therefore urgently needed" [14].

Although observational surveys are frequently performed as a substitute for the randomized clinical trial, the evidence from such surveys is frequently not convincing. The research reported here, which provides a methodologic approach that improves the quality of observational studies, is a comparison of only a single cohort study to a single randomized controlled trial. Clearly, more studies are needed to demonstrate the validity of the method and its generalizability.

If replicated in other studies, these improved methods should serve as welcome new strategies for assessing the efficacy of the many medical or surgical therapies that cannot or will not be tested with randomized clinical trials.

ACKNOWLEDGMENT

We are grateful to Drs. R. Makuch and C. Murray for review of the manuscript, and to Ms. L. Hurst and K. Anderson for extraction of the data.

REFERENCES

1. Horwitz RI. The experimental paradigm and observational studies of cause-effect relationships in clinical medicine. J Chronic Dis 1987; 40: 91–9.
2. Schwarz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. J Chronic Dis 1967; 20: 637–48.
3. Feinstein AR. An additional basic science for clinical medicine: II. The limitations of randomized trials. Ann Intern Med 1983; 99: 544–50.
4. Sacks H, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. Am J Med 1982; 72: 233–40.
5. Beta-Blocker Heart Attack Trial Research Group: A randomized trial of propranolol in patients with acute myocardial infarction. I. Mortality results. JAMA

1982; 247: 1707–14.

6. Davis KB, Fisher L, Gillespie MJ, Pettinger M. A test of the National Death Index using the Coronary Artery Surgery Study (CASS). Controlled Clin Trials 1985; 6: 179–91.

7. Curb JD, Ford CE, Pressel S, Palmer M, Babcock C, Hawkins CM. Ascertainment of vital status through the National Death Index and the Society Security Administration. Am J Epidemiol 1985; 121: 754–66.

8. Stampfer MJ, Willett WC, Speizer FE. Test of the National Death Index. Am J Epidemiol 1984; 119: 837–9.

9. Acquavella JF, Donaleski D, Hanis NM. An analysis of mortality follow-up through the National Death Index for a cohort of refinery and petrochemical workers. Am J Ind Med 1986; 9: 181–7.

10. Wentworth DN, Neaton JD, Rasmussen WL. An evaluation of the Social Security Administration Master Beneficiary Record File and the National Death Index in the ascertainment of vital status. Am J Public Health 1983; 73: 1270–4.

11. Byar DP. Why data bases should not replace randomized clinical trials. Biometrics 1980; 36: 337–42.

12. CASS Principal Investigators and Their Associates: Coronary Artery Surgery Study (CASS): A randomized trial of coronary artery bypass surgery. J Am Coll Cardiol 1984; 3: 114–28.

13. Hlatky MA, Califf RM, Harrell FE, Lee KL, Mark DB, Pryor DB. Comparison of predictions based on observational data with the results of randomized controlled clinical trials of coronary artery bypass surgery. J Am Coll Cardiol 1988; 11: 237–45.

14. Greenfield S. The state of outcome research: are we on target? N Engl J Med 1989; 320: 1142–3.