III - 2

Symposium

THE AMERICAN FEDERATION FOR CLINICAL RESEARCH

Presents a Public Policy Symposium

"THE SCIENTIFIC USES AND ABUSES OF THE CLINICAL TRIAL: TREATMENT OF CHRONIC STABLE ANGINA WITH SAPHENOUS VEIN BYPASS GRAFTING. RANDOMIZED VETERANS ADMINISTRATION COOPERATIVE STUDY"

Participants:

THOMAS C. **CHALMERS**. In Defense of the VA Randomized Control Trial of Coronary Artery Surgery.

WILLIAM L. **PROUDFIT**. Criticisms of the VA Randomized Study of Coronary Bypass Surgery.

ALVAN R. **FEINSTEIN**. The Scientific and Clinical Tribulations of Randomized Clinical Trials.

GERALD F. **DIBONA**, Moderator.

The purpose of this symposium is to explore current uses and possible abuses of the clinical trial. It focuses on the recent VA cooperative study of coronary artery surgery.

Dr. Chalmers emphasizes that acceptance of major clinical randomized trials may depend on whether or not such studies confirm prior clinical judgments. He strongly defends the quality of the VA study and demonstrates that survival following surgery is similar in this randomized study compared to other series whereas the survival rate is lower in "historical controls" compared to randomized patients receiving medical therapy.

Dr. Proudfit rejects the conclusions of the VA study citing problems of noncompliant patients, variable operative mortality rates among institutions, and the incidences of perioperative infarction and graft patency compared to results at a single institution.

Dr. Feinstein emphasizes the limitations of randomization *per se*, and proposes certain methodologic changes to improve randomized clinical trials including suitable selection of patients and identification of baseline status, equal performance of the two imposed treatments and ancillary therapy, and equal detection of outcome events after each treatment.

The Editors solicit your opinions on the issues and interpretations presented in this symposium.

In Defense of the VA Randomized Control Trial of Coronary Artery Surgery

THOMAS C. CHALMERS
HARRY SMITH, JR.
ALEXANDER AMBROZ
DINAH REITMAN
BIRUTA J. SCHROEDER

It is apparent that the acceptance of major randomized control trials (RCT) by physicians depends to a large extent on whether or not they confirm or deny prior clinical judgment. The observation that treatment of moderate hypertension with drugs prevented complications from vascular disease, reported by the VA Cooperative Study Group,[1] met with essentially no published criticism. It has resulted in vast national programs of case finding so that more and more patients may be treated. On the other hand, the report by the University Group Diabetes Program (UGDP),[2] that popular oral hypoglycemic agents were associated with increased rather than decreased deaths from cardiovascular disease, was met by scorn, derision, and continuous and expensive efforts to prevent a report of the findings from appearing in the package inserts of the drugs. Yet, the UGDP study was probably a better cooperative clinical trial than the VA study of hypertension because the patient material was less severely restricted, the blinding more complete, and the endpoint less subject to bias.

The report of the Veterans Administration randomized control trial of coronary bypass surgery and medical therapy in patients with chronic angina[3] has resulted in more published objections than any study since the UGDP. Are the outcries caused by the poor quality of the VA study, by its nonapplicability to the general population, or by the fact that its conclusion is contrary to conventional wisdom? It is the purpose of this paper to examine three possibilities.

From the Office of the Dean and Department of Biostatistics, Mount Sinai School of Medicine of The City University of New York, New York 10029.

Requests for reprints should be addressed to Dr. Chalmers at the above address.

Presented at the AFCR Public Policy Symposium, 35th Annual Meeting, San Francisco, April 30, 1978.

Before analyzing the quality of the VA study, it is pertinent to outline a system that has been developed at The Mount Sinai Medical Center for evaluating the quality of RCTs. Protocols as presented in the methods sections of RCTs have been searched for three categories of information which seem to be most pertinent to the quality of the studies: (1) the inclusion of essential descriptive material; (2) adequate blinding wherever possible; and (3) proper use of various testing procedures.

The individual items in each of the three categories are shown in Table 1. In scoring the quality of each trial, the five items in category 1 were given three points each, the three items in category 2 were given five points each, and the five items in category 3 were given three points each. A perfect trial would score a total of 45 points. Determination of the score for each item was made from a differentially photocopied portion of the published report that excluded the authors, the source, and any hints of the outcome. For comparative assessment, a quality score was derived by dividing the number of points achieved by the total number possible. Nonapplicable qualities were excluded from both numerator and denominator, as in the calculation of a batting average.

A comparison of the VA study[1] was made using 159 papers surveyed recently and published in the *New England Journal of Medicine* and *Lancet* during a four-year period from 1973 to 1976,[4] and with seven RCTs of coronary artery surgery, of which the VA study was one (Table 1). These data reveal the VA study to have satisfied almost all the criteria of quality. The quality score of .897 achieved by the VA study places it in the top 5% of all reviewed RCTs. The only item missing was evidence in the paper that there was prior determination of the endpoint, therapeutic effect of interest, and the numbers

Table 1. Percentage Frequency of Desirable Protocol Items in Randomized Control Trials.

	New England Journal of Medicine—Lancet	Total Coronary Artery Surgery	VA Study
Number of papers	159	7	1
Essential descriptive material			
Adequate selection description	98	86	Yes
Log of rejected patients kept	8	57	Yes
Withdrawals list	66	86	Yes
Adequate description of control regimen	97	100	Yes
Adequate description of therapeutic regimen	98	100	Yes
Blinding of:			
Randomization	40	71	Yes
Patients	67	N.A.	N.A.
Physicians	57	N.A.	N.A.
Testing procedures			
Evidence of prior estimate of numbers required	4	0	No
Validity of randomization	57	100	Yes
Validity of blinding	1	N.A.	N.A.
Compliance	20	N.A.	N.A.
Measurement of a biological equivalent of therapy	20	71	Yes
Mean quality score	.513	.684	.897

required to achieve reasonable Type 1 and 2 errors. This was done before the start of the VA study but was not mentioned in the paper because of space.* However, similar omissions may have occurred in the other reviewed papers. It is of interest that by this method the widely accepted VA cooperative study of antihypertensive therapy[1] received a lower quality score than the unpopular UGDP[2] demonstration that oral hypoglycemic drugs are hazardous—the score being .733 compared with .867.

A major objection to the VA coronary study has been the crossing over of patients from one treatment group to the other. Some patients randomized to surgery never received it. Some patients randomized to medicine were eventually operated on because they were doing poorly. It is rarely possible to accomplish an RCT of a surgical treatment in which this does not occur. The preferred solution is to analyze the data all possible ways: according to the original treatment assignment; according to the treatment actually received; with the crossovers removed from the study; and with the crossovers counted as lost to follow-up at the time treatment was changed. All four methods were employed by the VA investigators[3] and cumulative four-year survival rates were almost identical.

*Personal communication from K.M. Detre.

In the VA study the investigators searched for better ways to classify angiographic abnormalities and documented observer variability in the reading of preoperative coronary angiograms.[5] None of the advocates of widespread surgery have mentioned this problem, much less measured the error. Yet the VA study is castigated[6] because the investigators recognized potential defects and explored their implications.

A number of objections to the VA study are complaints that the high operative mortality was not reflective of that encountered elsewhere. It was emphasized that the surgical patients would have had a better three-year survival if more of them had survived for one month after operation. Many critics cited their own very low operative mortality as a reason for the surgery to continue to be done throughout the nation.

The case fatality rates of 16 New York City hospitals performing coronary artery surgery during 1976 are presented in Figure 1. These are compared with the operative mortality figures from 21 series reported in the literature during 1977. Although there are no statistically significant differences, the means reveal a higher death rate among unreported series, as might be expected, and they reveal that the operative mortality rate in the VA study was not inordinately different from either of these groups. In other words, isolated institutions with

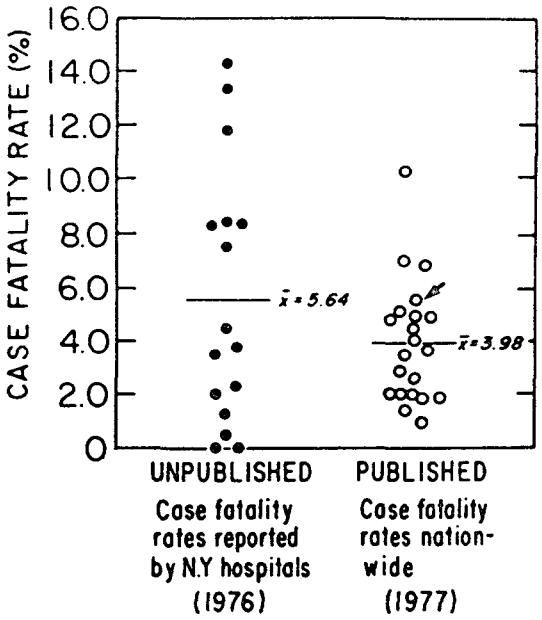


Figure 1. Operative fatality rates for coronary artery surgery in an unpublished series collected from different New York hospitals (1976), and from all series published from nationwide hospitals in 1977 (references available on request). — indicates VA cooperative study.

very low or very high case fatality rates do not accurately predict the mean operative mortality rates, since the operation is widely done throughout the country. There will always be some hospitals at either end of the distribution curve. The means and standard deviations from all available series are the only estimates that can be used to make this judgment.

Furthermore, without a randomized control group, the hospitals with high or low operative mortalities or three-year survivals will never be able to say whether their results are due to selection or therapy. The Cleveland Clinic[7] reports very different proportions of one- and three-vessel disease than does the VA study. Patients in the VA study had more severe coronary artery disease—52% of patients versus 13% had triple vessel disease (Table 2).

Table 2. Distribution of Coronary Surgery Patients by Severity of Disease.

	Cleveland*	VA[12]
Time period	1967-1970*	1972-1974
Number	741	332
Single vessel disease	56%	14%
Double vessel disease	31%	34%
Triple vessel disease	13%	52%

**These were operated upon earlier than the VA series, but are among the figures quoted in the papers which are cited in criticisms of the VA study.*[6,16,17]

It has often been assumed that operative mortality will be higher in VA hospitals, both because of the selection of patients with more complicating diseases such as alcoholism, and as a result of more surgery performed by resident staff. However, data from all published series that distinguished operative mortality for each degree of vessel disease illustrate that this is not the case (Table 3). When patients with one-, two- and three-vessel disease are combined from various papers, the VA hospitals have no higher mortality rates. The argument that the conclusions of an RCT carried out in VA hospitals are not generally applicable, because the mean operative mortality was too high, requires a comparison with operative mortalities in many non-VA hospitals, not simply those with the lowest rates.

The advantage of the RCT is that it allows for

Table 3. Coronary Bypass Surgery Operative Mortality by Number of Vessels Diseased: Comparison of VA and Non-VA Hospitals.

VA[3,18-20]			Non-VA[21-26]		
No. papers reviewed		4	No. papers reviewed		6
Total no. patients		1001	Total no. patients		3727
No. Vessels Diseased	No. Patients	Operative Mortality Percent	No. Vessels Diseased	No. Patients	Operative Mortality Percent
1	183	0.5	1	597	3.7
2	418	4.3	2	1564	5.2
3 or more	400	5.5	3 or more	1566	5.9
Totals	1001	4.1	Totals	3727	5.3

Table 4. Comparison of Therapeutic Trials Employing Randomized, (RCT), Nonrandom but Simultaneous (NRCT), and Historical (HCT) Controls.

Study Category	No. of Studies	No. of Studies Where $p < .05$ for Treated vs. Controls	Total No. Patients	% Efficacy Treated	% Efficacy Controls
RCT	19	3	13,609	70	70
NRCT	17	7	18,266	82	79
HCT	32	26	17,189	69	49

Therapies: stilbestrol in abortion prevention; anticoagulants in myocardial infarction; shunts in cirrhosis (prophylactic and therapeutic); gamma globulin for post-transfusion hepatitis.

separation of therapeutic effects from those that result from selection of patients. The overwhelming influence of selection, when compared with therapy, is illustrated by past experiences with portacaval shunt operations. A striking improvement in survival resulting from surgery was reported in 1956.[8] These effects of surgery have not been confirmed by RCTs.[9] The survival of patients selected for randomization, whether operated on or not, is much better than the "controls" not selected.[10]

The current dispute with regard to survival of patients with coronary artery disease results from differences in conclusions from trials employing randomized controls and those employing selected or historical controls. When the effects of several different treatments previously studied are combined and compared for these three types of controls, the data presented in Table 4 are obtained.[11] It is apparent that the therapeutic efficacy of the experimental groups is similar no matter what the source of the controls, while the historical control groups have a much lower therapeutic outcome. This phenomenon may result primarily from the fact that historical control groups are not selected for a hazardous procedure and are not required to give informed consent to enter the study no matter how well they are matched retrospectively. They are likely to be less healthy. That this principle applies to studies of coronary artery surgery is illustrated by survival curves redrawn from published data from the Cleveland Clinic[7] and the VA study[12] (Fig. 2). Survival rates of operated patients are similar in the randomized and unrandomized control trials, but the controls have a distinctly worse prognosis when they are not randomized. Thus, historically selected controls have the poorest outcome. These data also indicate that the Cleveland Clinic surgical group has a slightly better survival rate than the VA surgical group. However, when the VA groups are adjusted proportionately to the Cleveland Clinic groups according to relative frequency of one-, two-, and three-vessel disease (Fig. 3), there are no significant differences between the VA surgical, medical, and Cleveland Clinic surgical groups. This indicates that in these three groups, the only true differences are the result of the extent of disease of the patients allocated to the respective treatment groups. The poor showing of the Cleveland Clinic medical group remains and illustrates the powerful effects of selection when a control group is not assigned at random. The difference remains even when a correction is made for measurable indicators of severity. Obviously, there are other indicators of outcome that are less easily measured and can be distributed fairly only if the treatments are assigned at random. Even in the

Table 5. Relation of Published Reaction to VA Study to Professional Discipline.

	Generally: Approve of Study Conclusions	Generally: Object to Study Conclusions	Total
Cardiovascular surgeons	0	8	8
Others	6	6	12
Total	6	14	20

Fisher's Exact Test $p < .03$.

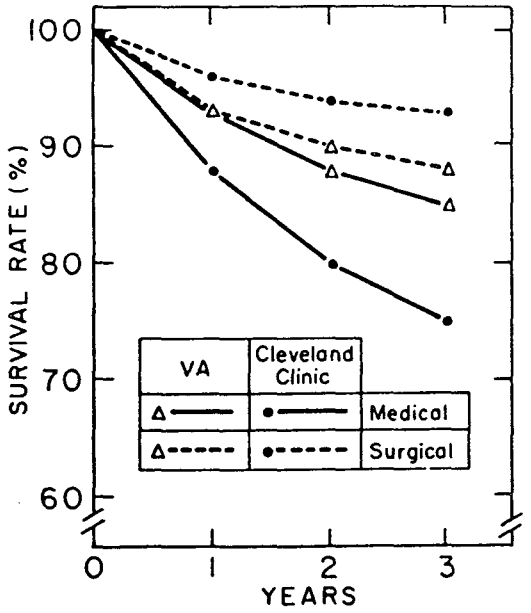
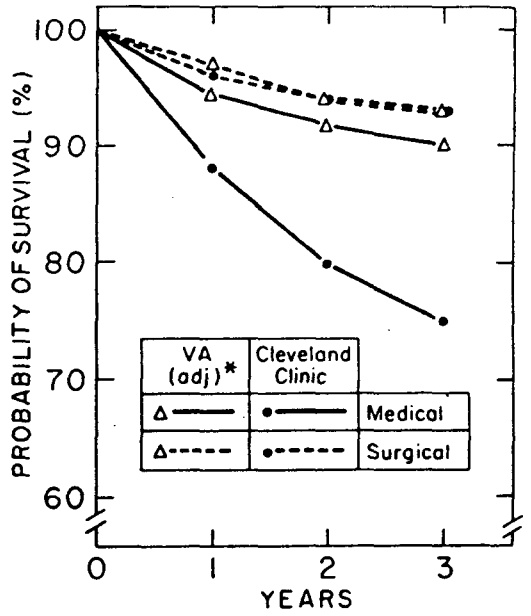


Figure 2. Three-year survival curves of surgical (broken lines) and medical (solid lines) from VA (triangles) and Cleveland Clinic (circles). These are composite curves combining one-, two-, and three-vessel disease according to the published frequencies presented in Table 2.



**Adjusted for relative frequency of 1, 2, and 3-vessel disease*

Figure 3. Curves in Figure 2 redrawn to adjust for different frequencies of one-, two-, and three-vessel disease as reported in references 7 and 12.

Duke study[13] careful matching did not rule out the possibility that subtle selection factors were contributing to outcome.

If the VA study was well designed and the outcome was not very different from that encountered in non-VA hospitals, why has it been badly received? Publication of the study resulted in a large number of letters of protest and praise to the editor[14] and also in a number of editorials in other journals, as well as comments in the daily press and medical news periodicals. In order to assess properly their meaning, each of these commentaries was photocopied so that the origin could not be detected, and a decision was made by two observers as to whether or not the opinion was generally in favor of or in opposition to the VA study. The data are presented in Table 5. It is clear that opinions about the study are related to the professional activities of the author. Opposition to the study seems to be universal among cardiac surgeons.

It should be pointed out that the VA study in question is a report of survival alone and that the data on relief of symptoms are yet to be reported. There is no doubt that a large group of patients with coronary artery disease receive dramatic symptomatic benefit from the operation. It also seems likely that obstruction of the left main coronary artery is responsive to surgery as shown by the data extracted early from the VA study.[15] The efficacy of surgery in patients with unstable angina is still unknown, but fortunately, there are several RCTs underway in this area. In the meantime, tens of thousands of patients without left main artery obstruction continue to be operated on with insufficient evidence that survival is prolonged. It is impossible to obtain data on what percentage of patients are undergoing surgery because they are crippled by coronary disease and probably should receive it, and what percentage have manageable symptoms but undergo surgery in the hope that their lives will be prolonged. The latter is almost surely the larger group in this country. Advocates of surgery in these patients might well consider conducting carefully designed trials themselves so as to assure themselves that their public criticism of the only large-scale, well done randomized control trial so far completed is valid.

References

1. Veterans Administration Cooperative Study Group: Effects of treatment on morbidity in hypertension—results in patients with diastolic blood pressures averaging 115 through 129 mm Hg. JAMA 202:116-122, 1967.
2. The University Group Diabetes Program: A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. Diabetes 19(suppl 2):747-830, 1970.
3. Murphy ML, Hultgren HN, Detre KM, et al: Treatment of chronic stable angina—a preliminary report of survival data of the Randomized Veterans Administration Cooperative Study. N Engl J Med 297:621-627, 1977.
4. Ambroz A, Chalmers TC, Smith H, et al: Deficiencies of randomized control trials. Clin Res 26:280A, 1978.
5. Detre KM, Wright E, Murphy ML, et al: Observer agreement in evaluating coronary angiograms. Circulation 52:979-986, 1975.
6. Loop FD, Proudfit WL, Sheldon WC: Coronary bypass surgery weighed in the balance. Am J Cardiol 42:154-156, 1978.
7. Sheldon WC, Rincon G, Pichard AD, et al: Surgical treatment of coronary artery disease: Pure graft operations, with a study of 741 patients followed 3-7 years. Prog Cardiovasc Dis 18:237-253, 1975.
8. Ellis DS, Linton RR, Jones C: Effect of venous-shunt surgery on liver function in patients with portal hypertension. N Engl J Med 254:931-936, 1956.
9. Chalmers TC: Randomized clinical trials in surgery. *In* Varco RL, Delaney JP (eds): Controversy in Surgery. Philadelphia, WB Saunders & Co, 1976, pp 3-11.
10. Garceau AJ, Donaldson RM, O'Hara ET, et al: A controlled trial of prophylactic portacaval-shunt surgery. N Engl J Med 270:496-500, 1964.
11. Chalmers TC, Smith H, Kunzler AM, et al: A comparison of therapeutic trials employing historical, alternate and randomized controls. Clin Res 24:475A, 1976.
12. Read RC, Murphy ML, Hultgren HN, et al: Survival of men treated for chronic stable angina pectoris: A cooperative randomized study. J Thorac Cardiovasc Surg 75:1-16, 1978.
13. McNeer JF, Starmer CF, Bartel AG, et al: The nature of treatment selection in coronary artery disease: experience with medical and surgical treatment of a chronic disease. Circulation 49:606-614, 1974.
14. Special Correspondence: A debate on coronary bypass. N Engl J Med 297:1464-1470, 1977.
15. Takaro T, Hultgren HN, Lipton MJ, et al: The VA cooperative randomized study of surgery for coronary arterial occlusive disease, II. Subgroup with significant left main lesions. Circulation 54(suppl 3):107-117, 1976.
16. Hart NJ, Sheldon WC: Medical vs surgical management of myocardial ischemia: Effect on survival. Pract Cardiol, March 1978, pp 74-90.
17. Sheldon WC, Loop FD, Proudfit WL: A critique of the VA cooperative study. Cleve Clin Qtly 45:225-230, 1978.
18. Aronow WS, Stemmer EA: Two-year follow-up of angina pectoris: Medical or surgical therapy. Ann Intern Med 82:208-212, 1975.
19. Kaiser GC, Barner HB, William VL, et al: Aortocoronary bypass grafting. Arch Surg 105:319-323, 1972.
20. Anderson RP, Rahimtoola SH, Bonchek LI: The prognosis of patients with coronary artery disease after coronary bypass operations: Time-related progress of 532 patients with disabling angina. Circulation 50:274-282, 1974.
21. Assad-Morrell JL, Connolly DC, Brandenburg RO, et al: Aortocoronary artery saphenous vein bypass grafts. J Thorac Cardiov Surg 69:841-850, 1975.
22. Cooley DA, Dawson JT, Hallman GL, et al: Aortocoronary saphenous vein bypass. Ann Thorac Surg 16:380-390, 1973.
23. Gann D, Colin C, Hildner FJ, et al: Coronary artery bypass surgery in patients seventy years of age and older. J Thorac Cardiovasc Surg 73:237-241, 1977.
24. Morris GC, Reul GJ, Howell JF, et al: Follow-up results of distal coronary artery bypass for ischemic heart disease. Am J Cardiol 29:180-185, 1972.
25. Reul GJ, Cooley DA, Wukasch DC, et al: Long-term survival following coronary artery bypass: Analysis of 4,522 consecutive patients. Arch Surg 110:1419-1424, 1975.
26. Winterscheid LC, Dillard DH, Hessel EA, et al: Outcome analysis of surgical therapy for patients with angina pectoris. Amer Surg 41:520-528, 1975.

Criticisms of the VA Randomized Study of Coronary Bypass Surgery

WILLIAM L. PROUDFIT

Many years ago, an editor asked Frank N. Wilson, the electrocardiographer, to review a paper, the title of which indicated that a certain reaction in dogs had been investigated. Dr. Wilson's admirable review consisted of inserting the word "some" before "dogs." It would be more descriptive if the words "in some patients" were included in the title of this symposium.

The advantage of a multicenter study of any disease is speed in accumulation of data. For rare diseases, this approach is essential, but in the case of common conditions, it is often preferable to carry out studies in a single institution in which high-quality work can be assured. Adherence to protocol is monitored more easily in single institutions and the effect of noncompliance on the part of investigators and patients is determined more readily. Nonuniformity of results of participating institutions has been a characteristic of previous, well-publicized multicenter studies that have been disputed later. The prospective randomized study is useful and may be essential in therapeutic research, but it is not the only route to wisdom. Actuarial (life table) analysis is another method of accelerating medical investigation. However, it introduces new problems as well as solutions. Rarely is the reader given sufficient information in an actuarial study to enable him to recalculate the data. In addition, noncompliant patients constitute a problem for which there is no easy answer.

The critical question is whether the VA study of nonsurgical and surgical treatment of chronic stable angina pectoris, excluding left main artery lesions, is flawed to such a degree that valid conclusions are impossible. To attempt to answer this question, the principal conclusions must be stated. The first conclusion is that previous studies of natural history of obstructive coronary disease have not paralleled recent experience with medical treatment. Second, coronary bypass surgery has not been shown to reduce the expected mortality, except for patients who have obstruction of the left main coronary artery.[1-3]

In the VA study, 5,538 patients were screened for chronic stable angina pectoris: 2,125 had arteriography; 1,392 met the arteriographic criteria for selection; and 1,015 entered the study.[4,5] Excluding the 1970-1971 period during which the operative mortality was 16%, and excluding patients who had left main artery lesions, there remained 596 in the study group. Randomization of patients into medical and surgical groups was done locally and the appropriate treatment initiated, but the classification of the catheterization findings used in data analysis was completed in December 1975, one year after patient-access to the study was terminated.[4,5] This is a defect in performance of a prospective study. If randomization is based on one set of catheterization data, correlation cannot be based on another.

Still a third review of some of the arteriograms was made two years ago, but the results of this study have not been reported.[4,5] It was stated that 34 of the 596 patients could not be categorized because "either the number of diseased vessels or the state of ventricular function was unknown."[1] However, these patients were not dropped from the calculation of the composite curves, nor was the distribution of these 34 patients between medical and surgical cases indicated.

It is possible that some surgical patients had inappropriate operations based on anatomical misinterpretations. There were 54 noncompliant medical patients and 18 surgical patients.[1] Medical patients may become noncompliant (have operations) at any time during a period of follow-up, but surgical patients must reject operations early. How were patients who were randomized to surgery, but died prior to an operation handled statistically? An early report indicated that there were eight such patients in the 1970-1974 period.[4]

From the Department of Cardiology, The Cleveland Clinic Foundation and The Cleveland Clinic Educational Foundation, Cleveland, Ohio 44106.

Requests for reprints should be addressed to Dr. Proudfit at the above address.

Presented at the AFCR Public Policy Symposium, 35th Annual Meeting, San Francisco, April 30, 1978.

The VA study was confined to men who had chronic stable angina pectoris for at least six months, a much more restrictive definition than that used in other studies of obstructive coronary disease. There was no uniform therapeutic protocol for the medical or surgical patients. Propranolol was known to have been taken by 39% of the randomized medical patients and 20% of the surgical patients.[1]

Careful review of previously reported large-scale prognostic studies of patients who had arteriographically proven obstructive coronary disease indicates that surgical candidacy for bypass operation was not considered in most studies nor were these investigations restricted to patients who had stable angina pectoris for at least six months. Only one large survey addressed the question of surgical candidacy, and problems in comparability to the VA study make comparison difficult.[6]

A report[7,8] published in 1973 seems to have been taken as a control for surgical treatment by some. It was not designed as such. In 1975, the 388 patients selected in this study were classified according to surgical candidacy by modern criteria and were followed until death, subsequent operation, or survival for at least 10 years. All deaths are counted in the following survival curves, not cardiac deaths only (Fig. 1). Thirty-six patients who had left main coronary artery lesions were excluded. For comparison with this composite curve, it would be best to use VA survival curves for 36 or 48 months.[1-3]

The VA curves show 310 medical patients, but some unknown number of these could not be classified anatomically on the basis of catheterization findings (34 total in medical and surgical groups), and should have been excluded. In addition, 54 had operations during the period of study. A recent correction of the paper of Murphy et al indicated that of patients followed 21 to 36 months, 72 died instead of the 77 reported; of these 37 were in the medical group.[1,10] Only 175 of the originally randomized 310 patients were followed for 36 months or until death. A life table calculation suggests that the three-year survival was 84.6%, not the 87% reported. The 48-month survival cannot be calculated because the number of deaths is not stated.[2,3] It is apparent that the percentage survival at 36 months was not greatly different from that of our study (80.8% for all patients).

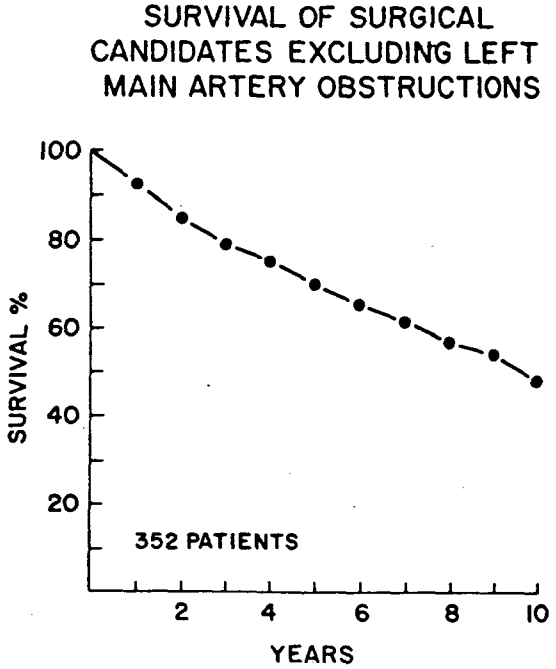


Figure 1. Composite survival curve of 352 medically treated surgical candidates, excluding those having left main artery obstruction.

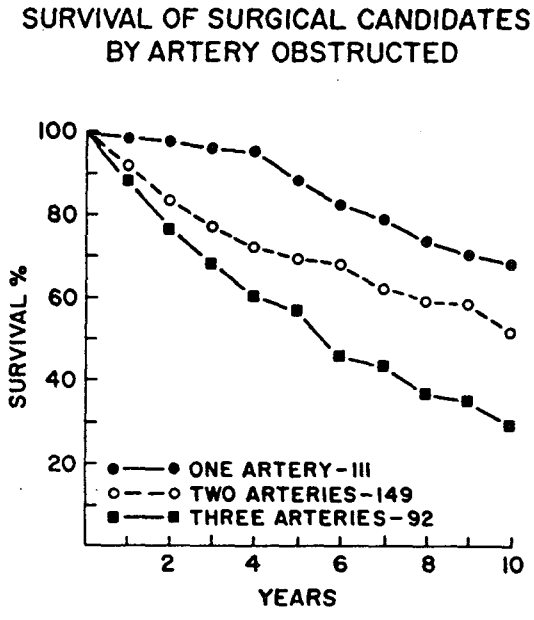


Figure 2. Survival curves for surgical candidates having one-, two-, and three-artery obstructions, excluding those with left main artery involvement.

These figures exclude patients who had left main artery lesions in both instances.

Composite curves are of limited value, because composite patients are not encountered. Figure 2 shows survival by number of arteries obstructed, for our surgical candidates, again excluding patients who had left main artery lesions. The method of presentation of the VA data precludes close comparison with our results. Obviously, patients with single-artery obstruction do well for four years before the slope of the survival curve changes. The VA curves show three- and four-year survival for only two- and three-artery obstructions with abnormal left ventricular function. The unique definition of the latter is unfortunate and comparable groups are not available. It is evident that the three-year survival was about 83% for obstruction of two arteries and 81% for three-artery involvement. However, other VA reports have given survival at one year in three-artery obstruction with abnormal left ventricular function of 86% to 93%, and at two years of 81% to 89%.[1-5,11,12]

The one common survival percentage for three long-term VA studies was 81% at three years. At four years, Read's graph seems to show 76% survival and Detre's figure was 79%.[2,3] We had 77% and 68% for two- and three-artery obstructions, respectively, at three years and 72% and 60% at four years. The number of deaths and compliant VA patients were not given for these subsets, so recalculation of percentages is not possible.

Considering that our study included patients who had high risk characteristics whereas such patients were excluded in the VA study, the differences in survival in the two investigations are not great. If studies were confined to surgical candidates who had moderate restriction of activity and obstruction of three arteries, our three- and four-year survival was 75% and 66%, respectively. Furthermore, a factor tending to improve the prognosis in the VA group is that 17.4% of the medical patients had operations and a large percentage of these patients had severe symptoms, leaving a relatively low-risk group of patients for survival study.[13] We have shown an adverse effect of severe symptoms on prognosis.[14]

Multicenter randomized prospective studies of diseases treated surgically require that patient selection, operative technique, and postoperative care be optimal. In 1939, Gross[15] reported closure of patent ductus arteriosus in four patients without a death. The apparent simplicity of the surgical technique tempted many surgeons to operate on this defect, but the operative mortality was considerable. If a multicenter prospective randomized study had been carried out, it would have been difficult to show that ductus closure influenced survival favorably, even in a long-term study. A delay in the tremendous developments in surgery of congenital heart disease might have resulted. In later years it became generally accepted that there should be almost no operative mortality in this condition, though it would still take a long-term study to show a beneficial effect on survival.

A valid multicenter study requires that the average operative mortality should be nearly as low as that reported by any major institution. The purpose of such investigation is not to test the skill of surgeons, but to determine the influence of the operation on the natural history of the disease. A poorly performed effective operation may affect survival in the same way as a well done ineffective one. Knowledge that a useful surgical procedure can be done at low risk is a powerful stimulus to surgeons to improve technique. This is the history of progress in surgery.

To achieve a low operative mortality, there can be little variability among participating institutions. The VA study shows a wide range of operative mortality. It has been stated that 10 of the 13 participating hospitals had an average mortality of 3.4%, which means that the remaining three must have had an average mortality of 12.9%. This is based on the assumption that there was an equal contribution of patients from each hospital, which was not the case.[13] If these three hospitals contributed less than an average number, the operative mortality would have been higher. One major participant had a 2% mortality.[16]

Such a wide range of operative mortality is unacceptable in a multicenter study. The low patient volume of many of the participating hospitals is certain to be associated with high operative mortality. Fifteen operative deaths were reported in 268 patients, but there was another "late operative death" which put the operative mortality at 6%.[17] In 1973, the middle

year of the VA investigation, our early and late operative mortality for 1,400 patients who had pure bypass operations (excepting left main coronary artery lesions, as in the VA study) was 0.6%.

A second index of the quality of bypass surgery is the incidence of perioperative myocardial infarction. The VA study reports 18%.[3] Our incidence was about 4%.

The third index is graft patency. The VA reports 69% patency whereas ours was 87%. In the VA series, 12% had no patent grafts; in our series, 6.4% of patients studied in 1974 had no patent grafts—catheterization being done a median interval of 13 months following operation. Excluding single grafts, failure of all grafts occurred in 10% in the VA study, compared with 1.7% in our series. If operative deaths were 6%, 12% had no patent grafts, and 2% were lost to follow-up, there remains only 80% that possibly could show increased longevity resulting from operation. Many of these may have had inadequate revascularization or perioperative infarction. Obstruction of single arteries is known to have a high survival for four years; consequently, the group having early potential for increased longevity is even smaller. Our low figures for operative mortality, perioperative infarction, and high graft patency are not unique. Others are reporting figures in the same range.

If operative mortality, perioperative infarction rate, and graft patency are important variables, improved rates for each should increase overall survival. In the VA study, the four-year survival for surgical patients (excluding patients who had obstructions of the left main artery) was 85%; our four-year survival was 95%.

Murphy et al[1] stated that there were 77 total deaths in the three-year VA survival study, but medical and surgical deaths were not separated. This figure was later corrected to 72 deaths.[10] There were 37 deaths in the medical group and 35 in the surgical group; 29 medical deaths and 15 late surgical deaths were thought to be cardiac or sudden.[17] In addition, 16 of the surgical deaths were operative, one medical death was of unknown cause, and the remaining deaths in both groups were considered noncardiac. Except for the operative deaths, cardiac deaths were about twice as common in the medical as in the surgical cases. A similar experience was reported by Mathur and Guinn[18] in a smaller prospective randomized study.

Obviously it should be easiest to demonstrate difference in survival in the highest risk medical group—obstruction of three arteries and abnormal left ventricle. Despite the adverse factors mentioned, it is apparent that this group is showing a widening gap between the VA medical and surgical survival curves. Apparently, according to Read's[3] report, this difference of 8% at four years favored surgical treatment. If this is true, the operation must have had an unfavorable effect of almost the same degree for patients who had obstructions of two arteries or of three arteries with normal left ventricular function because the other subsets had equal mortalities and the overall medical survival was slightly higher. Specific figures for the number of deaths and the number of surviving compliant patients at annual intervals would permit others to independently evaluate survival. The four-year surgical survival of three-artery obstruction in the VA series appears to be 84%, and ours is 93.5%.

The VA studies under discussion have excluded patients who had obstruction of the left main artery because a previous report[19] had indicated a clear superiority for surgical treatment. It would be peculiar if lesions of this artery behaved in a manner strikingly different from that of proximal lesions of its two major branches located 1 or 2 cm distally. The early prognosis is worse for left main artery lesions than for obstructions in three arteries, but the long-term prognosis is similar.[7,9]

Two VA papers have acknowledged the problems of operative mortality, perioperative infarction, and graft patency.[1,3] The suggested solution was institution of another prospective randomized study of medical and surgical treatment. It would not be possible to assemble a large series of patients if informed consent included the statement that surgical treatment decreases morbidity significantly as has been amply demonstrated in many studies. An even greater percentage of noncompliant medically randomized patients would be anticipated. The nonsurgical and surgical groups should be much larger if various subsets are to be analyzed in a

meaningful way. It is unlikely that such a study will be launched in a society in which freedom of choice of treatment is available. Finally, before another investigation is undertaken, it must be shown that the operative results are in the same range as those reported by the institutions having the best records.

A disturbing aspect of the VA study has been the rumor that one major participating hospital had an experience grossly different from the reported composite results. If this is correct, additional serious questions about the validity of the VA study would arise. Publication of variability of experience is an important part of any multicenter study. The mere breath of rumor that important nonuniformity exists requires detailed disclosure of the facts. Similar problems have plagued other multicenter studies. The question is too important to be dismissed with the statement that "no significant differences affecting survival have been noted to date."[13] Such a statement is not in conformity with the published fact that there was a wide variability in operative mortality alone.[13]

Returning to the two principal conclusions of the VA study, I believe that it has not been established that the natural history of coronary disease as defined in the study has changed in recent years, nor has it been shown that well performed surgical treatment is ineffective in altering the natural history of the disease. The work does demonstrate that multicenter randomized prospective studies of surgical treatment are difficult to monitor and control, especially if the operation being investigated affords striking symptomatic relief, and that excellent surgical results are an absolute necessity if the operation rather than the surgeon is to be evaluated.

References

1. Murphy ML, Hultgren HN, Detre K, et al: Treatment of chronic stable angina. N Engl J Med 297:621-627, 1977.
2. Detre K, Murphy ML, Hultgren H: Effect of coronary bypass surgery on longevity in high and low risk patients. Lancet 2:1243-1245, 1977.
3. Read RC, Murphy ML, Hultgren HN, et al: Survival of men treated for chronic stable angina pectoris. J Thorac Cardiovasc Surg 75:1-13, 1978.
4. Hultgren HN, Takaro T, Detre K: Medical versus surgical treatment of stable angina pectoris: Progress report of a large scale study. Postgrad Med J 52:757-764, 1976.
5. Detre K, Hultgren H, Takaro T: Veterans Administration cooperative study of surgery for coronary arterial occlusive disease. III. Methods and baseline characteristics, including experience with medical treatment. Am J Cardiol 40:212-224, 1977.
6. Webster J, Moberg C, Rincon G: Natural history of severe proximal coronary artery disease as documented by coronary arteriography. Am J Cardiol 33:195-200, 1974.
7. Bruschke AVG, Proudfit WL, Sones FM: Progress study of 590 consecutive nonsurgical cases of coronary disease followed 5-9 years. I. Arteriographic correlations. Circulation 47:1147-1153, 1973.
8. Bruschke AVG, Proudfit WL, Sones FM: Progress study of 590 consecutive nonsurgical cases of coronary disease followed 5-9 years. II. Ventriculographic and other correlations. Circulation 47:1154-1163, 1973.
9. Proudfit WL, Bruschke AVG, Sones FM: Natural history of obstructive coronary disease: 10-year study of 601 nonsurgical cases. Prog Cardiovasc Dis 21:53-78, 1978.
10. Murphy ML: Corrections. N Engl J Med 298:352, 1978.
11. Hultgren HN, Detre KM, Takaro T, et al: The VA cooperative study of coronary arterial surgery: Baseline characteristics of study population and survival in subgroups with medical versus surgical treatment. Prog Cardiol 6:67-79, 1977.
12. Hultgren HN, Takaro T, Detre K: Veterans Administration cooperative study of surgical treatment of stable angina: Preliminary results. Cardiovasc Clin 8:119-130, 1977.
13. Murphy ML, Hultgren HN, Detre K, et al: Letter to the editor. N Engl J Med 297:1470, 1977.
14. Proudfit WL, Shirey EK, Sones FM Jr: Distribution of arterial lesions demonstrated by selective cinecoronary arteriography. Circulation 36:54-62, 1967.
15. Gross RE: Surgical management of the patent ductus arteriosus. Ann Surg 110:321-356, 1939.
16. Hultgren HN: Controlled trials of coronary artery surgery: VA cooperative study. Presented at the 50th Scientific Session, American Heart Association, Miami Beach, Dec 1, 1977.
17. Murphy ML: Personal communication, 1978.
18. Mathur VS, Guinn GA: Sustained benefit from aortocoronary bypass surgery demonstrated for 5 years: A prospective randomized study. Circulation 56:abst 732, 1977.
19. Takaro T, Hultgren HN, Lipton MJ, et al: The VA cooperative randomized study of surgery for coronary arterial occlusive disease. II. Subgroup with significant left main lesions. Circulation 54(suppl 3):107-117, 1976.

The Scientific and Clinical Tribulations of Randomized Clinical Trials

ALVAN R. FEINSTEIN

The essence of tragedy has been defined as the destructive collision of two protagonists holding opposing positions, each of which is right. Most reasonable people will surely accept Dr. Chalmers'[1] contention that we need evidence rather than opinion to help settle the therapeutic dilemmas of modern medicine; and that the evidence is best obtained from well-designed controlled clinical trials in which the compared treatments have been assigned by randomization. On the other hand, it is also reasonable for Dr. Proudfit[2] to argue that the VA study of coronary artery surgery has not told us the effects of surgery performed with the very best of current techniques and skills; nor do the so-called life-table (or actuarial) analyses provide a clear account of what happened to whom and when; nor have the investigators successfully explained the inconsistencies found in results among the individual clinical centers that participated in the trial.

Rather than recapitulating the two sets of arguments that have been presented by Chalmers[1] and Proudfit,[2] I should like to return our focus to the larger theme of this symposium. The evaluation of coronary surgery is but one of several possible examples that illustrate the many general problems involved in "The Scientific Uses and Abuses of the Clinical Trial." I shall suggest that the problems are produced by inadequacies in our existing methods for designing large-scale clinical trials; and I shall propose some methodologic changes that may help improve some of the current imperfections.

If we want to determine the merits of a particular treatment, a demand of scientific reasoning is that Treatment A be compared concurrently against an alternative or control treatment, designated as B. The treatments are imposed upon two groups of patients, Group A and Group B, who are each in a baseline state. The outcome of treatment is then observed in both groups. Before this happens, the patients who are eligible for either treatment are evaluated and then assigned to get treatment A or B, thus forming the two treated groups. In ordinary clinical practice, treatment is chosen by the *ad hoc* judgment of the physician and by the patient's acceptance of that choice. These selective decisions can introduce major bias in the subsequent results. For example, if coronary patients in particularly poor clinical condition are excluded from surgery, the surgical group will regularly have better outcomes than the medically treated group, even if surgery accomplishes nothing. Conversely, if patients who are in particularly good and perhaps even asymptomatic clinical condition refuse to accept surgery, and are then counted in the medical group, a bias will occur in the opposite direction.

The main contribution of the randomized clinical trial is that treatment is assigned to the eligible patients in a formal random manner, thus removing any judgmental decisions that might bias the assignments. Before the eligible patients enter a randomized clinical trial, however, they must first be admitted—a process requiring them to fulfill a series of distinctive criteria, which make certain demands on the clinical, morphologic, co-morbid, or other characteristics of their baseline state. (The sequence of events is shown in Fig. 1.) The spectrum of eligible patients may be altered by these criteria into a different collection of patients, who then become admitted, randomized to receive the treatments, and studied to yield the results of the trial.

The selectivity of the admission process can create many problems in extrapolating the results, no matter what occurs in the trial itself. If the admitted patients do not suitably represent the eligible population, many clinicians may reject the results as not being pertinent to clinical practice in general or to their own patients in particular. Thus, people who disagree with the kind of coronary disease that was regarded as "operable" in the VA study, or who question the way the

Dr. Feinstein is Director of the Robert Wood Johnson Clinical Scholar Program, and Professor of Medicine and Epidemiology, Yale University School of Medicine, New Haven, Connecticut.

Requests for reprints should be addressed to Dr. Feinstein at the above address.

Presented at the AFCR Public Policy Symposium, 35th Annual Meeting, San Francisco, April 30, 1978.

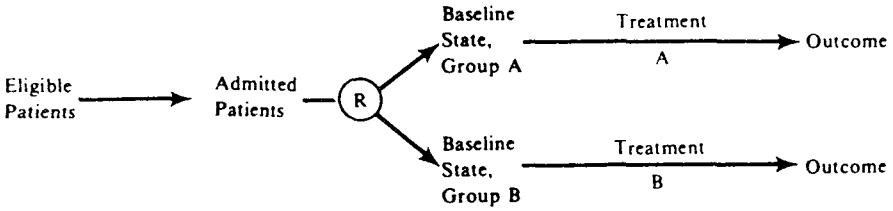


Figure 1. Architectural structure of a randomized controlled trial.

preadmission coronary arteriograms were performed and interpreted, may not want to accept the subsequent results, regardless of what happened after randomization. The problem of extrapolation is inevitable in any study of clinical therapy, with or without randomized assignment of treatment. Unless the admission criteria are acceptable, a reader may discard all the subsequent comparisons by saying, "You have not studied the kinds of patients I see in my practice."

For the rest of my comments, I shall concentrate on a separate set of problems that occur during the design, conduct, and subsequent analysis of the trial itself. These problems can divert the basic goals of the trial, distort its results, and create major controversies thereafter. Furthermore, the problems cannot be solved, and may even be worsened, by a devout belief that randomization and a large number of patients are a panacea for the scientific clinical challenges of therapy.

As shown in Table 1, for the results of a trial to be clinically pertinent for application to individual patients, the baseline state of the treated people must be suitably identified; the imposed treatment must be suitably proficient; and the outcome event must be suitably selected. For scientific validity in the comparison of treatment, the two groups at baseline must have equal susceptibility to the outcome event. The two imposed treatments should have an equal performance and the outcome events should be equally detected. Of these six clinical and scientific necessities, the act of randomization at best takes care of one. Within the limitations of chance, randomization may provide equal susceptibility in the baseline state of the two groups, but randomization alone does not deal with the other five challenges and problems.

Suitable identification of the baseline state requires attention to clinical and co-morbid distinctions, not just diagnosis alone or the paraclinical features found in technologic tests. Coronary artery disease, for example, has a diverse spectrum of severity in clinical and co-morbid patterns that are not suitably identified if patients are catalogued only according to hemodynamic measurements and the anatomic state of the coronary vasculature. Yet many important distinctions in clinical and co-morbid severity are customarily omitted from the baseline state identified in this and in other large-scale clinical trials; and the distinctions, if noted at all, are seldom used in specific analyses showing the outcomes for each of the distinctively different clinical and co-morbid groups.

To show that the compared patients have equal susceptibility to the outcome event, we must have appropriate determinations of what is meant by susceptibility—by demarcating strata (or subgroups) of patients with distinctly different prognoses (or risk). Furthermore, the post-therapeutic results should be analyzed according to strata with different prognoses.[3] For example, if patients with congestive heart failure or with major cardiac arrhythmias at baseline have worse prognoses than those who lack these features, the results should be presented separately for each treatment for the good-risk and poor-risk patients. Such separations according to well-chosen prognostic strata will improve both the comparison of results and their subsequent extrapolation. Prognostic stratifications have seldom been performed effectively because so many important features of clinical and co-morbid severity are omitted from the analyzed data.

Each of the treatments should be imposed with suitable proficiency, but the definition of *suitable* can be a major source of controversy. Does it mean *optimum* or *customary?* If optimum, each treatment should be given in the best way it can

Table 1. Clinical and Scientific Requirements in a Controlled Trial.

Requirement	Baseline State	Imposed Treatment	Outcome Event
Clinical pertinence for individual patients	Suitably identified	Suitably proficient	Suitably selected
Scientific validity for comparison of treatment	Equal susceptibility	Equal performance	Equal detection

be performed. If customary, the treatment is tested not in optimum circumstances, but in the way it might be given in ordinary, average clinical practice. The choice here is a matter of philosophy and has evoked many arguments, not just in the coronary surgery trial, but in many other studies as well. Critics of the VA trial of the coronary artery surgery study contend that the operations were not done with the best surgical skills currently available. Defenders of the study argue that those optimum skills are not available to all patients, and that the VA surgeons suitably represent the way the operation might be done at diverse institutions throughout the country.

Analogous controversies have arisen about clinical trials of oral agents such as anticoagulants, hypotensive drugs, and lipid-lowering treatment, where a suitable dosage of drug and regulation of a primary effect are needed to achieve a long-term goal. The controversy has been especially prominent for the UGDP study of diabetes, where the oral hypoglycemic agents were given in fixed, rather than flexible dosage, with no effort made to achieve an effective regulation of blood sugar.

Compliance in maintaining an oral agent is yet another aspect of its proficiency. For example, if blood pressure is not being effectively lowered, has the drug failed or has the patient failed to take it? The effects of different degrees of compliance with oral drugs, diet, exercise, and other regimens that must be self-maintained by the patient obviously have fundamental importance in evaluating those regimens, but the impact of compliance has only recently been well recognized,[4] and its scientific assessment is still relatively primitive in the quality of most methods used either to measure compliance or to analyze its results.

Regardless of what is meant by suitable proficiency, two principal treatments cannot be compared fairly unless they have been given with similar proficiency. If one treatment was maintained faithfully and with excellent regulation, we cannot properly compare its results against those of an alternative treatment that was maintained with ineffective regulation or that was not taken at all. This problem of achieving similar proficiency is particularly tricky when surgical and medical regimens are compared. If the surgery is performed in a suboptimal manner, what constitutes an equal reduction for optimal medical therapy? How do we arrange it quantitatively, and can we perform it ethically?

A separate problem arises from ancillary supportive treatment with additional drugs or other therapy beyond the principal agents under comparison. Any such ancillary treatment should have been given in a similar manner for the compared therapeutic groups. An important ancillary agent is the iatrotherapy of the physician himself. For example, if enthusiastic or charismatic support by the physician can have important effects on the patient's status, such support should be equally provided to patients receiving each of the compared treatments. Because neither the patients nor the doctors could be effectively "blinded" about the identity of therapy in the coronary surgery study, certain subjective improvements in the surgically treated group may be due to the personalities of the attending surgeons, who are often more enthusiastic, charismatic, and persuasive than their medical counterparts.

The event or events chosen as the main outcome of a clinical trial must be suitably selected to be clinically meaningful as well as statistically countable. Here we have the problem of "hard" and "soft" data.[5] Death is often chosen as *the* outcome event in large-scale trials because it is the hardest of data, easily measured and readily counted. Many patients and doctors, however,

may be primarily interested in the patient's changes in symptoms, functional capacity, and quality of life—but these entities are often dismissed as being "soft data," with no effort being made to establish effective observational techniques and criteria for identification that would "harden" the quality of the "soft" data and allow suitable analyses.

Even when symptoms, functional capacity, and quality of life are asked about, the inquiries are seldom conducted in a scientifically sophisticated manner. For example, when a patient with chronic angina pectoris says he has been able to go back to work after surgical treatment, we seldom determine whether he had not previously worked because of the pathophysiologic effects of the angina, because he was afraid to work, or because his doctor told him not to do so. A patient who is not working because of iatrogenic prophylaxis and who later returns to work because of surgical charisma may be falsely designated as improved due to a bypass graft whose main effect was to evoke enthusiastic iatrotherapy.

For equal detection of outcome events expressed in soft data, we can use double-blind observation for the results of oral agents that do not require regulation, and double-observer techniques, if dosage must be regulated. Although appealing, these tactics often may not work. From noting a physiologic side effect or directly tasting the drug, the patient or physician may be able to discern its identity. Here, too, we might be better off developing better ways to harden soft data and to improve precision in our observational methods, rather than relying on "blinded" observations performed with perforated masks.

For a fair comparison of the outcome events, the patients getting the compared treatments should receive similar surveillance, diagnostic test procedures, and interpretive criteria. For example, because silent gallstones are so common in an elderly population, any treatment that leads to gastric side effects and a GI workup with performance of cholecystography will be associated with an increased detection of these silent stones. The treatment may then be falsely accused of causing gallbladder disease unless patients getting the comparative treatment received cholecystography with the same intensity of surveillance. This type of quandary has arisen in another large-scale trial: the coronary drug project.

To help solve some of these problems, we must begin by noting that randomization, despite its major contribution to clinical trials, does not guarantee equal susceptibility at baseline, and cannot deal with the problems of extrapolating from the admission criteria and with the other five major difficulties in baseline state, performance, and outcome. Consequently, we shall need to give major attention to clinical, not just statistical, aspects of clinical trials. This activity will involve the development of what might be called "clinimetrics" for creating the indexes, scales, and criteria that will allow better identification of important clinical phenomena that are now omitted or inadequately managed in most statistical analyses. Finally, since a burgeoning technology will produce too many new therapeutic agents to allow each to be tested with a formal randomized trial, we shall need to make better use of the data acquired in natural therapeutic circumstances, and to get better data for those analyses.

The successful application of these principles might have reduced the amount of controversy that has developed about coronary artery surgery, and can eventually help provide a satisfactory answer to the questions that now remain incompletely resolved, despite the diligent work and laudable efforts of the VA investigators.

References

1. Chalmers TC: In defense of the VA randomized control trial of coronary artery surgery. Clin Res 26:230-235, 1978.
2. Proudfit WL: Criticisms of the VA randomized study of coronary bypass surgery. Clin Res 26:236-240, 1978.
3. Feinstein AR: Clinical Biostatistics. St. Louis, CV Mosby Co, 1977.
4. Sackett DL, Haynes RB: Compliance With Therapeutic Regimens. Baltimore, Johns Hopkins University Press, 1976.
5. Feinstein AR: Clinical biostatistics: XLI. Hard science, soft data, and the challenges of choosing clinical variables in research. Clin Pharmacol Ther 22:485-498, 1977.