

MEDICAL PRACTICE

Statistics in Medicine

Applying results of randomised trials to clinical practice: impact of losses before randomisation

MARY E CHARLSON, RALPH I HORWITZ

Abstract

The problem of generalisability in randomised clinical trials was highlighted by studies that entered only 10-14% of screened patients. To determine the magnitude and source of prerandomisation losses in clinical trials a survey was conducted of 41 trials listed in the 1979 inventory of the National Institute of Health. Two thirds of the trials maintained screening logs, but only half maintained any records of the number of patients who met the eligibility criteria but were not entered into the trial.

Among 21 trials (51%) that kept data on the number of patients who were eligible but not entered, losses of eligible subjects were attributable to refusals by patients in 25% and refusals by physicians in 29%. Other protocol requirements accounted for the remaining losses of eligible patients. Only a few trials documented the characteristics of patients who were eligible but not entered; in those trials the patients who were not entered were similar demographically but differed clinically from those enrolled. Thus minimising prerandomisation

losses of eligible patients requires the use of less restrictive criteria for entering patients.

Twenty four of the trials achieved 75% or more of their recruitment goals, eight between 25% and 74%, and six less than 25%. Among trials that screened less than twice their projected sample size, only three out of 13 (23%) achieved 75% or more of their recruitment goal. By contrast, 12 out of 16 trials (75%) that screened more than twice their projected sample size achieved 75% or more of their recruitment goal.

Screening large numbers of patients appears to be a pragmatic requirement for success in achieving recruitment goals; therefore, trials should not be criticised as lacking generalisability on that basis alone. The number and characteristics of eligible patients who were not entered, however, were documented by only a few trials; these data are critical in the assessment of generalisability. Additionally, the number of patients with the index disease who did not meet the eligibility criteria should also be documented. Together, these two types of data characterise the population to whom the trial results may be applied.

Introduction

Several published commentaries have urged investigators reporting the results of randomised clinical trials to include data describing the prerandomisation assembly of patients, especially the number screened but not randomised.^{1,2} These data were provided in the reports of two clinical trials, one on the effectiveness of timolol in patients with a recent myocardial infarction³ and the other on the use of aspirin and sulphapyrazone in patients with threatened stroke.⁴ In the timolol trial 17% (1884/11 125) of the screened population was randomised, while in the aspirin-sulphapyrazone trial 44% (585/1341)

Department of Medicine, Cornell University Medical College, New York, New York

MARY E CHARLSON, MD, assistant professor of medicine and Henry J Kaiser Family Foundation Faculty scholar in general internal medicine

Departments of Medicine and Epidemiology, Yale University School of Medicine, New Haven, Connecticut

RALPH I HORWITZ, MD, associate professor of medicine and Henry J Kaiser Family Foundation Faculty scholar in general internal medicine

Correspondence and requests for reprints to: Dr Mary E Charlson, Cornell Medical Center, 515 East 71st Street, New York, NY 10021, USA.

of the screened patients were randomised. As a result of these apparently large losses before randomisation, both trials were criticised as being insufficiently representative of all patients who had the study diseases.^{5, 6} Similar criticisms were directed at the Veterans Administration trial of medical versus surgical treatment of angina pectoris,⁷ in which 12% (685/5538) of the screened patients were randomised.^{8, 9}

The findings of these three trials are difficult to assess in the absence of data summarising the extent and the sources of prerandomisation losses in clinical trials. With the generalisability of these trials under attack, however, we were concerned that investigators may become reluctant to include data on prerandomisation losses in future trial reports. This study was therefore conducted to assess the process of patient assembly in recent randomised clinical trials and to devise a framework for evaluating subsequent trial reports.

Methods

ASSEMBLY OF TRIALS

A list of multicentre trials with coordinating centres was obtained from the 1979 inventory of clinical trials compiled by the National Institute of Health. Only randomised trials with control groups were studied. To be eligible for the survey the trial had to have a projected sample size of 250 patients or more, and three quarters or more of the projected study years must have been completed by January 1982. We also required that the purpose of the trial should be therapeutic or prophylactic—that is, not diagnostic—and that the study should have been conducted in North America.

Of the 156 multicentre trials with a coordinating centre, 105 were ineligible for inclusion in this study. The reasons for exclusion (listed in the order in which they were applied; only one reason listed for each trial) were: not randomised ($n=45$), no control (2), fewer than 250 patients (49), less than three quarters of the projected study years completed by January 1982 (6), conducted outside North America (2), and not therapeutic or prophylactic (1). A total of 51 trials therefore met the criteria for inclusion in this study.

DATA COLLECTION

A 12 item questionnaire was sent to the investigator listed as responsible at each of the 51 trial coordinating centres. (A copy of the questionnaire is available on request.) We assured the investigators that their trials would remain anonymous by undertaking not to identify the trials by name. There were 42 responses on 41 separate trials. In one instance two investigators provided data on the same trial; only the response from the person to whom the questionnaire was originally sent was used. Ten investigators did not respond despite repeated attempts to contact them. Of the 41 trials for which questionnaires were returned, three were still enrolling patients and data were therefore incomplete.

ANALYSIS OF DATA

Some of the questions were answered with percentages of patients rather than actual numbers—for example, randomised = 45% of screened. When percentages were provided the number of patients was calculated using the number of patients entered (which was equal to the number randomised) as the basis for the calculations.

Results

CHARACTERISTICS OF TRIALS AND SCREENING OF SUBJECTS

Table I gives the characteristics of the 41 trials. For each trial we recorded the principal disease or clinical disorder that was the focus of the trial, the secular period of the study, the projected sample size, and the number and percentage of patients finally randomised. Of the diseases included as the focus of the trial, cancer accounted for 13, cardiovascular disease for nine, neonatal disorders for four, and gastroenterological and pulmonary disorders for three each. Two

TABLE I—Characteristics of clinical trials

Trial No	Disease	Secular period of study	Projected sample size	No randomised	Projected sample size achieved (%)
1	Cancer	1978-81	250	?	?
2	Renal	1971-80	250	200	80
3	Infectious disease	1976-80	250	110	44
4	Cancer	1976-81	300	364	121
5	Pulmonary	1976-80	300	28	9
6	Neurological	1979-82	300	308	102
7	Cancer	1973-82	300	152	50
8	Pulmonary	1974-79	300	22	7
9	Cardiovascular	1972-81	300	45	15
10	Neonatal	1978-82	400	400	100
11	Cancer	1978-81	400	603	150
12	Cancer	1976-80	400	362	90
13	Cancer	1967-81	460	541	117
14	Renal	1967-81	500	450	90
15	Cancer	1976-82	500	43	9
16	Cancer	1979-82	500	204	41
17	Cancer	1973-82	500	654	131
18	Cancer	1979-82	500	870	174
19	Cancer	1974-82	600	324	54
20	Gastroenterological	1972-79	700	604	86
21	Cancer	1973-82	700	357	51
22	Neurological	1977-81	700	649	93
23	Neonatal	1976-81	800	694	87
24	Cardiovascular	1975-83	900	780	87
25	Cardiovascular	1973-80	1 000	?	?
26	Endocrine	1961-80	1 000	1 027	103
27	Gastroenterological	1973-83	1 000	916	92
28	Pulmonary	1976-83	1 000	985	99
29	Neonatal	1974-81	1 339	299	22
30	Cardiovascular	1977-81	1 500	644	43
31	Neonatal	1976-83	1 550	914	59
32	Gastroenterological	1979-82	1 600	2 225	139
33	Cancer	1974-82	1 600	400	25
34	Endocrine	1971-81	1 700	1 758	103
35	Cardiovascular	1973-84	3 810	3 810	100
36	Endocrine	1977-84	4 000	?	?
37	Cardiovascular	1974-80	4 524	4 524	100
38	Cardiovascular	1962-81	8 371	8 341	99
39	Cardiovascular	1973-83	10 940	11 386	104
40	Cardiovascular	1972-82	12 866	12 866	100
41	Radiological	1979-81	200 000	3 400	2

trials each studied disorders in nephrology, neurology, and endocrinology. The duration of the studies ranged from two to 19 years, over half of the trials lasting for more than seven years. The projected sample size varied from 250 to 200 000 patients; the median projected size was 700 patients. Overall the median number of patients randomised was 624.

The percentage of the projected sample size that was achieved (the number of patients randomised as a proportion of the planned sample size) varied from 2% (for a radiological trial with a projected sample size of 200 000) to 174% (for a therapeutic trial with a projected size of 500). Only 14 of the 41 trials enrolled 100% or more of their projected sample size; a further 10 trials entered at least 75%. Eight trials achieved between 25% and 74% of their projected size, and six randomised less than 25%. Thus there were 24 trials that enrolled at least 75% of their projected sample size and 14 that entered less than 75%. (In three trials data on enrolment of patients were incomplete at the time of the survey.)

RECORDS OF ASSEMBLY PROCESS

Of the 41 trials, 27 (66%) reported using a screening log, and two without such records provided estimates of the numbers of patients screened. Only 21 (51%) had recorded any data about the reasons that eligible patients were not entered. With the exclusion of the two multistage trials (trials 39 and 40), only 16 (39%) had complete data on patients eligible but not entered. Only 15 (37%) of the trials maintained both a screening log and detailed data on eligibility of patients. In eight of the trials investigators declared all patients refusing to participate as "ineligible"; these investigators also declared that 100% of the "eligible" patients were randomised. We think that this tactic is misleading.

ATTAINMENT OF SAMPLE SIZE

In most randomised trials the screening of potential candidates would continue until the projected sample size is attained, unless the trial is ended early either because the chosen level of alpha is reached before patient accrual is completed or because treatment is associated with a serious adverse side effect. Table II shows the relation between the number of patients screened and whether the trial reached its projected sample size. The analysis was limited to the 29 trials in

which complete data were available on the number of patients screened. So far as we know, none of these trials was stopped early.

Only two of the trials randomised more than 60% of the patients screened. One of them was a trial in neonates, in which 87% of the patients screened were randomised; the other trial, in neurological trauma, randomised 78% of the patients screened. This second trial (trial 6) was the only one in our survey that achieved its sample size without screening more than twice its projected sample size.

TABLE II—Relation of number of screened patients to achieving projected sample size ($n=29$)

Trial No	Sample size achieved (%)	Patients screened	Ratio of screened to projected sample size	Percentage of screened population who were entered
<i>>75% of projected sample size achieved</i>				
40	100	361 662	28.1	4
24	87	16 626	18.5	5
39	104	159 468	14.0	7
23	87	7 893	9.9	9
10	100	3 500	8.8	11
18	174	4 251	8.5	20
17	131	4 088	8.2	16
32	139	10 000	6.3	22
11	150	2 412	6.0	25
28	99	3 316	3.3	30
12	90	1 100	2.8	33
22	93	1 399	2.0	46
27	92	1 879	1.9	48
20	86	1 100	1.6	55
6	102	393	1.3	78
<i>25% to <75% of projected sample size achieved</i>				
16	41	3 078	6.2	7
21	51	4 184	6.0	9
30	43	5 073	3.4	13
19	54	1 204	2.0	27
7	50	567	1.9	27
3	44	258	1.0	42
33	25	1 100	0.7	36
31	59	1 050	0.7	87
<i><25% of projected sample size achieved</i>				
9	15	190	0.6	24
5	9	140	0.5	20
29	22	555	0.4	54
15	9	188	0.4	23
8	7	50	0.2	44
41	2	15 000	<0.1	22

TABLE III—Patient and physician sources of refusal (trials with complete data only; $n=16$)*

Trial No	Patients eligible	Patients randomised	Patients eligible not entered	Patient refusals	Patients withdrawn by physician
6	320†	308	12	12	0
8	22	22	0	0	0
9	83	45	38	?	?
10	750	400	350	50	10
11	844	603	241	241	0
12	402‡	362	40	40	0
14	474	450	24	12	0
23	1 066	694	372	235	125
24	2 066	780	1 286	360	887
28	1 933	985	948	531	0
29	398	299	99	19	10
30	1 122	644	478	146	127
31	1 038	914	124	28	0
32	4 018	2 225	1 793	626	36
33	444	400	44	22	22
41	7 500	3 400	4 100	410	1 640
Total	22 480	12 531	9 949	2 732	2 857

*Three trials with incomplete data and two multistage trials omitted.

†Actual number reported was 308.

‡Actual number reported was 400.

Sixteen trials screened more than twice their projected sample size. Of these, 12 achieved at least 75% of their projected size and none attained less than 40%. By contrast, of the 13 trials that screened less than two patients for each projected recruit, only three achieved at least 75% of their projected size, and six trials failed to reach even 25% of their sample size goal. The correlation between achieving the recruitment goals (number randomised/projected sample size) and the ratio of patients screened to the projected sample size (number screened/projected sample size) is mathematically obvious at the end of the trial. The screening strategy, however, is planned at the outset of the trial, before it is known whether the sample size is achieved.

Only trials that at the outset employed a strategy aimed at screening more than twice the number of patients that they wanted to randomise usually achieved their recruitment goals.

Overall, in the 29 trials included in this part of the analysis only 7% (41 244) of the screened patients were actually randomised. When the two trials with a multistage assembly process were excluded (trials 39 and 40), 19% (16 992) of the screened patients were entered into the trials. In order to assess the implications for the generalisability of the trial results the percentage of screened patients who were randomised must be interpreted in the context of the particular screening strategy. In particular, the issue is whether all the patients screened had the index disease. If large numbers of patients without the index disease were screened, then a small proportion of randomised patients would not pose any problem. If, however, most patients had the index disease, then a small proportion of patients randomised might impair the applicability of trial results. The question of how many screened patients had the index disease is often ignored in reports of randomised trials and was not addressed directly in this survey; none the less, it is a vitally important issue.

ELIGIBILITY

Characteristics of patients who were eligible but not entered—Seventeen trials collected data about the demographic characteristics of the patients who met all eligibility criteria for the trial but were not entered. Of these trials, 12 (71%) indicated that patients who were eligible but not entered were similar demographically to those entered into the trial. Two trials reported dissimilarities. Data describing the clinical characteristics of patients who were eligible but not entered were available for 14 trials. Only six of the 14 trials reported that the patients who were entered were clinically similar to those who were not entered, and in eight there were differences. In the three trials that provided specific information about the nature of the clinical differences, the patients who were entered had more severe illness than the patients who were eligible but not entered. Only six trials collected any follow up data about trial outcomes in patients who were eligible but not entered.

Reasons that eligible patients were not entered—Table III includes data from 21 of the trials that documented the specific reasons that eligible patients were not entered. In the 16 trials for which there were complete data, 22 480 patients fulfilled the stated eligibility criteria and 12 531 (56%) were entered. Of the 9949 subjects who were eligible but not entered, refusals by patients accounted for 2732 (27%) and by the patients' physicians for a further 2857 (29%). The remaining 41% of patients were "withdrawn" by the investigators for reasons unrelated to eligibility or refusal by the physician. In this analysis using patients eligible but not entered as the denominator, refusals by patients accounted for about one quarter of all losses before randomisation. We repeated the analysis using the eligible population as the denominator, excluding the trials with multistage assembly. In that analysis 12% (2732/22 480) of eligible patients refused to participate.

Discussion

The results of this survey provide new insights into the process of patient assembly for randomised clinical trials. We found that a large proportion of trials (66%) never achieved their projected sample size, and that the ratio of the number of patients screened to the number entered was strongly associated with whether the trial reached at least 75% of the original sample size objective. The responses to the survey also established that one third of the trials did not employ screening logs and that half failed to collect any data on eligible patients who were not entered. Finally, we were surprised to discover that refusals by patients accounted for many fewer non-participants in trials than has been suggested.¹⁰⁻¹² Rather, administrative requirements included in the protocols appeared to be a major reason why eligible patients were not entered into the trials.

The interpretation of the proportion of screened patients who were eligible depends in part on the actual screening strategy employed to identify potentially eligible patients.¹³ For example, a screening strategy that includes surveillance of all hospital admissions for potentially eligible patients will result in low proportions of screened patients who meet the criteria for

eligibility. In this setting many of the patients "screened" will not have the index disease under study. While it may be of interest to report the total number screened, studies that employ exhaustive strategies to identify eligible patients should not be unfairly judged as lacking generalisability.

With this consideration in mind, the proportion of patients with the index disease who are excluded from the trial by the eligibility criteria is clearly important, as are the reasons for the exclusions. For example, patients who have a definite indication for or contraindication to one of the treatments are necessarily excluded from trials. Data about the numbers of patients excluded for these reasons are useful in estimating the proportion of patients with the disease to whom the trial results may be extrapolated. A trial that excludes 75% of patients with any given disease is clearly quite different in its applicability from one that excludes 10%. Unfortunately, this is often not clear in published reports.

Aside from these mandatory exclusions, investigators may also exclude patients with comorbid diseases, with a poor overall prognosis, or patients with poor anticipated compliance. The purpose of such exclusions is to "enhance" the efficiency of the trial—that is, to increase the chance of finding a difference between the treatments if one exists. Trials that assess whether a treatment can work under ideal or restrictive conditions will have lower proportions of otherwise eligible patients with the index disease entered into the trial.¹⁴ Conversely, trials that assess all the clinical consequences, both good and bad, of treating an illness are customarily carried out as close to usual practice circumstances as possible. Such trials should have low proportions of patients with the index disease who are excluded. Although it is usually apparent whether a trial is planned with restrictive policies or with policies that replicate usual clinical practice, it is often difficult to assess whether a trial actually achieved its intended population, because the requisite data are often absent from the publication.

Although trial results are sometimes incorrectly extrapolated beyond the population actually studied, the investigators usually understand that the results can be applied only to patients described in the eligibility criteria. The patients actually entered into the trial, however, must be representative of the eligible patients. A critical question is whether the patients who were not randomised had similar susceptibility compared with subjects who were entered for all the outcome events under study. If not, the trial results may be difficult to extrapolate from the population outlined by the eligibility criteria. In this survey, while the eligible but not entered patients were similar demographically to those entered, they were less often similar with respect to clinical characteristics. Those not entered were less severely ill than those randomised in the trials that cited the specific differences. While it has been documented that participants differ from non-participants in surveys,¹⁵⁻¹⁷ only a few trials have documented the characteristics of patients refusing to participate.¹⁸⁻¹⁹

Although rarely done, follow up of patients eligible but not entered is especially important, because they may receive the treatment under study once the trial results are available. For example, if patients who refuse to participate differ from randomised patients in important prognostic features, then the trial results will have restricted applicability. If the numbers and characteristics of such patients are not documented, it may be impossible to decide to whom the results do apply.

Conceivably a trade off exists between refusals and drop outs,²⁰ depending on the qualitative aspects of the treatments under comparison. When the treatments differ qualitatively—for example, surgical versus medical—losses may be expressed as refusals. When the treatments are qualitatively similar losses may be expressed as drop outs. Investigators are generally aware of the potential for drop outs to bias the results and may make an effort to follow up such patients to the extent possible. Similar efforts need to be made to follow up patients who refuse to participate in the trial. While some patients may refuse absolutely to have any further contact with the study, a sub-

stantial proportion may consent to have their physicians supply some follow up information about the particular outcome events under study. Any data that document outcomes among this group of patients would be important in interpreting and applying the trial results.

Investigators who label patients who refuse to participate as ineligible may improve the appearance of their assembly data; however, this tactic significantly impairs our ability to understand the applicability of the results. The belief that substantial numbers of patients and their physicians refused to participate in clinical trials led one distinguished statistician to propose a new design for clinical trials,¹² in which only patients receiving new or experimental treatments are asked for their informed consent. The results of our study suggest that this new design may be addressing the wrong problem. We found that the largest losses before randomisation occurred as a result of the study criteria and not as a result of the refusal of patients or their physicians to participate. Half of the losses of eligible patients occurred because of the application of restrictive eligibility criteria. Hence the problem of impaired generalisability resulting from large losses before randomisation cannot be solved merely by efforts to encourage participation of patients and physicians through the use of alternative designs to limit the consent process. Minimising prerandomisation losses requires the use of less restrictive criteria for admitting patients. Improved strategies for identifying potentially eligible patients would also be important.

Finally, trials that do not screen more than twice as many patients as they require often do not achieve their projected sample size. Therefore, larger screened to projected sample size ratios are a pragmatic requirement for most trials. Criticising such trials as lacking generalisability may be unreasonable since the impact of the eligibility criteria may be far more important to the assessment of generalisability than the proportion of screened to entered patients. Entering a small proportion of eligible patients may impair the ability to apply the results to the population defined by the eligibility criteria. Failure to document the numbers and the clinical features of patients eligible but not entered poses a serious obstacle to interpreting the trial results.

References

- Gore S. Assessing clinical trials protocol and monitoring. *Br Med J* 1981;**283**:369-71.
- Hampton JR. Presentation and analysis of the results of clinical trials in the cardiovascular area. *Br Med J* 1981;**282**:1371-3.
- Norwegian Multicenter Study Group. Timolol induced reduction in mortality and reinfarction in patients surviving myocardial infarction. *N Engl J Med* 1981;**304**:801-7.
- Canadian Cooperative Study Group. A randomized trial of aspirin and sulfinpyrazone. *N Engl J Med* 1978;**299**:53-9.
- Mitchell JR. Timolol after myocardial infarction: an answer or a new set of questions? *Br Med J* 1981;**282**:1565-70.
- Whisnart JP. The Canadian trial of aspirin and sulfinpyrazone in threatened stroke. *Am Heart J* 1980;**99**:129-30.
- Detre K, Hultgren H, Takaro T, et al. Veterans Administration cooperative study of surgery for coronary arterial occlusive disease. III. Methods and baseline characteristics, including experience with medical treatment. *Am J Cardiol* 1977;**40**:212-24.
- Loop L, Proudfit W, Sheldon W. Coronary bypass surgery weighted in the balance. *Am J Cardiol* 1978;**42**:154-6.
- Proudfit W. Criticisms of the Veterans Administration randomized study of coronary artery bypass surgery. *Clinical Research* 1978;**26**:236-40.
- Fost N. Consent as barrier to research. *N Engl J Med* 1979;**300**:1272-3.
- Cancer Research Campaign Working Party in Breast Conservation. Informed consent: ethical, legal, and medical implications for doctors and patients who participate in randomised clinical trials. *Br Med J* 1983;**288**:1117-21.
- Zelen M. A new design for clinical trials. *N Engl J Med* 1979;**300**:1242-5.
- Lipid Research Clinics Program. Recruitment for clinical trials: the lipid research clinics coronary primary prevention trial experience. *Circulation* 1982;**66**:suppl.ii.
- Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979;**301**:1410-2.
- Bergstrand R, Vedin A, Wilhelmsson C, et al. Bias due to non-participation and heterogeneous sub-groups in population surveys. *J Chronic Dis* 1983;**36**:725-8.
- Greenlick M, Bailey J, Wild J, et al. Characteristics of men most likely to respond to an invitation to be screened. *Am J Public Health* 1979;**69**:1011-5.
- Criqui MH, Barrett-Connor E, Austin M. Differences between respondents and non-respondents in a population based cardiovascular disease study. *Am J Epidemiol* 1978;**108**:367-72.
- Jackson F, Perrin E, Smith A, et al. A clinical investigation of the portocaval shunt. *Am J Surg* 1968;**115**:22-42.
- Conn H, Lindenmuth W, May C, et al. Prophylactic portocaval anastomosis. *Medicine (Baltimore)* 1972;**51**:27-40.
- Greenland S. Responses and follow-up bias in cohort studies. *Am J Epidemiol* 1977;**106**:184-93.

(Accepted 17 August 1984)