ıl Commission for the Protection of Human Subjects of Biomedical and ehavioral Research (1977). *Report and Recommendations: Research ıvolving Children.* DHEW Pub. No. (OS) 77-0004.

ıl Commission for the Protection of Human Subjects of Biomedical and ehavioral Research (1978a). Appendix to *Report and Recommenda-ions: Institutional Review Boards.* DHEW Pub. No. (OS) 78-0009.

ıl Commission for the Protection of Human Subjects of Biomedical and ehavioral Research (1978b). *The Belmont Report: Ethical Principles ıd Guidelines for the Protection of Human Subjects of Research.* DHEW ıb. No. (OS) 78-0012.

al Commission for the Protection of Human Subjects of Biomedical and ehavioral Research (1978c). *Report and Recommendations: Institu-ional Review Boards.* DHEW Pub. No. (OS) 78-0008.

berg Code (1948). In *Experimentation with Human Beings,* J. Katz. Russell Sage Foundation, New York, 1972.

y, P. (1970). *The Patient as Person.* Yale, New Haven, Conn.

y, R. J., Alford, C. A., and the NIAID Collaborative Antiviral Study roup. (1979). Encephalitis and adenine arabinoside: An indictment ithout fact. *Hastings Center Rep.* 9(4):4, 44–46.

M. (1979). A new design for randomized clinical trials. *N. Engl. J. Med.* 00:1241–1245.

II-7

4
Exclusions, Losses to Follow-Up, and Withdrawals in Clinical Trials

Peter Armitage University of Oxford, Oxford, England

Criteria for Eligibility

If the outcome of a clinical trial is to provide any guidance for future medical practice, the investigators must describe clearly a number of important features. These include the types of patient under study, the precise nature of the treatments used, and the response variables measured. These topics should be covered in the protocol (see Chapter 6).

In particular, a careful description is needed of the clinical categories to be admitted into the trial. These will usually be a diagnostic classification, such as peptic ulcer, breast cancer, etc., although occasionally broader diagnostic categories may be permitted, such as patients with abdominal pain. More frequently, though, the classification is narrowed by the imposition of further restrictions. Thus, the disease may be qualified by site of action, stage of advance, pathological test results, duration of symptoms, and so on. The patients may be selected by exclusion of certain age groups, presence of other specified medical conditions, etc.

As Sackett points out (Chapter 2), the clinical criteria for inclusion in the trial must strike a balance between efficiency and generalizability. From the point of view of efficiency, one might try to approximate as closely as possible the conditions of a laboratory experiment with homogeneous material. This would involve tight restrictions on the clinical categories included, with

perhaps consequent difficulty in finding an adequate number of patients to provide a sufficiently precise comparison of treatments. Moreover, by striving for an efficient solution to a narrow question, the investigators may be unable to generalize the results to relate to a much broader class of patients. It is common experience to find that the results of a well-conducted trial are harshly criticized on the grounds that the protocol was too restrictive to be of general interest.

An alternative approach, then, is to relax the entry criteria, permitting the inclusion in the trial of patients with widely differing initial characteristics. This might seem to be a prescription for a poor experiment, since the "material" would be very heterogeneous. The answer to this criticism is that in the analysis of the data the heterogeneity can be allowed for, so that comparisons of treatments effectively become subject to the random variability found within subgroups defined by base-line variables of prognostic importance (see Chapters 1 and 7). If a comparison is required for a specific subgroup alone, it can be provided as if there had been a separately conceived minitrial. It follows that if such minitrials are to be based on adequate numbers of patients, the whole trial must involve many more patients than if a single subgroup were used. This consideration will lead to a discussion of the size of a trial which falls outside the scope of this chapter. If the total number of patients is strictly limited by available resources it may be wise to use narrowly defined criteria rather than the broader approach outlined earlier in this paragraph.

A further advantage of permitting a broad diagnostic range is that this may permit treatment to be started earlier in the course of the disease than would otherwise be so. In a trial for the secondary prevention of myocardial infarction in patients who have experienced a first infarction, the question of interest may be what treatment should be applied very soon after the first attack, say on admission to an intensive care unit. Full diagnostic details may not be available until a much later stage. A trial of early treatment could therefore encompass a variety of diagnostic conditions, but if it were desirable to examine specific subgroups separately this could be done in the subsequent analysis.

A similar choice between narrow and broad criteria may face the investigators in defining the treatments to be used. Should the dosages and frequency of administration of drugs be rigidly controlled, or should they be allowed to vary in some way that more closely simulates normal medical practice? Should an attempt be made to standardize other procedures that might be relevant—the use of other medicaments, type of nursing care, diet, etc.? When all these matters have been resolved, there will appear in the protocol descriptions of the types of patients and forms of treatment under study. However, it is extremely unlikely that, in any but the smallest trials, there will be complete adherence to the prescribed rules. It is the purpose of this chapter to describe some of the

reasons for these protocol departures, to discuss their likely effect on the results and to offer some suggestions for reducing the difficulties they cause.

The Patient Log

The patients entered into a trial will be selected from a larger group of patients who are initially considered for possible eligibility. This larger group will fall into the broad diagnostic category under study, although its precise boundaries may elude definition. Some subgroups of disease, or other categories such as male or female, may be known to be excluded from the trial so that it may be arbitrary whether they are said to have been considered or not. Nevertheless it is important that a record be kept of the wider group of patients considered for admission, and of the reasons for each instance of exclusion. These records will not be analyzed as part of the trial itself, but the information they contain may help the investigators and other workers understand the degree of selection involved in choosing the trial population, and they will provide a check on the extent to which patients who were eligible were in fact admitted (see Chapter 5). Moreover, the mere existence of a patient log will provide an incentive for the center to admit patients at the maximum rate, at times when enthusiasm may have waned or administrative difficulties arisen.

The crucial event, regarding the exclusion of patients, is the act of random assignment of treatment for a particular patient. If a patient is initially considered to be eligible but for some reason is withdrawn from consideration before randomization, that patient is merely added to the list of excluded patients on the log. Any questions of exclusions *after* randomization give rise to new problems, since such exclusions might detract from the benefits conferred by randomization. It is these postrandomization problems that have to be considered in the remainder of this chapter.

Protocol Deviations

It is doubtful whether there has ever been a large-scale trial in which all patients at all times, received treatment in full accordance with the protocol. As we shall see, protocol deviations create difficulties in the interpretation of results of trials, and should therefore be reduced to a minimum. It is usually advisable in the protocol to exclude from the trial patients who, for specific medical or social reasons or because of personality characteristics, seem likely to provoke protocol deviations. Even if this is done, circumstances will arise in which, for particular patients, some form of deviation cannot be avoided. It should be a major aim, in the drawing up of the protocol, to minimize the frequency of these future occurrences. Difficulties are particularly likely to arise if the clinical measures to be taken are so nonstandard or so rigid that few doctors will

perform them in the manner prescribed; if the administrative requirements are very complex; if unusual demands are made on the patient's willingness to cooperate; or if the diagnosis categories or therapeutic regimes are so imprecisely defined that misunderstandings arise. Simplification and clarification of these aspects of the protocol are likely to reduce the disparity between what should be done and what is done.

Some Types of Deviation

Incorrect Diagnosis. For some patients the diagnosis may be changed as a result of normal diagnostic procedures carried out after randomization. It is clearly undesirable for a trial to include patients not suffering from the disease ostensibly under study, and every effort should be made in construction of the protocol and in the execution of the trial to ensure that only patients with a clear diagnosis are entered. If, nevertheless, such diagnostic changes do occur, should the patients be omitted? The problem is that the opportunity to make diagnostic changes may differ appreciably from one treatment group to another. Consider, for instance, a trial like that of the Medical Research Council (1966) to compare surgery with radiotherapy for the treatment of operable lung cancer. The qualification "operable" is necessary here to ensure that any patient could receive whichever of the two treatments he or she might be allotted. It is likely that in the group assigned to receive surgery, a certain proportion will be found at an early stage of surgery to have an inoperable tumor which will therefore not be excised. In the group assigned to radiotherapy there will be no such opportunity for the misdiagnosis to be detected, although a similar but unidentifiable proportion will in fact have inoperable disease. If the inoperable cases were removed from the surgical group, the residual group would have a different distribution of characteristics from the total surgical group; the patients removed would be likely to have poorer prognosis than those remaining. It would, therefore, be unfair to compare the residual surgical group with the total group receiving radiotherapy. The only fair comparison is between the total groups assigned to the two treatments, including the wrongly diagnosed individuals.

We shall discuss the merits and drawbacks of this sort of comparison later; we merely note here the potential danger of following any other course.

Misdiagnosis is not always as serious a problem as in this example. A change of diagnosis may occur as a result of a test on a specimen taken shortly before randomization. The test result will be entirely unaffected by the choice of treatment, and no harm is done by excluding such cases. The problem could be avoided, of course, by delaying randomization, and therefore the start of the selected therapy, until after the test result becomes available, but this may often be inadvisable medically.

In trials involving varying lengths of survival of patients, a change of diagnosis may occur more often on one treatment than another if the first

treatment prolongs survival and therefore allows more information about the disease to become apparent. As Peto et al. (1976) point out, such patients can again be omitted if it is clear that differential rates of diagnostic change would be unlikely if the treatments were equally effective.

Other Reasons for Ineligibility. Similar arguments to those in the preceeding paragraphs apply when forms of ineligibility other than incorrect diagnosis are discovered after randomization. Since eligibility normally depends on factual information supposedly available at the time of randomization, a late verdict of ineligibility will usually be occasioned by the discovery of an error in the patient's records, by an oversight in the process of screening the patient, or by a later test result. It will usually be fairly clear that such events are uninfluenced by the choice of treatment. In such cases the patients can safely be excluded from the trial.

In some situations, though, the eligibility criteria in question may involve subjective judgments (such as whether there is a past history of certain rather ill-defined conditions, or whether the physician considers that the patient's condition will not deteriorate rapidly during the course of treatment). In these instances the judgment might well be influenced by a knowledge of the treatment to be applied, and the patients concerned should be retained in the trial. If there is any doubt on this question, the safest policy is to keep the patients in the trial.

Departure from Ideal Treatment Schedule. We now consider the most common form of protocol deviation, and that which causes the most difficult problems of analysis and interpretation: some patients may fail to receive their treatment in the precise form in which it is described in the protocol.

Such deviations from the ideal schedules of administration may occur for any of the following reasons:

1. A refusal by the patient to start or continue the assigned treatment, perhaps because of side effects or a belief that the treatment is ineffective.
2. A failure to comply with detailed instructions, for example as to drug dosage, or to attend for examination when requested to do so.
3. A change of treatment imposed by the physician for clinical reasons, usually occurrence of adverse effects, deterioration of the patient's illness, or the presence of additional disease.
4. An administrative error. In its most extreme form this may be the implementation of the wrong treatment altogether, but this should be, and is, rare. More commonly some detailed aspects of the schedule may be omitted; for instance, the doctor may fail to see the patient at the correct intervals during follow-up.

The general question that arises is whether such patients should be considered with the rest of their group for purposes of analysis, or whether they

should be omitted altogether or perhaps for that part of their treatment period when deviations occurred. If it happened that the patient received, instead of the treatment assigned, one of the rival treatments under comparison in the trial, it might even be argued that his or her records should be considered as part of the group assigned to this rival treatment.

To those accustomed to working under laboratory conditions, in which technical errors are eliminated as far as possible before an experiment is regarded as satisfactory, the natural course of action may seem to be to remove these instances of imperfect performance from further consideration. There are, unfortunately, serious drawbacks to such a proposal. The basic problem is that the patients for whom such protocol deviations occur are unlikely to be representative of the whole group of patients under study. Moreover, and this is the crucial point, the nature of the nonrepresentative selection may differ from one treatment group to another. For example, deviations of type 3 may be predominantly caused by adverse effects, and these may be more common in one treatment group than another. They may occur mainly on patients who suffer from a more severe form of the illness or who are older than the average, or they may occur in some other nonrandom way. To remove them would leave the groups unbalanced with respect to the initial characteristics of the patients: the benefits of randomization would be lost. Even if the number of patients and the types of deviation involved were similar in different treatment groups, the types of patients incurring these deviations might be quite different.

Some of these points are illustrated by the trial of coronary-artery bypass surgery in stable angina pectoris conducted by the European Coronary Surgery Group (1979) (see Table 1). Patients were assigned randomly to receive either medical or surgical treatment. Of 373 patients assigned to medical treatment, 50 in fact had surgery because unacceptable symptoms arose which could not be

Table 1 *Coronary-Artery Bypass Surgery in Stable Angina Pectoris: 2-Year Fatality Rates*

	Assigned surgical treatment			Assigned medical treatment		
	Actual medical treatment	Actual surgical treatment	Total	Actual medical treatment	Actual surgical treatment	Total
Survivals	20	354	374	296	48	344
Deaths	6	15	21	27	2	29
Total patients	26	369	395	323	50	373
Fatality rate (%)	23.1	4.1	5.3	8.4	4.0	7.8

Source: European Coronary Surgery Study Group (1979).

alleviated by medical therapy. Of 395 patients assigned to surgical treatment, 26 failed to receive surgery because of death before surgery could be started, medical contraindications of various sorts, or refusals. Thus, there were fewer changes of treatment in the group assigned to surgery, but the death rate within this small group of patients was high. It could certainly not be regarded as comparable in prognosis to the group switching from medical treatment to surgery. If, in the analysis of the data shown in Table 1, one concentrates on the selected subgroups with adherence to the assigned treatment (369 on surgery and 323 on medical treatment), the fatality rate would appear to be much lower on surgery (4.1%) than on medical treatment (8.4%). But this residual group on surgery has had removed from it a subgroup with particularly high mortality and is likely to be more favorable in prognosis than the residual medical group. The only safe comparison is on the two complete groups, the comparability of which is insured by randomization. The difference in fatality rates, between 7.8 and 5.3%, is much less striking and is not significant.

At first thought it may seem foolish, in a trial designed to compare treatments A and B, to contrast a group of patients mostly but not wholly receiving A with a group mostly but wholly on B. What we are doing in such a procedure is comparing the *policy* of using A where possible with the *policy* of using B where possible. If the intended treatments are always used there is, of course, no problem. If they are rarely used the trial will carry little information about the true effects of A and B but a great deal of information about the difficulties of attempting to use them. The usual situation is intermediate; the example above shows that even when the deviations are relatively infrequent their effect might be considerable.

This approach to therapeutic trials, of comparing policies or intentions rather than rigorously standardized regimes, is more than a mere verbal device for turning aside objections. It seems to many investigators to be a more realistic statement of the purposes of the investigation. It accords with what Schwartz and Lellouch (1967) called the pragmatic attitude to a clinical trial, as distinct from the explanatory attitude, and to what Cochrane (1972) called a measure of effectiveness rather than efficacy. As the latter two terms are almost synonymous it seems preferable to use the former pair. Sackett uses the term management trial to describe one in which pragmatic aims are foremost (Chapter 2). In this approach an attempt is made to reproduce the sort of choices which confront a medical doctor in the course of his or her regular practice. The doctor does not ask: "Would it be better to use A (regardless of the many types of difficulty which would normally lead me or the patient to change the therapy) or B (similarly applied)?" He or she asks, rather: "For patients of the type under consideration, would it be better to start with A (with the intention of continuing this therapy if possible but the willingless to be flexible) or to start with B (and have the same intentions)?"

In this sense, the pragmatic attitude of comparing the total groups formed by randomization represents a realistic clinical choice. It is true that the particular types of deviation from standard treatment, and their frequency, may be rather different in a trial and in uncontrolled practice. In particular, many doctors in ordinary practice would switch from one treatment to another, if progress was disappointing, more frequently than is usual in a trial. Nevertheless, the pure comparisons of completely standardized treatment regimes are artificial ideals, and there seems little point in regretting the near impossibility of achieving them in clinical trials.

Many of the points raised in this chapter are discussed in a useful paper by Gent and Sackett (1979); see also Chapter 2. The inclusion of the protocol deviants in the groups to be compared is safe in that it preserves the safeguard of randomization, but it will usually provide a biased estimate of the pure contrast which would have been achieved if there were no deviations. Often, but not always, the bias will be in the direction of reducing the difference between the mean responses to different treatments; the method could then be regarded as conservative. This would be so, for instance, if the comparison were between two treatments, A and B, of widely differing efficacy, and the deviations consisted of using the rival treatments in a certain proportion of cases. The difference between two-thirds A plus one-third B and two-thirds B plus one-third A would be likely to be less than between A and B. The contrast is diluted. In electing for a cautious, pragmatic approach, therefore, one might be losing sensitivity. A real difference in efficacy, if it exists, might reveal itself more clearly in a comparison of fully adherent subgroups than in the total comparison. This is quite true, but the observation is of limited value in view of the untrustworthy nature of the selective comparison.

Deviations from ideal, standardized treatment schedules need not always be regarded as protocol deviations, since in many trials the investigators will adopt the pragmatic attitude in writing the protocol and will explicitly permit specified variations in therapy to occur under defined circumstances. In a trial to compare penicillin with tetanus antitoxin in the prevention of tetanus among wounded patients, it might be recognized at the outset that some patients allocated to receive penicillin would be hypersensitive and would need to be given antitoxin instead; while some patients due to receive antitoxin would be hypersensitive to a test dose and would instead receive penicillin. Again, the comparison is between two strategies, of basing prophylaxis on either penicillin or antitoxin, and this precisely represents the choice confronting the doctor in the casualty department.

Losses to Follow-up. Many trials involve prolonged periods of follow-up, often requiring regular attendance of patients at clinics, when clinical observations and test measurements may be made. Even in the best-regulated trial some of the follow-up data may be missing because the patient missed one or

more of the scheduled visits. Or the visits may have taken place at times differing to some extent from those required by the protocol.

Suppose the investigators have decided to adopt the pragmatic approach and compare responses for the whole groups allotted to different treatments. There will still remain a problem in that, for certain measures of response, some of the required information will simply not be there. In many situations there is no alternative to leaving the patients with missing data out of the analysis, although the omission should be noted in any relevant tables. If the absence of data is caused by the death of the patient at an earlier time, or by a known exacerbation of the illness, it may be reasonable to count such patients as having extreme responses. It may then be useful to employ methods of statistical analysis in which the observations are ranked in order of magnitude. This approach permits these missing observations to be retained and to be recognized as being more extreme than the rest of the data, without the need to allocate a numerical value to them.

In some other situations, where the omission is thought to be due to reasons unconnected with the patient's clinical state, it may be possible to estimate the missing reading from those at neighboring points of time. This will usually be sensible only if the responses for a given patient are reasonably stable. There is clearly some risk in doing this, and common sense must be a guide. In any event, the use of such interpolations must be made clear in the report.

It is easy to overlook the serious bias that can result from the use of incomplete follow-up data. Patients who fail to provide information at any particular follow-up examination are very unlikely to be representative of all patients in the trial. They may be suffering from an acute exacerbation of their illness or they may have a chronic illness of greater severity than most of the other patients. In the former case, the technique of interpolation suggested in the last paragraph might be grossly misleading. If the number and type of patients absent on any one occasion were similar for each of the treatments, then the omission of their records would provide little or no bias in the treatment comparison, although it would reduce the precision of the comparison. However, one can rarely be sure that such similarity will prevail. It is only too likely that some absences are related to characteristics of the treatments, such as side effects. Indeed, if the treatments differ appreciably in their beneficial effects it will not be surprising to find a consequent difference in the proportion of seriously ill patients unable to attend for examination.

That this is not a trivial difficulty can be seen from a hypothetical but typical example. Suppose that two treatments A and B are being compared for their effect in preventing deterioration of patients suffering from a neurological illness. Hypothetical results at an examination 6 months after start of treatment are shown in Table 2. Of the patients who failed to report for examination a

Table 2 Hypothetical Results at an Examination 6 Months after Start of Treatment, in a Trial of Treatments to Prevent Neurological Deterioration (Proportions of Patients Showing Deterioration)

6-Month examination	Treatment A	Treatment B
Present	28/70 (40%)	36/90 (40%)
Absent	27/30 (90%)	9/10 (90%)
Total	55/100 (55%)	45/100 (45%)

high proportion experienced deterioration; there were many more such patients on A than on B. The overall comparison of the complete groups would have been in favor of B, yet in the subgroups actually seen at this examination the same proportion (40%) showed deterioration on A as on B.

Clearly, then, every attempt should be made, in planning the trial, to ensure that losses to follow-up are kept to a minimum. When the critical end point is death, it is possible in some countries (e.g., the United Kingdom) to make special arrangements for the national death registration system to notify the trial organizers of all deaths, and their causes, incurred by patients enrolled in the trial. This crucial information will then be available even for patients with whom contact has been lost.

Secondary Analyses

Potentially Irrelevant Events or Time Periods. The view taken in the preceding sections is that the main comparisons of results in clinical trials should normally be between the total groups of eligible patients assigned to different treatments. Naturally, the various types of protocol deviation should not be submerged entirely in their parent groups. The results for these subgroups should be displayed separately and scrutinized carefully for possible interpretations. These will normally be more speculative than interpretations based on the total groups, because of uncertainty about the comparability of treatments within subgroups. Nevertheless, the data should be presented as a secondary or subsidiary part of the analysis.

One question that may well arise is whether certain critical events should be excluded because it is known or suspected that their incidence would be quite unaffected by treatment. For example, if the trial concerns a specific disease with a high fatality rate, should one ignore deaths from other causes? The safe answer is that one should not do so, since one or more of the treatments might have an unsuspected effect on other causes of death. Moreover, an effect on the disease under study might so change the patient's life style as to modify or her risk of death from other causes. This view is purist but

sufficiently cogent to make it important to publish a comparison of deaths from all causes. However, the effect of treatments on other causes is likely to be small, and special interest will lie in the specific comparison which should certainly be made in a separate analysis. It should always be remembered in this connection that the attribution of causes of death is far from perfect, and this is an additional reason for comparing deaths from all causes, particularly when the allocation of causes is affected by a knowledge of the treatment used.

In many trials it is suspected that critical events occurring during certain time periods are very unlikely to be affected by choice of treatment. For example, it may be believed that a drug, the administration of which started shortly after randomization, could have no pharmacological effect for several days. If this were true for all the treatments under consideration, deaths (or other relevant events) during the first few days would be randomly distributed among the groups and could be disregarded in any comparison. Although it will usually be sensible to present secondary analyses omitting these early events, it will be unwise to base the principal analysis on such comparisons. There will usually be some degree of uncertainty about the early effects of drugs, and omission of early events might cause some bias. It is, of course, essential in any comparisons on the reduced data that the omitted time periods be the same for all treatments. If a placebo group is involved, the same initial period must be discounted for patients treated with placebo as for those receiving an active drug.

One may also wish to study the responses of patients who remain fully compliant with the protocol, for example by omitting critical events occurring during periods of withdrawal from the assigned therapy. For drug treatment there will usually be the possibility of a carry-over effect for a vaguely defined period after cessation of therapy, and events occurring during some defined period after withdrawal should be included. The potential dangers of omitting critical events will be clear to the reader (see Departure from Ideal Treatment Schedule under section Protocol Deviations). The patients in one group for whom certain time periods are omitted may be different in their initial characteristics from those similarly omitted from other groups. The comparison will then be biased. Nevertheless, such restricted comparisons are justifiable and perhaps valuable as secondary analyses, suggesting various hypotheses but not aspiring to provide conclusive evidence.

Some of these issues were illustrated by the reports of the Anturane Reinfarction Trial Research Group (1978, 1980), which involved a comparison of Anturane (sulfinpyrazone) and placebo in the prevention of secondary attacks of myocardial infarction in patients who had already had at least one infarction. The interim (1978) report paper was followed by correspondence in the *New England Journal of Medicine* [*298* (1978): 1257-1259], in which some of the points now under discussion were raised; the final report (1980) stimulated further discussion, including a paper by Temple and Pledger 80) explaining

the views of the Food and Drug Administration. The trial investigators decided at the outset to exclude events occurring either in the first seven days of treatment or more than seven days after withdrawal of assigned treatment. Some deaths occurred in patients who were regarded as ineligible, since the eligibility criteria had not been satisfied. Table 3 shows the numbers of patients and numbers of deaths in the various categories. The numbers of patients failing to complete their course of treatment were very similar in the two groups (220 on placebo and 195 on Anturane), and the report states that the reasons for discontinuation among the dropouts were similarly distributed in the two groups. These data exemplify the importance of deciding on the correct form of analysis. The 57 deaths among the excluded patients are crucial to the interpretation of the study. If they are included, the numbers of deaths in the two groups are 89 and 74; without them they are 62 and 44, a much more significant contrast. A safe approach to the data provided by this report would be to regard the comparison made after exclusions have taken place as a secondary analysis of considerable interest and suggestiveness, and to base the main conclusions of the trial on the total figures, relying on subsequent follow-up data and on other trials to confirm the apparent effect if it truly exists.

In the interpretation of secondary analyses, made after exclusion of some patients or events, much depends on the similarity or difference of the residual groups. This can sometimes be studied by the recording of known prognostic factors at the time of randomization, and by the comparison of these between the protocol deviants and the rest, between the deviants in different treatment groups, and between the residual members of these groups. Close similarity in the distribution of such factors between the deviants in different groups is no absolute guarantee of comparability (because the response will be affected

Table 3 Numbers of Patients in Anturane Reinfarction Trial, Showing Numbers of Deaths Included and Excluded

	Placebo			Anturane		
	Number of patients	Deaths		Number of patients	Deaths	
		Included	Excluded		Included	Excluded
Ineligible	33	–	4	38	–	10
Eligible	783	62	23	775	44	20
Total randomized	816	62	27	813	44	30

Source: Anturane Reinfarction Trial Research Group (1980). Reprinted by permission of the *New England Journal of Medicine*.

by unknown as well as by known factors), but it may provide some degree of confidence in the secondary analyses.

Analysis by Levels of Adherence. When the principal analysis suggests an appreciable difference between treatments in either therapeutic response or incidence of adverse effects, it may be useful in a secondary analysis to study the extent to which these differential effects appear to vary with the levels of adherence to the various treatments.

Some methods of conducting such analyses of adherence are outlined in a report of the Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents (Gilbert et al., 1975). This committee was set up to review the findings of various controlled trials using oral hypoglycemic agents, in particular the University Group Diabetes Program (UGDP, 1970). The UGDP trial for the treatment of adult-onset diabetes apparently disclosed an excess risk of cardiovascular death among patients taking the drug tolbutamide. In groups of similar size there were 26 such deaths among patients receiving the drug, and 10 in the group concurrently receiving placebo. However, there were considerable departures from the assigned schedules of administration, and it seemed worth asking whether the disparity in cardiovascular deaths could in some way be related to this marked noncompliance. (It is, of course, difficult to see how noncompliance could have been the sole cause for the contrast if the drug carried no excess risk.) The committee reanalyzed the data, taking account of the length of time that each patient was on each of the treatment schedules (of which there were two others besides tolbutamide and placebo). By a method they called relative allocation, deaths and periods of exposure were reassigned to different treatment groups in proportion to the fractions of the observation period that each patient spent on the different regimes. The results are summarized in Table 4. The cardiovascular death rate

Table 4 UGDP Data: Cardiovascular Fatality Rates with Relative Allocation of Deaths and Patients with Incomplete Adherence or Dose Modification

Treatment to which deaths and patients are allocated	Fatality rate during trial
Tolbutamide	
100% adherence to assigned dose	16/77 (21%)
Relative allocation to tolbutamide	7.6/95.8 (8%)
Other treatments	
Relative allocation to other treatments	37.4/650.2 (6%)
Total	61/823 (7%)

Source: Gilbert et al., *Journal of the American Medical Association*, *231*:583-608, © 1975, American Medical Association.

was particularly high among those patients who remained throughout the period on the assigned dose of tolbutamide. Those who moved to a modified dose, or to another treatment altogether, had a cardiovascular death rate nearer to that for the other patients. This is by no means conclusive proof of an adverse effect, since the patients with different levels of adherence are likely to vary in many respects relevant to prognosis. However, the analysis supports the view from the principal comparison, that tolbutamide increases the risk of cardiovascular death, by suggesting that those who took more of the drug were more highly at risk.

Conclusion

The main purpose of this chapter has been to emphasize the dangers inherent in the exclusion of patients or events after the act of randomization. The approach broadly recommended is to include in the principal analyses of trial results all such patients and events, except in the relatively rare situations when it is quite clear that no bias is incurred by their exclusion. The trial thus becomes a pragmatic study, designed to compare intentions to follow specified treatment regimes where these can be used. Comparison of groups from which more liberal exclusions have been allowed can form the subject of secondary analyses of the data.

Gent and Sackett (1979) take a more lenient view of postrandomization exclusions, and clearly there is room for further discussion. There will be little disagreement with their concluding recommendations, that investigators should:

1) increase [their] awareness of both the potential biases and the loss of sensitivity which can result from incomplete studies on individual patients,
2) reduce these joint consequences by increasing [their] efforts to:
   a) maximize patient drug compliance,
   b) minimize co-interventions,
   c) increase [their] effectiveness in keeping study clinicians and patients on the protocol,
3) make an explicit declaration of criteria for disqualification of patients and events a mandatory section of both the original protocol and the subsequent publications of such studies,
4) identify and resolve disagreements about decisions on the qualification and disqualification of patients and events in clinical trials.

References

Anturane Reinfarction Trial Research Group (1978). Sulfinpyrazone in the preve of cardiac death after myocardial infarction. *N. Engl. J. Med.* *298*:289–295.

Anturane Reinfarction Trial Research Group (1980). Sulfinpyrazone in the prevention of sudden death after myocardial infarction. *N. Engl. J. Med.* *302*:250–256.

Cochrane, A. L. (1972). *Effectiveness and Efficiency*. Nuffield Provincial Hospitals Trust, London.

European Coronary Surgery Study Group (1979). Coronary-artery bypass surgery in stable angina pectoris: Survival at two years. *Lancet 1*:889–893.

Gent, M., and Sackett, D. L. (1979). The qualifications and disqualifications of patients and events in long-term cardiovascular clinical trials. In *The Challenge of Clinical Trials in Thrombosis* (M. Verstraete, J. Vermylen, and H. Roberts, eds.). Schattauer, Stuttgart and New York, pp. 123–134.

Gilbert, J. P., Meier, P., Rümke, C. L., Saracci, R., Zelen, M., and White, C. (1975). Report of the Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents. *J. Am. Med. Assoc.* *231*:583–608.

Medical Research Council (1966). Comparative trial of surgery and radiotherapy for the primary treatment of small-celled or oat-celled carcinoma of the bronchus. *Lancet 2*:979–986.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J., and Smith, P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br. J. Cancer 34*:585–612.

Schwartz, D., and Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *J. Chron. Dis. 20*:637–648.

Temple, R., and Pledger, G. W. (1980). The FDA's critique of the Anturane Reinfarction Trial. *N. Engl. J. Med. 303*:1488–1492.

University Group Diabetes Program (1970). A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. I. Design, methods and baseline results. II. Mortality results. *Diabetes 19* (Suppl. 2):747–783, 787–830.