RANDOMISED CONTROLLED TRIALS: SAMPLING

P.C. ELWOOD

MRC Epidemiology Unit (South Wales), 4 Richmond Road, Cardiff CF 2 3AF

The randomised controlled trial is a most precise tool. It is probably the most exact strategy for the investigation of causes and effects in biological research, and if well conducted the analysis and interpretation of results is usually fairly straightforward. However the extrapolation of results from the trial groups to patients in general is quite another matter and it is largely with respect to this aspect that the nature of the selection of patients for a trial is highly important. That is, the results of a trial apply, strictly, only to the patients or subjects in the trial. If these had been selected at random from a large group of patients, and only if selection was random, then the results can be generalised to that larger group. Relevance of the results to other groups of patients does not necessarily follow and is dependent, largely, on how the original patients were selected for the trial.

On the other hand, it must be realised that the results of a trial represent an average effect, and cannot be expected to apply to every single patient who is subsequently seen in clinical practice. While this is always true, it is of particular importance in relation to a prophylactic or a treatment which is given in the hope of preventing a complication. In such cases the measure is used to reduce a risk, and although a group of patients may show benefit in terms of a reduced incidence, the measure has failed completely in those individuals who get the complication.

Sampling procedures are well described in a number of text books—for example, Snedecor (1956) and Cochran (1963). In brief, the 'population' to be sampled must first be defined. Ideally, a complete sampling frame should be constructed, that is, a list of every available subject who fulfils the criteria relevant to the trial. In practice this is usually unattainable in clinical research, and a sampling frame may have to be limited to, say, all the patients who attended a particular clinic, say, a diabetic, or a hypertension clinic, during the past year. Even this may be difficult and it may be preferred to sample from all patients who will attend a particular clinic during the next year. When a large sample is required a 'cluster' or 'multi-stage' sampling procedure can be used. For example, a random sample of hospitals might be drawn, and then random samples of patients drawn within each of these hospitals.

Sampling of patients from the sampling frame must always be random. This can probably best be achieved by the use of a table of random numbers. If, for example, a sample of one in two of the population was to be taken, then as each patient enters the clinic he would be allotted the next number from the table and included in the trial sample if the number was even and excluded if the number was odd. If a one in three sample is required then subjects allotted consecutive numbers from the table of random numbers, 0 would be ignored and subjects allotted, say, 3, 6 and 9 would be selected. Systematic sampling is often used, that is, say, every third or fifth patient is selected, but this procedure is hazardous if the researcher can in any way manipulate the order in which patients are seen.

In order to achieve a required balance within a sample either a stratified or a quota sampling procedure can be used. In the former, subjects are sampled separately within broad strata, such as sex, or age bands, or perhaps strata of disease defined by severity. As an example of this last point, one might decide to take separate random samples of mild and of severe hypertensives, or of insulin dependent and of other diabetic patients. In quota sampling fixed proportions of the total sample are taken from certain sub-groups of the total population. Thus one might fix the proportion of sample subjects within social class groups to correspond with the proportions of the total population in these social class groups. On the other hand, if one were interested in a social class effect, one might quota sample equal numbers within the social classes. Both quota and stratified sampling procedures have certain attractions but they can lead to difficulties in the interpretation of the results and, in particular, in the extrapolation of trial results to the general population of patients.

The estimation of the size of a sample needed for a particular trial is always difficult. Again this matter is well covered in text books (Snedecor, 1956; Cochran, 1963; Armitage, 1971) and certain aspects are covered in a later paper in this series. In short, an estimate of the size of the effect likely to be caused by the treatment under test must first be made. This is always a clinical decision and must take account of what would be judged as clinically worthwhile. For example, it might be hoped that a certain anti-

thrombotic drug would reduce mortality after a myocardial infarct, by, say, 20%. An estimate of the mortality in the untreated, or controlled patients, can easily be obtained from an earlier series of patients. In the case of myocardial infarction this might be about 15% in the year after infarction. From these data the sample size needed to demonstrate an effect in a trial with a specified power, and at a defined level of statistical significance, can be derived (Snedecor, 1956; Armitage, 1971). For example, in a comparative trial designed to detect a difference between two proportions 'p' and 'q' in two groups, each of 'n' subjects, then for a 5% significance level and a power of 95%

$$n > [13 (p_1 q_1 + p_2 q_2)] / (p_1 - p_2)^2 \text{ where } q = 1 - p$$

Once the required sample size is known, then division by the size of the total patient population enables the sampling ratio to be calculated.

However sampling from a total population of patients is usually impossible and there has to be some selection of patients. Even apart from the selection inherent in the basing of a trial in one clinic, or in a few collaborating clinics, it is unlikely that for any disease a fully representative group of patients could be obtained. Selection, in almost every case, is inevitable.

Firstly, some selection is inevitable in many serious conditions. For example, around one half of the patients who die after a heart attack die in the acute phase before they are seen by a doctor (Colling *et al.*, 1976). Therefore benefit from treatment of myocardial infarction can only be judged in the 40% or so of patients who survive long enough to be treated. Clearly the results of any trial based on these cannot be used as a reliable basis for preventive measures which might be used to reduce mortality from all myocardial infarction as the mechanisms in those who die early are likely to be different. Furthermore, the effectiveness of a treatment, assessed in a trial on those who survive should be judged in the total of affected patients. For example, a trial based on survivors who enter hospital might demonstrate a reduction in mortality from, say, 10% to 7%. The report of this trial could well speak of a 30% reduction in mortality by the treatment, though it might be better to speak of an increase in survival of about 3%. However an epidemiologist would wish to extrapolate such a result from a limited trial into the total spectrum of patients, including those who die early, and he would find it more appropriate to speak of a possible increase in the survival of patients who suffer a heart attack, were this treatment widely adopted, of only about 1% (for example, see Elwood & Sweetnam, 1980). Yet even this would certainly be an overestimate because it ignores patients treated outside hospital, in whom mortality is almost certainly lower than in those who might be admitted to a hospital based clinical trial.

This last is in fact a most important, though often inevitable, selection process. Most trials are probably based on patients identified in hospitals, and such patients are not representative of all affected patients. Neither the beneficial effects, nor the harmful effects detected in a hospital based trial can be extrapolated to all patients, as the factors which determine the place of treatment is likely to be relevant to the response to treatment. The answer is simple, no single trial is ever sufficient to begin to answer all questions, and there is always a need for a series of trials, the first being based perhaps on more seriously ill patients in hospital, followed, provided benefit is shown, by trials of less seriously affected patients, patients seen in general practice, etc.

Another important selection process arises for ethical reasons. For most diseases there is an accepted treatment and whether or not benefit from this has been established beyond all reasonable doubt, it will usually be judged to be unethical to withhold this treatment from the most severely affected patients. On the other hand, for all treatments there is a cost, both in terms of money and of risks, and so it is often unreasonable to give treatment to the most mildly affected patients.

These competing reasons for selection can be well illustrated by reference to trials of the effect of the treatment of blood pressure on the subsequent development of complications or death. There is a steeply rising, curvilinear relationship between blood pressure and stroke or death (Miall & Chinn, 1974). This means that the balance between the risks and the benefits of treatment will be very different at different levels of blood pressure. At the highest levels, hopes of benefit must be allowed maximum weight, but at lower levels, although there is still a positive regression, expectations of benefit must be low and fears of harmful side effects must weigh heavily. Therefore, above some level of blood pressure treatment must not be withheld, whereas below another level it would be unreasonable to give any treatment. This situation is summarised in Figure 1, and while of course every clinician must be aware of the issues involved it is probably only in discussions prior to the setting up of a trial that decisions about the limits of treatment become explicit. Of course there will be disagreement amongst clinicians as to the precise point above which all patients must be treated, and the precise point below which none should receive treatment. Furthermore other factors are relevant to the clinical situation, e.g. age, the occurrence of symptoms, etc. However there is likely to be, for every condition, however measured, an area of doubt, within which it is both ethical and reasonable to conduct a randomised trial.

However conclusions drawn from a trial based on patients within this 'area of doubt' cannot be extrapolated to all patients, with all severities of the disease. Even if one knew the proportionate distri-

Figure 1 Selection of patients for a trial of the effects of antihypertensive treatment.

bution of patients seen in general clinical practice, within the three bands indicated in Figure 1 (more severely affected, less severely affected and affected to a similar degree to the trial patients), one could not necessarily estimate the effect of the drug in unselected patients, because its effectiveness is likely to be very different in patients within the three bands.

Another example of these kinds of selective processes can be illustrated by considering the treatment of a condition such as peptic ulceration. Here it would be unreasonable to subject patients with, say, shallow early lesions to surgery, or to submit patients with deep chronic ulcers to medical treatment. However, between these extremes there is an area of doubt about the relative merits of medical and surgical treatment, and within this area a comparative randomised trial could be conducted. A similar situation for varicose vein treatment is illustrated in Figure 2.

It is most important to constantly bear these selective processes in mind when any attempt is made to generalise from the results of a trial. The Veterans Administration trial of anti-hypertensive treatment demonstrated a reduction in the proportion of

patients who developed a morbid event over a five year period from about 55% without treatment to about 18% with treatment (Veterans Administration, 1970). However it would clearly be quite unreasonable to bring these results into 'cost-benefit' considerations relevant to the treatment of hypertensive patients in a hospital out-patient clinic, or a general practice.

An example of a situation in which the limitations in trial results consequent on the selection of patients were ignored arises from the use of cimetidine. This drug was tested in a RCT based on selected hospital patients with bleeding from acute gastric erosions caused by drugs (Welch *et al.*, 1978). Unequivocal evidence of benefit emerged and this led to fairly widespread use of the drug throughout hospital and general practice. Unfortunately the fact that the original trial had been based on a highly selected group of patients was lost sight of, and treatment with this drug seems now to be used in a wide variety of patients with upper gastro-intestinal haemorrhage, in many of whom it may have very little, or no effect. Furthermore, although side effects were not a con-

Figure 2 Selection of patients for a comparative trial of two treatments for varicose veins.

cern in the initial trial, no long term evaluation of the therapy was done, and in view of the selection of patients none was then required. However, in these new, untested situations in which the drug is being used there are possible dangerous side-effects (Flind & Rowley-Jones, 1979) and it may cause obscuring of serious underlying lesions (Lancet, 1979). These developments are a consequence of physicians having ignored the limitations of the results of a trial consequent on the selection of patients for a trial. Ideally, in this kind of situation, further trials should have been set up to examine long term use in less highly selected groups of patients. Unfortunately, unless such further trials are set up at an early stage, widespread use of a drug may make their conduct very difficult.

This same kind of process probably leads to frequent disappointment with the effectiveness of drugs and other measures when used in general clinical practice. Drugs which have been tested in highly selective groups of patients are likely to have a very much smaller effect when used on all, or most, patients.

It is unfortunate that there is seldom any attempt made to adequately describe the selection of patients for a trial. In most reports the characteristics of patients admitted to a trial are well described (for example, the sex and age of the patients together with, say, their blood pressures or cholesterol levels, etc., as relevant). However little or no attempt is usually made to relate these factors in the trial patients to all patients with the disease. Clearly this is important in relation to general descriptive factors, such as age and sex, but it is even more important in relation to factors relevant to the effect being examined, particularly those factors which have been shown or can reasonably be assumed to have prognostic significance. Granted, to do this adequately may be very difficult, but it will be most helpful in the generalisation of the results of a trial, to state at the very least, the number of patients from which the trial group were selected and the reasons why patients were judged unsuitable. For example, it might be stated that during the period of recruitment to a hospital based trial of a drug, 100 patients were selected out of a possible total of 1,500 patients, the others being judged unsuitable on the grounds of severity (say, 200 too severely ill, 700 too mildly affected), complications of the disease (say, 150 patients), contraindications to the drug (say, 300 patients) and sensitivity to a test dose (say 50 patients). Such information will give an indication of the probable usefulness of the drug in general clinical practice. Thus in the example given, if the trial does suggest benefit, it is apparent that the drug may only be useful in the management of only about 7% of patients seen in a hospital setting. Knowledge of the pattern of hospital referral for the disease under consideration will then

enable some estimate of the likely usefulness of the drug in other situations, such as general practice, to be made.

It is to be hoped that more precise measures of selection will be made. Ideally, information will be given on every factor used to define the selection process, and hence the limitations in conclusions derived from the trial will be apparent. For example, a recent randomised controlled trial of the effect on growth of milk supplements was based on 520 children aged 7–9 years at schools in Mid Glamorgan. The report on this trial (DHSS, 1981) describes the selective process quite fully, viz: the records of a large number of defined schools were searched and all children aged 7, 8 and 9 years who came from families of 4 or more children were identified. Further descriptive items were noted for these children, viz: the proportions in the various social class, who came from single parent families, whose fathers were unemployed, etc. All these are valuable facts of relevance to the objective of the feeding trial. However they would not by themselves be adequate to enable any estimates to be made of the likely effect of a milk supplement on all children. To facilitate this the authors describe how, at the same time as the trial was in progress, they conducted a survey of a large representative group of children throughout the whole of Mid Glamorgan. In this they collected data on the factors which had been used in selection of the trial group, and they recorded height (and weight) to indicate the effect of the selection process on the growth measurements being used in the trial (Table 1). In summary, the trial group were considerably smaller (and incidentally, lighter) than all school children, and so it is certain that the effect of the milk

Table 1 The effect of selection of children for a randomised controlled growth trial

	<i>Children selected for the trial</i>	<i>Children in a random sample of all schools in the county</i>
Numbers	520	3337
Percentage in —		
Social class I	2%	15%
II		
III	41%	46%
IV		
V	25%	18%
Unemployed	31%	21%
Single parent family	8	9
Mean height at		
age 7 years	120.6 cm	123.4 cm
8 years	126.3 cm	128.3 cm
9 years	130.2 cm	133.5 cm

supplement on the growth of all children would be on average much less than the trial might suggest.

In the early examples given in this chapter selection was inevitable and beyond the control of the investigator. On the other hand, selection of the subjects for a trial is often intentional. For example, in the study just described, it was judged necessary to increase the sensitivity of the trial by selecting the most deprived children. In other cases the most severely diseased subjects might be selected. This is reasonable, if resources are limited and only a small treatment effect is expected, but it is highly relevant to subsequent attempts to generalise the results. On the other hand one may wish to gain information about a biological mechanism. In this case one may select a very carefully defined group of patients rather than identify a fully representative sample. For example, trials of aspirin and myocardial infarction could have very different objectives. The clinician and the epidemiologist are interested in the whole pattern of ischaemic heart disease in the community and so their trials should attempt to include a consecutive series of patients with myocardial infarction. A haematologist however may well be interested only in the effect of aspirin on some aspect of thrombosis so he will wish to base his trial on a highly selected group of patients, perhaps those in whom some change in a clotting factor has been detected. There are of course merits in both. The results of the first are likely to be of immediate relevance to general clinical practice, the results of the second are not. On the other hand, the first may not yield any precise answers about the relevance of a particular clotting mechanism to infarction.

The last aspect of selection which must be clearly understood by all who wish to set up a trial and by all who seem to evaluate the results of a trial, relates to selection during a trial. The only acceptable procedure to be followed is, having judged a patient to be suitable for admission to a trial, that is, one in whom there is no reasonable contraindication to either the active treatment, or a placebo (or either of two treatments, if appropriate) that patient is then allotted to a treatment group at random. Come what may, he is in that group and outcomes of the total groups (assessed blind of course) must be accepted and included in the analysis, whether or not the compliance of certain patients has been poor, or is believed to have been poor, or certain patients have been withdrawn from treatment for clinical or other reasons. Of course the inclusion of such patients will dilute out the effects of the treatment in the fully compliant patients. The justification for keeping deviators from allocated treatment in the trial is that one is comparing two policies or intentions to treat in a particular way in a non-ideal situation. This is a close approximation to the real clinical situation. The same problems with compliance will occur then, so that it is a necessary

factor to take into account when determining the effectiveness of a particular treatment. Every trial should be designed to allow for 'drop-outs', that is sufficient numbers should be included to maintain a fairly high power should the overall effect turn out to be reduced from that which can be reasonably predicted.

This is perhaps the most often debated aspect of clinical trials. On the one hand, a trial can be seen as an attempt to measure the effect of a decision to allocate patients to a particular treatment. Account will therefore be taken of patients who later refuse, or become unsuitable for the treatment, and comparisons between those treated and the controls will be based on the total groups of patients whatever the compliance of individual patients. On the other hand, a trial can be used in a strict attempt to measure the effect of a treatment and in this case attempts can be made to eliminate results from patients who did not fully comply, or worse, these results can be removed from the treated group and added to the control group. To introduce selection of this kind courts bias and misinterpretation, and as Sackett (1980) has put it, 'because highly compliant patients appear destined for rosey prognoses, a comparison with the less compliant cohorts becomes a test for bias, not efficacy'.

Of course the more knowledge one gains during a trial of compliance rates, collected blind with respect to treatment group, the better. Hence the value of tablet counts, urine or blood tests etc. in order to check on compliance. However such data can only be used to facilitate judgements of the degree of confidence which can be placed in the overall results of a trial, and not as a basis for selection of subjects in any way. Final conclusions about efficacy must be based on all results from all the patients ever admitted to the trial.

No great difficulty arises in this for the epidemiologist, nor, it is to be hoped, for the clinician, as they are concerned with the whole range of a disease in the community, and with the role of treatment in general clinical practice. The physiologist and biochemist, primarily interested in mechanisms, may be much more concerned about non-compliance, and unless compliance is very high they may regard a randomised controlled trial as a rather blunt instrument.

In the early days of randomised controlled trials a rigid design was not deemed necessary and often drop-outs were ignored, or perhaps characteristics such as age, severity of disease, etc. of the drop-outs in the two treatment groups were simply compared to judge whether or not bias might have been introduced by selective withdrawals. This is no longer judged to be adequate as such a comparison can never cover every relevant factor and so one can never be sure that the residual groups are adequately matched. The need to analyse results in all patients in whom there was an intention to treat lies in the elimination

of bias. Consider a trial of aspirin in the secondary prevention of death in patients who have had a myocardial infarct. If some patients have mild side effects from the aspirin, perhaps only vaguely identified as gastric, then it could happen that the equally vague prodromal symptoms which frequently seem to precede death for a few days in such patients would be attributed to the trial tablets and lead to their being stopped. Withdrawal of such patients because of inadequate compliance would therefore lead to a reduction in deaths in the group on the active treatment while inert placebo tablets would be less likely to be involved in such a sequence.

One further source of bias arising from selective omissions from a trial must be mentioned. This concerns the definition of 'ineligible' results, that is, outcomes which are considered as 'background noise' and ignored, because the treatment under trial is unlikely to affect them. Thus in a trial of cancer therapy a death due to a road traffic accident, or a myocardial infarct might be ignored. The hope is to increase efficiency but again the danger is of introducing bias. In the recent ART trial of sulphinpyrazone and mortality in patients who had had a myocardial infarct (The Anturane Reinfarction Trial Research Group, 1980) all deaths occurring within the first seven days of treatment were judged ineligible and were ignored. There are a number of ways in which this might have introduced bias, and led to an imbalance in the groups compared at the end of the

trial. In any case, if the drug is not effective until after seven days (and why seven?), there could only have been random differences in deaths during this period, and so bias could not have been introduced by their inclusion, and at worst, there would only have been trivial loss of power in the trial.

Conclusion

No trial can be based on a representative group of all affected patients. Selection is necessary for ethical and for practical reasons. This means that the results of a trial cannot necessarily be generalised to all patients. A trial is therefore best seen as one of a series, each based on a group of patients selected so as to represent a different part of the spectrum of the disease.

The selection of patients for a trial should always be described. Equally important, as much information as possible should be given about the total series of patients from which the trial group was selected, together with the reasons for rejection of the others. These data will facilitate generalisations of the results of a trial and will indicate the place of the treatment under test in general clinical practice.

Selection of patients, or of results, within a trial is intolerable. All results in all patients in whom there was an intention to treat must provide the basis for all conclusions drawn.

References

- ARMITAGE, P. (1971). *Statistical methods in medical research*. Blackwell Scientific Publications.
- COCHRAN, W.G. (1963). *Sampling techniques*. New York: Wiley.
- COLLING, A., DELLIPANI, A.W., DONALDSON, R.J. & MACCORMACK, P. (1976). Teesside coronary survey: an epidemiological study of acute attacks of myocardial infarction. *Br. med. J.*, **2**, 1169–1172.
- D.H.S.S (1981). A randomised controlled trial of the effect of the provision of free school milk on the growth of children. In *Sub-committee on Nutritional Surveillance: Second Report*, London: H.M.S.O.
- ELWOOD, P.C. & SWEETNAM, P.M. (1980). Aspirin and secondary mortality after M.I. *Circulation*, **62** (Suppl. V), V53–V58.
- FLIND, A.C. & ROWLEY-JONES, D. (1979). Mental confusion and cimetidine. *Lancet*, **i**, 379.
- LANCET (1979). Editorial. Cimetidine and gastric cancer. *Lancet*, **i**, 1013.
- MIALL, W.E. & CHINN, S. (1974). Screening for hypertension: some epidemiological observations. *Br. med. J.*, **2**, 595–600.
- SACKETT, D.L. (1980). The competing objectives of randomised trials. *New Engl. J. Med.*, **303**, 1059–1060.
- SNEDECOR, G.W. (1956). *Statistical methods*. Iowa U.S.A.: Iowa State University Press.
- THE ANTURANE REINFARCTION TRIAL RESEARCH GROUP (1980). Sulfinpyrazone in the prevention of sudden death after myocardial infarction. *New Engl. J. Med.*, **302**, 250–256.
- VETERANS ADMINISTRATION (1970). Effects of treatment on morbidity in hypertension. II. Results in patients with diastolic blood pressure averaging 90 through 114 mm Hg. *J. Am. med. Ass.*, **213**, 1143–1152.
- WELCH, R.W., BENTCH, H.L. & HARRIS, S.C. (1978). Reduction of aspirin-induced gastrointestinal bleeding with cimetidine. *Gastroenterology*, **74**, 459–463.