Medicine and Mathematics

Statistics and ethics in medical research

Study design

DOUGLAS G ALTMAN

The term "design" encompasses all the structural aspects of a study, notably the definition of the study sample, size of sample, method of treatment allocation, type of statistical design (randomised, cross-over, sequential, etc), and choice of outcome measures. The importance of this stage cannot be over-emphasised since no amount of clever analysis later will be able to compensate for major design flaws. In this article I will consider the relation between design and ethics in observational studies and clinical trials, but I will defer the problem of sample size until the next article.

Observational studies

In observational studies data from a sample of individuals are used, either implicitly or explicitly, to make inferences about the population of interest, such as men aged 20-65, hypertensives, or pregnant women. For this extrapolation to be valid, it is essential that the data obtained are as representative of the population as possible.

This usually entails some type of random sampling of subjects, for which a ready-made list of the whole population of interest (a sampling frame) is needed. Such lists, however, may be out of date (electoral registers) or inaccurate (doctors' lists of patients), in which case their use can lead to misleading results. Furthermore, it is often desirable to improve the representativeness of the sample by sampling separately from different subgroups—for example, by age and sex—but this additional information may not be available.

For many populations, such as the three examples above, no sampling frame exists, so that it may be impossible to obtain a representative sample. Consider, for example, trying to select a random sample of all the preschool children in an area to estimate the prevalence of vision or hearing defects. Yet for studies such as this, which set out to estimate the prevalence or incidence of some condition, the need for a truly representative sample is particularly great—otherwise the results are of uncertain value.

Even with a good selection procedure the study may be ruined by a poor response rate. Although deemed non-invasive, such studies may entail visiting people at home, expecting them to complete and return a questionnaire, or to attend a clinic, and thus may be liable to considerable non-cooperation.

Unfortunately, those who do not participate often tend to be somewhat different from those who do, both in respect of their medical condition (if this is relevant) and their social and demographic characteristics. This problem should be anticipated at the design stage, and plans made to "chase up" non-responders. It is generally advisable to keep questionnaires and other procedures short and simple to help reduce non-response. In the end, though, the response rate may largely depend on the subjects' perception of the importance of the study.

It is much less common in case-control studies to find researchers concerned about defining the subjects who will be eligible for a study, although Sackett¹ has described 22 biases that may arise at this stage. One of the most interesting is Berkson's bias, which Mainland² recently drew to the attention of readers of this journal. Case-control studies of hospital patients are often set up to study the relation between a specific disease and exposure to a suspected causal factor. If the hospital admission rates for exposed and unexposed cases and controls differ appreciably, then the observed association between the factor and the disease may be seriously biased (in either direction).²⁻³ Indeed, the choice of control group may affect the observed association between a disease and a suspected cause. A consequence of this is that such studies may need to be supported by prospective studies.

Another of Sackett's catalogue¹ is the membership (or "self-selection") bias. He cites the example of an apparent association between lack of exercise after myocardial infarction and the increased risk of recurrent attacks. This result was found in two observational studies where exercise was taken voluntarily, but was not substantiated by a prospective randomised study.

So the major problem of all observational studies is the selection of subjects for study. This aspect must be given considerable attention at the design stage, because if the sample is not representative of the population then the results will be unreliable and of dubious worth.

Clinical trials

Whatever one's view on the best type of design, clinical trials of some sort are clearly important for new treatments. As May⁴ says: "The ethical justification for such experimentation, which is outside the pure physician-patient relationship, is based on a judgment that in certain circumstances it is legitimate to put a subject at risk, with his or her consent, because of the overriding need of society for progress in combating certain diseases."

A revealing example concerns the epidemic of retrolental fibroplasia in the 1950s.⁵⁻⁶ The treatment of infants with early eye changes with adrenocorticotrophic hormone was thought

Division of Computing and Statistics, Clinical Research Centre, Harrow, Middx HA1 3UJ

DOUGLAS G ALTMAN, BSC, medical statistician (member of scientific staff)

to be a success as there was a cure rate of 75%. A clinical trial, however, would have shown that adrenocorticotrophic hormone was ineffective since 75% of such infants return to normal without treatment. The widespread use of this treatment meant that hundreds of infants were exposed to unnecessary risk, and that discovery of the cause of the epidemic (an oxygen-rich environment) was delayed.

The debate about the ethics of clinical trials is still very active. Some authors have suggested that it is unethical *not* to carry out a clinical trial on a new treatment, whereas others believe that such trials are unethical, at least in the way they are usually conducted.

IS IT ETHICAL TO RANDOMISE?

In most clinical trials subjects are allocated to the new treatment at random, others receiving either a standard treatment or a placebo. The main ethical problem is the balancing of the welfare of the individuals in the trial against the potential benefit to future patients.

It is the random allocation of subjects that comes in for most criticism. It is argued that even if at the beginning of a trial one may not know if a treatment is effective, as the study progresses it is unethical to continue to randomise ignoring the results so far.⁷ As Meier⁸ has observed, however, this attitude is based on the questionable premise "that it is unethical to deny an individual any expected benefit of treatment A over treatment B, regardless of how small that benefit may be or how uncertain."

Because of the difficulty in interpreting interim results of randomised studies, two types of non-randomised study have recently found some favour and deserve a closer look.

HISTORICAL CONTROLS

Is it really necessary to have a concurrent control group when carrying out a clinical trial? Cranberg⁹ has recently argued that instead one can use retrospective or "historical" controls—that is, previously collected data on patients who had received what would be the control treatment. Although widely practised, and perhaps of value in some circumstances,¹⁰ this can be extremely risky.

The main problem of studies using historical controls is their insensitivity to secular changes, most importantly in selection criteria.¹¹ The worst historical data to use are other people's published results, perhaps partly because of the publication bias towards positive results. Pocock¹² gives as an example 20 studies of fluorouracil for advanced cancer of the large bowel with reported success rates ranging from 8% to 85%. But data from a previous study in the same institution may also be unreliable. Pocock reports that in 19 instances where the same treatment was used in two consecutive trials of cancer chemotherapy in one organisation the changes in death rates from one trial to the next ranged from -46% to +24%, four of the differences being significant at the 2% level.

The use of historical controls is often advocated as being more ethical than using a concurrent randomised control group. The results of studies using historical controls are extremely unreliable, however, so that unless there is sound justification for their use such designs should themselves be rejected as unethical.

ADAPTIVE DESIGNS

Designs where the proportion of subjects allocated to each treatment depends on the accumulated results so far may appear preferable to randomised trials.⁷ It must be realised, however, that with such designs some subjects are still allocated to the treatment that is less successful so far, not so many as with randomised studies but still essentially at random. Further-

more, because of the unequal sample sizes for the two treatments, the study may require more subjects than an equal allocation study.^{8, 13}

Such designs require that the result for each individual is known quickly, which is often not the case. It is implicitly assumed that there is a single outcome of interest, whereas there may be several possible methods of assessment, as well as aspects such as side effects to be considered. They are also insensitive to any secular changes during the course of the study. For these reasons, although appealing in principle, adaptive designs have rarely, if ever, been used.⁸

SEQUENTIAL DESIGNS

Sequential designs¹⁴ may seem the best compromise in that they combine the many advantages of a randomised study with the desirable feature of taking account of the results so far in determining the length of the trial.

The main advantage over an ordinary randomised study is that the required sample size will be smaller if the treatment "effect" is larger. So the bigger the difference between treatments, the fewer subjects receive the less successful treatment.

Their main disadvantages are the same as for adaptive designs, especially the need for the results for each subject to be available quickly. Sequential designs are clearly of no value in long-term studies, where all the subjects will be recruited before any results are obtained. Nevertheless, in the right circumstances they can be useful and should probably be used more often.

CONSENT

Another problem of clinical trials is the need to obtain the "informed consent" of the subjects. In some cases this may be impossible because of the age or condition of the subjects, or because of the difficulty of explaining the scientific issues. Zelen¹⁵ has recently proposed a new design for comparing a new treatment with a standard one that neatly avoids the problem. He proposed that, of the subjects entering a trial, half are randomly assigned to receive the standard treatment (group 1). These subjects are treated as if they were not in the trial apart from the needs of standardised assessment and record keeping. The other half (group 2) are given a choice: they are offered the new treatment B, which is under investigation, but they may have the standard treatment A if they wish. The important point is that the subjects *choose*—this is quite different from agreeing to be randomised—so that the problems associated with informed consent do not arise.

If most of the second group elect to have the new treatment B, as is quite likely, then this design will probably be more efficient overall. It is of course essential to compare group 1 with group 2, not all those undergoing treatment A with those undergoing B. In this way two randomly selected groups will be compared. There will be some loss of efficiency because group 2 is "contaminated" by a minority undergoing treatment A, but this effect is likely to be outweighed by the advantage of having virtually no refusers.

This design, which seems perfectly ethical (Zelen¹⁵ discusses many of the issues), has two advantages over ordinary randomised trials—the ability to include all eligible subjects and the avoidance of the tricky problem of informed consent.

PLACEBOS

Too many studies compare a new treatment with a placebo rather than an existing treatment, and thus yield results that are of no practical importance. It is sometimes necessary to include placebos, but whenever possible they should be used

only when there is no appropriate treatment for comparison. Invasive placebo treatment is unlikely ever to be justified.

CONCLUSIONS

There is no one best design for all clinical trials. The choice for a specific trial must depend on the seriousness of the condition being treated, the nature of the treatments, the response time, the measures of outcome, and so on. The main ethical problem is balancing the interests of the individuals in the study with those of the much larger number who may benefit in the long term. But it is also vital that the research should provide useful results, and this may often be achieved best by a randomised study (double-blind if possible). If it is thought likely that highly favourable early results or a high incidence of side effects would argue in favour of premature termination of the study, then these considerations may be built in, using a sequential design.

The ethical difficulties associated with the widespread use of a new treatment without a trial are far greater than those associated with the trial itself. The importance of good design, however, is reflected in the many examples of conflicting results that may be found in series of case-control studies of the same topic.¹⁶ As a notable example, after 32 studies over 25 years there is still no consensus on the efficacy of anticoagulants following myocardial infarction.¹¹

References

- ¹ Sackett DL. Bias in analytic research. *J Chron Dis* 1979;**32**:51-63.
- ² Mainland D. Berkson's fallacy in case-control studies. *Br Med J* 1980;**280**:330.
- ³ Roberts RS, Spitzer WO, Delmore T, Sackett DL. An empirical demonstration of Berkson's bias. *J Chron Dis* 1978;**31**:119-28.
- ⁴ May WW. The composition and function of ethical committees. *J Med Ethics* 1975;**1**:23-9.
- ⁵ Silverman WA. The lesson of retrolental fibroplasia. *Sci Am* 1977;**236**(6):100-7.
- ⁶ Herbert V. Acquiring new information while retaining old ethics. *Science* 1977;**198**:690-3.
- ⁷ Weinstein MC. Allocation of subjects in medical experiments. *N Engl J Med* 1974;**291**:1278-85.
- ⁸ Meier P. Terminating a trial—the ethical problem. *Clin Pharmacol Ther* 1979;**25**:633-640.
- ⁹ Cranberg L. Do retrospective controls make clinical trials inherently fallacious? *Br Med J* 1979;**ii**:1265-6.
- ¹⁰ Gehan EA, Freireich EJ. Non-randomized controls in cancer clinical trials. *N Engl J Med* 1974;**290**:198-204.
- ¹¹ Doll R, Peto R. Randomised controlled trials and retrospective controls. *Br Med J* 1980;**280**:44.
- ¹² Pocock SJ. Allocation of patients to treatment in clinical trials. *Biometrics* 1979;**35**:183-97.
- ¹³ Byar DP, Simon RM, Friedewald WT, *et al*. Randomized clinical trials. Perspectives on some recent ideas. *N Engl J Med* 1976;**295**:74-80.
- ¹⁴ Armitage P. *Sequential medical trials*. 2nd ed. Oxford: Blackwell, 1975.
- ¹⁵ Zelen M. A new design for randomized clinical trials. *N Engl J Med* 1979;**300**:1242-5.
- ¹⁶ Horwitz BI, Feinstein AR. Methodologic standards and contradictory results in case-control research. *Am J Med* 1979;**66**:556-64.

*This is the second in a series of eight articles.
No reprints will be available from the author.*

Lesson of the Week

Endoscopic assessment of oesophageal disease

G LITTLE, H R MATTHEWS

The widespread availability of fiberoptic endoscopy has led to significant changes in the investigation of the upper gastrointestinal tract. We have noted in particular a tendency for doctors to dispense with a radiological examination completely and rely solely on endoscopic findings for both the diagnosis and the management of oesophageal disorders. This has led to several serious errors in the treatment of patients with oesophageal disease, and we describe three cases. We want to emphasise that radiological contrast examination must precede endoscopy in the investigation of cases of oesophageal disease if similar errors are to be avoided.

Case reports

Case 1—A 76-year-old man presented with a five-month history of progressive dysphagia and a history of heartburn.

Queen Elizabeth Hospital, Birmingham B15 2TH

G LITTLE, FRCS, senior surgical registrar

East Birmingham Hospital, Bordesley Green East, Birmingham B9 5ST

H R MATTHEWS, FRCS, consultant thoracic surgeon

Endoscopic findings alone do not tell the whole story in the investigation of oesophageal disease.

No barium studies were done, but endoscopy showed a stricture of the middle third of the oesophagus. The cause could not be determined, but when dilatation was attempted (Eder-Puestow technique) the oesophagus was perforated. The patient was transferred to our unit and the underlying pathology was unknown.

A Gastrografin swallow failed to outline the oesophagus, all the contrast medium having passed directly into the pleural cavity. Because of the possibility of a benign stricture thoracotomy was performed, but an unresectable carcinoma of the middle third of the oesophagus was discovered. Despite palliative intubation the patient died. Had a radiological diagnosis been made on this patient first a useless and painful thoracotomy could have been avoided.

Case 2—A woman aged 68 presented with retrosternal discomfort and vomiting. Her general practitioner had requested that barium studies be arranged. Instead, endoscopy was performed and a possible paraoesophageal hernia diagnosed.