

EDITORIAL

Clinical trials in psychiatry¹

Clinical trials were introduced as a method of scientific investigation before 1945 (Bull, 1959) and have been used extensively since, but it is only recently that publications devoted entirely to their management (design, conduct and analysis) have appeared (Controlled Clinical Trials, 1980; Friedman *et al.* 1981; Harris & Fitzgerald, 1970; Johnson & Johnson, 1977; Schwartz *et al.* 1980). Indeed, the methodology of clinical trials has developed to such an extent that their management is now highly complex (Peterson & Fisher, 1980), and the number of criteria to be satisfied in their design is formidable (Chalmers *et al.* 1981; Gore, 1982). In psychiatry many clinical trials conducted during the past twenty years have incorporated such important features of design as the inclusion of adequate controls, random and blind allocation of treatments, and blind assessment of 'objective' response. This editorial considers other equally important, though neglected, aspects of the design of clinical trials in psychiatry and re-emphasizes some notable aspects of their conduct and analysis.

OBJECTIVES

The objectives of a clinical trial, first, should be simple and specific and, secondly, should seek to answer a question of importance to medical science or patient welfare within a reasonable period. Clinical trials, which should be designed for the efficient comparison of treatments, are not suitable vehicles for other investigations such as constructing, testing or validating rating scales or biochemical studies of the effects of a particular drug treatment on key metabolites. Identifying subgroups of patients who respond favourably to treatment, for example, those with particular combinations of features characteristic of illness in a clinical trial comparing treatments for neurotic disorders or for whom a particular treatment is indicated or contra-indicated, for example, one anti-depressant in a clinical trial comparing response to several, is difficult and nebulous, and such vague objectives should be avoided (see below – Analysis). Of course, a comparison of response to treatment in two separate subgroups of patients (for example, unipolar depressives and bipolar manic-depressives) may be made, but preferably in a trial specifically designed with such an objective.

PROTOCOL

A detailed protocol should be prepared and piloted as part of the design of a clinical trial. Ideally, this document should provide a complete specification of the management of the trial from its objectives to the reporting of results (Chaput de tonge, 1977), together with specimen copies of forms and rating scales. Instructions for the completion and scoring of rating scales should also be included. It is particularly important in psychiatry that the protocol specifies the methods of data handling since large quantities of both baseline and response data are frequently collected, and imperative that it describes the methods of analysis. The preparation of such protocols can be tedious and exacting, and may take several months of discussion but it does focus thought about the trial wonderfully. Obviously, it is better to spend time preparing and piloting a protocol than to be forced into abandoning or fundamentally changing the management of a trial before its scheduled termination.

¹ Address for correspondence: Dr Anthony L. Johnson, Medical Research Council Biostatistics Unit, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH.

CRITERIA FOR ENTRY

Patients entered in a clinical trial should preferably be representative of a much larger patient population so that the results from the trial are of potential application to many patients. In psychiatry, where diagnosis may be both difficult and diffuse, the criteria for patient entry to a clinical trial should be assessed critically. They may require standardization – for example, in terms of the presence of specific Schneiderian first-rank symptoms in schizophrenia, the attainment of a minimal score on the Hamilton or some other rating scale for depression, or particular categorization by an instrument such as the Present State Examination – so that the implications for treatment deduced from a clinical trial in one centre or region can be extrapolated to the treatment of patients elsewhere.

PATIENT ACCRUAL

Overestimation of patient entry to a clinical trial occurs so frequently that it has been (whimsically) expressed in Lasagna's Law: once a trial begins the number of suitable patients dwindles to a tenth of what was calculated before the trial began (Harris & Fitzgerald, 1970). Consequently, it is important to monitor the overall rate of patient accrual, especially during the early stages of recruitment, and compare it with the rate estimated during planning. Serious discrepancy between the two may then be corrected either by extending the recruitment period of the trial or by extending the trial itself to other clinicians and/or hospitals; alternatively, the trial may be abandoned before the commitment of many patients and extensive resources.

Patients treated at a particular location or by a particular clinician will fall into two groups: namely, those who satisfy the clinical and demographic conditions for entry to a clinical trial and are therefore 'eligible' for entry, and those who are ineligible; some of the 'eligible' patients will enter the trial, others will not. It is important to record demographic and clinical details of *all* 'eligible' patients, together with reasons for non-entry of those who are not included; these may indicate that some factor important in the logistics of the trial such as increased workload on busy psychiatrists or other clinical staff, or inadequate clerical assistance has been overlooked. This is particularly true of clinical trials recruiting patients at more than one clinic or hospital (multi-centre trials). The monitoring of non-eligible patients may also be useful both to highlight unnecessary restrictions in entry criteria at the pilot stage and to indicate the appropriate proportion of some target population to which the results of the trial are applicable.

TYPE OF CLINICAL TRIAL

In psychiatry clinical trials may be broadly divided into those comparing the therapeutic efficacy of two or more treatments (drugs, therapies, doses) in relieving symptoms, and those comparing prophylactic treatments which prevent the recurrence of symptoms or episodes; these are discussed separately. In addition, clinical trials comparing different durations of therapy, especially prophylaxis, are also considered for, although such trials have been rarely performed, they are likely to become of increasing importance.

(a) Trials of therapeutic efficacy

Trials of therapeutic efficacy are usually of short duration (less than two months) and frequently conducted at a single centre; patients either receive one of several treatments in a fixed-size comparative trial, sometimes called a parallel group study, or, less frequently, two treatments in a crossover trial. In psychiatry, crossover trials with more than two treatments are rare. Symptoms of the disease or psychiatric state are recorded by observers and/or the patients themselves, and the response to treatment is assessed by the changes in symptom rating on appropriate scales.

An example of the fixed-size comparative trial is the Northwick Park trial of electroconvulsive therapy (ECT) (Johnstone *et al.* 1980), in which 70 patients with endogenous depression were allocated randomly to a course of 8 simulated ECTs or a course of 8 real ECTs, all other aspects

of ECT being identical in the two groups; depression was assessed using rating scales by a psychiatrist, the nursing staff and the patients themselves at weekly intervals up to 4 weeks from trial entry and again at 1 month and 6 months following the course of ECT. The principal aim of this trial was to assess the role of the convulsion in the efficacy of ECT. A second example is the comparison of flupenthixol and amitriptyline in the treatment of mild or moderately severe depression not requiring ECT in 60 out-patients (Young *et al.* 1976); the two treatments were prescribed using flexible dose schedules for a period of 6 weeks and assessed using observer-rated and self-rated scales at initial examination and after 1, 3 and 6 weeks of treatment. The aim of the study was a controlled comparison of flupenthixol against a standard tricyclic compound in depressed out-patients.

Two-period crossover trials (Grizzle, 1965; Wallenstein & Fisher, 1977; Hills & Armitage, 1979; Barker *et al.* 1982) in which patients receive two treatments, sometimes separated by a 'washout period', in random order are used in the study of acute conditions in chronic illness, for example Parkinsonism (Godwin-Austen *et al.* 1970; Hughes *et al.* 1971); the self-limiting nature of many psychiatric episodes restricts their use in psychiatry. They have been used in studies of the alleviation of side-effects from anti-psychotic or neuroleptic medication – for example, the effects of lithium carbonate on established tardive dyskinesia were compared with placebo in 11 psychiatric patients who had been treated with neuroleptics for at least two years (Mackay *et al.* 1980). Patients received 5 weeks' treatment with either lithium or placebo followed by a 6 week 'washout' and 5 weeks of the alternative treatment; dyskinesia was rated by observers, nurses and the patients themselves.

Studies of the types described above may be complicated by patients who do not comply with the treatment regimes under investigation – for example, because of failure to tolerate side-effects of treatment, failure to respond to treatment (usually styled 'withdrawn because of treatment failure'), or simply failure to 'cooperate'. In addition, 'ratings' may be missing from patients who are too ill to be evaluated by the chosen scales, who have been discharged well, who are under treatment elsewhere for concurrent illnesses, or who simply decide not to attend. The exclusion from analysis of patients in any of these categories *may* lead to a biased comparison of treatments. In trials where the amount of 'missing' information is too large to permit legitimate substitution by standard statistical techniques the analysis of the rating scale data themselves may have to be abandoned. An analysis of the proportions of patients who recover – measured in a clinical trial of anti-depressant therapy by a score on the Hamilton Rating Scale below a pre-defined threshold, for example, or of the times from randomization to recovery may be used instead.

(b) Trials of prophylaxis

Trials of prophylaxis are usually conducted over a long period (greater than 6 months) and frequently involve the participation of several hospitals in a multi-centre trial. On recovery from a psychiatric episode patients are randomly allocated to treatment and are followed up until relapse (the 'event' of interest), death, or the end of the stipulated 'follow-up' period, whichever is the shortest. All patients are followed up *irrespective* of additional treatment and side-effects of, or non-compliance with, prescribed trial treatment. The times from randomization to relapse are summarized for each treatment using actuarial 'life-table' methods which take account of the (censored) times in those patients who die without relapse or who complete the follow-up period, and the overall patterns of relapse on each treatment are compared. Some patients, inevitably, will be 'lost to follow-up' but the numbers of such patients may be reduced to a very small proportion of the total by special follow-up of patients who do not keep clinic appointments and by intensive tracing of patients who have migrated. Compliance with drug treatment regimes in such studies may be monitored through counts of tablets when relevant, the issue of prescriptions and the routine estimation of serum drug concentrations. The aim of such studies is usually to compare the *policy* of giving one treatment with the *policy* of giving another, in other words to compare the usefulness of treatments in routine clinical practice. An example is the trial of continuation therapy in unipolar depressive illness (Medical Research Council Drug Trials Subcommittee, 1981) in which 136 patients were randomly allocated to prophylactic treatment with lithium, amitriptyline or placebo on

recovery from a depressive episode, and followed up for 3 years or until relapse. Only one patient who emigrated was lost to follow-up.

The management of trials involving the prolonged observation of each patient with special reference to cancer therapy has been described in two classic papers by Peto *et al.* (1976, 1977). In clinical trials comparing survival under different chemotherapy regimes in cancer the 'event of interest', namely death, is unambiguous, can be dated easily and, if necessary, monitored by national agencies such as the Office for Population Censuses and Surveys. By contrast, in clinical trials in psychiatry patients may be followed up until psychotic or depressive relapse or some other 'event' which cannot be diagnosed unambiguously and whose onset is difficult to date. Such complications are unimportant in the sense that they do not bias the comparison of treatments, provided that a practical and objective definition of the 'event of interest' can be devised. For example, in the MRC trial cited above, relapse was defined as an episode of affective illness requiring treatment other than the administration of a benzodiazepine and dated by the prescription of such treatment. The assessment of such 'events' should be made by a clinician who is unaware of the treatment randomly allocated to the patient on entry to the trial.

(c) Trials of duration of therapy

Clinical trials comparing different durations of therapy (especially prophylaxis) have rarely been conducted in psychiatry (Paykel *et al.* 1975), although they have been used in other areas of medicine such as the study of long-term (6 months, 1 year and 2 years) chemotherapy in the treatment of chronic pulmonary tuberculosis (Medical Research Council, 1962). Concern about dependence and other adverse effects of long-term use of psychotropic drugs (Williams *et al.* 1982) and lithium (*British Medical Journal*, 1977), for example, suggests that once prophylactic efficacy has been demonstrated, it is pertinent to enquire whether such treatment should be continued indefinitely or whether it can be discontinued without the risk of relapse after some further interval. In practice, decisions to terminate continuation therapy may involve careful balancing of the risks of relapse and the risks of adverse effects, especially if the latter involve cardiovascular, oncogenic or other severe toxic reactions. While the risks of adverse effects associated with continuation therapy can only be assessed from large monitoring or epidemiological studies, the risks of relapse associated with different durations of therapy can be compared in clinical trials similar to those used to study prophylaxis ((b) above). Thus, for example, patients may be randomly allocated to receive continuation therapy for one or two years and followed up from randomization to relapse, death or the end of the follow-up period (say 3 years). The aim of such a study would be to compare the policy of 1 year's continuation therapy with that of 2 years; in other words, to determine whether therapy for an additional year was efficacious in routine clinical use. While non-compliance with the prescribed treatment will reduce the difference in relapse rates between the two groups, especially when therapy is maintained over long periods, the increasing availability of depot and sustained release preparations should make such studies feasible.

CLINICAL TRIALS IN GENERAL PRACTICE

Following the development of effective psychotropic drugs and the increased availability of other forms of treatment such as supportive therapy, many patients who might otherwise have been referred to a psychiatrist are not treated in general practice. As a result, clinical trials of psychiatric treatment in general practice, which have been conducted only rarely (for example, Porter, 1970), are likely to become more common; some guidelines for their conduct have been published recently (*Drug Trials in General Practice*, 1981). However, a treatment known, from other studies, to be efficacious in the treatment of a particular disorder need not be re-evaluated by a controlled 'scientific' study in general practice. Essentially, the general practitioner or out-patient clinician will want to know which is the best treatment or dose in routine clinical use. Since lifestyle and attitude may interfere with the treatment prescribed, these effects should not be 'designed out' of clinical trials undertaken in general practice. For example, with two drugs, A and B, both known

to be effective in the relief of anxiety, but with A more effective than B, it may nonetheless be better policy to prescribe B if it is associated with a lower non-compliance rate than A.

CLINICAL TRIAL SIZE

Many clinical trials are too small (*British Medical Journal*, 1978; Freiman *et al.* 1978; Tate *et al.* 1979) because they are conducted simply on the basis of the number of patients available at a particular clinic or to a particular clinician, instead of by anticipating the differences between treatments likely to be found and the methods of analysis to be employed. Subjecting patients to clinical trials which are likely to produce inconclusive results may be 'unethical' (Altman, 1980). Given any difference between treatments and a measure of its variability, then it is always possible to specify the number of patients required to be, say, 80 or 90% confident of detecting significance at a specified level (whether 5%, 1% or 0.1%). These numbers may, of course, run into hundreds, thousands or even more, especially if the specified treatment difference is extremely small. Essentially, clinical trials should be designed with a good chance, at least 75% and preferably more than 80%, of detecting an important clinical difference between treatments, so that a statistically non-significant result (at some specified level) is of clinical relevance and does not imply that the trial merely had little chance of detecting anything other than a large treatment difference. Quantifying the 'important clinical difference' is a difficult (and partly subjective) task but it must be done if a clinical trial is to be more than a serendipitous excursion into the unknown. Perhaps the greatest fallacy in scientific inference in the past twenty years has been the assumption that failure to detect a significant difference at 5% or any other level automatically implies no difference whatsoever.

In psychiatry the failure to consider clinical trial size at the design stage has led to a large number of trials with little or no prospect of detecting even a moderate difference between treatments (*Lancet*, 1981). Further, a significant difference between treatments reported in a small-sized trial must be interpreted cautiously since such results are often produced by selective effects or chance and are not reproducible. In clinical trials which employ rating scales the power of a statistical analysis, that is, its chance of detecting a difference between treatments, is dependent upon: first, the magnitude of the difference between treatments as measured by the rating scale; secondly, the variability of the rating scale itself; thirdly, the level of statistical significance at which this difference is to be detected; and, lastly, the number of patients entered in the trial. In designing such trials the number of patients necessary to achieve a given power, preferably greater than 80%, may be calculated from standard tables (Cohen, 1977; Lachin, 1981). While it is difficult to define sensible guidelines without resort to specific examples, experience suggests that fixed-size comparative trials with fewer than 30 patients per treatment are likely to be inconclusive. In crossover trials each patient receives both treatments and the treatment comparison is made within rather than between patients, as in a parallel group study. Since variation within patients is usually less than variation between patients, crossover trials are often thought to require fewer patients than comparable parallel group studies. Such a view is too simple; crossover trials should be conducted only with the anticipation that it may be necessary to use either an analysis of the data collected during the first treatment period only (Hills & Armitage, 1979) or an analysis based on multivariate methods (Zimmermann & Rahlfs, 1980; Poloniecki & Daniel, 1981).

As well as using rating scales to assess psychiatric state, clinical trials in psychiatry frequently elicit information on the side-effects of treatments, sometimes with the idea that, even if there is no demonstrable difference in efficacy, there may at least be a difference in the prevalence of undesirable side-effects. However, a clinical trial which is optimally designed to detect a clinically important difference between treatments using sensitive rating instruments is usually not well designed to detect differences in the prevalence of side-effects. Any trial using more than one method of assessing treatment differences should therefore be designed either to meet the size criterion determined by the least sensitive of these methods or, since side-effects are usually a secondary consideration, to exclude or reveal only gross disadvantages of treatment. No clinical trial, however

large, can be sure of detecting an unexpected and rare side-effect; these should be sought through careful drug monitoring once efficacy has been established.

In clinical trials where more than a small proportion of patients do not complete the scheduled programme of assessment it may be necessary to abandon the original intention to analyse the rating scale data themselves and, instead, adopt an analysis based, for example, upon the proportion of patients who recover, the proportion of patients who achieve a specific target rating, or the times to such 'events'. Such changes to the methods of analysis should be anticipated at the design stage and followed by appropriate adjustment to the size of the clinical trial. In particular, the power of an analysis of times to some 'event', such as recovery from an episode in a trial of therapeutic efficacy or relapse as in a trial of prophylaxis, is dependent not so much upon the size of the trial but on the total number of 'events' that has occurred (for example, the number of patients who have recovered or the number of relapses) (Peto *et al.* 1976). The number of patients required to achieve a given power can again be calculated from standard tables (Lachin, 1981; Freedman, 1982). Analyses based on the observation of less than 70 'events' are insensitive to even moderate treatment differences.

ANALYSIS

The method of analysing data collected from a clinical trial should be anticipated during the design of the trial and should be outlined in the protocol. It must be appropriate to the type of data collected and should incorporate (in the initial stages at least) all features of the trial design, including matching and more complicated schemes of stratification.

It should be emphasized that for all clinical trials, irrespective of type or design, the only treatment comparison known to be unbiased is that based on the analysis of *all* randomized patients. To quote Friedman *et al.* (1981): 'If subjects are withdrawn (from analysis), the burden rests with the investigator to convince the scientific community that the analysis has not been biased.' The fate of all patients entered in a trial should therefore be summarized with special mention of the numbers lost to follow-up and deviating from prescribed treatment; reasons for the latter should also be summarized, and death and other untoward incidents reported in detail. An analysis based on intention to treat comparing treatment groups as randomized should always be given, irrespective of any other analyses of selected subgroups. The exclusion of patients who do not comply with, or respond to, treatment or for other reasons may lead to serious bias in treatment comparisons and inconclusive analyses (Gore, 1981 *a*). In clinical trials recording the time to some 'event', and especially those involving long-term follow-up of patients (such as in prophylaxis), the analysis should be based on actuarial techniques and not on the presentation of results for patients followed for some arbitrary interval.

To aid the sensible interpretation of significance tests, and especially those reported as 'not significant', the power of the tests to detect appropriate differences between treatments should be quoted. Alternatively, a 95% confidence interval should be given (Gore, 1981 *b*). Such information allows a realistic assessment of the usefulness of any clinical trial; this cannot be deduced from the reporting of the results of significance tests alone.

The objectives of many clinical trials in psychiatry include non-specific statements about identifying the characteristics of subgroups of patients who respond either favourably or unfavourably to a particular treatment. While it is difficult to curb the enthusiasm for such 'data fishing' expeditions, it must be stressed that they are of extremely limited value and may be counter-productive. In many clinical trials it is possible to identify specific subgroups of patients for whom one treatment is 'significantly' better than another as well as subgroups of patients who respond 'significantly' better than other subgroups to treatment. Since such 'significant' differences frequently arise by chance or result from selective effects, the conclusions drawn from analyses of subgroups of patients should be regarded with scepticism, and investigated further in other trials only when they are scientifically or clinically plausible.

MULTI-CENTRE TRIALS

As greater understanding of psychiatric disorders leads to improvements in the efficacy of, and compliance with, treatment, it is likely that clinical trials will be required to detect smaller differences between new and standard treatments. Unless there are simultaneous refinements in assessing the effects of treatment, the size of clinical trials must inevitably increase. When no single centre can recruit sufficient patients in a reasonable period, the collaboration of several centres in a multi-centre trial is desirable. Such trials are already used in the study of prophylaxis, where large numbers of patients are entered over an extended period.

Multi-centre trials are complex and require detailed planning, coordination and control both by central and local management. Experience with such trials in other areas of medicine, especially cancer, suggests that many multi-centre trials are conducted over long periods, entry plus follow-up in excess of 5 years (Pocock, 1978). This situation has arisen mainly because entry rates during the early stages of trials have been much too low; indeed, many trials may have been stopped not by a definite policy decision but simply because patient entry has declined. There is no reason to expect that the situation is any different in psychiatry. In the recent trial of continuation therapy with lithium and amitriptyline (Medical Research Council Drug Trials Subcommittee, 1981) 136 patients were recruited in a period of 54 months with one centre entering 53 patients (39% of the total) and four other centres less than 10 patients each. If these four centres had been replaced by other centres recruiting patients at the rate of one per month then this trial could either have been completed in half the time or could have achieved greater power to detect differences between treatments.

The organization of multi-centre trials in psychiatry through the collaboration of a small number of major treatment centres which are able to maintain a steady and appreciable recruitment rate would allow comparatively rapid assessment of therapies and would achieve a great advance in clinical trial methodology.

CONCLUSION

The management of any clinical trial involves an elaborate and complex interaction of several disciplines between which failure to establish or maintain adequate communication can be nothing but a recipe for disaster. Many difficulties arise in the conduct of trials, the collection of data and interpretation of results. Careful consideration of the issues discussed above may anticipate and resolve some of these problems before the recruitment of patients to clinical trials in psychiatry.

ANTHONY L. JOHNSON

I thank Laurence Freedman, Sheila Gore and Ian Sutherland for their helpful comments and suggestions.

REFERENCES

- Altman, D. G. (1980). Statistics and ethics in medical research. Misuse of statistics is unethical. *British Medical Journal* *ii*, 1182–1184.
- Barker, N., Hews, R. J., Huitson, A. & Poloniecki, J. (1982). The two period cross over trial. *Bulletin in Applied Statistics* *9*, 67–116.
- British Medical Journal* (1977). Leading article: Adverse effects of lithium treatment; *ii*, 346–347.
- British Medical Journal* (1978). Leading article: Interpreting clinical trials; *ii*, 1318.
- Bull, J. P. (1959). The historical development of clinical therapeutic trials. *Journal of Chronic Disease* *10*, 218–248.
- Chalmers, T. C., Smith, H., Blackburn, B., Silvermann, B., Schroeder, B., Rettman, D. & Ambroz, A. (1981). A method of assessing the quality of a randomized clinical trial. *Controlled Clinical Trials* *2*, 31–49.
- Chaput de tonge, D. M. (1977). Aide-memoire for preparing clinical trial protocols. *British Medical Journal* *i*, 1323–1324.
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioural Sciences* (second edn). Academic Press: New York and London.
- Controlled Clinical Trials. Design and Methods (1980). *Official Journal of the Society for Clinical Trials*. Elsevier/North Holland: Amsterdam.
- Drug Trials in General Practice (1981). *Drug and Therapeutics Bulletin* *19*, 97–99.
- Freedman, L. S. (1982). Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine* *1*, 121–129.
- Freiman, J. A., Chalmers, T. C., Smith, H. & Kuebler, R. R. (1978). The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 'negative' trials. *New England Journal of Medicine* *299*, 690–694.
- Freedman, L. M., Furberg, C. D. & DeMets, D. L. (1981). *Fundamentals of Clinical Trials*. John Wright: Bristol.

- Godwin-Austen, R. B., Frears, C. C., Bergmann, S., Parkes, J. D. & Knill-Jones, R. P. (1970). Combined treatment of Parkinsonism with L-dopa and amantadine. *Lancet* *ii*, 383–385.
- Gore, S. M. (1981*a*). Assessing clinical trials—rash adventures. *British Medical Journal* *ii*, 426–428.
- Gore, S. M. (1981*b*). Assessing methods—confidence intervals. *British Medical Journal* *ii*, 660–662.
- Gore, S. M. (1982). Statistics in question: assessing clinical trials. In *Statistics in Practice* (by S. M. Gore and D. G. Altman), pp. 27–63. British Medical Association: London.
- Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics* *21*, 467–480.
- Harris, E. L. & Fitzgerald, J. D. (eds.) (1970). *The Principles and Practice of Clinical Trials*. Livingstone: Edinburgh and London.
- Hills, M. & Armitage, P. (1979). The two-period crossover clinical trial. *British Journal of Clinical Pharmacology* *8*, 7–20.
- Hughes, R. C., Polgar, J. G., Weightman, D. & Walton, J. N. (1971). L-Dopa in Parkinsonism and the influence of previous thalamotomy. *British Medical Journal* *i*, 7–13.
- Johnson, F. N. & Johnson S. (eds.) (1977). *Clinical Trials*. Blackwell: Oxford.
- Johnstone, E. C., Deakin, J. F. W., Lawler, P., Frith, C. D., Stevens, M., McPherson, K. & Crow, T. J. (1980). The Northwick Park Electroconvulsive Therapy Trial. *Lancet* *ii*, 1317–1320.
- Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials* *2*, 93–113. *Lancet* (1981). Leading article: Multicentre depression; *ii*, 563–564.
- Mackay, A. V. P., Sheppard, G. P., Saha, B. K., Motley, B., Johnson, A. L. & Marsden, C. D. (1980). Failure of lithium treatment in established tardive dyskinesia. *Psychological Medicine* *10*, 583–587.
- Medical Research Council (1962). A report by the Tuberculosis Chemotherapy Trials Committee. Long term chemotherapy in the treatment of chronic pulmonary tuberculosis with cavitation. *Tubercle* *43*, 201–267.
- Medical Research Council Drug Trials Subcommittee (1981). Continuation therapy with lithium and amitriptyline in unipolar depressive illness: a controlled clinical trial. *Psychological Medicine* *11*, 409–416.
- Paykel, E. S., DiMascio, A., Haskell, D. & Prusoff, B. A. (1975). Effects of maintenance amitriptyline and psychotherapy on symptoms of depression. *Psychological Medicine* *5*, 67–77.
- Peterson, A. V. & Fisher, L. D. (1980). Teaching the principles of clinical trials design and management. *Biometrics* *36*, 687–697.
- Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. & Smith, P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer* *34*, 585–612.
- Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. & Smith, P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *British Journal of Cancer* *35*, 1–39.
- Pocock, S. J. (1978). Size of cancer clinical trials and stopping rules. *British Journal of Cancer* *38*, 757–766.
- Poloniecki, J. & Daniel, D. (1981). Further analysis of the Hills and Armitage data. *The Statistician* *30*, 225–229.
- Porter, A. M. W. (1970). Depressive illness in a general practice. A demographic study and a controlled trial of imipramine. *British Medical Journal* *i*, 773–778.
- Schwartz, D., Flamant, R. & Lellouch, J. (1980). *Clinical Trials* (transl. M. J. R. Healy). Academic Press: London.
- Tate, H. C., Rawlinson, J. B. & Freedman, L. S. (1979). Randomized comparative studies in the treatment of cancer in the United Kingdom: room for improvement. *Lancet* *ii*, 623–625.
- Wallenstein, S. & Fisher, A. C. (1977). The analysis of the two-period repeated measurements crossover design with application to clinical trials. *Biometrics* *33*, 261–269.
- Williams, P., Murray, J. & Clare, A. (1982). A longitudinal study of psychotropic drug prescription. *Psychological Medicine* *12*, 201–206.
- Young, J. P. R., Hughes, W. C. & Lader, M. H. (1976). A controlled comparison of flupenthixol and amitriptyline in depressed out-patients. *British Medical Journal* *i*, 1116–1118.
- Zimmermann, H. & Rahlfs, W. (1980). Model building and testing for the change-over design. *Biometrical Journal* *22*, 197–210.