Trials: the next 50 years

Large scale randomised evidence of moderate benefits

Over the past half century there has been a vast proliferation first of randomised trials and now of meta-analyses, both of which (if appropriately analysed) can avoid bias. But to get medically reliable answers to previously unanswered questions about life or death treatment decisions it isn't enough just to avoid bias. We must also ensure that we are not seriously misled by the play of chance, and often the only way to do this reliably is to get appropriate analyses of really large scale randomised evidence.[1]

At present, many wrong, or at least unreliable, therapeutic answers are being generated by non-randomised "outcomes research," by small randomised studies, by small meta-analyses, and by statistically inappropriate analyses. Moreover, even when large scale randomised evidence is available, wrong conclusions can be drawn from unduly selective emphasis on particular trials or subgroups—and such "selection biases" can cause even greater errors when there is only a limited amount of evidence to review.

Over the past 50 years randomisation has already delivered reliable answers to some important questions and it offers the promise of reliable answers to many more. For that promise to be properly realised over the next 50 years, however, medical research needs to find practicable ways of greatly increasing the size of randomised studies; otherwise moderate but worthwhile benefits will continue to be missed. One important step towards larger size is the recent emphasis on meta-analyses:[2 3] when many different trials have all addressed similar therapeutic questions a synthesis of all of their results not only avoids selective biases but also helps avoid random error.

But it often happens that there are no really large trials and that even a meta-analysis of all the trials in the world isn't big enough to give statistically reliable answers about major outcomes. The key question then is how, in practice, is it possible to randomise a really large number of patients? For if one is trying to decide how millions of future patients should be treated it may often be appropriate to randomise at least many thousands—as is now becoming possible in breast and intestinal cancer—or even tens of thousands, as has occasionally been possible in stroke and heart disease.

Generally the only practicable way to achieve this is to design trials that are extremely simple and flexible: simplify the entry criteria by use of the "uncertainty principle" (see box), simplify the treatments, and simplify enormously the data requirements. Using the uncertainty principle should allow the process of providing information and gaining consent to become much closer to what is appropriate in normal medical practice. Collecting less information may mean bigger numbers and hence better science: many trials still collect ten or a hundred times too much information per patient, often at the behest of study sponsors or their committees. Requirements for large amounts of defensive documentation imposed on trials by well intentioned guidelines on good clinical practice (or good research practice) or excessive audits may, paradoxically, substantially reduce the reliability with which therapeutic questions are answered, if their indirect effect is to make randomised trials smaller or even to prevent them starting.

The uncertainty principle

A patient can be entered if, and only if, the responsible clinician is substantially uncertain which of the trial treatments would be most appropriate for that particular patient. A patient should not be entered if the responsible clinician or the patient are for any medical or non-medical reasons reasonably certain that one of the treatments that might be allocated would be inappropriate for this particular individual (in comparison with either no treatment or some other treatment that could be offered to the patient in or outside the trial).

To argue the need for some large, simple randomised trials is not, of course, to argue that all other trials are useless: indeed, many small (or complex) trials will continue to be needed for certain purposes, as will many other types of clinical research. But for many important questions about practicable therapeutic improvements in controlling the common causes of death or serious disability there is no reliable alternative to large scale randomised evidence.

The reason for this is simple: when it comes to major outcomes it is generally unrealistic to hope for large therapeutic effects. Moreover, if a particular treatment did produce a really large effect on survival then we might well be able to recognise this reliably without any randomised trials. The efficacy of penicillin, for example, was so great that it was recognised before the introduction of randomisation. Likewise, the main hazards of tobacco are so great that they were recognised without randomisation. Hence, if substantial uncertainty remains about the effects of some particular treatment on survival then these effects are likely to be small or only moderate. For example, it might be reasonable to hope that a new treatment for acute stroke or acute myocardial infarction could reduce recurrent stroke or death in hospital from 10% to 9% or 8% (as aspirin does,[4 5] preventing 10 000 or 20 000 deaths per million treated), but not to hope that it could halve in-hospital mortality. Many lives could, however, be saved by moderate reductions in the common causes of death—and if, eventually, several moderate benefits are reliably demonstrated their combined effects may be substantial.[5]

Thus, those who sponsor, perform, and regulate therapeutic research need to find ways of making trials much simpler and much larger. Otherwise the next 50 years of randomised evidence will not fulfil the promise of 50 years ago, when a properly[6] randomised clinical trial was first published,[6 7] transforming medical research by its method of generating unbiased answers to many therapeutic questions.

Richard Peto *Professor of medical statistics and epidemiology*
Colin Baigent *Medical Research Council scientist*

Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Radcliffe Infirmary, Oxford OX2 6HE

BMJ 1998;317:1170–1

1 Collins R, Peto R, Gray R, Parish S. Large-scale randomised evidence: trials and overviews. In: Maynard A, Chalmers I, eds. *Non-random reflections on health services research*. London: BMJ Books, 1998:197-230.
2 The Cochrane Library. *Cochrane Collaboration; Issue 3*. Oxford: Update Software, 1998.
3 Clarke M. Ovarian ablation in breast cancer, 1896 to 1998: milestones along hierarchy of evidence from case report to Cochrane Review. *BMJ* 1998;317:1246-8.
4 CAST (Chinese Acute Stroke Trial) Collaborative Group. CAST: randomised placebo-controlled trial of early aspirin use in 20,000 patients with acute ischaemic stroke. *Lancet* 1997;349:1641-9.
5 ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction. *Lancet* 1988;ii:349-60.
6 Doll R. Controlled trials: the 1948 watershed. *BMJ* 1998;317:1217-20.
7 MRC Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment for pulmonary tuberculosis. *BMJ* 1948;ii:769-82.

Cluster randomised trials: time for improvement

The implications of adopting a cluster design are still largely being ignored

Cluster randomised trials, where groups of patients rather than individuals are randomised, are increasingly being used in health services research. Randomisation by individual is inappropriate for evaluating some interventions, such as organisational changes, where it may not be feasible to randomise at the patient level. In such cases cluster randomisation at the level of the health professional or organisation is necessary. Such randomisation can also minimise the potential for contamination between treatments when trial patients are managed within the same setting.

The main consequence of adopting a cluster design is that the outcome for each patient can no longer be assumed to be independent of that for any other patient (which is the case in an individually randomised trial). Patients within any one cluster are more likely to have similar outcomes. For example, the management of patients within a single general practice is more likely to be consistent than management across several practices.

This lack of independence has implications for the design and analysis of these trials.[1] The statistical power of a cluster randomised trial is greatly reduced in comparison with a similar sized individually randomised trial. Therefore standard sample size estimates have to be inflated to take account of the cluster design.

The impact on sample size can be substantial and depends on the size of the clustering effect and the number of clusters available. The clustering effect would be high if, for example, management of patients within individual hospitals was very consistent but there was wide variation across hospitals. Consider a trial of an educational intervention to implement a clinical guideline. A patient randomised trial would require 194 patients to detect a change from 40% to 60% in the proportion of patients who are managed appropriately (with 80% power and 5% significance). However, this design would be inappropriate because of the potential for contamination. For a cluster randomised trial with a moderate clustering effect and 10 available patients per cluster the equivalent sample size adjusting for clustering is 38 clusters or 380 patients—that is, almost double.[1]

The analysis of cluster randomised trials must also take into account the clustered nature of the data. Standard statistical techniques are no longer appropriate, unless an aggregated analysis is performed at the level of the cluster,[2] as they require data to be independent. If the clustering effect is ignored P values will be artificially extreme, and confidence intervals will be over-narrow, increasing the chances of spuriously significant findings and misleading conclusions.

Although an aggregated analysis can be performed at the cluster level using standard statistical tests, this approach is statistically inefficient. Furthermore, it does not allow variation at the patient level to be explored—for example, it cannot take account of patient characteristics such as disease severity. More advanced techniques have now been developed to analyse patient level data arising from a clustered design, which allow the hierarchical nature of the data to be modelled appropriately.[3] They essentially allow variation to be modelled at each level of the data—for example, at both the practice and the patient level.

Despite the increased use of cluster randomised trials, the implications of adopting such a design continue to be largely ignored. For example, a review by Devine et al which examined studies of physicians' patient care practices observed that 70% of studies identified had not appropriately accounted for the clustered nature of their study data.[4]

Many trials do not take cluster randomisation into account when calculating the required sample size, resulting in studies which are underpowered. This may, in part, be explained by the lack of published information on the likely size of the clustering effect, known as the intracluster correlation coefficient.[5] Reliable estimates of this coefficient are required to ensure robust sample size calculations, yet publication of intracluster correlation coefficients in trial reports is rare.

We need a standardised approach to the reporting of all aspects of cluster randomised trials, including intracluster correlation coefficients. The CONSORT statement for standards of reporting of patient randomised trials[6] has improved the reporting of such trials, and a similar approach for the reporting of cluster randomised trials would be beneficial. A letter to the *BMJ* highlighted the need for such an approach,[7] and a proposed amendment to the CONSORT statement is being formulated. We believe that this will aid the appraisal of published trial reports and the planning of future research. It deserves the full support of the medical research community.

Marion K Campbell *Senior statistician*
Jeremy M Grimshaw *Programme director*

Health Services Research Unit, University of Aberdeen, Aberdeen AB25 2ZD (m.k.campbell@abdn.ac.uk)

BMJ 1998;317:1171–2