Randomised controlled trials in primary care: scope and application

Aziz Sheikh, Liam Smeeth and Richard Ashcroft

SUMMARY

There is now widespread acknowledgement of the absence of a sound evidence base underpinning many of the decisions made in primary care. Randomised controlled trials represent the methodology of choice for determining efficacy and effectiveness of interventions, yet researchers working in primary care have been reluctant to use intervention studies, favouring observational study designs. Unfamiliarity with the different trial designs now available, and the relative advantages and disadvantages conferred by each, may be one factor contributing to this paradox. In this paper, we consider the principal trial designs available to primary care researchers, discussing the contexts in which a particular design may prove most useful. This information will, we hope, also prove useful to primary care clinicians attempting to interpret trial findings.

Keywords: *randomised controlled trial; evidence-based medicine; research methodology.*

A Sheikh, BSc, MRCP, MRCGP, DCH, DRCOG, DFFP, NHS R&D national primary care training fellow; and R Ashcroft, MA, PhD, Leverhulme senior lecturer in medical ethics, Department of Primary Health Care and General Practice, Imperial College School of Science, Technology and Medicine. L Smeeth, MSc, MRCGP, DCH, DRCOG, DFFP, clinical lecturer, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine.

Address for correspondence

Dr Aziz Sheikh, Department of Primary Health Care, ICSTM, Charing Cross Campus, St Dunstan's Road, London W6 8RP. E-mail: aziz.sheikh@ic.ac.uk

Submitted: 7 June 2001; Editor's response: 11 September 2001; final acceptance: 14 December 2001.

Introduction

THE absence of a sound evidence base underpinning many of the clinical decisions made in routine general practice is now widely acknowledged.[1-3] Randomised controlled trials (RCTs) represent the gold standard in study design, primarily because of their ability to control for (known and unknown) confounding factors. However, not all the evidence required to inform health care can come from randomised trials. For example, qualitative methods are needed to investigate practitioners' and patients' attitudes, beliefs, and preferences. Even when assessing the effectiveness of an intervention, randomised trials may not be ethical or feasible, necessitating recourse to observational study designs.[4] The paper by Thomas *et al* was, however, the first to draw attention to primary care researchers' neglect of RCTs as an investigative tool and how this was one of the reasons contributing to the dearth of high quality evidence available.[5]

Traditional double-blind designs represent the method of choice for reducing risk of bias.[6] However, they are not necessarily the most suitable trial design for answering the complex clinical scenarios facing primary health care professionals in their day-to-day work. To increase the external validity (generalisability) of research there is, we believe, a need for a greater willingness to choose trial designs that are most appropriate for the question or questions being posed. In this paper, we aim to discuss the advantages and disadvantages of different trial designs available and consider their application in relation to primary care research. The intricately related conceptual issue of identifying and choosing outcome measures most suitable for use in primary care-based trials is also explored.

Efficacy and effectiveness

The possible impact of bias needs always to be considered, both by researchers and by those interpreting research findings. One of the key advantages offered by randomisation is that, if performed correctly, the risk of selection (allocation) bias is greatly reduced. Details of the practicalities of randomisation fall beyond the scope of this paper; however, this important subject has recently been reviewed by Altman and Bland.[7] So-called 'triple-blind' trials, in which subjects, researchers, and those responsible for outcome assessments are all unaware of assigned treatment, further minimise the possible impact of performance and detection bias and, if coupled with an intention-to-treat analysis, this also minimises the impact of attrition bias.[8] Blinding is thus clearly of great importance when conducting research into the efficacy of new interventions; these are often described as 'explanatory trials' (Box 1).[9]

It is common practice to test novel interventions, initially in small-scale trials on highly selective patient groups. Once

Efficacy

The extent to which a specific intervention, procedure, regimen, or service produces a beneficial result under ideal circumstances.

Effectiveness

A measure of the extent to which a specific intervention, procedure, regimen, or service, when deployed in the field of routine circumstances, does what it is intended to do for a specified population.

Explanatory trial

A study that aims to provide explanation typically achieved by isolating the effects of specific variables and understanding the mechanism of action.

Pragmatic trial

A study that aims to improve health status or health care to a specified population, provide a basis for decisions about health care, or evaluate previous actions.

Box 1. Definition of key terms.[9]

efficacy has been established, there is often a need to conduct larger trials in more representative patient groups. To determine the effectiveness of the intervention, it is also important that, as far as possible, such trials mirror routine care. Clinical care is delivered in an 'open' manner, in which both providers and recipients of care are aware of assigned treatment. Increasingly, treatment decisions are made taking into account patient preference (see later). For decisions on the value of a new intervention, it is important for policy makers and health professionals to have available results from both explanatory (efficacy) and pragmatic (effectiveness) trials. While the scientific community may be more interested in the former, clinicians will invariably be more interested in the latter.

Trial designs

Patient preference trials

Incorporating patient preferences into the clinical decision-making progress is central to the provision of patient-centred care.[10] But, in addition to respecting patient autonomy, it is increasingly believed that accommodating preference may confer clinical benefit.[11] Although direct evidence for the existence of preference effects is lacking, it is thought that these advantages may be conferred through improved adherence (concordance) by those who indicate a preference for a particular treatment option. More controversially, some have argued that taking preference into account may improve outcomes *per se*, these benefits being mediated possibly through psychobiological pathways.[12] Clear answers to such questions are obviously important if we are to be able to correctly interpret the implications of research findings and apply them meaningfully in individual patients.

Studying the effect of preference is difficult for several reasons. First, we have as yet no validated tool for measuring preference; secondly, preferences by their very nature are frequently transient; thirdly, preferences are very likely to be intervention-specific; and finally, those with strong preferences may be systematically different from those not expressing preferences.

A number of trial designs have been suggested that may overcome at least some of these difficulties. These range from, most simply, enquiring about preference prior to randomisation and exploring the interaction between preferences for treatment and clinical outcome during data analysis.[13] Trials employing this method have shown that eliciting preferences need not impact on recruitment since the majority of subjects are still willing to be randomised. The principal difficulty with this approach is that to have sufficient power to detect reliably interactions of this sort, sample sizes will typically need to be at least quadrupled.[14]

The comprehensive cohort design proposed by Brewin and Bradley is another alternative.[15] In this model, patients are offered randomisation; those unwilling to be randomised are given their preferred treatment option and followed up throughout the study period, thereby allowing comparison to be made between those randomised and those choosing a particular treatment. This trial design has recently been employed to investigate treatment strategies for the primary care management of depression.[16,17] The main difficulty with this approach is that systematic differences may exist between those willing to be randomised and those expressing a preference.

RCTs in single patients: n*-of-1 trials*

In an *n*-of-1 trial, a treatment is compared with placebo (or treatments are compared with each other) in a single patient.[18-20] The treatments being compared are used one at a time in random order and, where possible, both clinician and patient are blinded to the order of treatments. A single period of each of the treatments being compared may be used, but more often each treatment is repeated more than once; patient diaries are often used to monitor outcomes.

The *n*-of-1 trial is particularly useful when the results of large RCTs are not able to adequately guide the best treatment for an individual patient. There are at least three reasons why this might be true. First, trials of more than one patient provide summary results for all the patients in a trial. In a trial of a new intervention; for example, the observation that '60% of patients benefited' provides no means of differentiating the 60% who benefited from the 40% that did not. Secondly, we may have reason to believe that a particular individual, or the conditions under which treatment will be given, are different to those in the large trials. Thirdly, and perhaps most commonly, there may be insufficient trial evidence to guide treatment decisions. In such situations, clinicians will often perform a 'therapeutic trial'. However, there are many sources of possible error in a therapeutic trial, most of which favour wrongly concluding that a treatment is useful, when in fact it is not. These possible biases include**:** the effect of time producing a change in the condition which is wrongly assumed to be owing to treatment; the patient (and/or clinician) wanting to see an improvement and this influencing the conclusions drawn; and the placebo effect. The *n*-of-1 trial aims to overcome many of these problems; such trials have now been reported for a range of conditions including osteoarthritis (Box 2)[21,22] and treatments for chronic fatigue syndrome,[23] hyperemesis gravidarum,[24] and attention deficit hyperactivity disorder.[25-27]

A major problem with *n*-of-1 trials in primary care is the time involved in their planning and execution. Other potential difficulties include problems with blinding treatments

Considerable uncertainty exists about the likely effectiveness of these two treatments for individuals troubled with osteoarthritis.

Double-blind *n*-of-1 trials were undertaken in 25 patients in a general practice setting. Patients received three treatment cycles with two weeks' supply each of paracetamol (1 g twice daily) and diclofenac (50 mg twice daily) prepared in identical gelatine capsules. Main outcome measures were a diary of pain and stiffness, function, and side effects.

Fifteen patients completed the study; five withdrew early but had made a therapeutic decision, and a further five dropped out very early. Several patterns of response evolved. There was no difference in response to treatments in eight of the 20 patients. Five indicated a clear preference for diclofenac and two other peoples' symptoms were controlled within the first two weeks of receiving diclofenac with no relapse with subsequent treatment changes. After three months, nine of the 20 patients had adequate symptom control with paracetamol alone.

Box 2. Example of individual patient trial: paracetamol versus diclofenac for osteoarthritis.[21]

(because of, for example, different dosing schedules or characteristic side effects), difficulties in obtaining placebos and drugs with long half-lives or prolonged effects requiring 'wash-out' periods. Randomisation of treatments can simply be decided by the toss of a coin, although more sophisticated approaches can also be used.[7] Although large differences in effect may not require formal analysis, simple statistical analysis is usually helpful. The increasingly widespread availability of personal computers in primary care makes practice-based analyses feasible.

Whether utilisation of *n*-of-1 trials produces better health outcomes is not yet clear. There have been two small studies in which patients were randomised to either an *n*-of-1 trial or to usual care. Both studies were of theophylline for chronic airflow limitation, included around 30 patients, and were undertaken by the same research group.[28] In the first, patients randomised to an *n*-of-1 trial had lower levels of drug use with no reduction in exercise capacity or quality of life. However, in a second very similar study, both drug use and outcome measures were no different between the two groups. There is thus an absence of evidence of benefit from the use of *n*-of-1 trials; this should not be confused with evidence of absence of benefit. Ongoing areas of research in *n*-of-1 methodology are the idea of pooling a series of *n*-of-1 trials to produce results of wider generalisability[29] and using *n*-of-1 trials to estimate cost-effectiveness of treatments in individual patients.[30]

Factorial trials

In a factorial trial, the effects of different interventions are assessed both individually and when used in combination.[31] Factorial trials most commonly include two interventions and either a placebo or no intervention, but larger numbers of different interventions can be assessed.[32-35] The design is most clearly illustrated by looking at a specific example (Box 3).[36]

The factorial design is particularly suitable if it can be assumed that there is no interaction between the different treatments; or, in other words, the effects of different interventions are independent of each other. In the example given in Box 3, this means that the effects of aspirin compared with placebo are the same as the effects of aspirin plus warfarin compared with warfarin alone.[36] If we are

A factorial trial, comparing warfarin and low-dose aspirin for the primary prevention of ischaemic heart disease, was undertaken in general practice in the United Kingdom. Participants were randomised to four different treatments: aspirin plus warfarin; placebo aspirin plus warfarin; aspirin plus placebo warfarin; and placebo aspirin plus placebo warfarin.

Both aspirin and warfarin were effective in reducing the risk of ischaemic heart disease. Combined treatment was more effective than either treatment on its own, but was also associated with a higher risk of haemorrhagic stroke.

There was no evidence of an interaction between the different treatments: the effect of aspirin was independent of the effect of warfarin and vice versa.

Box 3: Example of a factorial trial. Comparison of warfarin and aspirin for prevention of ischaemic heart disease.[36]

happy that this condition is met, then factorial designs allow us to compare more than two intervention strategies without the need for a larger trial. In the present example, although four treatment strategies were compared, half of the participants received aspirin and half received warfarin. Because the effects of the treatments were independent, the factorial trial therefore had the same power to compare aspirin with warfarin as a conventional trial the same size that *only* compared aspirin with warfarin. The advantage of the factorial design was that it allowed the investigators to assess the effects of the two treatments combined and to compare each treatment with placebo.

Unfortunately, we will often not be able to assume that the effects of two or more interventions are independent. Statistical tests for interaction between interventions are of little help because they lack power. This means that failure to detect an interaction may not mean the effects of the interventions are independent. An analysis that ignores such an interaction will be incorrect. The existence of interaction between the different interventions does not preclude a factorial design; however, the study must have sufficient power to be able to quantify the interaction. This generally requires very large studies, negating one of the key advantages of this design. The factorial design has probably been underutilised in the past. However, the need for either prior knowledge of interaction between treatments or the ability to reliably detect or exclude such an interaction in the new study are limiting factors in the use of this trial design.

Crossover trials

In a crossover trial, two (or more) interventions are given in random sequence to all participants.[37,38] The comparison is made *within* individuals rather than *between* individuals. Unlike *n*-of-1 trials, which aim to identify the best treatment for an individual patient, crossover trials — like all the other designs described in this paper — aim to produce more generalisable conclusions about the likely effects of treatments. One advantage of crossover trials is that within-person comparisons are generally more statistically powerful than between-person comparisons, hence smaller sample sizes are needed. Another advantage is that the comparison of treatments is more direct, with potential problems of between-participant differences removed (Box 4).[39]

Other examples of interventions assessed in crossover trials undertaken in primary care include: sumatriptan for migraine,[40-41] parental administration of systemic pred-

A crossover trial was undertaken in a general practice setting to compare the effects of bisoprolol and bendrofluazide on blood pressure and quality of life.

The 81 participants all had a mean diastolic blood pressure of 95–120 mmHg after receiving placebo for four to six weeks. Participants were given bisoprolol (5 mg once daily) or bendrofluazide (2.5 mg once daily) for eight weeks, with the order of treatments randomised. Both participants and researchers were blinded to the order of treatments.

There were no significant differences in quality of life or in the systolic/diastolic blood pressure reduction achieved (bisoprolol 10/13 mmHg and bendrofluazide 9/11 mmHg).

The crossover trial could not assess the clinical outcomes resulting from the blood pressure reductions and only short-term effects on quality of life could be measured. A factorial trial of the two treatments could have overcome these problems and could in addition have allowed assessment of the effect of both drugs given together.

Box 4. Example of a crossover trial: comparison of bisoprolol and bendrofluazide for hypertension.[39]

nisolone in acute asthma,[42] and drug treatments for nocturia,[43] osteoarthritis,[44] and dysmenorrhoea.[45]

There are a number of important disadvantages and limitations of crossover trials. It is only possible to study the short-term effects of interventions. The disease needs to be chronic and stable. For this reason, a 'run-in' period is often employed, in which participants are observed over a period of time prior to the trial starting and those people with 'unstable' disease are excluded. This clearly has implications for the generalisability of the trial results to less selected groups. Treatments that aim to 'cure' a disease are clearly unsuitable for crossover trials. Some degree of carry-over effect from one treatment period into the next is likely. One strategy to reduce this problem is the use of 'wash-out' periods of either no treatment or placebo after each treatment period. However, carry-over effects can still be problematic and therefore crossover trials are best used to measure effects that wear off quickly once a treatment is stopped. Finally, participants who do not complete a trial are a particular problem in crossover trials because of the likelihood that they will not be exposed to all the interventions being assessed. They can therefore contribute no data to the comparison between treatments.

Crossover trials are widely used by the pharmaceutical industry, particularly in preliminary assessments of the effects of new drugs (phase II trials). The crossover design is a powerful method of measuring short-term biological efficacy in selected participants in an experimental setting. It is less useful in measuring the effectiveness in unselected populations of interventions as they are used in day-to-day practice.

Cluster randomised trials

A study in which the unit of randomisation and allocation consists of clusters (such as whole communities, organisations or geographical areas), rather than individuals, is referred to as a cluster randomised trial.[46-48] One of the most commonly used cluster units is the primary care clinic (Box 5).[49-58] There are a number of reasons to randomise at the cluster rather than at the individual level, including:

- The intervention occurs at a cluster level; for example, a practice-wide smoking cessation intervention.[59]
- It may be considered unethical or clinicians may feel uncomfortable about offering an intervention to some patients in a clinic while not offering the intervention to others; for example, a new screening test.[60]
- The risk of contamination between the allocation groups; for example, in a trial of safety advice provision at child health surveillance consultations, randomising families within the same practice could have led to some families in the control group being inadvertently exposed to elements of safety advice intervention.[57] Randomising at the level of the general practice aims to avoid contamination between the randomised groups, but this may not always be the most appropriate strategy.[61]

A general practice-based trial aimed to compare a nurse-led health promotion strategy for men aged 40 to 59 years and their partners with usual care. The intervention was designed to be implemented at a practice level; it therefore needed to be evaluated at a practice level.

Randomly allocating individuals would have required the same nurse to undertake health promotion with some patients but not with others; this may have been unacceptable to the nurses, practices or patients involved and may have led to contamination between the randomised groups.

The preferred solution in this case was to perform a cluster trial in which practices were randomised either to the intervention or usual care arms.

Box 5. Example of cluster randomised trial. The British Family Heart Study.[49,50]

The unit of analysis in a cluster RCT can be either the cluster level or the individual participants, provided appropriate corrections are made for any clustering effects. The choice of the unit of analysis should be dictated by the primary research question of interest.[62] Cluster level analyses are most appropriate when the intervention is aimed at the practice as a whole. For example, in a trial evaluating the effect of guidelines on radiological referral, the practice was the experimental unit and the outcome measure was the percentage of appropriate practice x-ray requests.[63] In other cluster trials, the primary target of the intervention may be the individual participant, with the choice of practice as the unit of randomisation largely made for ethical and logistical reasons. In this case, analysis at the level of individual participants would be more appropriate.

Individuals within the same cluster may be more similar than randomly selected individuals; that is, observations may be correlated.[64] The intra-class correlation coefficient is a commonly used measure of this correlation.[65] Failure to take the clustering effect into account is likely to lead to a spuriously low *P*-value and narrow confidence intervals and produce misleading results. All analyses arising from cluster trials must therefore take the clustering effect into account. The traditional approach to the analysis of cluster trials has been to compare summary measures for whole clusters. However, more powerful random effects or multi-level analyses that take into account variation at both the individual and cluster level effects are increasingly being used. Problems with many cluster RCTs undertaken in the past have been inadequate sample size and the use of inappropriate methods of statistical analyses that do not take into account the

clustered design.[66]

Choice of appropriate outcome measures

The appropriate choice of outcome measures is an integral aspect of the planning of intervention studies. Outcome measures should aim to satisfy some obvious requirements: they should be clinically significant (or be predictive of the values of 'hidden' clinically relevant variables); easy to record; easy to interpret in the study; make the results of the study easy to apply to patient care; and make the study comparable with other studies of the same or similar interventions (i.e. in meta-analysis).

An outcome measure may be a direct measure of harm or benefit (death, for instance), or it can be a surrogate endpoint. Surrogate endpoints have been defined as follows: 'A variable that is relatively easily measured and that predicts a rare or distant outcome of either a toxic stimulus (for example, a pollutant) or a therapeutic intervention (for example, a surgical procedure, or a piece of advice), but which is not itself a direct measure of either harm or clinical benefit.[67]

Surrogate endpoints have to be selected with care. They should either be clinically relevant in their own right, or reliable predictors of the important but distant or rare clinical outcomes for which they are notional surrogates. For instance, an anti-hypertensive drug might be good at reducing blood pressure, but in fact make little or no difference to the incidence of strokes or myocardial infarcts, or a difference which is small relative to the increased burden of other health problems (for instance, impotence).[68]

Outcomes can be either objective or subjective. An objective outcome is a directly or indirectly observable natural quantity or state, such as death, CD_4 count, or rate of attendance at a clinic. A subjective outcome is a state or quantity relating to a personal or psychological state. Subjective outcomes include quality of life, anxiety or depression, satisfaction with or preferences among services. Subjective outcomes can be just as reliable and generalisable as more traditional objective outcome measures. However, they do present special challenges in development and use, to ensure that they are internally and externally valid, and that they are recorded in an unbiased way.[69]

Selection of outcome measures can have a great impact on the design, size and duration of a trial. For this reason, outcome measures are sometimes chosen for convenience rather than theoretical merit, researchers often preferring short-term outcome measures rather than long-term outcome measures because of funding constraints, concerns regarding dropouts, and the pressure to publish. Research evidence shows that the 'relevance' of the outcome measures used in some clinical areas is assessed differently by patients, professionals, and sponsors.[70] This underlines the importance of involving consumers in the design of research relevant to their care. A real ethical challenge in selecting outcome measures is the tension between the use of apparently objective, easily measurable outcomes, which have little or no relevance to patient experience, and the use of measures that are relevant to patient experience, but which require more effort to validate and make generalisable.[71]

Conclusions

When both ethical and feasible, randomised trials are the best study design for assessing the effects of an intervention because of their unique ability to reduce confounding and bias.[72] A number of alternatives to the traditional parallel group design exist, and use of these alternative designs is likely to increase in the future. This paper has summarised some of the designs available and considered their roles in primary care research. There is a clear need to improve the evidence base that underpins many interventions used within primary care. To achieve this requires more sufficiently large high quality randomised trials that use the most appropriate methodology for the specific interventions and outcomes being assessed. There is also a need for better reporting of trials. The recently revised CONSORT statement aims to improve the quality of reporting of RCTs and thus facilitate interpretation and the use of the results of such trials.[74] The CONSORT statement is aimed at first reports of two group parallel designs, but modified guidelines for reporting other trial designs are in preparation.[75]

References

1. Medical Research Council. *Primary health care*. [Topic review.] London: MRC, 1997.
2. Department of Health. *R&D in primary care.* [National Working Group Report.] London: The Stationery Office, 1997.
3. Horton R. Evidence and primary care. *Lancet* 1999; **353:** 609-610.
4. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; **312:** 1215-1218.
5. Thomas T, Fahey T, Somerset M. The content and methodology of research papers published in three United Kingdom primary care journals. *Br J Gen Pract* 1998; **48:** 1229-1232.
6. Robson C. *Real world research.* Oxford: Blackwell, 1993.
7. Altman DG, Bland JM. How to randomise. *BMJ* 1999; **319:** 703-704.
8. Last JM. *Dictionary of epidemiology*. Oxford: OUP, 2001.
9. General Medical Council. *Duties of a doctor.* London: GMC, 1995.
10. Torgerson DJ, Sibbald B. Understanding controlled trials: what is a patient preference trial? *BMJ* 1998; **316:** 360.
11. Britton A, McPherson K, McKee M, *et al.* Choosing between randomised and non-randomised studies: a systematic review. *Health Technology Assessment* 1998; **2:** 37-46.
12. Klaber Moffett J, Torgerson D, Bell-Syer S, *et al.* Randomised controlled trial of exercise for low back pain: clinical outcomes, costs, and preferences. *BMJ* 1999; **319:** 279-283.
13. Torgerson DJ, Klaber-Moffett J, Russell IT. Patient preferences in randomised trials**:** threat or opportunity? *J Health Serv Res Policy* 1996; **1:** 194-197.
14. Smith PG, Day NE. The design of case-control studies: the influence of confounding and interaction effects. *Int J Epidemiol* 1984; **13:** 356-365.
15. Brewin CR, Bradley C. Patient preferences and randomised clinical trials. *BMJ* 1989; **299:** 313-315.
16. Ward E, King M, Lloyd M, *et al.* Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. I: Clinical effectiveness. *BMJ* 2000; **321:** 1383-1388.
17. Chilvers C, Dewey M, Fielding K, *et al.* Antidepressant drugs and generic counselling for treatment of major depression in primary care: randomised trial with patient preference arms. *BMJ* 2001**;** **322:** 772-775.
18. Guyatt G, Sackett D, Adachi J, *et al*. A clinician's guide for conducting randomized trials in individual patients. *CMAJ* 1988; **139:** 497-503.
19. Sackett DL, Haynes RB, Guyatt GH, *et al.* Deciding on the best therapy. In: Sackett DL, Haynes RB, Guyatt GH, Tugwell P (eds). *Clinical epidemiology*. Boston: Little, Brown & Co., 1991: 187-248.
20. Nikles CJ, Glasziou PP, Del Mar CB, *et al.* N of 1 trials. Practical tools for medication management. *Aust Fam Physician* 2000; **29:** 1108-1112.
21. March L, Irwig L, Schwarz J, *et al.* N of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. *BMJ* 1994; **309:** 1041-1045.
22. Nikles CJ, Glasziou PP, Del Mar CB, *et al.* Preliminary experiences with a single-patient trials service in general practice. *Med J Aust*

2000; **173:** 100-103.
23. Wiebe E. N of 1 trials. Managing patients with chronic fatigue syndrome: two case reports. *Can Fam Physician* 1996; **42:** 2214-2217.
24. Magee LA, Redman CW. An N-of-1 trial for treatment of hyperemesis gravidarum. *Br J Obstet Gynaecol* 1996; **103:** 478-480.
25. Kamien M. The use of an N-of-1 randomised clinical trial in resolving therapeutic doubt. The case of a patient with an 'attention disorder'. *Aust Fam Physician* 1998; **27:** S103-S105.
26. Kent MA, Camfield CS, Camfield PR. Double-blind methylphenidate trials**:** practical, useful, and highly endorsed by families. *Arch Pediatr Adolesc Med* 1999; **153:** 1292-1296.
27. Duggan CM, Mitchell G, Nikles CJ, *et al.* Managing ADHD in general practice. N of 1 trials can help! *Aust Fam Physician* 2000; **29:** 1205-1209.
28. Mahon JL, Laupacis A, Hodder RV, *et al*. Theophylline for irreversible chronic airflow limitation: a randomized study comparing n of 1 trials to standard practice. *Chest* 1999; **115:** 38-48.
29. Zucker DR, Schmid CH, McIntosh MW, *et al.* Combining single patient (N-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. *J Clin Epidemiol* 1997; **50:** 401-410.
30. Karnon J, Qizilbash N. Economic evaluation alongside n-of-1 trials: getting closer to the margin. *Health Econ* 2001; **10:** 79-82.
31. Pocock SJ. *Clinical trials. A practical approach.* Chichester: Wiley, 1983.
32. Imperial Cancer Research Fund General Practice Research Group. Effectiveness of a nicotine patch in helping people stop smoking: results of a randomised trial in general practice. *BMJ* 1993; **306:** 1304-1308.
33. Dexter PR, Wolinsky FD, Gramelspacher GP, *et al*. Effectiveness of computer-generated reminders for increasing discussions about advance directives and completion of advance directive forms. A randomized, controlled trial. *Ann Intern Med* 1998; **128:** 102-110.
34. Peveler R, George C, Kinmonth AL, *et al.* Effect of antidepressant drug counselling and information leaflets on adherence to drug treatment in primary care: randomised controlled trial. *BMJ* 1999; **319:** 612-615.
35. McBride P, Underbakke G, Plane MB, *et al*. Improving prevention systems in primary care practices: the Health Education and Research Trial (HEART). *J Fam Pract* 2000; **49:** 115-125.
36. The Medical Research Council's General Practice Research Framework. Thrombosis prevention trial: randomised trial of low-intensity oral anticoagulation with warfarin and low-dose aspirin in the primary prevention of ischaemic heart disease in men at increased risk. *Lancet* 1998; **351:** 233-241.
37. Woods JR, Williams JG, Tavel M. The two-period crossover design in medical research. *Ann Intern Med* 1989; **110:** 560-566.
38. Altman DG. *Practical statistics for medical research*. London: Chapman & Hall, 1991.
39. Vanmolkot FH, de Hoon JN, van de Ven LL, *et al.* Impact of antihypertensive treatment on quality of life: comparison between bisoprolol and bendrofluazide. *J Hum Hypertens* 1999; **13:** 559-563.
40. Russell MB, Holm-Thomsen OE, Rishoj NM, *et al.* A randomized double-blind placebo-controlled crossover study of subcutaneous sumatriptan in general practice. *Cephalalgia* 1994; **14:** 291-296.
41. Jensen K, Tfelt-Hansen P, Hansen EW, *et al.* Introduction of a novel self-injector for sumatriptan. A controlled clinical trial in general practice. *Cephalalgia* 1995; **15:** 423-429.
42. Grant CC, Duggan AK, DeAngelis C. Independent parental administration of prednisolone in acute asthma: a double-blind, placebo-controlled, crossover study. *Pediatrics* 1995; **96:** 224-229.
43. Pedersen PA, Johansen PB. Prophylactic treatment of adult nocturia with bumetanide. *Br J Urol* 1988; **62:** 145-147.
44. Gostick N, James IG, Khong TK, *et al.* Controlled-release indomethacin and sustained-release diclofenac sodium in the treatment of osteoarthritis: comparative controlled clinical trial in general practice. *Curr Med Res Opin* 1990; **12:** 135-142.
45. Tilyard MW, Dovey SM. A comparison of tiaprofenic acid, mefenamic acid and placebo in the treatment of dysmenorrhoea in general practice. *Aust NZ J Obstet Gynaecol* 1992; **32:** 165-168.
46. Murray DM. *Design and analysis of community trials*. Oxford: Oxford University Press, 1998.
47. Ukoumunne OC, Gulliford MC, Chinn S, *et al.* Methods in health service research. Evaluation of health interventions at area and organisation level. *BMJ* 1999; **319:** 376-379.
48. Donner A, Klar N. *Design and analysis of cluster randomization trials in health research*. London: Arnold, 2000.
49. Family Heart Study Group. British Family Heart Study: its design and method, and prevalence of cardiovascular risk factors. *Br J Gen Pract* 1994; **44:** 62-67.
50. Wood DA, Kinmonth AL, Davies GA, *et al.* Randomised controlled trial evaluating cardiovascular screening and intervention in general practice**:** principal results of British family heart study. *BMJ* 1994; **308:** 313-320.
51. Oakeshott P, Kerry SM, Williams JE. Randomized controlled trial of the effect of the Royal College of Radiologists' guidelines on general practitioners' referrals for radiographic examination. *Br J Gen Pract* 1994; **44:** 197-200.
52. Kendrick T, Burns T, Freeling P. Randomised controlled trial of teaching general practitioners to carry out structured assessments of their long term mentally ill patients. *BMJ* 1995; **311:** 93-98.
53. Roderick P, Ruddock V, Hunt P, *et al.* A randomized trial to evaluate the effectiveness of dietary advice by practice nurses in lowering diet-related coronary heart disease risk. *Br J Gen Pract* 1997; **47:** 7-12.
54. Alexander FE, Anderson TJ, Brown HK, *et al.* Fourteen years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* 1999; **353:** 1903-1908.
55. Kerse NM, Flicker L, Jolley D, *et al.* Improving the health behaviours of elderly people: randomised controlled trial of a general practice education programme. *BMJ* 1999; **319:** 683-687.
56. Steptoe A, Doherty S, Rink E, *et al.* Behavioural counselling in general practice for the promotion of healthy behaviour among adults at increased risk of coronary heart disease: randomised trial. *BMJ* 1999; **319:** 943-947.
57. Kendrick D, Marsh P, Fielding K, *et al.* Preventing injuries in children: cluster randomised controlled trial in primary care. *BMJ* 1999; **318:** 980-983.
58. Montgomery AA, Fahey T, Peters TJ, *et al.* Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomised controlled trial. *BMJ* 2000; **320:** 686-690.
59. Butler C, Bachmann M. Design and analysis of studies evaluating smoking cessation interventions where effects vary between practices or practitioners. *Fam Pract* 1996; **13:** 402-407.
60. Edwards SJ, Braunholtz DA, Lilford RJ, *et al.* Ethical issues in the design and conduct of cluster randomised controlled trials. *BMJ* 1999; **318:** 1407-1409.
61. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ* 2001; **322:** 355-357.
62. Altman DG, Bland JM. Statistics notes. Unit of analysis. *BMJ* 1997; **314:** 1874.
63. Oakeshott P, Kerry SM, Williams JE. Randomized controlled trial of the effect of the Royal College of Radiologists' guidelines on general practitioners' referrals for radiographic examination. *Br J Gen Pract* 1994; **44:** 197-200.
64. Donner A, Klar N. Statistical considerations in the design and analysis of community intervention trials. *J Clin Epidemiol* 1996; **49:** 435-439.
65. Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. *BMJ* 1998; **316:** 1455.
66. Ukoumunne OC, Gulliford MC, Chinn S, *et al.* Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technology Assessment* 1999; **3:** 376-379.
67. Greenhalgh T. *How to read a paper: the basics of evidence based medicine.* London: BMJ Books, 1997: 90.
68. Chalmers I. What do I want from health research and researchers when I am a patient. *BMJ* 1995; **310:** 1315-1318.
69. Bowling A. *Measuring health: a review of quality of life measuring scales.* Milton Keynes: Open University Press, 1997.
70. Fitzpatrick R, Davey C, Buxton MJ, *et al.* Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; **2:** 1-69.
71. Tallon D, Chard J, Dieppe PA. Relation between agendas of the research community and the research consumer. *Lancet* 2000; **355:** 2037-2040.
72. World Medical Association. *Declaration of Helsinki: Ethical principles for medical research involving human subjects, 2000*. http**:** //www.wma.net/e/policy/17-c_e.html (Accessed 24 October 2001).
73. Altman DG, Schulz KF, Moher D, *et al.* The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals Int Med* 2001; **134:** 663-694.
74. Elbourne DR, Campbell MK. Extending the CONSORT statement to cluster randomized trials: for discussion. *Stat Med* 2001; **20:** 489-496.

Acknowledgements

We thank the three anonymous referees for their constructive comments on an earlier draft of this paper. Dr Aziz Sheikh is supported by an NHS R&D National Primary Care Award. Dr Liam Smeeth is funded by a research fellowship from London NHS Executive. Dr Richard Ashcroft is partly supported by the Leverhulme Trust.