Age and Ageing 2002; **31:** 227–232

COMMENTARY

N-of-1 randomized controlled trials ('N-of-1 trials'): singularly useful in geriatric medicine

Conventional evidence-based medicine is concerned with what happens on average, but clinicians and their patients are more concerned with knowing what is going to happen in a particular case [1]. All too often, large randomized controlled trials (RCTs) are carried out on populations so ill-characterized that clinicians cannot be sure whether it is appropriate to extrapolate the results to individual patients confronting them. Subgroup analysis to generate hypotheses about indicators of individual outcome should be the logical development following large RCTs. Although performed in some instances with distinction [2], *post hoc* subgroup analyses are actively discouraged by statisticians for fear that doctors may not distinguish between hypothesis-generation and hypothesis-testing. The problem is compounded by a widespread but misguided perception of medical services as a public health tool rather than a service for individuals. Widespread confusion of 'equity' with 'equality' encourages the imposition of uniformity of prescription regardless of variations in individual appropriateness.

With chronic illnesses, and treatment outcomes more subtle than death or survival, N-of-1 randomized controlled trials ('N-of-1 trials') can provide an objective basis for identifying the best treatment for an individual patient. In essence, each trial consists of a random sequence of different treatments that may include placebo, administered in double-blind protocol, with regular and standardized measurements of relevant effects. Relevant effects might be physiological or functional such as blood pressure or exercise tolerance, but more often will be patient-generated measurements such as ratings of pain or quality of life.

Experimental studies of single subjects have long been a part of psychological research [3], and most research designs have been developed in that context. N-of-1 trials have, however, been recognized as relevant to general medical practice since their introduction into mainstream medical literature 15 years ago [4]. Given their utility and their potential for integration with normal clinical practice, it is surprising that they are not more commonly used [5], particularly as a means of optimizing prescription for older patients with chronic, multiple and interacting illnesses.

In this review we discuss the utility, methodology and evidence for N-of-1 trials in geriatric medicine.

Special relevance to geriatric medicine

Geriatricians are frequently confronted by a lack of evidence to guide their treatment of individual older patients. Most Cochrane reviews observe that the general quantity and quality of evidence about specific clinical questions is poor. Large-scale randomized controlled trials have not been or cannot be carried out for many clinical disorders, particularly those seen as less important by decision-makers, but of common occurrence in later life. The problem has been compounded for geriatricians by the exclusion of older people from trials, observed at the level of ethics committee approval [6] and publication [7], and condemned by the Medical Research Council [8].

Increasingly, licensing agencies demand the recruitment of older people into trials, but usually patients with the comorbidity and medical instability common in a geriatric or general medical clinic are excluded. Where new drugs are being evaluated, and interactions and adverse effects have yet to be more fully explored, caution on the part of trialists is understandable. But disasters such as that associated with benoxaprofen (Opren) [9] underline the need for patients recruited into trials to be representative of those for whom the treatment is likely to be prescribed.

Outcome measurements that are methodologically convenient and typical for contemporary large trials may not be appropriate or important for individual older patients. 'Hard' outcome measurements such as fatality or 'clinical events' may be less relevant for older people than measurements reflecting pain, quality of life or well-being. The values ascribed to elements of functional status scores are known to differ markedly between patients and healthcare professionals [10].

In addition, the increasing variance, both intra-individual and inter-individual, in physiological and psychological functioning that comes with age needs to be taken into account. If maximum therapeutic benefit is to be combined with minimal adverse effects, the choice of drugs and their appropriate dosages are likely

to be more critical for older than for younger people; N-of-1 trials can be used in the comparison of drug doses during drug development and in determining the optimal dose for individuals in subsequent clinical practice [11].

No treatment is effective in all patients. Even in instances where there is no evidence of counterbalancing ill effects of a treatment, large trials teach us not to expect a predictable physiological response, merely an improvement in the odds of a beneficial effect. Usually, only a modest proportion of patients show a good response. An important benefit of N-of-1 trials in the service clinic is to provide reliable identification of individual non-response or harm. This avoids the costs of ineffective treatment as well as sparing the patient from adverse effects. These potential benefits are greater for treatments that are expensive, prolonged or where adverse effects may be uncommon although severe or irreversible.

The causality of an association between an intervention and an adverse event that happens to occur at the same time may be in doubt, and may be evaluated by means of an N-of-1 trial. This issue of *Age and Ageing* includes a description of a trial in which a significant symptom that might have led to the discontinuation of an active treatment was found later to have occurred during a placebo phase [12].

Methods and methodology

The informal 'trial of treatment' for an individual, employed commonly in clinical practice, is susceptible to a number of forms of bias [4]. These include:

- Placebo response. Although perhaps less powerful than once thought [13], this remains an important confounder in trials with subjective endpoints.
- Secondary effects. For example, the general sense of well-being often induced by corticosteroids may obscure assessment of the intended specific benefit.
- Natural history of the illness. Many diseases improve naturally with time. Patients with relapsing-remitting conditions will seek medical attention more commonly during relapses, so improvement may occur during treatment simply through reversion towards the mean.
- Expectations. Patient and clinician may have positive expectations about the treatment effect and/or may not wish to disappoint each other.
- Uncontrolled interactions. The clinician may not be aware of all the factors that may influence the effect of a prescribed treatment. In particular, use of over-the-counter or previously prescribed remedies is not uncommon in geriatric practice.
- Outcome assessments with significant measurement error may mislead if there are insufficient replications.

Several design features of N-of-1 trials can contribute to reducing bias and ensuring the validity of results [14, 15]. The detailed design of a particular trial will depend on the nature of the clinical question and intervention, the views of the patient (and carer) and local circumstances including availability of trial support services.

N-of-1 trials require the rigour of double-blinding, or as a minimum, blind assessment of outcome by the patient or a third party. Randomization of treatment sequences is essential, against control by either placebo or comparison drug. Treatments (A and B) may be randomized (r) within sequential pairs (eg, r<u>AB</u> r<u>BA</u> r<u>BA</u> . . .) or *en bloc (*eg, r<u>AABABB</u> . . .); alternatively, systematic alternation of treatments within pairs may follow a single, initial randomization (eg, r<u>ABABAB</u> . . .), although this is more vulnerable to carry-over effects and inadvertent unblinding.

The patient and clinician are then presented with a series of treatment periods. Usually only two treatments (or treatment and placebo) are compared but only analytical complexity inhibits the evaluation of more. Usually, three or more replications are performed, in order to reduce the chance that non-drug-related trends will lead to false conclusions. The number of periods may be predetermined by convenience or by a power calculation; more pairs are required where a confident conclusion is demanded or where treatment effect is expected to be modest in comparison with the underlying fluctuation of the condition or the sensitivity of the outcome measurements. It is salutary to note that within a trial consisting of a series of three pairs and where outcome is assessed only qualitatively as 'better' or 'worse', a consistent preference for treatment over placebo will occur by chance alone with a probability of $1/2^3 = 0.125$; quantitative evaluation enhances the sensitivity (power) of a trial. Alternatively, comparisons may be continued under the guidance of an independent unblinded data monitor who terminates the trial when one treatment emerges as more effective or associated with more adverse effects than another, or it becomes apparent that there is no meaningful difference between treatments [16].

Outcome criteria should be reliable, valid, defined before the trial begins and assessed at time points that are appropriate to the intervention. Treatment intervals must allow for known delays in onset of drug action or for washout periods. Ill effects from stopping and starting treatments must not be overlooked, as might occur if reviews were restricted to each period's end.

Where possible, outcome criteria should include objective measurements of function, as well as more subjective assessment of symptoms or well-being. Asking the patient to identify the most troubling symptoms may help determine a suitable treatment outcome measurement. Respondent-generated measurements such as the Patient-Generated Index [17] and Schedule for the Evaluation of Individual Quality of Life (SEIQoL) [18] have face validity, especially when applied to individuals

or groups with characteristics distinct from those on whom researcher-generated instruments have been validated. Where symptoms fluctuate, the use of a diary or a daily self-completed questionnaire may enable a quantitative estimate of time spent with or without symptoms. Severity of symptoms may be assessed by visual analogue scales, but not all patients find these easy to understand. In some circumstances, for example in testing for an effect of a drug aimed at enhancing cognitive function, practice effects of at least two types may occur. A general effect due to growing familiarity with the test procedure may be avoided by the inclusion of a run-in training period. Specific learning of answers to repeated test items can be obviated by the use of equivalent but non-identical versions of a test. The Hopkins Verbal Learning Test, with its six validated versions provides an example [19].

There is no consensus on how the results of an N-of-1 trial may best be appraised. Simple informal visual assessment of raw or graphical data by clinician and patient has been widely advocated, as it is swift, straightforward, and has face validity in recognising clinically meaningful benefit. Sceptics may argue that such an approach is subjective, open to bias, and does not quantify the likelihood of a false positive result (Type I error). Reassuringly, the results of visual inspection may approximate those of a statistical approach [20].

The potential diversity of N-of-1 trial designs is matched by the spectrum of suggested statistical analysis techniques and the debate around their appropriate application and validity [4, 11, 14, 21–24]. Data from N-of-1 trials often violate the strict assumptions underlying parametric tests such as the t-test, although usually not to a degree that threatens validity [4, 11]. Fewer assumptions are made in applying non-parametric methods, the simplest (although least powerful) of which is the sign test. Bayesian methods generate distinct effectiveness estimates for each patient, incorporating information obtained from others [25]. In the individual case, statistical evaluation is likely to be tempered by other inputs to the decision-making process. These may include the relative costs and benefits of false positive and false negative outcomes as determined by factors such as illness severity and implications, treatment characteristics and the preferences of patient, carer and clinician.

N-of-1 trials may present practical difficulties. They can be time-consuming for clinician and patient, and there may be difficulties in obtaining matching drug and control formulations although sometimes, particularly with newer drugs, matching placebos remain from earlier trials. Successful blinding of placebo-controlled trials of drugs with characteristic effects such as dry mouth may be impossible, although adverse effects are not confined to active treatments [12] and in many trials the comparison is between two active drugs with similar side effects. Pharmacy support is helpful in design, randomization, preparation of placebo, supply and packaging, in holding the randomization code to maintain blinding, and in interpretation of results [4].

Rigour versus pragmatism: when to perform an N-of-1 trial

It is useful—given the heterogeneity of physiology, pathology, therapeutic response, and patient values—to consider all therapeutic interventions in chronic conditions as single patient experiments or 'trials of treatment'. Routinely, clinical effectiveness will be assessed through the careful assessment of (often surrogate) outcomes. This assessment may be quantitative (eg. blood pressure fall following antihypertensive therapy), qualitative (eg. reduction in pain following analgesia) or, less commonly, both quantitative and qualitative (eg. the assessment of response to cholinesterase inhibitors in dementia, utilizing both clinician's global impression and a validated questionnaire).

When to adopt the greater rigour of the more formal N-of-1 trial depends on a consideration of the balance of costs and benefits in the individual case. Costs include those of the clinicians and patient's time, of drug and placebo preparation and dispensing and, importantly, the possibility of delay in introduction of an effective treatment.

N-of-1 trials are more appropriate when several of the following criteria coexist:

- There are significant doubts about treatment effectiveness for an individual.
- The potential exists for important treatment benefit or harm.
- Prolonged or expensive drug treatment is being considered.
- The disease is chronic and relatively stable, or frequently recurring, so that modest but clinically important treatment or preventive effects can be detected.
- Relevant outcomes are measurable.
- The patient is enthusiastic and is likely to comply through a sufficiently prolonged trial.
- The clinician has adequate time and expertise, or has access to a trials support service.

Ethics

N-of-1 trials use some of the tools developed for research, but are undertaken to improve treatment of the participating patient. They are more appropriately viewed, therefore, as contributing to optimal clinical care rather than as research in a conventional sense [14, 26]. Indeed it can be argued that they should be a part of routine clinical practice, and encouraged by a cost-aware healthcare system. It is therefore desirable that 'class approval' for standard designs of N-of-1 trials should be agreed with Local Research Ethics Committees so that it is not necessary for each trial to be submitted separately. If a new intervention is being

examined, or a new indication for an existing treatment, then Research Ethics Committee approval will need to be obtained.

The choice of placebo or an alternative potentially active treatment as comparator will need thought, both in the light of the question being asked, and in acceptability to the patient. The general requirement for informed consent by the participating patient, and the principles of partnership in care and individualized evidence-based medicine [1] are integral to the ethical conduct of N-of-1 trials. Local Research Ethics Committees vary in their requirements for documentation, but a written description of the trial should be available, as well as explicit data monitoring procedures to initiate unblinding of the trial design if the patient suffers a possibly relevant intercurrent illness.

The existence of parallel group evidence of the efficacy of a treatment does not in itself raise ethical problems for the use of a placebo control in an N-of-1 trial. Except in the unlikely situation of a 100% response rate, individuals vary in the benefit they will obtain from a treatment, and they will also vary in the incidence and severity of side-effects. The purpose of an N-of-1 trial is to find how the average effects identified in parallel group studies apply to the individual. It is not only ethical but arguably obligatory to undertake a placebo-controlled N-of-1 trial to find if a patient is a responder to a treatment known to be of benefit to only a proportion of those who receive it. Into the ethical equation may also go the potential costs of not identifying that a patient is a non-responder.

Examples

It is likely that only a small proportion of N-of-1 trials performed in service clinics are submitted for publication and that only a modest proportion of these are accepted, risking publication bias and impairing the dissemination of good clinical and methodological practice. A few reports of summation of N-of-1 trials have been published, describing trials of inhaled therapy in chronic obstructive pulmonary disease [27], amitriptyline in fibromyalgia [28], non-steroidal anti-inflammatory drugs in osteoarthritis [29, 30], enalapril in essential hypertension [31] and of drugs for some other conditions [23]. Therapy was judged to be ineffective in 19–60% of patients, often despite apparent benefit in previous open trials, and clinician confidence in treatment plans was increased [23].

A single, parallel group RCT compared N-of-1 trials with usual clinical practice in the evaluation of theophylline in chronic airflow limitation [32]; drug prescription was reduced in the N-of-1 group, without detriment to the patients' functional status. There are few reports of N-of-1 trials in the geriatrics literature, although there is a recent small study of methylphenidate in depression or dementia [33].

Comment

Conventional individual therapeutic trials are quick, inexpensive and straightforward. Clinicians will continue to rely on them at times, despite their disadvantages. However, N-of-1 trials have greater value in guiding clinical practice, especially for older people, where the evidence base for treatment may be weak or of uncertain relevance, or where an established treatment carries high risks, high costs, will be used for extended periods, or has apparently been poorly tolerated by a patient. In the case of new, expensive, therapies restricted by constraints on funding, N-of-1 trials can furnish a powerful evidence base for provision on an individual basis [34], allaying managerial fears of the cost of infrequently effective therapies being applied to an expanding elderly population.

In view of the complexities of assessment and treatment of older people, it is surprising and disappointing that N-of-1 trials are not more frequently offered to them. Ironically, the complexity that renders clinical assessment uncertain and strengthens the case for more objective evaluation may also deter application of the N-of-1 trial. The infrequency of N-of-1 trials may also reflect the passivity of some older people in their dealings with health professionals. Practitioners should be more active in identifying areas of clinical doubt and ready to discuss the possibility of a trial with their patients. More or less sophisticated models of decision analysis [35] can be employed in structuring the dialogue.

Clinicians' perceptions of N-of-1 trials as time-consuming and complex may have deterred their application. Comprehensive pharmacy support for N-of-1 trials existed elsewhere [4] and facilitated their use. Unfortunately such help may be unavailable or judged unacceptably expensive. As N-of-1 trials should form a part of good clinical practice, and have been demonstrated to reduce ineffective prescribing, support should form a part of the package of services provided by hospital pharmacies.

Within the context of research, the differing purpose of parallel group and N-of-1 trials must be borne in mind. In particular, the results of a single N-of-1 trial cannot be generalized to other patients (indeed, heterogeneity of patient response is often the reason for such a trial). However, summation of results from a series of trials could supplement estimates of the proportion of responders, particularly if entry criteria, treatments and outcome measurements were homogeneous. A case can be made for national initiatives for encouraging and registering the results of N-of-1 trials, perhaps linked to genetic or other possible identifiers of responder status. N-of-1 trials can also be deployed to explore extensions of indications for drugs, for example the role of cholinesterase inhibitors for memory impairment due to conditions other than Alzheimer's disease [12]. Geriatricians are well placed to promote such initiatives,

through information-sharing research networks, and establishing an obligation on clinicians to develop the collectivist evidence from parallel group trials into the individualized knowledge necessary for good medical care.

James D. Price
John Grimley Evans
Division of Clinical Geratology, Nuffield Department of Clinical Medicine, University of Oxford, Radcliffe Infirmary, Oxford OX2 6HE, UK
Fax: (+44) 1865 224815
Email: james.price@geratology.oxford.ac.uk

References

1. Grimley Evans J. Evidence-based and evidence-biased medicine. Age Ageing 1995; 24: 461–3.

2. Rothwell PM, Warlow CP. Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. Lancet 1999; 353: 2105–10.

3. Martin M. Single subject designs for assessment of psychotropic drug effects in children. Child Psychiat Hum Dev 1971; 2: 102–15.

4. Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy – randomized trials in individual patients. N Engl J Med 1986; 314: 889–92.

5. Porta M, Bolumar F, Hernandez I, Vioque J. N of 1 trials. Research is needed into why such trials are not more widely used. Br Med J 1996; 313: 427.

6. Bayer A, Tadd W. Unjustified exclusion of elderly people from studies submitted to research ethics committee for approval: descriptive study. Br Med J 2000; 321: 992–3.

7. Bugeja G, Kumar A, Banerjee AK. Exclusion of elderly people from clinical research: a descriptive study of published reports. Br Med J 1997; 315: 1059.

8. Medical Research Council. The health of the UK's elderly people. London: Medical Research Council, 1994.

9. Anonymous. Opren scandal. Lancet 1983; 1: 219–20.

10. Kane RL, Rockwood T, Philp I, Finch M. Differences in valuation of functional status components among consumers and professionals in Europe and the United States. J Clin Epidemiol 1998; 51: 657–66.

11. Guyatt GH, Heyting A, Jaeschke R, Keller J, Adachi JD, Roberts RS. N of 1 randomized trials for investigating new drugs. Control Clin Trials 1990; 11: 88–100.

12. Price JD and Grimley Evans J. An N-of-1 Randomized Controlled Trial ('N-of-1 trial') of donepezil in the treatment of non-progressive amnestic syndrome. Age Ageing 2002; 31: 307–9.

13. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med 2001; 344: 1594–602.

14. Guyatt G, Sackett D, Adachi J *et al*. A clinician's guide for conducting randomized trials in individual patients. Clin Med Ass J 1988; 139: 497–503.

15. Anonymous. Randomized controlled trials in single patients (editorial). Drug Ther Bull 1998; 36: 40.

16. Armitage P. Sequential medical trials. Oxford: Blackwell Scientific, 1975.

17. Ruta DA, Garratt AM, Leng M, Russell IT, MacDonald LM. A new approach to the measurement of quality of life. The Patient-Generated Index. Med Care 1994; 32: 1109–26.

18. Browne JP, O'Boyle CA, McGee HM *et al*. Individual quality of life in the healthy elderly. Qual Life Res 1994; 3: 235–44.

19. Brandt J. The Hopkins Verbal Learning Test: development of a new memory test with six equivalent forms. Clin Neuropsychol 1991; 5: 125–42.

20. Bobrovitz CD, Ottenbacher KJ. Comparison of visual inspection and statistical analysis of single-subject data in rehabilitation research. Am J Phys Med Rehabil 1998; 77: 94–102.

21. Wulff HR. Single-case studies. An introduction. Scand J Gastroenterol 1988; 147: (suppl.) 7–10.

22. Spiegelhalter DJ. Statistical issues in studies of individual response. Scand J Gastroenterol 1988; 147: (suppl.) 40–5.

23. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The N-of-1 randomized controlled trial: clinical usefulness. Our three-year experience. Ann Intern Med 1990; 112: 293–9.

24. Senn S. Statistical issues in drug development. Wiley: Chichester, 1995; pp. 249–55.

25. Zucker DR, Schmid CH, McIntosh MW, D'Agostino RB, Selker HP, Lau J. 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: 401–10.

26. Irwig L, Glasziou P, March L. Ethics of N-of-1 trials. Lancet 1995; 345: 469.

27. Patel A, Jaeschke R, Guyatt GH, Keller JL, Newhouse MT. Clinical usefulness of N-of-1 randomized controlled clinical trials in patients with nonreversible chronic airflow limitation. Am Rev Respir Dis 1991; 144: 962–4.

28. Jaeschke R, Adachi J, Guyatt G, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. J Rheumatol 1991; 18: 447–51.

29. March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. N of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. Br Med J 1994; 309: 1041–5.

30. Nikles CJ, Glasziou PP, Del-Mar CB, Duggan CM, Clavarino A, Yelland MJ. Preliminary experiences with a single-patient trials service in general practice. Med J Aust 2000; 173: 100–3.

31. Chatellier G, Day M, Bobrie G, Menard J. Feasibility study of N-of-1 trials with blood pressure self-monitoring in hypertension. Hypertension 1995; 25: 294–301.

32. Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice. Br Med J 1996; 312: 1069–74.

33. Jansen IH, Olde-Rikkert MG, Hulsbos HA, Hoefnagels WH. Toward individualized evidence-based medicine: five 'N of 1' trials of methylphenidate in geriatric patients. J Am Geriatr Soc 2001; 49: 474–6.

34. Bollert FGE, Paton JY, Marshall TG, Calvert J, Greening AP, Innes JA. Recombinant DNase in cystic fibrosis: a protocol for targeted introduction through n-of-1 trials. Eur Resp J 1999; 13: 107–13.

35. Sarasin FP. Decision analysis and the implementation of evidence-based medicine. QJM 1999; 92: 669–71.