

The history and development of N-of-1 trials

RD Mirza¹, S Punja², S Vohra² and G Guyatt¹

¹Faculty of Health Sciences, McMaster University, Hamilton, ON L8S 4L8 Ontario, Canada

²Department of Pediatrics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB T6G 2R3 Alberta, Canada

Corresponding author: S Vohra. Email: svohra@ualberta.ca

Introduction

‘Trials of therapy’, in which physicians ‘try out’ treatments and assess patients’ responses, are long-established, common elements of routine medical practice. Because ‘trials of therapy’ are usually informal, they may only be reported if treatments are associated with dramatic changes in a patient’s condition – whether by improvement or deterioration.

Our understanding of bias suggests that informal ‘trials of therapy’ – comparisons of patients’ condition before and after treatment – do not provide a trustworthy basis for inferring treatment effects. More sophisticated comparisons are usually needed: for example, comparing a patient’s responses when treatments are given or withheld (‘crossed over’) and conducting formal assessment of outcomes.

In 1676, Richard Wiseman (a surgeon to King Charles II) reported an unplanned experiment. He had prescribed a pair of laced stockings for a patient suffering from leg oedema. The stockings had reduced the oedema to the extent that the patient ‘was able to walk to his closet, and take the air in his coach, and was well pleased with them’.¹ However, someone suggested to the patient that the stockings might do him harm and persuaded him to remove them. His legs swelled up, he became confined to bed again and developed leg ulcers. Dr Wiseman waited six weeks for the ulcers to heal, restored the laced stockings, with the result that the patient recovered.

A century after Wiseman’s crude crossover trial of laced stockings, Caleb Parry,^{2,3} a doctor in Bath, England, published a more formal, planned use of between two and six crossover periods of variable duration in 13 patients, to compare the purgative effects of three varieties of rhubarb. Parry was unable to find any advantage of the more costly Turkish rhubarb compared with English rhubarb.

Parry’s ‘trials of therapy’ were important in having used at least two crossovers, but he took no steps to ensure that his and his patients’ assessments of the treatment effects were not influenced by his or the patients’ knowledge of the type of rhubarb being given. Fourteen years later, also in Bath, John Haygarth⁴ compared the

effects on rheumatism of a metal ‘tractor’ with a matched wooden (placebo) tractor. This demonstrated that the assumed treatment effects of the metal tractor resulted from patients’ imagination.⁵

Haygarth’s study made clear that informal ‘trials of therapy’ can be plagued by false positives (due to placebo effects, physicians’ and patients’ desires to please, the pre-existing expectations of both parties and natural history). And they can also result in false negatives (patients destined to deteriorate and the intervention resulting in them remaining stable). Although more than a century passed after Haygarth before Paul Martini set out principles for designing unbiased crossover trials in his 69-page book,^{6,7} it appears that it was not until 1953 that serious scientific consideration was given to how controlled trials in individual patients could complement traditional parallel group trials. Hogben and Sim⁸ recognised that:

The now current recipe for a clinical trial based on group comparison sets out a balance sheet in which individual variability with respect both to nature and to previous nurture does not appear as an explicit item in the final statement of the account; but such variability of response to treatment may be of paramount interest in practice.

Trialists conducting parallel group trials using alternate or random allocation had been trying for half a century to deal with the challenge of deducing how to treat individual patients by using estimates of effects in subgroups of participants, but this was only a partial way of addressing the fundamental underlying issue – ascertaining individual responses.⁹

The experiment reported by Hogben and Sim is a methodological landmark (see Appendix 1 for a list of N-of-1 trials completed to date), celebrated more than half a century later by republication and commentaries in the *International Journal of Epidemiology*.^{10–12} One of the commentaries¹² summarises the features of the study:

Because they used patient’s self-reported symptoms, they put a particular emphasis on careful blinding:

the use of a placebo and keeping both clinical and patient unaware of the sequence of treatments. They were also concerned about the non-specific response to prostigmine so they used two comparators: dexamphetamine (a stimulant) and lactose (as an inert placebo). Their weighted analysis, based on concerns about wash-out and wash-in effects, also appears to be novel. Finally, with a minimum of eight periods for each treatment, they seemed to have set a new record for the number of crossovers in any crossover trial in an individual patient.

Hogben's and Sim's paper does not appear to have had an impact – possibly because it was published in a non-clinical journal. Glasziou¹² identified only 12 citations, and only one of those reported a replication of their methods (in 30 patients in a neurosis unit).¹³ Thereafter, these two studies and developments in the application of single subject design methodology in the social sciences¹⁴ appear to have gone unnoticed in the medical community until 1986.

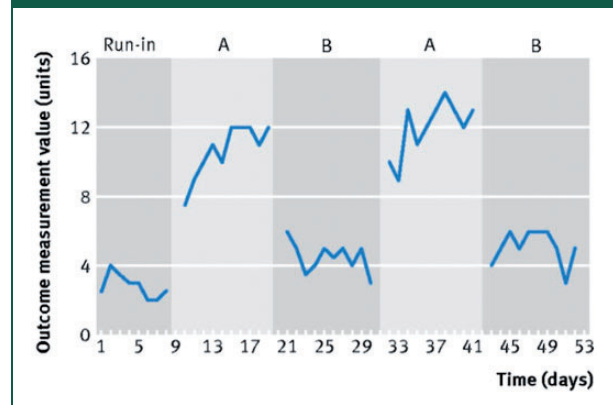
Baskerville *et al.*¹⁵ were the first to apply principles of adaptive design to the N-of-1 model. Instead of fixed treatment periods, length was determined by adverse events, clinical deterioration, and patient preference. Their model was further expanded to account for typical crossover features, including carry-over effects.¹⁶

N-of-1 trials come of age

In 1986, in the *New England Journal of Medicine*, a group of clinical investigators at McMaster University, Canada, published a paper entitled 'Determining optimal therapy – randomized trials in individual patients', in which they labelled such studies 'N of 1 randomized control trials'.¹⁷ Their interest had been prompted by a poorly controlled asthmatic patient treated with inhaled beta agonists, theophylline and prednisone. The N-of-1 trial they designed addressed the utility of the theophylline the patient was using. After the second paired block of theophylline and placebo, the patient ended the trial early: the results were clear to him, and, from the symptom diary he had been keeping, to the clinician who instituted the trial. When the blind was broken, it was clear that during the periods when the patient had been using theophylline his symptoms were much worse. Improvement was sustained when theophylline was withheld after the trial ended, with much better asthma control despite a reduced dose of steroids. The trial proved spectacularly helpful: improved symptom control, reduced drug burden and decreased costs.

Among the class of single patient/person study designs,^{18–20} N-of-1 trials are unique as rigorously

Figure 1. Depiction of N-of-trial. Modified from Shamseer *et al.*²¹



controlled intervention studies that can provide a basis for inferring cause and effect. Though many variations exist, the work that originated at McMaster University focused on single patient trials with two or more pairs of treatment periods, one for the intervention and one for the comparator, ideally with blinding of both patients and healthcare providers (Figure 1). The outcome measures in such trials are the experiences of the patients, recorded using individualised, patient-reported outcomes.

Clinicians have now formally reported on hundreds, if not thousands, of N-of-1 trials, exploring their utility in avoiding unnecessary treatment and improving patient outcomes, and also in facilitating drug development (See Appendix 1). Despite these reports, and the enormous potential that the originators saw for use of N-of-1 trials, their uptake has remained limited in the decades since 1986, although there have been recent signs of renewed interest.^{22–25}

The N-of-1 niche

The N-of-1 trial identifies whether an intervention is likely to benefit or cause unwanted effects in an individual patient. The design is most suited to assessing interventions that act and cease to act quickly. It is particularly useful in clinical contexts in which variability in patient responses is large, when the evidence is limited, and/or when the patient differs in important ways from the people who have participated in conventional randomised controlled trials. Examples include conditions with quickly acting symptomatic treatment, in which variability in response is large (e.g. chronic pain, obstructive lung disease); conditions with a prevalence too low for large, parallel group randomised controlled trials; medically complex patients who differ substantially from patients who have participated in existing

trials; and patients who have been treated over a long time when there is uncertainty about ongoing need for treatment (e.g. proton pump inhibitors in longstanding dyspepsia). Indeed, the applicability of the results of parallel group randomised clinical trials to individual patients (i.e. external validity) may sometimes be limited by narrow inclusion criteria and the exclusion of patients with co-morbidities and/or concurrent treatment. Reviews of randomised controlled trials have found average exclusion rates of 73% and recruitment of less than 10% of patients with the primary diagnosis.²⁶ These concerns, however, should be tempered by knowledge that true subgroup effects are very unusual.²⁷ The real issue of importance to N-of-1 trials is the likelihood, in many instances, of large variability in responses among patients.²⁸

N-of-1 trial services

The result of their first N-of-1 trial inspired the team at McMaster to develop a full N-of-1 referral service to address patient dilemmas that met criteria for our N-of-1 designs: therapeutic impact was uncertain, the treatment target was to reduce daily or otherwise frequent symptoms, the intervention (typically a drug) worked quickly, and it quickly ceased acting. Within two years, the group had completed 57 N-of-1 trials. Results had provided a definite therapeutic answer in 88% of the patients studied and these results prompted 39% of physicians to change their prior-to-trial treatment plan. This experience led the McMaster team to offer guides for clinicians wishing to apply the N-of-1 concept in their own practice.²⁹ Ultimately, however, the clinical communities interest in conducting N-of-1 trials diminished and the service was terminated.

Eric Larson was in the audience at a presentation of the McMaster work at the American Federation for Clinical Research.³⁰ Appreciating the utility of the design, Larson developed an N-of-1 clinical service at the University of Washington. Over two years, Larson's group completed 34 trials, again demonstrating that N-of-1 trials could provide physicians with useful treatment guidance in uncertain cases and improve patient satisfaction.³¹ Unfortunately, funding for the service ran dry and it was discontinued.

In 1999, the University of Queensland in Australia created the first national N-of-1 research service, referred to as a 'single patient trial service'.³² The service was designed to acquaint general practitioners with research methodology and to introduce research-derived data into clinical decision-making for conditions where treatment effectiveness was uncertain. Physicians could refer their patients to the service, which was centrally located, and so used mail and telephone communication only. The service managed all

major components of trial management: randomisation, preparing tablets, sending all materials to patients, following up, and relaying results to clinicians. Of the N-of-1 trials carried out by this service and which had available data, post-trial management decisions were consistent with trial results at 12 months in approximately 70% of attention deficit hyperactivity disorder trials,³³ 45% of osteoarthritis trials,³⁴ and 32% of neuropathic pain trials.³⁵ This is a successful example of how N-of-1 trials can be implemented at a national level, though, again, only as a temporary research initiative.

Another example of the versatility of N-of-1 trials began when the Complementary and Alternative Research and Education (CARE) programme at the University of Alberta established the first academic paediatric integrative medicine programme in Canada. In 2006, as part of this programme, a paediatric N-of-1 service responded to the increased use of complementary therapies in children with chronic conditions. The goal of this service is to offer an objective, evidence-based approach to assessing whether a given complementary therapy is effective for a specific patient. The service is designed to assist patients, their parents and referring physicians throughout all stages of the N-of-1 trial, including the design and implementation of the N-of-1 evaluation. For example, this service has assessed natural health products (e.g. melatonin, probiotics, micronutrients) and acupuncture for conditions including attention deficit hyperactivity disorder, eczema, sleep disturbances, chemo-induced nausea and vomiting, irritable bowel syndrome and autism.

N-of-1 in drug development

The McMaster group speculated that drug development might also benefit from use of the N-of-1 methodology. The reasoning was that pre-approval drug development costs are high (average \$479–936 million USD^{36,37} and rising³⁸). Conducting N-of-1 trials before a costly large-scale randomised controlled trial could (a) help to assess early efficacy, (b) be less expensive than traditional approaches, and (c) identify predictors of response.³⁹

The idea of applying the N-of-1 approach to early drug development arose from experience with multiple N-of-1 trials in specific conditions. For instance, when what is now termed myofascial pain syndrome was labelled fibrositis and there had been one apparently positive randomised controlled trial of amitriptyline, the condition provided a framework for N-of-1 trials in early drug development. The McMaster team conducted 14 N-of-1 trials which demonstrated substantial benefit from amitriptyline at doses far lower than had been used for the primary indication

for the drug, depression.³⁹ The McMaster team also demonstrated the utility of multiple N-of-1 trials in Alzheimer's disease⁴⁰ and in the use of home oxygen in patients with chronic obstructive pulmonary disease.⁴¹ In each of these situations the process appeared to be efficient, requiring limited cost and time investment. Nevertheless, subsequent attempts to apply the reasoning in drug development have been sporadic and unsuccessful.

Failure to revolutionise clinical practice: were N-of-1 trials ahead of their time?

Early experience was disappointing, shattering the initial optimism that N-of-1 trials would quickly revolutionise clinical practice. There had been some tantalising results,⁴² but randomised controlled trials in which patients were randomised to conventional care or to N-of-1 trials generally failed to show dramatically convincing benefits of participation in the N-of-1 trials.^{43,44}

At McMaster University, despite educating local clinicians, playing cheerleader, succeeding in conducting 73 N-of-1 trials over three years,⁴⁵ and inspiring other 'N of 1 services', interest still faded. An attempt to use venture capital to create an efficient, marketable service went nowhere. Thirty years after our initial publication, few clinicians have even heard of N-of-1 trials.

Sporadic reports of success with N-of-1 continue. For instance, Joy *et al.*⁴⁶ reported findings consistent with 'the nocebo phenomenon' – patients sometimes report side effects to placebo:⁴⁷ in seven patients with suspected but uncertain statin-associated myalgia, N-of-1 trials failed to detect any statin-related symptoms in any of the patients, allowing patients to continue the drugs. Despite such isolated reports of successes, clinicians seldom use N-of-1 trials and most remain unaware of the design.

Renewed interest in N-of-1 trials

At the University of Alberta, recent efforts have focused on methodological issues related to N-of-1 trial design and reporting. For example, N-of-1 trials have been criticised for their lack of generalisability. The Alberta group recently partnered with the *Journal of Clinical Epidemiology* to publish a series dedicated to N-of-1 trials and included papers to address this concern. A comprehensive systematic review of the design, analysis and meta-analysis of N-of-1 trials found that the majority (60%) of published N-of-1 trials are published as a series (i.e. one report publishing N-of-1 trial data about more than one participant for the same condition-intervention

pair), suggesting their value beyond assessing individual treatment effects and their potential to provide more generalisable treatment effects. Indeed, the Oxford Centre for Evidence-Based Medicine⁴⁸ has classified N-of-1 trials as Level 1 evidence, comparable to systematic reviews of randomised controlled trials.

By virtue of their methods (i.e. use of randomisation, blinding, formal outcome assessment), the meta-analysis of N-of-1 trials may provide a valuable source of population data for conditions that have little to no randomised controlled evidence, and to help refine evidence when parallel group randomised controlled trials may exist.

Given the large number of published N-of-1 trials in attention deficit hyperactivity disorder, the condition may serve as a clinical model to explore the applicability of N-of-1 trials beyond the individual patient.⁴⁹ Investigators at the University of Alberta conducted a systematic review and meta-analysis of N-of-1 trials and demonstrated the use of traditional randomised controlled trial meta-analysis methods in N-of-1 trials.⁴⁹ In another study, Punja *et al.*⁵⁰ demonstrated the value of N-of-1 trials in meta-analyses by conducting a combined meta-analysis of N-of-1 trial data with randomised controlled trial data. The inclusion of N-of-1 data in randomised controlled trial meta-analyses improved the precision of population treatment effects, suggesting their potential to provide a rich source of data allowing for more powerful and reliable assessments of treatment effects.⁵⁰ This example also highlights the relevance of N-of-1 trials in conditions for which there is also traditional randomised controlled evidence.

Range of conditions assessed in N-of-1 literature	n
Diseases of the nervous system	27
Diseases of the musculoskeletal system and connective tissue	20
Mental and behavioural disorders	17
Diseases of the digestive system	11
Diseases of the respiratory system	09
Diseases of the circulatory system	04
Endocrine, nutritional, metabolic diseases	02
Infections and parasitic diseases	02
Other (non-specific)	08

N=100; number of published N-of-1 studies that have assessed treatments for the respective condition category (adapted from Punja *et al.*⁵¹).

Challenges and future directions

Methodological considerations for N-of-1 trials differ from those for standard, parallel group randomised controlled trials. When considering N-of-1 trials as a research endeavour, investigators have proposed solutions to three major limitations among reported N-of-1 trials: incomplete reporting, marked variability in quality, and unacceptably high rates of prospective protocol registration.

First, as is the case with parallel group randomised controlled trials, lack of complete and transparent reporting is a problem in the N-of-1 trial literature. The Alberta group⁵¹ found that authors of N-of-1 trials failed to report on a number of critical design and conduct elements: trial registration (97%), whether individuals with co-morbid conditions (77%) or on concurrent therapies (69%) were included, and whether adverse events were assessed (64%). Another review confirmed that the quality of reporting of published N-of-1 trials was highly variable.⁵² The Alberta group led the development of the CONSORT Extension for N of 1 Trials (CENT) in response to the limitations and heterogeneity in reporting,^{53,54} serving as a minimum checklist for reporting N-of-1 trials.

Second, careful development and reporting of N-of-1 protocols is necessary for researchers, ethics review boards and funders. The Alberta group is currently developing a SPIRIT Extension for N of 1 Trials (SPENT). This will recommend essential elements in N-of-1 trial protocols, in the expectation that this will help to improve the quality of published reports of N-of-1 trials and promote the inclusion of N-of-1 trial protocols in trial registries.

Third, only 3% of published N-of-1 trials are reported as having registered protocols prospectively. It is certain that not all N-of-1 trials are published and readily available (nor, for those conducted as part of optimal routine clinical practice, should they be) – unpublished trials begun as part of the research endeavour may create a risk of bias for future systematic reviews and meta-analyses. One way of capturing these trials would be to establish an electronic repository (as is done for conventional randomised controlled trials with clinicaltrials.gov) and encourage authors to register their N-of-1 trial protocols. This would help reviewers to identify selective outcome reporting and publication biases.

Beyond these challenges, emerging methodologies may facilitate optimal use of N-of-1 principles. Bayesian and adaptive designs have potential applicability to N-of-1 trials. Trials can be designed with preset points based on adverse effects or patient preferences to crossover, change dose or discontinuation. These methods can be used both to analyse and to

meta-analyse N-of-1 trials.^{55,56} The strength of Bayesian approaches lies in their ability to maximise the use of reliable available information from each participant, as well as the use of reliable prior information for incorporation in the statistical model so that each N-of-1 trial can inform the next. Zucker *et al.*⁵⁶ have demonstrated the use of Bayesian methods to aggregate N-of-1 trials to yield estimates of population treatment effects. Combining Bayesian approaches with adaptive designs may prove to be a useful combination for future N-of-1 trials.

Discussion

What explains the failure to adopt and sustain N-of-1 trials? The obstacles to conducting N-of-1 trials as an element of routine clinical practice have been too great. For many pharmacists, preparing identical drug and placebo combinations proved too labour-intensive. For clinicians, N-of-1 trials take too much time, even with easy-to-use guidance:²⁹ preparing questionnaires, instructing patients and examining the results all require clinician commitment. By comparison, the simple question, ‘did the treatment help’ is too easy, and has too much face validity, compared to the more onerous substitution of a formal N-of-1 trial. The late Professor Charles Bridge-Webb proposed a workaround to the expensive, time-consuming process of arranging placebo.⁵⁷ He suggested a simplified N-of-1, The Single Patient Open Trials (SPOTs), substituting the blinded trial for an open one. This trial trades pragmatism for rigour, particularly useful for independent practitioners without access to N-of-1 services.

The advent of technological advances may help to overcome the operational complexity and costs that have hindered the uptake of the N-of-1 methodology. The emergence of mobile electronic health devices makes it easier than ever for patients to engage in their own healthcare. The creation of an IT-based N-of-1 trial platform would help clinicians and patients to collaborate in designing their own N-of-1 trials, track health outcomes and produce a report of results for patients and clinicians to discuss. Researchers from the University of California, Davis, have developed a mobile application called the ‘Triallist’ specifically to facilitate the conduct of N-of-1 trials in clinical settings. They are testing the feasibility and efficacy of this application in a randomised controlled trial comparing the effects on patient outcomes of participating in a mobile N-of-1 trial versus usual care.⁵⁸

This potential for N-of-1 trials as a way of providing clinical care differs from its use as a research endeavour. The distinction comes down to the intent behind conducting an N-of-1 trial. If the objective is to

inform treatment decisions for an individual patient, the trial is optimal clinical care and should therefore not require formal ethics approval⁵⁹ nor regulatory oversight from agencies monitoring clinical research. When choices from among two or more alternative treatments are being considered, patients should be informed about genuine uncertainties about their relative merits and how treatment should be selected in these circumstances.⁶⁰ Random allocation within formal treatment comparisons is one of the options that should be offered to patients.

If the primary purpose of N-of-1 trials is to produce generalisable knowledge to inform treatment decisions for future patients, these N-of-1 trials are more properly regarded as research. In these circumstances, compliance with methodological and ethics standards will be expected. In 2014, the Agency for Healthcare Research and Quality commissioned a user's guide to N-of-1 trials, which clarifies this distinction.²⁴

N-of-1 trials may have a future, both as a research endeavour complementing standard trials and as a strategy for improving clinical care outside of the research setting. Unlike conventional parallel group randomised controlled trials, which assess what is best on average for a given population, N-of-1 trials assess what is best for an individual patient.⁶¹ They are thus particularly well suited to emerging interests in patient-centred research and 'precision' or 'personalised' medicine. N-of-1 trials support the evolution of patient-centred research by offering an evidence-based approach for personalising care. They help to answer, for example, which treatment options are most effective through a process that strengthens the clinician-patient relationship and ultimately empowers the patient to be more engaged with their healthcare. Furthermore, with the advent of 'big data', and its hoped-for potential to inform care, N-of-1 trials can provide opportunities to learn how to improve care. The potential exists. The extent to which it will be realised remains uncertain.

Declarations

Competing Interests: None declared

Funding: None declared

Ethics approval: Not required

Guarantor: SV

Contributorship: RM wrote the initial draft. The remaining authors (SP, SV, GG) provided major contributions, edits, and final approval.

Acknowledgements: The authors are grateful to Liam Smeeth, Paul Glasziou and Chris Del Mar for comments on earlier drafts.

Provenance: Invited article from the James Lind Library

References

1. Wiseman R. *Eight Chirurgical Treatises*. London: R Royston, 1676.
2. Rolls R. Caleb Hillier Parry (1755–1822). JLL Bulletin: Commentaries on the history of treatment evaluation. See <http://www.jameslindlibrary.org/articles/caleb-hillier-parry-1755-1822/> (2003, last accessed 10 November 2016).
3. Parry C. Experiments relative to the medical effects of Turkey Rhubarb, and of the English Rhubarbs, No I and No. II made on patients of the Pauper Charity. *Lett Pap Bath Soc* 1786; 11: 431–453.
4. Booth CC. John Haygarth FRS (1740–1827). JLL Bulletin: Commentaries on the history of treatment evaluation, 2002. See <http://www.jameslindlibrary.org/articles/john-haygarth-frs-1740-1827/>.
5. Haygarth J. *Of the Imagination, as a Cause and as a Cure of Disorders of the Body: Exemplified by Fictitious Tractors, and Epidemical Convulsions*. Bath: R. Crutwell, 1800.
6. Martini P. *Methodenlehre der Therapeutischen Untersuchung [Methodological Principles for Therapeutic Investigations]*. Berlin: Springer, 1932.
7. Stoll S. Paul Martini's Methodology of therapeutic investigation. JLL Bulletin: Commentaries on the history of treatment evaluation, 2004. See <http://www.jameslindlibrary.org/articles/paul-martini-methodology-of-therapeutic-investigation/>.
8. Hogben L and Sim M. The self-controlled and self-recorded clinical trial for low-grade morbidity. *Br J Prev Soc Med* 1953; 7: 163–179.
9. Hamburg MA and Collins FS. The path to personalized medicine. *N Engl J Med* 2010; 363: 301–304.
10. Ebrahim S. Hogben on speed, paradoxes and strain. *Int J Epidemiol* 2011; 40: 1429–1430.
11. Tabery J. Hogben vs the tyranny of averages. *Int J Epidemiol* 2011; 40: 1454–1458.
12. Glasziou P. The history and place of n-of-1 trials: a commentary on Hogben and Sim. *Int J Epidemiol* 2011; 40: 1458–1460.
13. Hare EH. Comparative efficacy of hypnotics: a self-controlled, self-recorded clinical trial in neurotic patients. *Br J Prevent Soc Med* 1955; 9: 140–146.
14. Kazdin EA. *Single-Case Research Designs: Methods for Clinical and Applied Settings*, 2nd edn. New York: Oxford University Press, 2011.
15. Baskerville JC, Toogood JH, Mazza J and Jennings B. Clinical trials designed to evaluate therapeutic preferences. *Stat Med* 1984; 3: 45–55.
16. Lindsey JK and Jones B. A model for cross-over trials evaluating therapeutic preferences. *Stat Med* 1996; 15: 443–447.
17. Guyatt G, Sackett D, Taylor DW, Ghong J, Roberts R and Pugsley S. Determining optimal therapy –

- randomized trials in individual patients. *N Engl J Med* 1986; 314: 889–892.
18. Tate RL, McDonald S, Perdices M, Togher L, Schultz R and Savage S. Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the single-case experimental design (SCED) scale. *Neuropsychol Rehabil* 2008; 18: 385–401.
 19. Perdices M and Tate R. Single-subject designs as a tool for evidence-based clinical practice: are they unrecognized and undervalued? *Neuropsychol Rehabil* 2009; 20: 939.
 20. Gagnier J, Kienle G, Altman D and Moher D. The CARE guidelines: consensus-based clinical case reporting guideline development. *BMJ Case Rep*, 2013. See <http://casereports.bmj.com/content/2013/bcr-2013-201554.full> (last checked 24 November 2016).
 21. Shamseer L, Sampson M, Bukutu C, et al. CONSORT extension for reporting N of 1 trials (CENT) 2015: explanation and elaboration. *BMJ* 2015; 350: 18–46.
 22. Scuffham PA, Nikles J, Mitchell GK, Yelland MJ, Vine N, Poulos CJ, et al. Using N-of-1 trials to improve patient management and save costs. *J Gen Intern Med* 2010; 25: 906–913.
 23. Lillie EO, Patay B, Diamant J, Issell B, Topol EJ and Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Per Med* 2011; 8: 161–173.
 24. Kravitz RL, Duan N (eds) and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, et al.). *Design and Implementation of N-of-1 Trials: A User's Guide*. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality. See <http://www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm> (2014, last accessed 7 December 2016).
 25. Schork NJ. Time for one-person trials. *Nature* 2015; 520: 609–611.
 26. Rothwell PM. External validity of randomized controlled trials: “to whom do the results of this trial apply?” *Lancet* 2005; 365: 82–93.
 27. Sun X, Briel M, Walter S and Guyatt G. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010; 340: c117.
 28. Larson EB. N-of-1 clinical trials: a technique for improving medical therapeutics. *West J Med* 1990; 152: 52–56.
 29. Guyatt G, Sackett D, Adachi J, Roberts R and Chong J. A clinician's guide for conducting randomized trials in individual patients. *CMAJ* 1988; 139: 497–503.
 30. Kravitz RL, Duan N, Niedzinski EJ, Hay MC, Subramanian SK and Weisner TS. What ever happened to N-of-1 trials? Insiders' perspectives and a look to the future. *Milbank Q* 2008; 86: 533–555.
 31. Larson E, Ellsworth A and Oas J. Randomized clinical trials in single patients during a 2-year period. *JAMA* 1993; 270: 2708–2712.
 32. Nikles C, Glasziou PP, Del Mar CB, Duggan CM and Mitchell G. N of 1 trials. Practical tools for medication management. *Aust Fam Physician* 2000; 29: 1108–1112.
 33. Nikles C, Mitchell G, Del MC, Clavarino A and McNairn N. An n-of-1 trial service in clinical practice: testing the effectiveness of stimulants for attention-deficit/hyperactivity disorder. *Pediatrics* 2006; 117: 2040–2046.
 34. Yelland M, Nikles C, McNairn N, Mar C, Schluter P and Brown R. Celecoxib compared with sustained-release paracetamol for osteoarthritis: a series of n-of-1 trials. *Rheumatology* 2007; 46: 135–140.
 35. Yelland MJ, Poulos CJ, Pillans PI, Bashford GM, Nikles CJ, Sturtevant JM, et al. N-of-1 randomized trials to assess the efficacy of gabapentin for chronic neuropathic pain. *Pain Med* 2009; 10: 754–761.
 36. DiMasi J, Hansen R and Grabowski H. The price of innovation: new estimates of drug development costs. *J Health Econ* 2003; 22: 151–185.
 37. Adams C and Brantner VV. Estimating the costs of new drug development: is it really \$802m? *Health Aff* 2006; 25: 420–428.
 38. Dickson M and Gagnon J. Key factors in the rising cost of new drug discovery and development. *Nat Rev Drug Discov* 2004; 3: 417–429.
 39. Guyatt GGH, Heyting A, Jaeschke R, Keller J, Adachi JD and Roberts RS. N of 1 randomized trials for investigating new drugs. *Control Clin Trials* 1990; 11: 88–100.
 40. Molloy DW, Guyatt GH, Wilson DB, Duke R, Rees L and Singer J. Effect of tetrahydroaminoacridine on cognition, function and behaviour in Alzheimer's disease. *CMAJ* 1991; 144: 29–34.
 41. Nonoyama ML, Brooks D, Guyatt GH and Goldstein RS. Effect of oxygen on health quality of life in patients with chronic obstructive pulmonary disease with transient exertional hypoxemia. *Am J Respir Crit Care Med* 2007; 176: 343–349.
 42. Mahon J, Laupacis A, Donner A and Wood T. Randomised study of n of 1 trials versus standard practice. *BMJ* 1996; 312: 1069–1074.
 43. Pope J, Prashker M and Anderson J. The efficacy and cost effectiveness of N of 1 studies with diclofenac compared to standard treatment with nonsteroidal anti-inflammatory drugs in osteoarthritis. *J Rheumatol* 2004; 31: 140–149.
 44. Mahon JL, Laupacis A, Hodder RV, McKim DA, Paterson NA, Wood TE, et al. Theophylline for irreversible chronic airflow limitation: a randomized study comparing n of 1 trials to standard practice. *Chest* 1999; 115: 38–48.
 45. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD and Newhouse MT. The n-of-1 randomized controlled trial: clinical usefulness. Our three-year experience. *Ann Intern Med* 1990; 112: 293–299.
 46. Joy T, Monjed A, Zou G, Hegele R and McDonald C. N-of-1 (single-patient) trials for statin-related myalgia. *Ann Intern Med* 2014; 161: 531–532.
 47. Barsky AJ, Saintfort R, Rogers MP and Borus JF. Nonspecific medication side effects and the nocebo phenomenon. *JAMA* 2002; 287: 622–627.
 48. OCEBM. The Oxford Levels of Evidence, 2011. See <http://www.cebm.net/ocbm-levels-of-evidence/>.
 49. Punja S, Xu D, Schmid CH, Hartling L, Urichuk L, Nikles CJ, et al. N-of-1 trials can be aggregated to

- generate group mean treatment effects: a systematic review and meta-analysis. *J Clin Epidemiol* 2016; 76: 65–75.
50. Punja S, Schmid CH, Hartling L, Urichuk L, Nikles CJ and Vohra S. To meta-analyze or not to meta-analyze? A combined meta-analysis of N-of-1 trial data with RCT data on amphetamines and methylphenidate for pediatric ADHD. *J Clin Epidemiol* 2016; 76: 76–81.
 51. Punja S, Bukutu C, Shamseer L, Sampson M, Hartling L, Urichuk L, et al. N-of-1 trials are a tapestry of heterogeneity. *J Clin Epidemiol* 2016; 76: 47–56.
 52. Li J, Gao W, Punja S, Ma B, Vohra S, Duan N, et al. Reporting quality of N-of-1 trials published between 1985 and 2013: a systematic review. *J Clin Epidemiol* 2016; 76: 57–64.
 53. Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *BMJ* 2015; 350: h1738.
 54. Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *J Clin Epidemiol* 2016; 76: 9–17.
 55. Zucker DR, Ruthazer R, Schmid CH, Feuer JM, Fischer PA, Kieval RI, et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. *J Rheumatol* 2006; 33: 2069–2077.
 56. Zucker DR, Ruthazer R and Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *J Clin Epidemiol* 2010; 63: 1312–1323.
 57. Smith J, Yelland M and Del Mar C. Single patient open trials (SPOTs). In: Nikles J and Mitchell G (eds) *The Essential Guide to N-of-1 trials in Health*. Dordrecht: Springer, 2015, pp.195–209.
 58. Barr C, Marois M, Sim I, Schmid CH, Wilsey B, Ward D, et al. The PREEMPT study – evaluating smart-phone-assisted n-of-1 trials in patients with chronic pain: study protocol for a randomized controlled trial. *Trials* 2015; 16: 67.
 59. Irwig L, Glasziou P and March L. Ethics of n-of-1 trials. *Lancet* 1995; 345: 469.
 60. Chalmers I and Lindley R. Double standards on informed consent to treatment. In: Doyal L and Tobias JS (eds) *Informed Consent in Medical Research*. London: BMJ Publications, 2001, pp.266–275.
 61. Zucker DR, Schmid CH, McIntosh MW, D’Agostino RB, Selker HP and 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: 401–410.
 62. Guyatt G, Haynes RB, Jaeschke RZ, Cook DJ, Naylor CD, Wilson MC, et al. Users’ guide to the medical literature: XXV. Evidence-based medicine: principles for applying the users’ guides to patient care. Evidence-Based Medicine Working Group. *JAMA* 2000; 284: 1290–1296.
 63. Guyatt G, Zhang Y, Jaeschke R and McGinn T. *Therapy and Validity: N of 1 Randomized Controlled Trials. Users’ Guides to the Medical Literature: A manual for Evidence-Based Clinical Practice*. Chicago, IL: American Medical Association, 2002.
 64. Gabler NB, Duan N, Vohra S and Kravitz RL. N-of-1 trials in the medical literature: a systematic review. *Med Care* 2011; 49: 761–768.
 65. Tate RL, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, et al. Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: The 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychol Rehabil* 2013; 23: 619–638.
 66. Tate RL, Perdices M, McDonald S, Togher L and Rosenkoetter U. The design, conduct and report of single-case research: Resources to improve the quality of the neurorehabilitation literature. *Neuropsychol Rehabil* 2014; 24: 315–331.
 67. Tate RL, Perdices M, Rosenkoetter U, Shadish R, Vohra S, Barlow DH, et al. The Single-Case Reporting Guideline In Behavioural Interventions (SCRIBE) 2016 Statement. *Aphasiology* 2016; 6: 862–876.
 68. OCEBM Levels of Evidence Working Group. *The Oxford Levels of Evidence 2*. Oxford Centre for Evidence-Based Medicine. See <http://www.cebm.net/index.aspx?o=5653> (2011, last accessed 7 December 2016).

Appendix 1

N-of-1 timeline

Author, year	Citation	Description/significance
Hogben and Sim, 1953	Hogben L, Sim M (1953). The self-controlled and self-recorded clinical trial for low-grade morbidity. <i>Br J Prev Soc Med</i> 7:163–179.	
Hare, 1955	Hare EH (1955). Comparative efficacy of hypnotics: a self-controlled, self-recorded clinical trial in neurotic patients. <i>British Journal of Preventive and Social Medicine</i> 9:140–146.	

(continued)

Continued.

Author, year	Citation	Description/significance
Guyatt, 1986	Guyatt G, Sackett D, Taylor DW, Ghong J, Roberts R, Pugsley S (1986). Determining Optimal Therapy – – Randomized Trials in Individual Patients. <i>N Engl J Med</i> 314:889-892.	N-of-I trials are first brought to the attention of a wide medical readership
Guyatt, 1988	Guyatt G, Sackett D, Adachi J, Roberts R, Chong J (1988). A clinician's guide for conducting randomized trials in individual patients. <i>CMAJ</i> 139: 497–503.	The practical approach presented encourages clinicians to conduct N-of-I trials
Guyatt, 1990	Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT (1990b). The n-of-I randomized controlled trial: clinical usefulness. Our three-year experience. <i>Ann Intern</i> 112:293–299.	Results of the first N-of-I trial clinical service
Guyatt, 1990	Guyatt GGH, Heyting A, Jaeschke R, Keller J, Adachi JD, Roberts RS (1990a). N of I randomized trials for investigating new drugs. <i>Control Clin Trials</i> 11:88–100.	A proposal that individual N-of-I randomized controlled trials could be used to elucidate drug effects at an early stage of development
Molloy, 1991	Molloy DW, Guyatt GH, Wilson DB, Duke R, Rees L, Singer J (1991). Effect of tetrahydroaminoacridine on cognition, function and behaviour in Alzheimer's disease. <i>CMAJ</i> 144:29–34.	Early use of multiple N-of-I trials in a particular condition to address drug effectiveness
Zucker, 1997	Zucker DR, Schmid CH, McIntosh MW, D'Agostino RB, Selker HP, Lau J (1997). Combining single patient (N-of-I) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. <i>J Clin Epidemiol</i> 50:401–410.	Using Bayesian technique to combine N-of-I trials
Guyatt et al., 2000	Guyatt G, Haynes RB, Jaeschke RZ, Cook DJ, Naylor CD, Wilson MC, Richardson WS (2000). Users; guide to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. Evidence-Based Medicine Working Group. <i>JAMA</i> 284:1290–96.	N-of-I trial described as top of methodological hierarchy for informing treatment decisions
Guyatt, 2002	Guyatt G, Zhang Y, Jaeschke R, McGinn T (2002). Therapy and Validity: N of I Randomized Controlled Trials. Users' Guides to the Medical Literature: A manual for Evidence-Based Clinical Practice. Chicago, IL: American Medical Association 275–290.	Users' guide to N-of-I trials

(continued)

Continued.

Author, year	Citation	Description/significance
Nikles, 2006	Nikles C, Mitchell G, Mar C Del, Clavarino A, McNairn N (2006). An N of 1 trial service in clinical practice: Testing the effectiveness of stimulants for attention-deficit/ hyperactivity disorder. <i>Pediatrics</i> 2006;117:2040–2046.	One of the largest series of N-of-1 trials carried out in children with ADHD - carried out by a National Service to support the conduct of N-of-1 trials (the first ever) in Australia
Zucker, 2006	Zucker DR, Ruthazer R, Schmid CH, Feuer JM, Fischer PA, Kieval RI, Mogavero N, Rapoport RJ, Selker HP, Stotsky SA, Winston E, Goldenberg DL (2006). Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. <i>J Rheumatol</i> 33:2069–2077.	Using Bayesian technique to combine N-of-1 trials
Tate, 2008	Tate RL, McDonald S, Perdices M, Togher L, Schultz R, Savage S (2008). Rating the methodological quality of single-subject designs and N of 1 trials: Introducing the single-case experimental design (SCED) scale. <i>Neuropsychol Rehabil</i> 18:385–401.	Single-Case Experimental Design (SCED) Scale
Zucker, 2010	Zucker DR, Ruthazer R, Schmid CH (2010). Individual (N of 1) trials can be combined to give population comparative treatment effect estimates: methodological considerations. <i>Journal of Clinical Epidemiology</i> 63:1312–1323.	Using Bayesian technique to combine N-of-1 trials
Gabler, 2011	Gabler NB, Duan N, Vohra S, Kravitz RL (2011). N-of-1 Trials in the Medical Literature: a systematic review. <i>Med Care</i> 49:761–768.	Systematic overview of the N-of-1 literature
OCEBM, 2011	OCEBM Levels of Evidence Working Group*. “The Oxford Levels of Evidence 2”. Oxford Centre for Evidence-Based Medicine. http://www.cebm.net/index.aspx?o=5653	N-of-1 trials are considered level 1 evidence by the Oxford Centre for Evidence-Based Medicine
Tate, 2013	Tate RL, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, McDonald S (2013). Revision of a method quality rating scale for single-case experimental designs and N of 1 trials: The 15-item Risk of Bias in N of 1 Trials (RoBiNT) Scale. <i>Neuropsychological Rehabilitation</i> 23:619–38.	15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale
DEClDE Methods Center N of 1 Guidance Panel, 2014	Kravitz RL, Duan N, eds, and the DEClDE Methods Center N of 1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S) (2014). Design and Implementation of N of 1 Trials: A User’s Guide. AHRQ	N-of-1 AHRQ Publication

(continued)

Continued.

Author, year	Citation	Description/significance
	Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014. www.effectivehealthcare.ahrq.gov/N-of-I-Trials.cfm .	
Tate, 2014	Tate RL, Perdices M, McDonald S, Togher L, Rosenkoetter U (2014). The design, conduct and report of single-case research: Resources to improve the quality of the neurorehabilitation literature. <i>Neuropsychological Rehabilitation</i> 2014;24:315–31.	Reporting guidelines: Risk of Bias in <i>N of I</i> Trials (RoBiNT) Scale; Single-Case Reporting guideline In BEhavioural interventions (SCRIBE)
Shamseer, 2015	Shamseer L, Sampson M, Bukutu C, et al. CONSORT extension for reporting <i>N of I</i> trials (CENT) 2015: Explanation and elaboration. <i>BMJ</i> 2015;350:18–46	CENT explanation and elaboration
Vohra, 2015	Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, Nikles J, Zucker DR, Kravitz R, Guyatt G, Altman DG, Moher D, CENT group (2015). CONSORT extension for reporting <i>N-of-I</i> trials (CENT) 2015 Statement. <i>BMJ</i> 350:h1738	CENT checklist
Punja, 2016	Punja S, Bukutu C, Shamseer L, Sampson M, Hartling L, Urichuk L, Vohra S (2016c). <i>N-of-I</i> trials are a tapestry of heterogeneity. <i>J Clin Epidemiol</i> 76:47–56	A systematic review of the methods of design, analysis, and meta-analysis used in published <i>N-of-I</i> trials
Punja, 2016	Punja S, Schmid CH, Hartling L, Urichuk L, Nikles CJ, Vohra S (2016). To meta-analyze or not to meta-analyze? A combined meta-analysis of <i>N of I</i> trial data with RCT data on amphetamines and methylphenidate for pediatric ADHD. <i>J Clin Epidemiol</i> 76:76–81.	Combined meta-analysis of <i>N-of-I</i> trials with randomised controlled trial data
Punja, 2016	Punja S, Xu D, Schmid CH, Hartling L, Urichuk L, Nikles CJ, Vohra S (2016a). <i>N of I</i> trials can be aggregated to generate group mean treatment effects: A systematic review and meta-analysis. <i>J Clin Epidemiol</i> 76:65–75.	Meta-analysis of <i>N-of-I</i> trials
Tate, 2016	Tate RL, Perdices M, Rosenkoetter U, Shadish W, Vohra S, Barlow DH, Horner R, Kazdin A, Kratochwill T, McDonald S, Sampson M, Shamseer L, Togher L, Albin R, Backman C, Douglas J, Evans JJ, Gast D, Manolov R, Mitchell G, Nickels L, Nikles J, Ownsworth T, Rose M, Schmid CH and Wilson B (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 Statement. <i>J Clin Epidemiol</i> 73:142–152.	Provides authors, readers and reviewers of single case design studies with a tool to maximise clarity, transparency and completeness in reports of such trials