



Published in final edited form as:

Soc Personal Psychol Compass. 2014 August ; 8(8): 408–421. doi:10.1111/spc3.12121.

Personalizing Behavioral Interventions Through Single-Patient (N-of-1) Trials

Karina W. Davidson, PhD¹, James Peacock, MD^{1,2}, Ian M. Kronish, MD, MPH¹, and Donald Edmondson, PhD, MPH¹

¹Center for Behavioral Cardiovascular Health, Department of Medicine, Columbia University Medical Center, New York

²Division of Cardiology, Columbia University Medical Center, New York

Abstract

Behavioral interventions are typically studied with the use of a conventional between-subject randomized controlled trial (RCT) design. In this design, the effect of an intervention on one group of patients is compared with the effect of a control condition on another group of patients, such that a between-subject change is tested. A between-subject design has an underlying assumption that there is a homogenous treatment effect for a behavioral intervention, drug or psychotherapy, and that the way the intervention operates in the study that will tend to operate in the same way in many other patients. We review some of the philosophical and practical problems with the use of this design when a clinician is attempting to decide on a course of behavioral treatment aimed at *within-subject* change in patients who are likely to have heterogeneous or unique responses to behavioral treatment. We also review the biases inherent in our current clinical practice model, which does not use any empirical data collection or design for testing if a treatment is useful, and also in the conventional between-subject personalized medicine RCT designs. We propose increased use of single-patient (also known as N-of-1) trials that employ within-subject designs, in cases where treatment response is heterogeneous—as is the case for most psychological and behavioral treatments. Limitations of such designs include that they can only be used when the treatment is potentially reversible, the patient can act as their own control, and the outcome can be measured repeatedly. Increased use of within-subject trials may address in many more instances the more clinically relevant question of how a specific patient will respond to a specific treatment, and could introduce a more harmonious scientific approach into the way we treat our patients. We have incorporated a case presentation that illustrates the complexities of applying evidence drawn from these different designs to selecting and evaluating treatments for the behavioral issues commonly faced by clinicians and patients.

Overview

There is a public health crisis in the United States because of a failure of health behavior change and lifestyle management (Mokdad, Marks, Stroup, & Gerberding, 2004; Spring et al., 2013). The methodological approaches to behavior change science that we have used

thus far have not succeeded in improving this situation (Norman, 2008). Evidence-based management of health behaviors presents a unique set of challenges. Between-subject randomized controlled trials (RCTs), considered the gold standard trial design for evaluating whether a treatment or intervention is useful, may not be well suited to account for the within-subject variations of multifactorial, dynamic processes such as health behaviors. Thus, perhaps by examining other study designs that could be employed when testing psychological and behavioral change interventions, we will be better equipped to understand the effect of health behavior change interventions on individual patients.

In this review, we present the current conventional between-subject RCT study designs used to evaluate behavior change interventions and, more importantly, one of the assumptions underlying those designs. We then present alternative study designs that are being proposed by the emerging field of personalized medicine and some of the associated assumptions that underlie them. Finally, and somewhat ironically, we present an old study design paradigm—the *single-patient (also known as N-of-1) within-subject trial*—that takes us back to designs that were first used by behaviorists and clinical psychology leaders. We have incorporated a case presentation that illustrates the complexities of applying evidence drawn from these different designs to the behavioral issues faced by clinicians and patients. After careful assessment of the strengths and weaknesses of each design, we argue that N-of-1 designs may in many circumstances be well-suited for identifying the interventions that will help us tackle our large, seemingly intractable public health behavioral problems such as obesity and sedentary lifestyle.

In this era of evidence-based psychology (Kazdin, 2008), behavioral medicine (Davidson et al., 2003; Spring et al., 2005), and medicine (Guyatt et al., 2002), conventional between-subject RCTs are considered the gold standard for obtaining relatively bias-free evidence (Luscher, 2013). However, between-subject RCTs are often not well suited to the realities of clinical practice. First, clinicians are faced with a unique patient who has unique comorbidities and complexities of care. Such complexity is rarely represented in patients selected for participation in our typical RCTs (Dzewaltowski, Estabrooks, Klesges, Bull, & Glasgow, 2004; Jilcott, Ammerman, Sommers, & Glasgow, 2007) because of the premium placed on determining the treatment's efficacy and desire to increase internal validity, leading to a lack of generalizability from between-subject RCTs. More importantly, conventional between-subject RCTs only estimate between-subject treatment response and thus are able to answer the question of how much within-subject treatment response will occur under the rare (or fantasy) circumstances where minimal differences among individuals exist. Thus, between-subject RCTs frequently are unable to answer the more clinically relevant question of how a *specific individual* will respond to that specific treatment or intervention (Figure 1). Indeed, individual participants in conventional between-subject RCTs often show no benefit from the statistically successful experimental intervention, or have adverse reactions to that intervention, or even benefit from the control intervention that was shown, on average, to be statistically inferior (Sedgwick, 2012). In general, it is difficult (if not impossible) to know whether an individual patient is more similar to the average participant in a conventional between-subject RCT or is more similar to one of the participants for whom the experimental treatment was not helpful—or worse.

Below, we provide a true clinical example that we return to throughout this review to illustrate our points about the challenges in identifying evidence-based, effective behavior change interventions for our patients when we only employ between-subject experimental designs.

Clinical Scenario

You are a clinical psychology intern on the inpatient pediatric floor. You are asked to take the lead for a recent inpatient admission. Richard is 7.5 years old, weighs 34 lb, and has been diagnosed as having attention-deficit/hyperactivity disorder (ADHD), nonorganic failure to thrive, and oppositional defiant disorder. He is taking 60 mg/d of methylphenidate (Ritalin). He has many extreme behavioral problems, including suicide attempts, homicidal ideations, and disruptive classroom behavior, which resulted in a recent school suspension. Richard is already taking the Food and Drug Administration (FDA)-approved maximum dosage for his age (not his weight), and any further dose increase would be an off-label use. The pediatrician knows that in some children methylphenidate can have the off-target effect of appetite suppression (which he wants to avoid). Nevertheless, the attending pediatrician has admitted Richard to the hospital because he wishes to monitor him closely while increasing the methylphenidate dose. As an intern, you wonder whether increasing the methylphenidate dose is the best course of action. You also wonder how to best evaluate the response to the increased dose. Although you understand that there is some urgency in controlling Richard's behavior immediately, before he or someone else is injured, you also wonder if there are other pharmacologic or behavioral treatments that you could recommend.

An Intern's Perspective on How to Proceed

What should an intern do? You have been taught to use evidence to inform your behavioral practice, so you first go to PubMed to review (between-subject) RCTs that have been published in peer-reviewed literature. However, when you review the articles from RCTs that are available to you, you discover that there are no data germane to determining the best course of treatment for this specific pediatric patient. No RCT has examined dosages outside the FDA-approved range or included a pediatric patient with so many diverse behavioral issues, including the nonorganic failure to thrive. You learn from your clinical supervisors that in the absence of a directly relevant research evidence base, clinicians base treatment decisions on intuition, which in turn is composed of prior experience, an understanding of the patient's (or in the case of minors – family's) expectations and preferences, and a careful consideration of alternative diagnoses and treatments (Spring et al., 2005). This intuitive approach is often iterative and tailored to the patient but is potentially biased and has neither a control condition nor a formal assessment of the effectiveness of each iteratively tested treatment (Graber et al., 2012; Nendaz & Perrier, 2012). Thus, conventional between-subject RCTs do not allow clinicians to test their intuition, and clinical practice as conventionally operationalized does not provide objectively obtained data from an a priori planned experiment on which to judge the usefulness of the proposed treatment for that specific patient. In this review, we discuss the various scientific approaches that could be used to deal with this and other behavioral issues in individual patients. We also review the logic

used to link the evidence derived from these approaches with an individualized approach to patient care.

Conventional RCTs: Definition and Assumptions

A conventional between-subject RCT is the design considered the gold standard for evaluating whether a treatment or intervention is beneficial relative to a control or comparator condition (Sackett, Sharon, & Richardson, 2000). It is a specific type of experiment in which participants, if found eligible for the trial, are randomly allocated to the intervention of interest or to a control intervention or condition(s). The most important advantage of proper randomization is that it minimizes allocation bias, balancing both known and unknown prognostic factors, in the assignment of interventions. Well-designed conventional RCTs ensure that all assessments and contacts are identical for those who receive either the intervention or the control condition, so that if any difference is found for the outcome, it can be attributed to the difference in intervention to which they were assigned. Well-designed conventional RCTs also mask the assessors, the participants, and the treatment providers, as much as possible, to the condition to which the participant was assigned. In this way, measurement bias or expectations will be minimized in their influence on the conventional or between-subject RCT results. An excellent introduction to the planning, execution, and appraisal of conventional RCT methods can be found in the article by Friedman, Furberg, and DeMets (Friedman, Furberg, & DeMets, 2010).

To provide context concerning why the evidence-based movement accords conventional RCT data such high credibility, consider several unexpected results that have arisen from major conventional RCTs in medicine and that stood in stark contrast to previously obtained, observational study results, as well as clinical intuition. Recent findings about hormone replacement therapy (HRT) offer one example. A substantial body of observational evidence from decades of survey research indicated that postmenopausal women who received HRT, compared with their counterparts who did not, had a markedly lower incidence of coronary heart disease. A seemingly logical inference was that the risk of coronary heart disease could be nearly halved by making it standard of care to routinely prescribe HRT to postmenopausal women (Stampfer & Colditz, 1991). That bit of clinical wisdom was so firmly accepted that it was widely considered a foolish waste of resources to conduct an RCT to test the cardioprotective effects of HRT. Nevertheless, the National Institutes of Health eventually launched the Women's Health Initiative (WHI) (Wassertheil-Smoller et al., 2003), an RCT that evaluated HRT under the better controlled conditions afforded by random assignment and outcome assessment masking. The surprising results indicated that, for the postmenopausal women enrolled in the trial, HRT was not a panacea to stave off the full roster of chronic diseases associated with advancing age. In fact, HRT actually significantly increased the risk of some life-threatening conditions (eg, stroke). With the wisdom afforded by hindsight, it became apparent that the early observational studies of HRT were subject to a self-selection sample bias. At least before the WHI, postmenopausal women who elected to take HRT differed from their peers in being better educated, more affluent, and generally healthier (Grimes & Lobo, 2002). Random assignment to alternative intervention arms in the WHI reduced markedly these sources of

bias (Kunz, Khan, & Neumayer, 2000), yielding RCT results that starkly contradicted the findings of observational studies.

These unexpected results from the WHI and many other interventions illustrate why conventional between-subject RCTs currently remain the gold standard approach for collecting the evidence needed to identify effective interventions, including effective behavioral interventions. Being able to demonstrate that health behavior interventions pass the test of rigorous, consensually accepted scientific methods is indispensable for making decisions that pertain to reimbursement and allows our trainees to have a body of unbiased evidence available to guide their practices. Having recourse to the rigorous methods afforded by conventional between-subject RCTs also protects health behavior science from offering interventions that are ineffective or even harmful. Although health behavior interventions might appear immune from producing adverse effects of the kind detected in the WHI, some presumably innocuous behavioral interventions, such as cognitive behavioral therapy for depression or social isolation, have demonstrated the potential for significant harms with respect to serious outcomes, including increased heart disease or death (Berkman et al., 2003; Frasure-Smith et al., 2002; Hall, Tuskell, Vila, & Duffy, 1992). If we continue to follow the conventional path that has been laid down by evidence-based medicine, psychology, and behavioral medicine, and we accept the underlying assumptions of this approach, logic dictates that we must continue to conduct conventional RCTs of those behavioral interventions found promising in observational studies. We turn now to the assumptions underlying the conventional RCT methodological approach.

In applying the results of conventional between-subject RCTs to practice, an underlying assumption is that between-subject treatment change will be roughly equivalent to within-subject treatment change. The between-subjects RCT study design falls under a theoretical approach to knowledge that Kant described as a tendency to generalize and explain an objective phenomenon that was later coined by Kantian philosopher Wilhelm Windelband as 'nomothetic'. The complementary approach to knowledge in Kantian philosophy is an 'idiographic' approach that has the tendency to specify and explain more unique and often subjective phenomena. Conventional RCTs provide between-subject differences, but we apply this knowledge to predict a specific patient's response to treatment over time, which means we are interested in a within-subject difference. Applying results from an RCT design that obtained and reported on between-subject differences has the assumption that the proposed treatment response is sufficiently generalizable that between-subject variation will be equitable to within-subject variation. This assumption may not be valid for phenomena as complex as human behaviors. For example, in an examination of how often heterogeneous treatment effects are found, Fernandez and others point out that although there may be methodological reasons for the differing treatment effect sizes, there can also be patient subgroups, who differ from the 'average patient group' for whom a different treatment response may be found. (Fernandez, Nguyen, Duan, Gabler, & Kravitz, 2009)

In our clinical scenario started above, the intern would look up methylphenidate and ADHD in PubMed and find the conventional double-blind RCT that reports that methylphenidate was superior to placebo in reducing ADHD symptoms in children 6 to 12 years old (Wolraich et al., 2001). Although the aggregated data from two groups of individuals show a

mean improvement in the methylphenidate group on a masked assessment of ADHD behavior, the reality is some individuals within the group will improve with methylphenidate, some will not improve, and some will actually be harmed by adverse effects of methylphenidate without any significant improvement in symptoms. Conventional between subject RCTs cannot always determine how a specific individual will respond to a treatment. Given that the patient, Richard, has not improved with the already prescribed methylphenidate, which is the treatment supported by best evidence, the intern has to think of another approach to this patient. Perhaps, Richard is different from the sample population in the published study. Could there be another study with a sample population that more closely resembles Richard?

Conventional Between-Subject Personalized Medicine: Definitions and Assumptions

The term *personalized medicine* was first coined in the context of advances in genetics that promised a tailored drug approach based on an understanding of genetic variability among individuals. One of the first significant discoveries from a personalized medicine approach was the development of imatinib (Gleevec), a tyrosine kinase inhibitor for chronic myelogenous leukemia (CML). A *subset* of patients develop CML because of a reciprocal translocation between chromosomes 9 and 22 (i.e., Philadelphia chromosome) that led to a chimeric, constitutively active Bcr-Abl tyrosine kinase that has been implicated in the pathogenesis of CML. Imatinib is a tyrosine kinase inhibitor that selectively inhibits this abnormal chimeric tyrosine kinase. The effectiveness of the drug was tested in an RCT that selected the subgroup of CML patients with the Philadelphia chromosome and ultimately showed a high response rate and low relapse rate that led *TIME* to deem this drug the “magic bullet” for cancer treatment (Lemonick M, 2001). Because this drug targeted a specific putative cause of CML, a conventional RCT on all patients with leukemia may not have demonstrated a statistically significant effect because most patients without the Philadelphia chromosome would not be expected to improve while taking the medication. Clearly, this personalized genetic approach to the development and testing of this drug led to an important discovery in CML therapy.

The term *personalized medicine* has since been extended beyond genetics and now encompasses many disciplinary studies. Although conventional between-subject RCTs systematically attempt to neutralize individual differences through randomization, a personalized medicine approach groups populations into subgroups to better characterize individual differences and then randomizes participants to treatment or control groups within the smaller subgroup. We implicitly use this subgrouping approach in behavior science and practice when we specify that women’s assertiveness behaviors differ from men’s or that the group who has depressive symptoms due to a sleep disorder should be treated differently than a group who has depressive symptoms due to a recent loss. In both cases, we are subgrouping first and then considering which behavioral treatment to test, rather than trying one treatment on everybody and assuming that if it does not work for most then it must not be useful.

In conventional between-subject personalized medicine approaches, personalization comes from creating smaller and smaller subgroups of people based on genetic, physiologic, or behavioral or suspected etiologic similarities. It assumes that if a population is divided and subdivided into smaller and smaller subgroups based on some important similarity then individuals will become interchangeable and we will be able to assume that between-subject change or response to treatment will be roughly equivalent to within-subject change or response to treatment. In cases in which an intervention is not effective in a sample of people, then at one level, conventional between-subject personalized medicine approaches hold out the possibility that the right subgroup has not yet been identified. However, conventional personalized medicine approaches do not help identify the best way forward for identifying the best subgroup of patients to apply their approach to.

In our clinical scenario, Richard did not do well with FDA-approved doses of methylphenidate. The intern may next take a conventional between-subject personalized medicine approach and search for studies that describe other treatments that may work for a subgroup of patients with behavioral problems similar to Richard's. Organic causes, including infections and nutritional deficiencies, have been ruled out as the cause of Richard's behavioral and weight problems. Richard does not have any family history of psychosis or behavioral problems that could indicate a genetic basis for his behavior. Despite the intern's best efforts, she is unable to find a distinguishable subgroup difference in Richard that would explain his lack of improvement (and possible harm with respect to his weight loss) with methylphenidate or that would identify an evidence-based alternative to methylphenidate for treatment of ADHD.

Conventional Clinical Practice: Individual Patient Decision-Making with Only Between-Subjects Evidence as a Guide

The situation our intern now faces—one in which a patient does not improve with treatment based on best evidence from conventional between-subject RCTs and in which there is no appropriate subgroup for alternative approaches identified by a subgrouping or personalized medicine approach—is a common one for clinicians. The intern, without an evidence-based option for Richard, would now take a best-guess approach in picking a treatment based on past experience and an understanding of Richard as an individual. The intern would formulate a differential diagnosis for the underlying cause of Richard's difficult behavior and might draw on different psychological, behavioral, biological, and other conceptualizations to propose treatment options for Richard. The possible treatment targets might include the following: lack of sufficient psychiatric medication (a biological or neuropsychological dysregulation), exposure to unstable mothering (an object-relations dysregulation), lack of structure (a behavioral reinforcement dysregulation), exposure to failure (unconditional positive regard dysregulation), and lack of sleep or lack of food (physiological dysregulations). Many of these putative causes are based on an understanding of Richard's unique circumstances. Without any further evidence, the intern could suggest increasing his methylphenidate dose beyond its FDA-approved dose. The intern would wait several weeks and then determine whether Richard's behavior and weight improved. If these did not improve, the intern would conclude that increasing the medication did not work and

attempt another treatment based on the remaining putative causes. The intern would continue this approach until Richard showed improvement. Although this iterative trial-and-error approach seems experimental, it is susceptible to multiple biases. For example, without an appropriate masked control, the clinician cannot determine whether any perceived improvement is due to the treatment, or to secular trends, or to characteristics that influence the retrospective self-report about the behaviors, or to other known biases that can influence this type of clinical treatment approach (Graber et al., 2012; Nendaz & Perrier, 2012). Conventional between-subject, double-blind RCTs are better at eliminating potential biases and showing causality but can be limited in their ability to predict to a single subject response. Conventional clinical practice is focused on the individual but is biased in its lack of a priori design, objective and masked outcome measurement, and randomization such that it cannot be considered a true scientific experiment in the age of evidenced-based medicine. Another methodologic approach is needed that is unbiased and yet focused on an individual patient.

Unconventional Single-Patient (N-of-1 or Within-Subject) RCTs: Definition and Assumptions

Ironically, Guyatt (one of the pioneers of evidence-based medicine) proposed a potential solution to the discrepancy between the results provided by conventional between-subject RCTs and the needs of a clinician to treat a specific patient more than three decades ago (Guyatt et al., 1988; Guyatt et al., 1986). Surprisingly, psychologists were performing this type of single, within-subject experimental approach on psychological and behavioral phenomena for many decades before that (Angell & Pierce, 1891–92; Ebbinghaus, 1913).

Importantly, an N-of-1 within-subject RCT does not evaluate what is best for the population but what is the optimum treatment for a single patient (Sedgwick, 2012). In this approach, a single patient is randomized to receive one or more reversible interventions with a control condition, in a random order, and with a suitable consideration of a washout period (i.e., a period in a clinical study during which subjects receive no treatment for the indication under study and the effects of a previous treatment are eliminated) to determine what is best for that patient. These are essentially RCTs that are multiple crossover trials conducted on single individuals (Guyatt et al., 1986). This approach entails making a priori decisions about when and how outcomes will be measured; measuring those outcomes objectively; randomizing the individual patient to two or more conditions; keeping the patient and the treatment provider masked, whenever possible, to which condition is offered; and detailing a statistical approach and a decision rule for using or rejecting the tested treatment before the trial is initiated. The primary advantage of this type of within-subject, single-patient design is that it offers both the clinician and the patient direct, objective evidence about the usefulness of a particular treatment for that patient.

There are many assumptions and requirements that must be present to consider such an N-of-1 approach (Tables 1 and 2). The symptoms or behaviors must occur across time and must have some variation. For most behaviors (e.g., dieting and exercising), these conditions are met, but extreme examples (e.g., successfully committing suicide) exist where these conditions are not met. Furthermore, the intervention to be tested in the N-of-1 design must

be able to be withdrawn and so must be reversible. If exposure to a psychodynamic therapy is not considered reversible when it is stopped, then it is not a treatment that can be tested in this type of design. One must also have some understanding of the washout period, or the required time between exposures to treatments, in order to set the time for each crossover period. Finally, in most cases, having some credible placebo treatment or alternative sham treatment that has similar expectancy effects is desirable, although not necessary. The final important assumption that each patient is sufficiently unique to warrant his/her own test of the usefulness of a specific treatment. Estimates of variance in the treatment effect of a large RCT with sufficient heterogeneity of sample characteristics are a good indicator of whether this assumption is likely to be met. If this assumption is not met (i.e., Figure 2), then designing and implementing N-of-1 trials would be an enormously inefficient process because we could instead conduct one large conventional between-subject RCT and then simply apply the results from that single trial to each patient with that problem. However, in settings when the above assumptions are met, then N-of-1 design approaches can be applied in many ways.

We provide details pertaining to two important N-of-1 trial approaches below. There are also N-of-1 trials that consider how to test irreversible interventions, such that a crossover design is not used, but a serial exposure design is used (and the trial is stopped when any of the treatments appears efficacious). There are also N-of-1 studies that simply wait for exposure to hypothesized causes and then determine whether these triggers or exposures seem to precede the symptom. These N-of-1 designs are observational and not a trial design but nevertheless are important for identifying, for the individual patient, possible causes on which to focus future behavioral treatment efforts.

N-of-1 Dosing RCT

An N-of-1 dosing RCT measures changes to the outcome within a single patient after escalating a dose of a single intervention across time. The N-of-1 dosing RCT is studying a single intervention of a putative cause, and this within-subject approach measures within-subject instead of between-subject differences. Like the clinical practice model, the N-of-1 dosing RCT is measuring within-subject differences based on clinical intuition as to the underlying cause, but unlike the clinical practice model, it can be both masked and randomized, and so designed to eliminate some of the biases of the clinical practice approach listed above.

In our clinical scenario, the intern initially was asked to increase the dose of methylphenidate because the pediatrician thought that this increase would improve Richard's ADHD symptoms without leading to further weight loss. If the intern chose to use an N-of-1 randomized dosing RCT, instead of simply increasing Richard's methylphenidate dose, she would work with a pharmacist to instead randomize Richard to his current methylphenidate dose, the proposed doubled dose of methylphenidate, or a placebo in a crossover design. To avoid bias in outcome assessments, only the pharmacist would know which dose (or placebo) Richard was receiving. The observation period for each crossover period would be set up to ensure that a washout period was long enough between doses. Teachers could be asked to report on attentional behaviors at the end of each period, and a

school nurse could weigh Richard daily. In this way, any positive or negative response to the increased dose of methylphenidate would be free of the placebo effect and observer bias.

This N-of-1 dosing RCT was conducted with Richard, and after 90 days it was determined that his increased dose of methylphenidate was associated with a 6% further weight loss and with no change in the teachers' ratings of his attentional behavior in the classroom. Interestingly, there was also no difference between the placebo drug and his original methylphenidate dose for either his weight or attentional behavior, suggesting that methylphenidate was not helpful to this patient and that a further increase in dose would not ameliorate his behavioral issues. This example demonstrates how the N-of-1 dosing RCT approach is best suited to test the dosing efficacy of a single intervention for a single putative cause.

What if the clinical question posed had been whether a more stringent behavioral reinforcement and punishment of attentional problems was the better behavioral treatment for Richard's attentional issues? An N-of-1 dosing RCT could be designed in which a teaching assistant used a token economy approach (in which Richard won or lost tokens, which could be exchanged later for toys) as the intervention. The token economy system could be designed with varying reinforcement schedules set up, and the teaching assistant could be told at the beginning of each day which schedule to follow. The mother could be kept masked to exactly which token economy system was used and could then report on daily attentional behavior. Once an N-of-1 dosing RCT design is introduced to clinicians (and to their interns), it becomes an enticing design to consider as a way to objectively collect data on the best level of treatment for a single patient based on the clinical intuition about what treatment might serve this patient best.

N-of-1 Comparative RCTs

An N-of-1 comparative RCT assesses within a single patient many different interventions. All of the assumptions and conditions listed above apply to each putative treatment; that is, each treatment has to be reversible and must have a reasonable washout period. Compared with the N-of-1 dosing study, this type of design allows many putative treatments to be tested sequentially, and a longer time frame is generally required to obtain sufficient data for each possible treatment. A more complicated version of this N-of-1 comparative RCT would posit that combinations of treatments might be optimal for the patient, and so the trial would be set up to test single treatments and combinations of treatments to determine the optimal treatment approach for that particular patient.

In our clinical scenario, the intern would not assume that one treatment would be the only one to test for Richard, but instead she would choose a series of potential interventions and test them in a crossover design with an appropriate washout period between each treatment or combination of treatments. In this approach, treatments are still tested in the context of carefully masked, objective, and repeatedly assessed outcomes.

Methodological, Analytic, and Ethical Barriers to Conducting N-of-1 Randomized Controlled Trials

There are a number of methodological, statistical, and ethical considerations associated with N-of-1 trials. The Agency for Healthcare Research and Quality (AHRQ)(Duan N et al., 2014) recently published a guide to N-of-1 trials that gives an excellent review of these considerations, and we include a brief list in Tables 1 and 2.

When is an N-of-1 Approach Appropriate?

N-of-1 trials are not a panacea, and should not replace traditional between-subjects RCTs for many (or most) clinical questions. Some of the contraindications and barriers inherent to the N-of-1 design are given in Tables 1 and 2.

Conclusions

Conventional between-subject RCTs and, more recently, conventional between-subject personalized medicine approaches have allowed enormous improvements in the health and well-being of the public. Nevertheless, primarily because of limitations in the generalizability of conventional approaches to individual patients, particularly in the area of behavior change, there is still a need for different approaches to deriving evidence to inform individual treatment decisions. Given the heterogeneity or patient-unique issues underlying many behavioral problems, N-of-1 approaches may provide a way forward. Ironically, the N-of-1 within-subject RCT approach brings back into psychology and behavioral medicine the focus on the individual person that existed before the convention of pooling across many persons (Molenaar, 2004). Each person is initially viewed as the unit of interest.

Although N-of-1 RCTs apply scientific rigor to an approach that most closely mimics clinical practice, remarkably little progress has been made in the development of the N-of-1 RCT method and its related statistical approaches during the last three decades (Barlow & Nock, 2009; Hamaker, 2012; Molenaar & Campbell, 2009; Ridenour, Pineo, Maldonado Molina, & Hassmiller Lich, 2013). In the health care industry, this is purported to be due to the research and development model of pharmaceutical companies, reimbursement practices, and governmental regulatory procedures (Aspinall & Hamermesh, 2007). The pharmaceutical industry has historically focused on developing and marketing blockbuster drugs that can reach a broad audience and is not incentivized to further refine the potential market base for each drug to exclude individuals who may not respond to it. Physicians are not offered the time and resources to use rigorous testing of their often trial-and-error approach to treatment. Furthermore, regulatory practices have put a strong emphasis on large conventional between-subject RCTs for drug approval but little on longitudinal safety and efficacy after FDA approval.

Recent changes in health care financing coupled with advances in health technology are creating new opportunities for implementing N-of-1 or within-subject RCTs. With the cost of health care rapidly outstripping corresponding improvements in public health, the health care landscape is undergoing a necessary paradigm shift toward a reimbursement model

based on performance instead of providing services. This shift will create a need to systematically monitor individual health performance. At the same time, technological advances in mobile health are making it possible to remotely, unobtrusively, and objectively measure many health behaviors, which should facilitate the implementation of many more N-of-1 RCT designs. To date, our ability to provide objective information about which therapy or intervention is best for any given patient by relying on clinical intuition or large conventional RCT results has remained modest at best (Woolf, 2008). A sound N-of-1 RCT methodology could change that.

References

- Angell J, Pierce A. Experimental research upon the phenomena of attention. *American Journal of Psychology*. 1891–92; 4:528–541.
- Aspinall MG, Hamermesh RG. Realizing the promise of personalized medicine. *Harv Bus Rev*. 2007; 85(10):108–117. 165. [PubMed: 17972499]
- Barlow D, Nock M. Why can't we be more idiographic in our research? *Perspectives on Psychol Sci*. 2009; 4:19–21.
- Berkman LF, Blumenthal J, Burg M, Carney RM, Catellier D, Cowan MJ, Schneiderman N. Effects of treating depression and low perceived social support on clinical events after myocardial infarction: The Enhancing Recovery in Coronary Heart Disease Patients (ENRICHD) Randomized Trial. *Journal of the American Medical Association*. 2003; 289(23):3106–3116. [PubMed: 12813116]
- Davidson KW, Goldstein M, Kaplan RM, Kaufmann PG, Knatterud GL, Orleans CT, Whitlock EP. Evidence-based behavioral medicine: what is it and how do we achieve it? *Annals of Behavioral Medicine*. 2003; 26(3):161–171. [PubMed: 14644692]
- Duan, N.; Eslick, I.; Kaplan, GN.; Kravitz, HC.; Larson, RL.; Pace, EB.; Vohra, WD. Design and Implementation of N-of-1 Trials: A User's Guide. Kravitz, RL.; DN, editors. Rockville, MD: Agency for Healthcare Research and Quality; 2014. p. 94
- Dzewaltowski DA, Estabrooks PA, Klesges LM, Bull S, Glasgow RE. Behavior change intervention research in community settings: how generalizable are the results? *Health Promotion International*. 2004; 19(2):235–245. [PubMed: 15128715]
- Ebbinghaus, M. *Memory*. New York: Columbia Teacher's College; 1913.
- Fernandez YGE, Nguyen H, Duan N, Gabler NB, Kravitz RL. Assessing Heterogeneity of Treatment Effects: Are Authors Misinterpreting Their Results? *Health Services Research*. 2009; 10.1111/j.1475-6773.2009.01064.x
- Frasure-Smith N, Lespérance F, Gravel G, Masson A, Juneau M, Bourassa M. Long-term survival differences among low-anxious, high-anxious and repressive copers enrolled in the Montreal heart attack readjustment trial. *Psychosomatic Medicine*. 2002; 64:571–579. [PubMed: 12140346]
- Friedman, L.; Furberg, C.; DeMets, D., editors. *Fundamentals of Clinical Trials*. 4. New York: Springer; 2010.
- Grabner ML, Kissam S, Payne VL, Meyer AN, Sorensen A, Lenfestey N, Singh H. Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf*. 2012; 21(7):535–557. 10.1136/bmjqs-2011-000149
- Grimes DA, Lobo RA. Perspectives on the Women's Health Initiative trial of hormone replacement therapy. *Obstet Gynecol*. 2002; 100(6):1344–1353. [PubMed: 12468183]
- Guyatt G, Haynes B, Jaeschke R, Cook D, Greenhalgh T, Mead M, Richardson W. Introduction: the philosophy of evidence-based medicine. *Users' guides to the medical literature*. 2002:3–11.
- Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, Keller J. A clinician's guide for conducting randomized trials in individual patients. *CMAJ*. 1988; 139:497–503. [PubMed: 3409138]
- Guyatt G, Sackett D, Taylor D, Chong J, Roberts R, Pugsley S. Determining optimal therapy - randomized trials in individual patients. *New England Journal of Medicine*. 1986; 314:889–892. [PubMed: 2936958]

- Hall SM, Tuskell CD, Vila KL, Duffy J. Weight gain prevention and smoking cessation: Cautionary findings. *American Journal of Public Health*. 1992; 82:799–803. [PubMed: 1585959]
- Hamaker E. Why researchers should think “within-person”. *Handbook of Research Methods for Studying Daily Life*. 2012:43–61.
- Jilcott S, Ammerman A, Sommers J, Glasgow RE. Applying the RE-AIM Framework to Assess the Public Health Impact of Policy Change. *Annals of Behavioral Medicine*. 2007; 34(2):105–114. [PubMed: 17927550]
- Kazdin AE. Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *Am Psychol*. 2008; 63(3):146–159. 2008-03389-001 [pii]. 10.1037/0003-066X.63.3.146 [PubMed: 18377105]
- Kunz R, Khan KS, Neumayer HH. Observational studies and randomized trials. *New England Journal of Medicine*. 2000; 343(16):1194–1195. [PubMed: 11041757]
- Lemonick, M.; PA. *Time Magazine Web*. 2001. New Hope For Cancer. (May 28, 2001 ed.)
- Luscher TF. The bumpy road to evidence: why many research findings are lost in translation. *European Heart Journal*. 2013.10.1093/eurheartj/eh396
- Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA*. 2004; 291(10):1238–1245.10.1001/jama.291.10.1238 [PubMed: 15010446]
- Molenaar P. A Manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever. *Measurement*. 2004; 24:201–218.
- Molenaar P, Campbell C. The new person-specific paradigm in psychology. *Cur Dir Psychol Sci*. 2009; 18:112–117.
- Nendaz M, Perrier A. Diagnostic errors and flaws in clinical reasoning: mechanisms and prevention in practice. *Swiss Medical Weekly*. 2012; 142:w13706.10.4414/smw.2012.13706 [PubMed: 23135902]
- Norman GJ. Answering the “What Works?” Question in Health Behavior Change. *Am J Prev Med*. 2008; 34(5):449–450. [PubMed: 18407014]
- Ridenour TA, Pineo TZ, Maldonado Molina MM, Hassmiller Lich K. Toward rigorous idiographic research in prevention science: comparison between three analytic strategies for testing preventive intervention in very small samples. *Prev Sci*. 2013; 14(3):267–278.10.1007/s11121-012-0311-4 [PubMed: 23299558]
- Sackett, DL.; Sharon, ES.; Richardson, WS. *Evidence-based medicine : how to practice and teach EBM*. 2. London, England: Churchill Livingstone; 2000.
- Sedgwick P. “n of 1” trials. *BMJ*. 2012; 34410.1136/bmj.e844
- Spring B, Ockene JK, Gidding SS, Mozaffarian D, Moore S, Rosal MC, Stroke N. Better Population Health Through Behavior Change in Adults: A Call to Action. *Circulation*. 2013.10.1161/01.cir.0000435173.25936.e1
- Spring B, Pagoto S, Kaufmann PG, Whitlock EP, Glasgow RE, Smith TW, Davidson KW. Invitation to a dialogue between researchers and clinicians about evidence-based behavioral medicine. *Annals of Behavioral Medicine*. 2005; 30(2):125–137. [PubMed: 16173909]
- Stampfer MJ, Colditz GA. Estrogen replacement therapy and coronary heart disease: A quantitative assessment of the epidemiologic evidence. *Preventive Medicine*. 1991; 20(1):47–63. [PubMed: 1826173]
- Wassertheil-Smoller S, Hendrix SL, Limacher M, Heiss G, Kooperberg C, Baird A, Mysiw WJ. Effect of estrogen plus progestin on stroke in postmenopausal women: the Women’s Health Initiative: A randomized trial. *Journal of the American Medical Association*. 2003; 289(20):2673–2684. [PubMed: 12771114]
- Wolraich ML, Greenhill LL, Pelham W, Swanson J, Wilens T, Palumbo D, August G. Randomized, controlled trial of oros methylphenidate once a day in children with attention-deficit/hyperactivity disorder. *Pediatrics*. 2001; 108(4):883–892. [PubMed: 11581440]
- Woolf SH. The Meaning of Translational Research and Why It Matters. *JAMA*. 2008; 299(2):211–213.10.1001/jama.2007.26 [PubMed: 18182604]

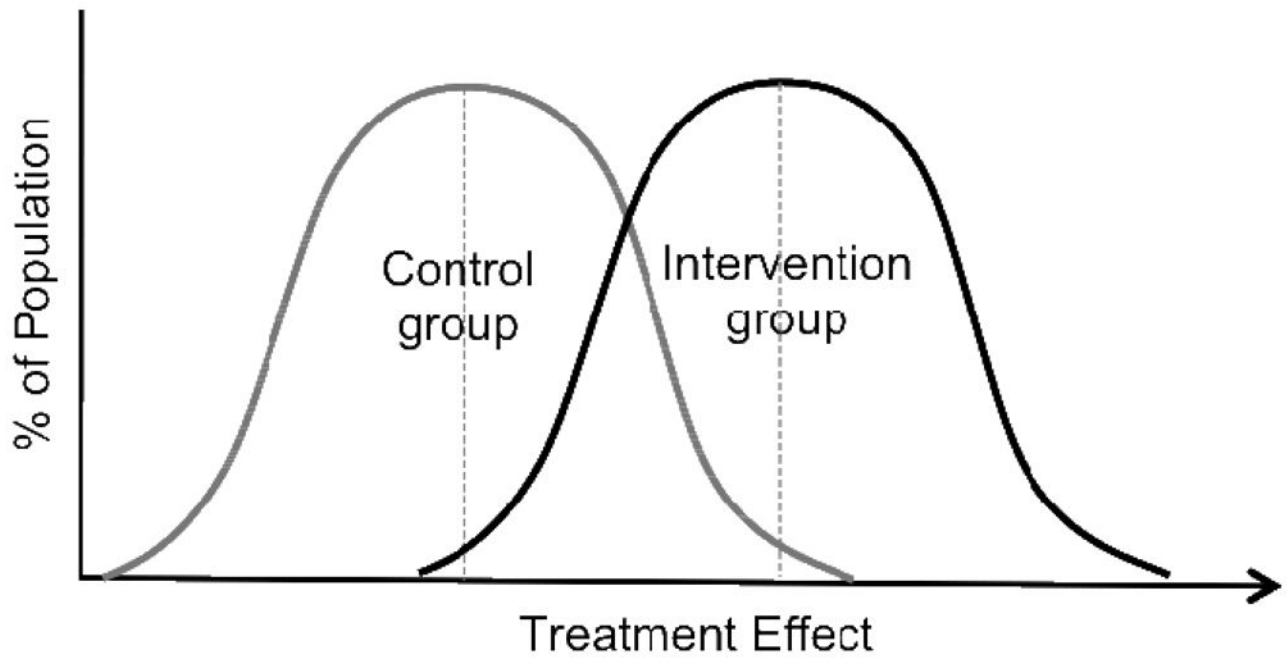


Figure 1. Distribution of Responses to Treatment and Control in Setting of a Heterogeneous Treatment Effect. Here, some *individuals* in treatment group (those in the left tail of the treatment curve) will have a worse treatment response than the *average* individual in the control group whereas other *individuals* in the treatment group (those in the right tail of the treatment curve) will have much larger response to treatment than the *average* individual in the control group. Due to this large between subject difference in treatment response, it is impossible to provide individual patients with advice as to whether the treatment is right for them.

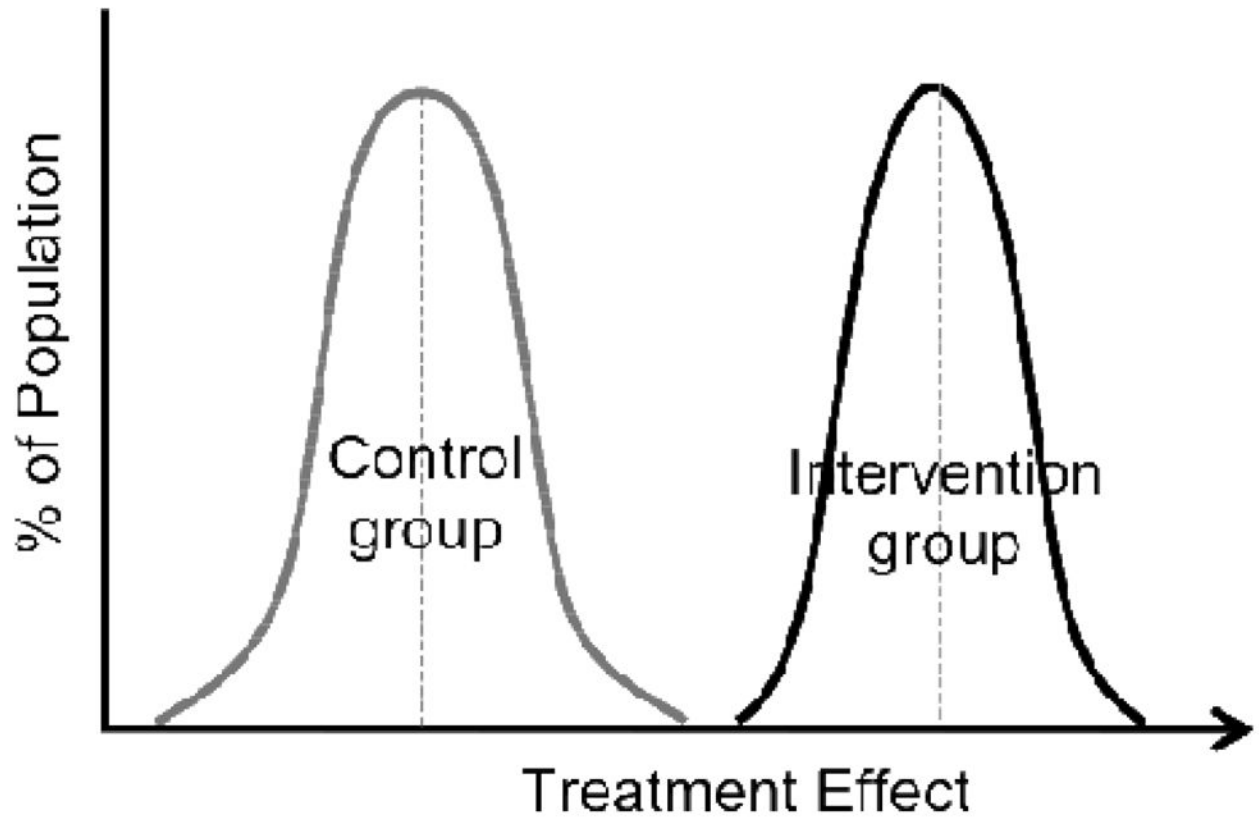


Figure 2.

Distribution of Responses to Treatment and Control in Setting of a Homogenous Treatment Effect. Here, all of the individuals in the treatment group have a greater response to treatment than individuals in the control group and there are no large differences in the magnitude of this effect between individuals in the treatment as compared to the control group. As all individuals in treatment group have better response to the treatment than do the individuals in the control condition, there is no need to conduct additional trials to best understand the treatment needed for an individual patient.

Table 1

Considerations for Conducting N-of-1 Trials

	N-of-1 appropriate	N-of-1 not appropriate	Notes
State of knowledge	Clinical uncertainty or equipoise	Clear benefit of one treatment for the patient	
Nature of the disorder	Chronic stable or slowly progressing, frequently recurring symptoms Washout period between treatments is safe	Rapid progression Patient harm is possible if active treatment is discontinued	Time trends for symptoms should be considered in N-of-1 design (randomization vs counterbalancing)
Nature of treatments	Rapid efficacy Minimal carryover across time Significant individual differences in treatment response are expected	Slow efficacy onset Substantial carryover across time (long washout) Too complex/requires constant adjustment Small individual differences in treatment response are expected	Blinding of treatment assignment is ideal but not always necessary Consider different treatment efficacy onset periods in analyses
Outcome assessment	Valid measures of outcome can be assessed multiple times	Primary outcome is assessed at a single time point	Repeated assessments within treatment condition periods are ideal
Willingness of Stakeholders	Patient, physician, pharmacist, and statistician willing to expend effort	One of the stakeholders not available/willing	All items must be conducted <i>a priori</i> , including obtaining IRB approval and patient consent
Availability of Financial Resources	Measurement devices for outcome, cost of compounding drug, and multiple visits with physician can be procured	Resources not available	Most previous N-of-1 RCT clinical services have only existed where external grant support was available to cover these costs

Table 2

Barriers to N of 1 Trials

Cost	<i>Unreimbursed costs include:</i> Developing data collection tools Developing randomized treatment plans Developing blinding strategies Designing and preparing medications/interventions Database preparation and analysis Developing communication strategies for trial enrollment and results dissemination
Physician time	Trials may require multiple appointments for transition between conditions and outcome assessment
Patient time	Multiple assessments and outpatient visits can be onerous
Pharmacist availability to create compounds and drugs	For N-of-1 drug trials, a pharmacist may need to create identical placebo and active pills for blinding, and may need to dispense according to randomization schedule
Statistician availability to set up randomization, time period, conduct analyses	To maintain blinding and randomization, a statistician must maintain and communicate the randomization schedule to physicians and/or pharmacists
Physician or patient lack of equipoise	In many cases, the physician, patient, or both may be convinced of a particular treatment's effectiveness for that patient without clear empirical evidence to support the belief
IRB approval frequently needed	An infrastructure is necessary for obtaining IRB approval and ongoing communication with IRB, as results are used for both clinical and research purposes