

# Optimizing behavioral health interventions with single-case designs: from development to dissemination

Jesse Dallery, Ph.D.,<sup>1</sup> Bethany R Raiff, Ph.D.<sup>2</sup>

<sup>1</sup>Department of Psychology,  
University of Florida, P. O. box  
112250, Gainesville, FL 32611, USA

<sup>2</sup>Department of Psychology,  
Rowan University, Glassboro, USA

Correspondence to: J Dallery  
dallery@ufl.edu

Cite this as: *TBM* 2014;4:290–303  
doi: 10.1007/s13142-014-0258-z

## Abstract

Over the past 70 years, single-case design (SCD) research has evolved to include a broad array of methodological and analytic advances. In this article, we describe some of these advances and discuss how SCDs can be used to optimize behavioral health interventions. Specifically, we discuss how parametric analysis, component analysis, and systematic replications can be used to optimize interventions. We also describe how SCDs can address other features of optimization, which include establishing generality and enabling personalized behavioral medicine. Throughout, we highlight how SCDs can be used during both the development and dissemination stages of behavioral health interventions.

## Keywords

Generality, Optimization, Research designs, Single-case designs

Research methods are tools to discover new phenomena, test theories, and evaluate interventions. Many researchers have argued that our research tools have become limited, particularly in the domain of behavioral health interventions [1–9]. The reasons for their arguments vary, but include an overreliance on randomized controlled trials, the slow pace and high cost of such trials, and the lack of attention to individual differences. In addition, advances in mobile and sensor-based data collection now permit real-time, continuous observation of behavior and symptoms over extended durations [3, 10, 11]. Such fine-grained observation can lead to tailoring of treatment based on changes in behavior, which is challenging to evaluate with traditional methods such as a randomized trial.

In light of the limitations of traditional designs and advances in data collection methods, a growing number of researchers have advocated for alternative research designs [2, 7, 10]. Specifically, one family of research designs, known as single-case designs (SCDs), has been proposed as a useful way to establish the preliminary efficacy of health interventions [3]. In the present article, we recapitulate and expand on this proposal, and argue that they can be used to optimize health interventions.

## Implications

**Practitioners:** practitioners can use single-case designs in clinical practice to help ensure that an intervention or component of an intervention is working for an individual client or group of clients.

**Policy makers:** results from a single-case design research can help inform and evaluate policy regarding behavioral health interventions.

**Researchers:** researchers can use single-case designs to evaluate and optimize behavioral health interventions.

We begin with a description of what we consider to be a set of criteria, or ideals, for what research designs should accomplish in attempting to optimize an intervention. Admittedly, these criteria are self-serving in the sense that most of them constitute the strengths of SCDs, but they also apply to other research designs discussed in this volume. Next, we introduce SCDs and how they can be used to optimize treatment using parametric and component analyses. We also describe how SCDs can address other features of optimization, which include establishing generality and enabling personalized behavioral medicine. Throughout, we also highlight how these designs can be used during both the development and dissemination of behavioral health interventions. Finally, we evaluate the extent to which SCDs live up to our ideals.

## AN OPTIMIZATION IDEAL

During development and testing of a new intervention, our methods should be efficient, flexible, and rigorous. We would like efficient methods to help us establish preliminary efficacy, or “clinically significant patient improvement over the course of treatment” [12] (p. 137). We also need flexible methods to test different parameters or components of an intervention. Just as different doses of a drug treatment may need to be titrated to optimize effects, different parameters or components of a behavioral treatment may need to be titrated to

optimize effects. It should go without saying that we also want our methods to be rigorous, and therefore eliminate or reduce threats to internal validity.

Also, during development, we would like methods that allow us to assess replications of effects to establish the reliability and generality of an intervention. Replications, if done systematically and thoughtfully, can answer questions about for whom and under what conditions an intervention is effective. Answering these questions speaks to the generality of research findings. As Cohen [13] noted in a seminal article: “For generalization, psychologists must finally rely, as has been done in all the older sciences, on replication” (p. 997). Relying on replications and establishing the conditions under which an intervention works could also lead to more targeted, efficient dissemination efforts.

During dissemination, when an intervention is implemented in clinical practice, we again would like to know if the intervention is producing a reliable change in behavior for a particular individual. (Here, “we” may refer to practitioners in addition to researchers.) With knowledge derived from development and efficacy testing, we may be able to alter components of an intervention that impact its effectiveness. But, ideally, we would like to not only alter but verify whether these components are working. Also, recognizing that behavior change is idiosyncratic and dynamic, we may need methods that allow ongoing tailoring and testing. This may result in a kind of personalized behavioral medicine in which what gets personalized, and when, is determined through experimental analysis.

In addition, during both development and dissemination, we want methods that afford innovation. We should have methods that allow rapid, rigorous testing of new treatments, and which permit incorporating new technologies to assess and treat behavior as they become available. This might be thought of as systematic play. Whatever we call it, it is a hallmark of the experimental attitude in science.

#### INTRODUCTION TO SINGLE-CASE DESIGNS

SCDs include an array of methods in which each participant, or case, serves as his or her own control. Although these methods are conceptually rooted in the study of cognition and behavior [14], they are theory-neutral and can be applied to any health intervention. In a typical study, some behavior or symptom is measured repeatedly during all conditions for all participants. The experimenter systematically introduces and withdraws control and intervention conditions, and assesses effects of the intervention on behavior across replications of these conditions within and across participants. Thus, these studies include repeated, frequent assessment of behavior, experimental manipulation of the independent variable (the intervention or components of the intervention), and replication of effects within and across participants.

The main challenge in conducting a single-case experiment is collecting data of the same behavior or symptom repeatedly over time. In other words, a time series must be possible. If behavior or symptoms cannot be assessed frequently, then SCDs cannot be used (e.g., on a weekly basis, at a minimum, for most health interventions). Fortunately, technology is revolutionizing methods to collect data. For example, ecological momentary assessment (EMA) enables frequent input by an end-user into a handheld computer or mobile phone [15]. Such input occurs in naturalistic settings, and it usually occurs on a daily basis for several weeks to months. EMA can therefore reveal behavioral variation over time and across contexts, and it can document effects of an intervention on an individual’s behavior [15]. Sensors to record physical activity, medication adherence, and recent drug use also enable the kind of assessment required for single-case research [10, 16]. In addition, advances in information technology and mobile phones can permit frequent assessment of behavior or symptoms [17, 18]. Thus, SCDs can capitalize on the ability of technology to easily, unobtrusively, and repeatedly assess health behavior [3, 18, 19].

SCDs suffer from several misconceptions that may limit their use [20–23]. First, a single case does not mean “*n* of 1.” The number of participants in a typical study is almost always more than 1, usually around 6 but sometimes as many as 20, 40, or more participants [24, 25]. Also, the unit of analysis, or “case,” could be individual participants, clinics, group homes, hospitals, health care agencies, or communities [1]. Given that the unit of analysis is each case (i.e., participant), a single study could be conceptualized as a series of single-case experiments. Perhaps a better label for these designs would be “intrasubject replication designs” [26]. Second, SCDs are not limited to interventions that produce large, immediate changes in behavior. They can be used to detect small but meaningful changes in behavior and to assess behavior that may change slowly over time (e.g., learning a new skill) [27]. Third, SCDs are not quasi-experimental designs [20]. The conventional notions that detecting causal relations requires random assignment and/or random sampling are false [26]. Single-case experiments are fully experimental and include controls and replications to permit crisp statements about causal relations between independent and dependent variables.

#### VARIETIES OF SINGLE-CASE DESIGNS

The most relevant SCDs to behavioral health interventions are presented in Table 1. The table also presents some procedural information and advantages and disadvantages for each design. (The material below is adapted from [3]) There are also a number of variants of these designs, enabling flexibility in tailoring the design based on practical

**Table 1** | Several single-case designs, including general procedures, advantages, and disadvantages

Design	Procedure	Advantages	Disadvantages
Reversal (ABA, ABAB)	Baseline conducted, treatment is implemented, and then treatment is removed	Within-subject replication; clear demonstration of an intervention effect in one subject	Not applicable if behavior is irreversible, or when removing treatment is undesirable
Multiple baseline (interrupted time series, stepped wedge)	Baseline is conducted for varying durations across participants, then treatment is introduced in a staggered fashion	Treatment does not have to be withdrawn	No within-subject replication. Potentially more subjects needed to demonstrate intervention effects than when using reversal design
Changing criterion	Following a baseline phase, treatment goals are implemented. Goals become progressively more challenging as they are met	Demonstrates within-subject control by levels of the independent variable without removing treatment; Useful when gradual change in behavior is desirable	Not applicable for binary outcome measures—must have continuous outcomes
Combined	Elements of any treatment can be combined	Allows for more flexible, individually tailored designs	If different designs are used across participants in a single study, comparisons across subjects can be difficult

or empirical considerations [27, 28]. For example, there are several variants to circumvent long periods of assessing behavior during baseline conditions, which may be problematic if the behavior is dangerous, before introducing a potentially effective intervention [28].

Procedural controls must be in place to make inferences about causal relations, such as clear, operational definitions of the dependent variables,

reliable and valid techniques to assess the behavior, and the experimental design must be sufficient to rule out alternative hypotheses for the behavior change. Table 2 presents a summary of methodological and assessment standards to permit conclusions about treatment effects [29, 30]. These standards were derived from Horner et al. [29] and from the recently released What Works Clearinghouse (WWC) pilot standards for evaluating single-

**Table 2** | Quality indicators for single-case research [29]

Dependent variable	<ul style="list-style-type: none"> <li>• Dependent variables are described with operational and replicable precision</li> <li>• Each dependent variable is measured with a procedure that generates a quantifiable index</li> <li>• Dependent variables are measured repeatedly over time</li> <li>• In the case of remote data capture, the identity of the source of the dependent variable should be authenticated or validated [3]</li> </ul>
Independent variable	<ul style="list-style-type: none"> <li>• Independent variable is described with replicable precision</li> <li>• Independent variable is systematically manipulated and under the control of the experimenter</li> <li>• Overt measurement of the fidelity of implementation of the independent variable is highly desirable</li> </ul>
Baseline	<ul style="list-style-type: none"> <li>• The majority of single-case research will include a baseline phase that provides repeated measurement of a dependent variable and establishes a pattern of responding that can be used to predict/compare against the pattern of future performance, if introduction or manipulation of the independent variable did not occur</li> <li>• Baseline conditions are described with replicable precision</li> </ul>
Experimental control/internal validity	<ul style="list-style-type: none"> <li>• The design provides at least three demonstrations of experimental effect at three different points in time</li> <li>• The design controls for common threats to internal validity (e.g., permits elimination of rival hypotheses)</li> <li>• There are a sufficient number of data points for each phase (e.g., minimum of five) for each participant</li> <li>• The results document a pattern that demonstrates experimental control</li> </ul>
Social validity	<ul style="list-style-type: none"> <li>• The dependent variable is socially important</li> <li>• The magnitude of change in the dependent variable resulting from the intervention is socially important</li> <li>• The methods are acceptable to the participant</li> </ul>

case research to inform policy and practice (hereafter referred to as the SCD standards) [31].

All of the designs listed in Table 1 entail a baseline period of observation. During this period, the dependent variable is measured repeatedly under control conditions. For example, Dallery, Glenn, and Raiff [24] used a reversal design to assess effects of an internet-based incentive program to promote smoking cessation, and the baseline phase included self-monitoring, carbon monoxide assessment of smoking status via a web camera, and monetary incentives for submitting videos. The active ingredient in the intervention, incentives contingent on objectively verified smoking abstinence, was not introduced until the treatment phase.

The duration of the baseline and the pattern of the data should be sufficient to predict future behavior. That is, the level of the dependent variable should be stable enough to predict its direction if the treatment was not introduced. If there is a trend in the direction of the anticipated treatment effect during baseline, or if there is too much variability, the ability to detect a treatment effect will be compromised. Thus, stability, or in some cases a trend in the direction opposite the predicted treatment effect, is desirable during baseline conditions.

In some cases, the source(s) of variability can be identified and potentially mitigated (e.g., variability could be reduced by automating data collection, standardizing the setting and time for data collection). However, there may be instances when there is too much variability during baseline conditions, and thus, detecting a treatment effect will not be feasible. There are no absolute standards to define what “too much” variability means [27]. Excessive variability is a relative term, which is typically determined by a comparison of performance within and between conditions (e.g., between baseline and intervention conditions) in a single-case experiment. The mere presence of variability does not mean that a single-case approach should be abandoned, however. Indeed, identifying the sources of variability and/or assessing new measurement strategies can be evaluated using SCDs. Under these conditions, the outcome of interest is not an increase or a decrease in some behavior or symptom but a reduction in variability. Once accomplished, the researcher has not only learned something useful but is also better prepared to evaluate the effects of an intervention to increase or decrease some health behavior.

#### REVERSAL DESIGNS

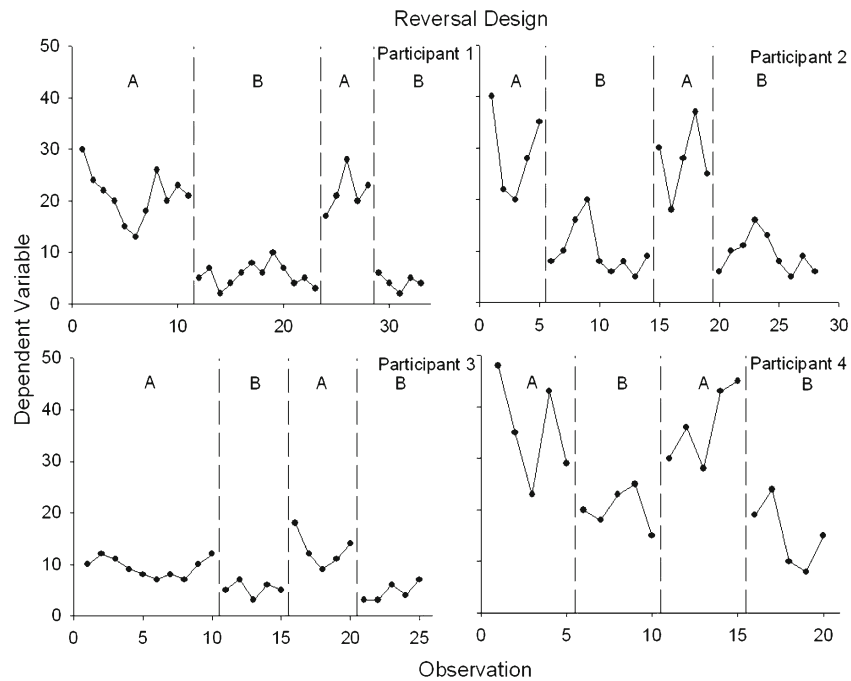
In a reversal design, a treatment is introduced after the baseline period, and then a baseline period is re-introduced, hence, the “reversal” in this design (also known as an ABA design, where “A” is baseline and “B” is treatment). Using only two conditions, such as a pre-post design, is not considered sufficient to demonstrate experimental control because other sources of influence on behavior cannot be ruled

out [31, 32]. For example, a smoking cessation intervention could coincide with a price increase in cigarettes. By returning to baseline conditions, we could assess and possibly rule out the influence of the price increase on smoking. Researchers also often use a reversal to the treatment condition. Thus, the experiment ends during a treatment period (an ABAB design). Not only is this desirable from the participant’s perspective but it also provides a replication of the main variable of interest—the treatment [33].

Figure 1 displays an idealized, ABAB reversal design, and each panel shows data from a different participant. Although all participants were exposed to the same four conditions, the duration of the conditions differed because of trends in the conditions. For example, for participant 1, the beginning of the first baseline condition displays a consistent downward trend (in the same direction as the expected text-message treatment effects). If we were to introduce the smoking cessation-related texts after only five or six baseline sessions, it would be unclear if the decrease in smoking was a function of the independent variable. Therefore, continuing the baseline condition until there is no visible trend helps build our confidence about the causal role of the treatment when it is introduced. The immediate decrease in the level of smoking for participant 1 when the treatment is introduced also implicates the treatment. We can also detect, however, an increasing trend in the early portion of the treatment condition. Thus, we need to continue the treatment condition until there is no undesirable trend before returning to the baseline condition. Similar patterns can be seen for participants 2–4. Based on visual analysis of Fig. 1, we would conclude that treatment is exerting a reliable effect on smoking. But, the meaningfulness of this effect requires additional considerations (see the section below on “[Visual, Statistical, and Social Validity Analysis](#)”).

Studies using reversal designs typically include at least four or more participants. The goal is to generate enough replications, both within participants and across participants, to permit a confident statement about causal relations. For example, several studies on incentive-based treatment to promote drug abstinence have used 20 participants in a reversal design [24, 25]. According to the SCD standards, there must be a minimum of three replications to support conclusions about experimental control and thus causation. Also, according to the SCD standards, there must be at least three and preferably five data points per phase to allow the researcher to evaluate stability and experimental effects [31].

There are two potential limitations of reversal designs in the context of behavioral health interventions. First, the treatment must be withdrawn to demonstrate causal relations. Some have raised an ethical objection about this practice [11]. However, we think that the benefits of demonstrating that a



**Fig 1 |** Example of a reversal design showing experimental control and replications within and between subjects. Each panel represents a different participant, each of whom experienced two baseline and two treatment conditions

treatment works outweigh the risks of temporarily withdrawing treatment (in most cases). The treatment can also be re-instituted in a reversal design (i.e., an ABAB design). Second, if the intervention produces relatively permanent changes in behavior, then a reversal to pre-intervention conditions may not be possible. For example, a treatment that develops new skills may imply that these skills cannot be “reversed.” Some interventions do not produce permanent change and must remain in effect for behavior change to be maintained, such as some medications and incentive-based procedures. Under conditions where behavior may not return to baseline levels when treatment is withdrawn, alternative designs, such as multiple-baseline designs, should be used.

**MULTIPLE-BASELINE DESIGNS**

In a multiple-baseline design, the durations of the baselines vary systematically for each participant in a so-called staggered fashion. For example, one participant may start treatment after five baseline days, another after seven baseline days, then nine, and so on. After baseline, treatment is introduced, and it remains until the end of the experiment (i.e., there are no reversals). Like all SCDs, this design can be applied to individual participants, clusters of individuals, health care agencies, and communities. These designs are also referred to as interrupted time-series designs [1] and stepped wedge designs [7].

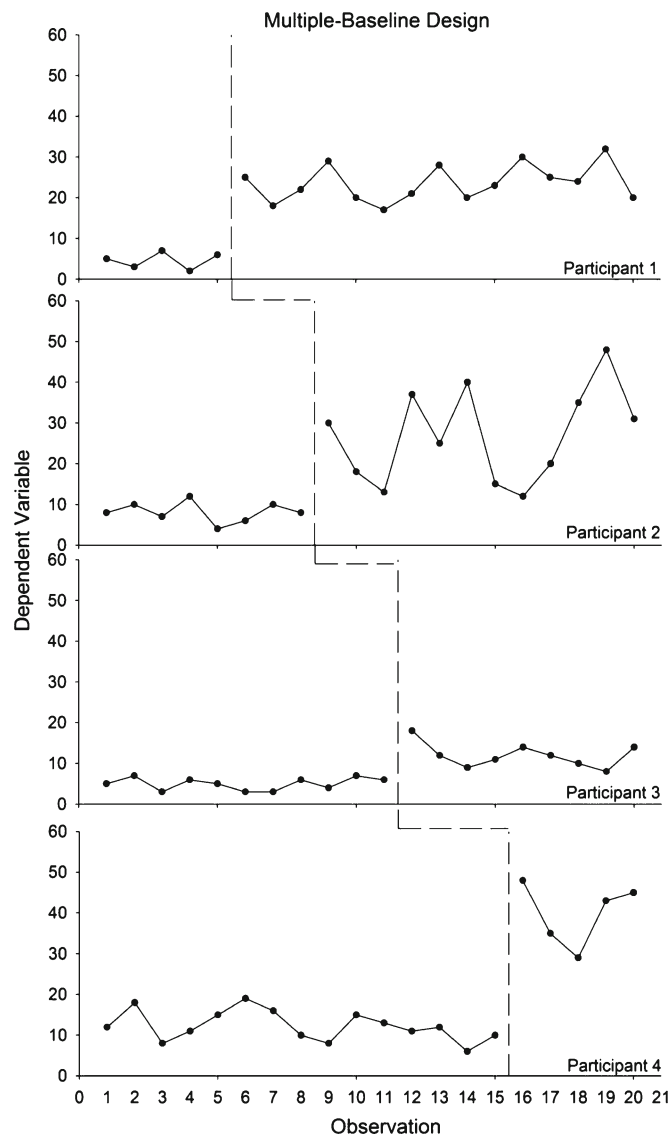
The utility of these designs is derived from demonstrating that change occurs when, and only when, the intervention is directed at a particular

participant (or whatever the unit of analysis happens to be [28]). The influence of other factors, such as idiosyncratic experiences of the individual or self-monitoring (e.g., reactivity), can be ruled out by replicating the effect across multiple individuals. A key to ruling out extraneous factors is a stable enough baseline phase (either no trends or a trend in the opposite direction to the treatment effect). As replications are observed across individuals, and behavior changes when and only when treatment is introduced, confidence that behavior change was caused by the treatment increases.

As noted above, multiple-baseline designs are useful for interventions that teach new skills, where behavior would not be expected to “reverse” to baseline levels. Multiple-baseline designs also obviate the ethical concern about withdrawing treatment (as in a reversal design) or using a placebo control comparison group (as in randomized trials), as all participants are exposed to the treatment with multiple-baseline designs.

Figure 2 illustrates a simple, two-condition multiple-baseline design replicated across four participants. As noted above, the experimenter should introduce treatment only when the data appear stable during baseline conditions. The durations of the baseline conditions are staggered for each participant, and the dependent variable increases when, and only when, the independent variable is introduced for all participants. The SCD standards requires at least six phases (i.e., three baseline and three treatment) with at least five data points per phase [31]. Figure 2 suggests reliable increases in





**Fig 2 |** Example of a multiple-baseline design showing experimental control and replications between subjects. Each row represents a different participant, each of whom experienced a baseline and treatment. The baseline durations differed across participants

behavior and that the treatment was responsible for these changes.

**CHANGING CRITERION DESIGN**

The changing criterion design is also relevant to optimizing interventions [34]. In a changing criterion design, a baseline is conducted until stability is attained. Then, a treatment goal is introduced, and goals are made progressively more difficult. Behavior should track the introduction of each goal, thus demonstrating control by the level of the independent variable [28]. For example, Kurti and Dallery [35] used a changing criterion design to increase activity in six sedentary adults using an internet-based contingency management program to promote walking. Weekly step count goals were gradually increased across 5-day blocks. The step counts for all six

participants increased reliably with each increase in the goals, thereby demonstrating experimental control of the intervention. This design has many of the same benefits of the multiple-baseline design, namely that a reversal is not required for ethical or potentially practical reasons (i.e., irreversible treatment effects).

**VISUAL, STATISTICAL, AND SOCIAL VALIDITY ANALYSIS**

Analyzing the data from SCDs involves three questions: (a) Is there a reliable effect of the intervention? (b) What is the magnitude of the effect? and (c) Are the results clinically meaningful and socially valid [31]? Social validity refers to the extent to which the goals, procedures, and results of an intervention are socially acceptable to the client, the researcher or health care practitioner, and society [36–39]. The first two

questions can be answered by visual and statistical analysis, whereas the third question requires additional considerations.

The SCD standards prioritizes visual analysis of the time-series data to assess the reliability and magnitude of intervention effects [29, 31, 40]. Clinically significant change in patient behavior should be visible. Visual analysis prioritizes clinically significant change in health-related behavior as opposed to *statistically* significant change in group behavior [13, 41, 42]. Although several researchers have argued that visual analysis may be prone to elevated rates of type 1 error, such errors may be limited to a narrow range of conditions (e.g., when graphs do not contain contextual information about the nature of the plotted behavioral data) [27, 43]. Furthermore, in recent years, training in visual analysis has become more formalized and rigorous [44]. Perhaps as a result, Kahng and colleagues found high reliability among visual analysts in judging treatment effects based on analysis of 36 ABAB graphs [45]. The SCD standards recommends four steps and the evaluation of six features of the graphical displays for all participants in a study, which are displayed in Table 3 [31]. As the visual analyst progresses through the steps, he or she also uses the six features to evaluate effects within and across experimental phases.

In addition to visual analysis, several regression-based approaches are available to analyze time-series data, such as autoregressive models, robust regression, and hierarchical linear modeling (HLM) [46–49]. A variety of non-parametric statistics are also available [27]. Perhaps because of the proliferation of statistical methods, there is a lack of consensus about which methods are most appropriate in light of different properties of the data (e.g.,

the presence of trends and autocorrelation [43, 50], the number of data points collected, etc.). A discussion of statistical techniques is beyond the scope of this paper. We recommend Kazdin’s [27] or Barlow and colleague’s [28] textbooks as useful resources regarding statistical analysis of time-series data. The SCD standards also includes a useful discussion of statistical approaches for data analysis [31].

A variety of effect size calculations have been proposed for SCDs [13, 51–54]. Although effect size estimates may allow for rank ordering of most to least effective treatments [55], most estimates do not provide metrics that are comparable to effect sizes derived from group designs [31]. However, one estimate that provides metrics comparable to group designs has been developed and tested by Shadish and colleagues [56, 57]. They describe a standardized mean difference statistic (*d*) that is equivalent to the more conventional *d* in between-groups experiments. The *d* statistic can also be used to compute power based on the number of observations in each condition and the number of cases in an experiment [57]. In addition, advances in effect size estimates has led to several meta-analyses of results from SCDs [48, 58–61]. Zucker and associates [62] explored Bayesian mixed-model strategy to combining SCDs using, which allowed population-level claims about the merits of different intervention strategies.

Determining whether the results are clinically meaningful and socially valid can be informed by visual and most forms of statistical analysis (i.e., not null-hypothesis significance testing) [42, 63]. One element in judging social validity concerns the clinical meaningfulness of the magnitude of behav-

**Table 3** | Four steps and six outcome measures to evaluate when conducting visual analysis of time-series data

Four steps to visual analysis of single-case research designs	
Step	Description
Step 1: document a stable baseline	Data show a predictable and stable pattern over time
Step 2: identify within-phase patterns of responding	Examine data paths within each phase of the study. Examine whether there is enough data within each phase and whether the data are stable and predictable
Step 3: compare data across phases	Compare data within each phase to the adjacent (or similar) phase to assess whether manipulating the independent variable is associated with an effect
Step 4: integrate information from all phases	Determine whether there are at least three demonstrations or replications of an effect at different points in time
Six outcome measures	
Name	Definition
Level	Average of the outcome measures within a phase
Trend	The slope of the best-fitting line of the outcome measures within a phase
Variability	Range, variance, or standard deviation of the best-fitting line of the outcome measures within a phase, or the degree of overall scatter
Immediacy of the effect	Change in level between the last three data points of one phase and the first three data points in the next
Overlap	Proportion of data from one phase that overlaps with data from the previous phase
Consistency of data patterns	Consistency in the data patterns from phases with the same conditions

ior change. This judgment can be made by the researcher or clinician in light of knowledge of the subject matter, and perhaps by the client being treated. Depending on factors such as the type of behavior and the way in which change is measured, the judgment can also be informed by previous research on a minimal clinically important difference (MCID) for the behavior or symptom under study [64, 65]. The procedures used to generate the effect also require consideration. Intrusive procedures may be efficacious yet not acceptable. The social validity of results and procedures should be explicitly assessed when conducting SCD research, and a variety of tools have emerged to facilitate such efforts [37]. Social validity assessment should also be viewed as a process [37]. That is, it can and should be assessed at various time points as an intervention is developed, refined, and eventually implemented. Social validity may change as the procedures and results of an intervention are improved and better appreciated in the society at large.

#### OPTIMIZATION METHODS AND SINGLE-CASE DESIGNS

The SCDs described above provide an efficient way to evaluate the effects of a behavioral intervention. However, in most of the examples above, the interventions were held constant during treatment periods; that is, they were procedurally static (cf. [35]). This is similar to a randomized trial, in which all components of an intervention are delivered all at once and held constant throughout the study. However, the major difference between the examples above and traditional randomized trials is efficiency: SCDs usually require less time and fewer resources to demonstrate that an intervention can change behavior. Nevertheless, a single, procedurally static single-case experiment does not optimize treatment beyond showing whether or not it works.

One way to make initial efficacy testing more dynamic would be to conduct a series of single-case experiments in which aspects of the treatment are systematically explored. For example, a researcher could assess effects of different frequencies, timings, or tailoring dimensions of a text-based intervention to promote physical activity. Such manipulation could also be conducted in separate experiments conducted by the same or different researchers. Some experiments may reveal larger effects than others, which could then lead to further replications of the effects of the more promising intervention elements. This iterative development process, with a focus on systematic manipulation of treatment elements and replications of effects within and across experiments, could lead to an improved intervention within a few years' time. Arguably, this process could yield more clinically useful information than a procedurally static randomized trial conducted over the same period [5, 17].

To further increase the efficiency of optimizing treatment, different components or parameters of an

intervention can be systematically evaluated within and across single-case experiments. There are two ways to optimize treatment using these methods: parametric and component analyses.

#### PARAMETRIC ANALYSIS

Parametric analysis involves exposing participants to a range of values of the independent variable, as opposed to just one or two values. To qualify as a parametric analysis, three is the minimum number of values that must be evaluated, as this number is the minimum to evaluate the function form relating the independent to the dependent variable. One goal of a parametric analysis is to identify the optimal value that produces a behavioral outcome. Another goal is to identify general patterns of behavior engendered by a range of values of the independent variable [26, 63].

Many behavioral health interventions can be delivered at different levels [66] and are therefore amenable to parametric analysis. For example, text-based prompts can be delivered at different frequencies, incentives can be delivered at different magnitudes and frequencies, physical activity can occur at different frequencies and intensities, engagement in a web-based program can occur at different levels, medications can be administered at different doses and frequencies, and all of the interventions could be delivered for different durations.

The repeated measures, and resulting time-series data, that are inherent to all SCDs (e.g., reversal and multiple-baseline designs) make them useful designs to conduct parametric analyses. For example, two doses of a medication, low versus high, labeled B and C, respectively, could be assessed using a reversal design [67]. There may be several possible sequences to conduct the assessment such as ABCBCA or ABCABCA. If C is found to be more effective of the two, it might behoove the researcher to replicate this condition using an ABCBCAC design. A multiple baseline across participants could also be conducted to assess the two doses, one dose for each participant, but this approach may be complicated by individual variability in medication effects. Instead, the multiple-baseline approach could be used on a within-subject basis, where the durations of not just the baselines but of the different dose conditions are varied across participants [68].

Guyatt and colleagues [5] provide an excellent discussion about how parametric analysis can be used to optimize an intervention. The intervention was amitriptyline for the treatment of fibrositis. The logic and implications of the research tactics, however, also apply to other interventions that have parametric dimensions. At the time that the research was conducted, a dose of 50 mg/day was the standard recommendation for patients. To determine whether this dose was optimal for a given individual, the researchers first exposed participants to low doses, and if no response was noted relative



to placebo, then they systematically increased the dose until a response was observed, or until they reached the maximum of 50 mg/day. In general, their method involved a reversal design in which successively higher doses alternated with placebo. So, for example, if one participant did not respond to a low dose, then doses might be increased to generate an ABCD design, where each successive letter represents a higher dose (other sequences were arranged as well). Parametrically examining doses in this way, and examining individual subject data, the researchers found that some participants responded favorably at lower doses than 50 mg/day (e.g., 10 or 20 mg/day). This was an important finding because the higher doses often produced unwanted side effects. Once optimal doses were identified for individuals, the researchers were able to conduct further analyses using a reversal design, exposing them to either their optimal dose or placebo on different days.

Guyatt and colleagues also investigated the minimum duration of treatment necessary to detect an effect [5]. Initially, all participants were exposed to the medication for 4 weeks. Visual analysis of the time-series data revealed that medication effects were apparent within about 1–2 weeks of exposure, making a 4-week trial unnecessary. This discovery was replicated in a number of subjects and led them to optimize future, larger studies by only conducting a 2-week intervention. Investigating different treatment durations, such as this, is also a parametric analysis.

Parametric analysis can detect effects that may be missed using a standard group design with only one or two values of the independent variable. For example, in the studies conducted by Guyatt and colleagues [5], if only the lowest dose of amitriptyline had been investigated using a group approach, the researchers may have incorrectly concluded that the intervention was ineffective because this dose only worked for some individuals. Likewise, if only the highest dose had been investigated, it may have been shown to be effective, but potentially more individuals would have experienced unnecessary side effects (i.e., the results would have low social validity for these individuals). Perhaps most importantly, in contrast to what is typically measured in a group design (e.g., means, confidence intervals, etc.), optimizing treatment effects is fundamentally a question about an *individual's* behavior.

#### COMPONENT ANALYSIS

A component analysis is “any experiment designed to identify the active elements of a treatment condition, the relative contributions of different variables in a treatment package, and/or the necessary and sufficient components of an intervention” [69]. Behavioral health interventions often entail more than one potentially active treatment element. Determining the active elements may be important

to increase dissemination potential and decrease cost. Single-case research designs, in particular the reversal and multiple-baseline designs, may be used to perform a component analysis. The essential experimental ingredients, regardless of the method, are that the independent variable(s) are systematically introduced and/or withdrawn, combined with replication of effects within and/or between subjects.

There are two main variants of component analyses: the dropout and add-in analyses. In a dropout analysis, the full treatment package is presented following a baseline phase and then components are systematically withdrawn from the package. A limitation of dropout analyses is when components produce irreversible behavior change (i.e., learning a new skill). Given that most interventions seek to produce sustained changes in health-related behavior, dropout analyses may have limited applicability. Instead, in add-in analyses, components can be assessed individually and/or in combination before the full treatment package is assessed [69]. Thus, a researcher could conduct an ABACAD design, where A is baseline, B and C are the individual components, and D is the combination of the two B and C components. Other sequences are also possible, and which one is selected will require careful consideration. For example, sequence effects should be considered, and researchers could address these effects through counterbalancing, brief “washout” periods, or explicit investigation of these effects [26]. If sequence effects cannot be avoided, combined SCD and group designs can be used to perform a component analysis. Thus, different components of a treatment package can be delivered between two groups, and within each group, a SCD can be used to assess effects of each combination of components. Although very few component analyses have assessed health behavior or symptoms per se as the outcome measure, there are a variety of behavioral interventions that have been evaluated using component analysis [63]. For example, Sanders [70] conducted a component analysis of an intervention to decrease lower back pain (and increase time standing/walking). The analysis consisted of four components: functional analysis of pain behavior (e.g., self-monitoring of pain and the conditions that precede and follow pain), progressive relaxation training, assertion training, and social reinforcement of increased activity. Sanders concluded that both relaxation training and reinforcement of activity were necessary components (see [69] for a discussion of some limitations of this study).

Several conclusions can be drawn about the effects of the various components in changing behavior. The data should first be evaluated to determine the extent to which the effects of individual components are independent of one another. If they are, then the effects of the components are additive. If they are not, then the effects are multiplicative, or the effects of one component

depend on the presence of another component. Figure 3 presents simplified examples of these two possibilities using a reversal design and short data streams (adapted from [69]). The panel on the left shows additive effects, and the panel on the right shows multiplicative effects. The data also can be analyzed to determine whether each component is necessary and sufficient to produce behavior change. For instance, the panel on the right shows that neither the component labeled X (e.g., self-monitoring of health behavior) nor the component labeled Y (e.g., counseling to change health behavior) is sufficient, and both components are necessary. If two components produce equal changes in behavior, and the same amount of change when both are combined, then either component is sufficient but neither is necessary.

The logic of the component analyses described here is similar to new methods derived from an engineering framework [2, 9, 71]. During the initial stages of intervention development, researchers use factorial designs to allocate participants to different combinations of treatment components. These designs, called fractional factorials because not all combinations of components are tested, can be used to screen promising components of treatment packages. The components tested may be derived from theory or working assumptions about which components and combinations will be of interest, which is the same process used to guide design choices in SCD research. Just as engineering methods seek to isolate and combine active treatment components to optimize interventions, so too do single-case methods. The main difference between approaches is the focus on the individual as the unit of analysis in SCDs.

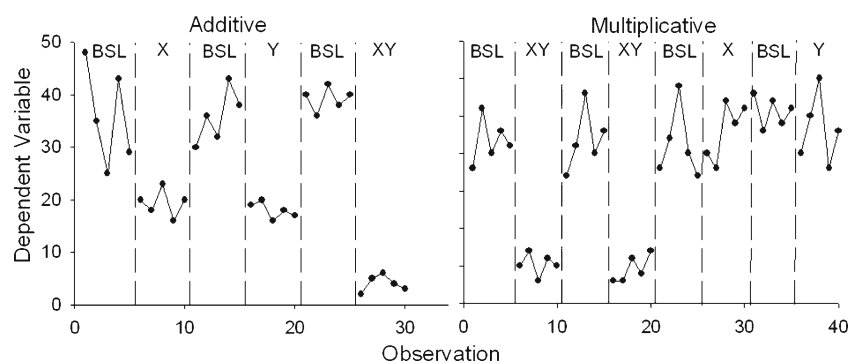
**OPTIMIZING WITH REPLICATIONS AND ESTABLISHING GENERALITY**

Another form of optimization is an understanding of the conditions under which an intervention may be successful. These conditions may relate to particular characteristics of the participant (or whatever the unit of analysis happens to be) or to different

situations. In other words, optimizing an intervention means establishing its generality.

In the context of single-case research, generality can be demonstrated experimentally in several ways. The most basic way is via direct replication [26]. Direct replication means conducting the same experiment on the same behavioral problem across several individuals (i.e., a single-case experiment). For example, Raiff and Dallery [72] achieved a direct replication of the effects of internet-based contingency management (CM) on adherence to glucose testing in four adolescents. One goal of the study was to establish experimental control by the intervention and to minimize as many extraneous factors as possible. Overall, direct replication can help establish generality across participants. It cannot answer questions about generality across settings, behavior change agents, target behaviors, or participants that differ in some way from the original experiment (e.g., to adults diagnosed with type 1 diabetes). Instead, systematic replication can answer these questions. In a systematic replication, the methods from previous direct replication studies are used in a new setting, target behavior, group of participants, and so on [73]. The Raiff and Dallery study, therefore, was also a systematic replication of effects of internet-based CM to promote smoking cessation to a new problem and to a new group of participants because the procedure had originally been tested with adult smokers [24]. Effects of internet-based CM for smoking cessation also were systematically replicated in an application to adolescent smokers using a single-case design [74].

Systematic replication also occurs with parametric manipulation [63]. In other words, rather than changing the type of participants or setting, we change the value of the independent variable. In addition to demonstrating an optimal effect, parametric analysis may also reveal boundary conditions. These may be conditions under which an intervention no longer has an effect, or points of diminishing returns in which further increases in some parameter produce no further increases in efficacy. For example, if one study was conducted



**Fig 3 |** Two examples of possible results from a component analysis. *BSL* baseline, *X* first component, *Y* second component. The panel on the left shows an additive effect of components X and Y, and the panel of the right shows a multiplicative effect of components X and Y

showing that 30 min of moderate exercise produced a decrease in cigarette cravings, a systematic replication, using parametric analysis, might be conducted to determine the effects of other exercise durations (e.g., 5, 30, 60 min) on cigarette craving to identify the boundary parameters (i.e., the minimum and maximum number of minutes of exercise needed to continue to see changes in cigarette craving). Boundary conditions are critical in establishing generality of an intervention. In most cases, the only way to assess boundary conditions is through experimental, parametric analysis of an individual's behavior.

By carefully choosing the characteristics of the individuals, settings, or other relevant variables in a systematic replication, the researcher can help identify the conditions under which a treatment works. To be sure, as with any new treatment, failures will occur. However, the failure does not detract from the prior successes: "...a procedure can be quite valuable even though it is effective under a narrow range of conditions, as long as we know what those conditions are" [75]. Such information is important for treatment recommendations in a clinical setting, and scientifically, it means that the conditions themselves may become the subject of experimental analysis.

This discussion leads to a type of generality called scientific generality [63], which is at the heart of a scientific understanding of behavioral health interventions (or any intervention for that matter). As described by Branch and Pennypacker [63], scientific generality is characterized by knowledgeable reproducibility, or knowledge of the factors that are required for a phenomenon to occur. Scientific generality can be attained through parametric and component analysis, and through systematic replication. One advantage of a single-case approach to establishing generality is that a series of strategic studies can be conducted with some degree of efficiency. Moreover, the data intimacy afforded by SCDs can help achieve scientific generality about behavioral health interventions.

#### PERSONALIZED BEHAVIORAL MEDICINE

Personalized behavioral medicine involves three steps: assessing diagnostic, demographic, and other variables that may influence treatment outcomes; assigning an individual to treatment based on this information; and using SCDs to assess and tailor treatment. The first and second steps may be informed by outcomes using SCDs. In addition, the clinician may be in a better position to personalize treatment with knowledge derived from a body of SCD research about generality, boundary conditions, and the factors that are necessary for an effect to occur. (Of course, this information can come from a variety of sources—we are simply highlighting how SCDs may fit in to this process.)

In addition, with advances in genomics and technology-enabled behavioral assessment prior to treatment (i.e., a baseline phase), the clinician may further target treatment to the unique characteristics of the individual [76]. Genetic testing is becoming more common before prescribing various medications [17], and it may become useful to predict responses for treatments targeting health behavior. Baseline assessment of behavior using technology such as EMA may allow the clinician to develop a tailored treatment protocol. For example, assessment could reveal the temporal patterning of risky situations, such as drinking alcohol, having an argument, or long periods of inactivity. A text-based support system could be tailored such that the timings of texts are tied to the temporal pattern of the problem behavior. The baseline assessment may also be useful to simply establish whether a problem exists. Also, the data path during baseline may reveal that behavior or symptoms are already improving prior to treatment, which would suggest that other, non-treatment variables are influencing behavior. Perhaps more importantly, compared to self-report, baseline conditions provide a more objective benchmark to assess effects of treatment on behavior and symptoms.

In addition to greater personalization at the start of treatment, ongoing assessment and treatment tailoring can be achieved with SCDs. Hayes [77] described how parametric and component analyses can be conducted in clinical practice. For example, reversal designs could be used to conduct a component analysis. Two components, or even different treatments, could be systematically introduced alone and together. If the treatments are different, such comparisons would also yield a kind of comparative effectiveness analysis. For example, contingency contracting and pharmacotherapy for smoking cessation could be presented alone using a BCBC design (where B is contracting and C is pharmacotherapy). A combined treatment could also be added, and depending on results, a return to one or the other treatment could follow (e.g., BCDCB, where D is the combined treatment). Furthermore, if a new treatment becomes available, it could be tested relative to an existing standard treatment in the same fashion. One potential limitation of such designs is when a reversal to baseline conditions (i.e., no treatment) is necessary to document treatment effects. Such a return to baseline may be challenging for ethical, reimbursement, and other issues.

Multiple-baseline designs also can be used in clinical contexts. Perhaps the simplest example would be a multiple baseline across individuals with similar problems. Each individual would experience an AB sequence, where the durations of the baseline phases vary. Another possibility is to target different behavior in the same individual in a multiple-baseline across behavior design. For example, a skills training program to improve social behavior

could target different aspects of such behavior in a sequential fashion, starting with eye contact, then posture, then speech volume, and so on. If behavior occurs in a variety of distinct settings, the treatment could be sequentially implemented across these settings. Using the same example, treatment could target social behavior at family events, work, and different social settings. It can be problematic if generalization of effects occurs, but it may not necessarily negate the utility of such a design [27].

Multiple-baseline designs can be used in contexts other than outpatient therapy. Biglan and associates [1] argued that such designs are particularly useful in community interventions. For example, they described how a multiple baseline across communities and even states could be used to assess effects of changes in drinking age on car crashes. These designs may be especially useful to evaluate technology-based health interventions. A web-based program could be sequentially rolled out to different schools, communities, or other clusters of individuals. Although these research designs are also referred to as interrupted time series and stepped wedge designs, we think it may be more likely for researchers and clinicians to access the rich network of resources, concepts, and analytic tools if these designs are subsumed under the category of multiple-baseline designs.

The systematic comparisons afforded by SCDs can answer several key questions relevant to optimization. The first question a clinician may have is whether a particular intervention will work for his or her client [27]. It may be that the client has such a unique history and profile of symptoms, the clinician may not be confident about the predictive validity of a particular intervention for his or her client [6]. SCDs can be used to answer this question. Also, as just described, they can address which of two treatments work better, whether adding two treatments (or components) together works better than either one alone, which level of treatment is optimal (i.e., a parametric analysis), and whether a client prefers one treatment over another (i.e., via social validity assessment). Furthermore, the use of SCDs in practice conforms to the scientist-practitioner ideal espoused by training models in clinical psychology and allied disciplines [78].

#### OPTIMIZING FROM DEVELOPMENT TO DISSEMINATION

We are now in a position to evaluate whether SCDs live up to our ideals about optimization. During development, SCDs may obviate some logistical issues in using between-group designs to conduct initial efficacy testing [3, 8]. Specifically, the costs and duration needed to conduct a SCD to establish preliminary efficacy would be considerably lower than traditional randomized designs. Riley and colleagues [8] noted that randomized trials take approximately 5.5 years from the initiation of enrollment to publication, and even longer from

the time a grant application is submitted. In addition to establishing whether a treatment works, SCDs have the flexibility to efficiently address which parameters and components are necessary or optimal. In light of traditional methods to establish preliminary efficacy and optimize treatments, Riley and colleagues advocated for “rapid learning research systems.” SCDs are one such system.

Although some logistical issues may be mitigated by using SCDs, they do not necessarily represent easy alternatives to traditional group designs. They require a considerable amount of data per participant (as opposed to a large number of individuals in a group), enough participants to reliably demonstrate experimental effects, and systematic manipulation of variables over a long duration. For the vast majority of research questions, however, SCDs can reduce the resource and time burdens associated with between group designs and allow the investigator to detect important treatment parameters that might otherwise have been missed.

SCDs can minimize or eliminate a number of threats to internal validity. Although a complete discussion of these threats is beyond the scope of this paper (see [1, 27, 28]), the standards listed in Table 1 can provide protection against most threats. For example, the threat known as “testing” refers to the fact that repeated measurement alone may change behavior. To address this, baseline phases need to be sufficiently long, and there must be enough within and/or between participant replications to rule out the effect of testing. Such logic applies to a number of other potential threats (e.g., instrumentation, history, regression to the mean, etc.). In addition, a plethora of new analytic techniques can supplement experimental techniques to make inferences about causal relations. Combining SCD results in meta-analyses can yield information about comparative effects of different treatments, and combining results using Bayesian methods may yield information about likely effects at the population level.

Because of their efficiency and rigor, SCDs permit systematic replications across types of participants, behavior problems, and settings. This research process has also led to “gold-standard,” evidence-based treatments in applied behavior analysis and education [29, 79]. More importantly, in several fields, such research has led to scientific understanding of the conditions under which treatment may be effective or ineffective [79, 80]. The field of applied behavior analysis, for example, has matured to the extent that individualized assessment of the causes of problem behavior must occur before treatment recommendations.

Our discussion of personalized behavioral medicine highlighted how SCDs can be used in clinical practice to evaluate and optimize interventions. The advent of technology-based assessment makes SCDs much easier to implement. Technology could propel a “super convergence” of SCDs and clinical practice



[76]. Advances in technology-based assessment can also promote the kind of systematic play central to the experimental attitude. It can also allow testing of new interventions as they become available. Such translational efforts can occur in several ways: from laboratory and other controlled settings to clinical practice, from SCD to SCD within clinical practice, and from randomized efficacy trials to clinical practice.

**CONCLUSION**

Over the past 70 years, SCD research has evolved to include a broad array of methodological and analytic advances. It also has generated evidence-based practices in health care and related disciplines such as clinical psychology [81], substance abuse [82, 83], education [29], medicine [4], neuropsychology [30], developmental disabilities [27], and occupational therapy [84]. Although different methods are required for different purposes, SCDs are ideally suited to optimize interventions, from development to dissemination.

**Acknowledgments:** We wish to thank Paul Soto for comments on a previous draft of this manuscript. Preparation of this paper was supported in part by Grants P30DA029926 and R01DA023469 from the National Institute on Drug Abuse.

**Conflict of interest:** The authors have no conflicts of interest to disclose.

1. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prev Sci.* 2000; 1(1): 31–49.
2. Collins LM, Murphy SA, Nair VN, et al. A strategy for optimizing and evaluating behavioral interventions. *Ann Behav Med.* 2005; 30(1): 65–73. PMID: 16009093.
3. Dallery J, Cassidy RN, Raiff BR. Single-case experimental designs to evaluate novel technology-based health interventions. *J Med Internet Res.* 2013; 15(2). doi:10.2196/jmir.2227.
4. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' guides 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(10): 1290–1296. PMID: 10979117.
5. Guyatt GH, Heyting A, Jaeschke R, et al. N of 1 randomized trials for investigating new drugs. *Control Clin Trials.* 1990; 11(2): 88–100. PMID: 2161315.
6. Williams BA. Perils of evidence-based medicine. *Perspect Biol Med.* 2010; 53(1): 106–120. PMID: 20173299.
7. Mercer SL, DeVinney BJ, Fine LJ, et al. Study designs for effectiveness and translation research: identifying trade-offs. *Am J Prev Med.* 2007; 33(2): 139–154. PMID: 17673103.
8. Riley WT, Glasgow RE, Etheredge L, et al. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. *Clin Transl Med.* 2013; 2(1): 10-1326-2-10. PMID: 23663660.
9. Rivera DE, Pew MD, Collins LM. Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction. *Drug Alcohol Depend.* 2007; 88(Suppl 2): S31–S40. PMID: 16009093.
10. Kumar S, Nilsen WJ, Abernethy A, et al. Mobile health technology evaluation: the mhealth evidence workshop. *Am J Prev Med.* 2013; 45(2): 228–236. PMID: 23867031.
11. Riley WT, Rivera DE, Atienza AA, et al. Health behavior models in the age of mobile interventions: are our theories up to the task? *Transl Behav Med.* 2011; 1(1): 53–71. PMID: 21796270.
12. Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: getting started and moving on from stage I. *Clin Psychol: Sci Pract.* 2001; 8(2): 133–142.
13. Cohen J. The earth is round ( $p < .05$ ). *Am Psychol.* 1994; 49(12): 997–1003.
14. Morgan DL, Morgan RK. Single-participant research design: bringing science to managed care. *Am Psychol.* 2001; 56(2): 119–127.

15. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Ann Rev Clin Psychol.* 2008; 4: 1–32.
16. Marsch LA, Dallery J. Advances in the psychosocial treatment of addiction: the role of technology in the delivery of evidence-based psychosocial treatment. *Psychiatr Clin North Am.* 2012; 35(2): 481–493.
17. Lillie EO, Patay B, Diamant J, et al. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Per Med.* 2011; 8(2): 161–173. PMID: 21695041.
18. Goodwin MS, Velicer WF, Intille SS. Telemetric monitoring in the behavior sciences. *Behav Res Methods.* 2008; 40(1): 328–341. PMID: 18411557.
19. Dallery J, Raiff BR. Contingency management in the 21st century: technological innovations to promote smoking cessation. *Subst Use Misuse.* 2011; 46(1): 10–22.
20. Aeschleman SR. Single-subject research designs: some misconceptions. *Rehabil Psychol.* 1991; 36(1): 43–49.
21. Dermer ML, Hoch TA. Improving descriptions of single-subject experiments in research texts written for undergraduates. *Psychol Rec.* 1999; 49(1): 49–66.
22. Dixon MR. Single-subject research designs: dissolving the myths and demonstrating the utility for rehabilitation research. *Rehabil Educ.* 2002; 16(4): 331–343.
23. Kravitz RL, Paterniti DA, Hay MC, et al. Marketing therapeutic precision: potential facilitators and barriers to adoption of n-of-1 trials. *Contemp Clin Trials.* 2009; 30(5): 436–445. PMID: 19375521.
24. Dallery J, Glenn IM, Raiff BR. An internet-based abstinence reinforcement treatment for cigarette smoking. *Drug Alcohol Depend.* 2007; 86(2–3): 230–238. PMID: 16930854.
25. Silverman K, Higgins ST, Brooner RK, et al. Sustained cocaine abstinence in methadone maintenance patients through voucher-based reinforcement therapy. *Arch Gen Psychiatry.* 1996; 53(5): 409–415.
26. Sidman M. *Tactics of Scientific Research.* Oxford: Basic Books; 1960.
27. Kazdin AE. *Single-Case Research Designs: Methods for Clinical and Applied Settings.* 2nd ed. New York: Oxford University Press; 2011. ISBN 978-0-19-534188-1.
28. Barlow DH, Nock MK, Hersen M. *Single Case Experimental Designs: Strategies for Studying Behavior Change.* 3rd ed. Boston: Allyn & Bacon; 2009.
29. Horner RH, Carr EG, Halle J, et al. The use of single-subject research to identify evidence-based practice in special education. *Except Child.* 2005; 71(2): 165–179.
30. Tate RL, McDonald S, Perdices M, et al. Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the single-case experimental design (SCED) scale. *Neuropsychol Rehab.* 2008; 18(4): 385–401.
31. Kratochwill TR, Hitchcock JH, Horner RH, et al. Single-case intervention research design standards. *Remedial Spec Educ.* 2013; 34(1): 26–38.
32. Risley TR, Wolf MM. Strategies for analyzing behavioral change over time. In: Nesselroade J, Reese H, eds. *Life-Span Developmental Psychology: Methodological Issues.* New York: Academic; 1972: 175.
33. Barlow DH, Hersen M. Single-case experimental designs: uses in applied clinical research. *Arch Gen Psychiatry.* 1973; 29(3): 319–325.
34. Hartmann D, Hall RV. The changing criterion design. *J Appl Behav Anal.* 1976; 9(4): 527–532.
35. Kurti AN, Dallery J. Internet-based contingency management increases walking in sedentary adults. *J Appl Behav Anal.* 2013; 46(3): 568–581.
36. Francisco VT, Butterfoss FD. Social validation of goals, procedures, and effects in public health. *Health Promot Pract.* 2007; 8(2): 128–133.
37. Foster SL, Mash EJ. Assessing social validity in clinical treatment research: issues and procedures. *J Consult Clin Psychol.* 1999; 67(3): 308–319.
38. Schwartz IS, Baer DM. Social validity assessments: is current practice state of the art? *J Appl Behav Anal.* 1991; 24(2): 189–204.
39. Wolf MM. Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart. *J Appl Behav Anal.* 1978; 11(2): 203–214.
40. Parsonson BS, Baer DM. The visual analysis of data, and current research into the stimuli controlling it. In: *Single-Case Research Design and Analysis: New Directions for Psychology and Education.* Hillsdale: Lawrence Erlbaum Associates, Inc; 1992: 15–40.
41. Hubbard R, Lindsay RM. Why P values are not a useful measure of evidence in statistical significance testing. *Theory Psychol.* 2008; 18(1): 69–88.
42. Lambdin C. Significance tests as sorcery: science is empirical—significance tests are not. *Theory Psychol.* 2012; 22(1): 67–90.



43. Fisher WW, Kelley ME, Lomas JE. Visual aids and structured criteria for improving inspection and interpretation of single-case designs. *J Appl Behav Anal.* 2003; 36(3): 387–406.
44. Fisher JD, Fisher WA, Amico KR, et al. An information-motivation-behavioral skills model of adherence to antiretroviral therapy. *Health Psychol.* 2006; 25(4): 462–473.
45. Kahng S, Chung K, Gutshall K, et al. Consistent visual analyses of intrasubject data. *J Appl Behav Anal.* 2010; 43(1): 35–45.
46. Van D, Onghena P. The aggregation of single-case results using hierarchical linear models. *Behav Analyst Today.* 2007; 8(2): 196–209.
47. Gorman BS, Allison DB, et al. Statistical alternatives for single-case designs. In: Franklin RD, Allison DB, Gorman BS, eds. *Design and Analysis of Single-Case Research.* Hillsdale: Lawrence Erlbaum Associates, Inc; 1996: 159–214.
48. Jenson WR, Clark E, Kircher JC, et al. Statistical reform: evidence-based practice, meta-analyses, and single subject designs. *Psychol Sch.* 2007; 44(5): 483–493.
49. Ben-Zeev D, Ellington K, Swendsen J, et al. Examining a cognitive model of persecutory ideation in the daily life of people with schizophrenia: a computerized experience sampling study. *Schizophr Bull.* 2011; 37(6): 1248–1256.
50. Huitema BE. Autocorrelation in applied behavior analysis: a myth. *Behav Assess.* 1985; 7(2): 107–118.
51. Wilkinson L. Statistical methods in psychology journals: guidelines and explanations. *Am Psychol.* 1999; 54(8): 594–604.
52. Manolov R, Solanas A, Sierra V, et al. Choosing among techniques for quantifying single-case intervention effectiveness. *Behav Ther.* 2011; 42(3): 533–545.
53. Parker RI, Vannest KJ, Davis JL. Effect size in single-case research: a review of nine nonoverlap techniques. *Behav Modif.* 2011; 35(4): 303–322.
54. Mason LL. An analysis of effect sizes for single-subject research: a statistical comparison of five judgmental aids. *J Precis Teach Celeration.* 2010; 26: 3–16.
55. Duan N, Kravitz RL, Schmid CH. Single-patient (*n*-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. *J Clin Epidemiol.* 2013; 66(8 Suppl): S21–S28. PMID: 23849149.
56. Hedges LV, Pustejovsky JE, Shadish WR. A standardized mean difference effect size for single case designs. *Res Synth Methods.* 2012; 3(3): 224–239.
57. Shadish WR, Hedges LV, Pustejovsky JE, et al. A *d*-statistic for single-case designs that is equivalent to the usual between-groups *d*-statistic. *Neuropsychol Rehabil.* 2013 doi:10.1080/09602011.2013.819021.
58. Wang S, Cui Y, Parrila R. Examining the effectiveness of peer-mediated and video-modeling social skills interventions for children with autism spectrum disorders: a meta-analysis in single-case research using HLM. *Res Autism Spectr Disord.* 2011; 5(1): 562–569.
59. Davis JL, Vannest KJ. Effect size for single case research a replication and re-analysis of an existing meta-analysis. *Rem Spec Educ.* In press.
60. Ganz JB, Parker R, Benson J. Impact of the picture exchange communication system: effects on communication and collateral effects on maladaptive behaviors. *Augment Altern Commun.* 2009; 25(4): 250–261.
61. Vannest KJ, Davis JL, Davis CR, et al. Effective intervention for behavior with a daily behavior report card: a meta-analysis. *Sch Psychol Rev.* 2010; 39(4): 654–672.
62. Zucker DR, Ruthazer R, Schmid CH. Individual (*N*-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *J Clin Epidemiol.* 2010; 63(12): 1312–1323. PMID: 20863658.
63. Branch MN, Madden GJ, Hackenberg T. Generality and Generalization of Research Findings. In: Madden GJ, Hackenberg T, Lattal KA, eds. *APA Handbook of Behavior Analysis.* Washington: American Psychological Association; 2011.
64. Jaeschke R, Singer J, Guyatt GH. Measurement of health status ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989; 10(4): 407–415. PMID: 2691207.
65. Wright A, Hannon J, Hegedus EJ, et al. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther.* 2012; 20(3): 160–166. PMID: 23904756.
66. Kaplan RM, Stone AA. Bringing the laboratory and clinic to the community: mobile technologies for health promotion and disease prevention. *Annu Rev Psychol.* 2013; 64: 471–498.
67. Gulley V, Northup J, Hupp S, et al. Sequential evaluation of behavioral treatments and methylphenidate dosage for children with attention deficit hyperactivity disorder. *J Appl Behav Anal.* 2003; 36(3): 375–378. PMID: 14596580.
68. Rapport MD, Murphy HA, Bailey JS. Ritalin vs. response cost in the control of hyperactive children: a within-subject comparison. *J Appl Behav Anal.* 1982; 15(2): 205–216.
69. Ward-Horner J, Sturmey P. Component analyses using single-subject experimental designs: a review. *J Appl Behav Anal.* 2010; 43(4): 685–704.
70. Sanders SH. Component analysis of a behavioral treatment program for chronic low-back pain. *Behav Ther.* 1983; 14(5): 697–705.
71. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent health interventions. *Am J Prev Med.* 2007; 32(5): S112–S118.
72. Raiff BR, Dallery J. Internet-based contingency management to improve adherence with blood glucose testing recommendations for teens with type 1 diabetes. *J Appl Behav Anal.* 2010; 43(3): 487–491.
73. Valentine JC, Biglan A, Boruch RF, et al. Replication in prevention science. *Prev Sci.* 2011; 12(2): 103–117. PMID: 17210552.
74. Reynolds B, Dallery J, Shroff P, et al. A web-based contingency management program with adolescent smokers. *J Appl Behav Anal.* 2008; 41(4): 597–601.
75. Johnston JM, Pennypacker Jr. HS. *Strategies and Tactics of Behavioral Research (3rd ed).* New York: Routledge/Taylor & Francis Group; 2009.
76. Topol EJ. *The Creative Destruction of Medicine How the Digital Revolution Will Create Better Health Care.* New York: Basic Books; 2012.
77. Hayes SC. Single case experimental design and empirical clinical practice. *J Consult Clin Psychol.* 1981; 49(2): 193–211.
78. Blampied NM. In: Madden GJ, Dube WV, Hackenberg TD, eds. *Single-Case Research Designs and the Scientist-Practitioner Ideal in Applied Psychology.* Washington: American Psychological Association; 2013: 177–197.
79. Cooper JO, Heron TE, Heward WL. *Applied Behavior Analysis. Upper Saddle River.* NJ: Prentice Hall; 2007.
80. Hanley GP. Functional assessment of problem behavior: dispelling myths, overcoming implementation obstacles, and developing new lore. *Behav Anal Pract.* 2012; 5(1): 54–72.
81. Parker RI, Brossart DF. Phase contrasts for multiphase single case intervention designs. *Sch Psychol Q.* 2006; 21(1): 46–61.
82. Silverman K, Wong CJ, Higgins ST, et al. Increasing opiate abstinence through voucher-based reinforcement therapy. *Drug Alcohol Depend.* 1996; 41(2): 157–165.
83. Dallery J, Raiff BR. Delay discounting predicts cigarette smoking in a laboratory model of abstinence reinforcement. *Psychopharmacology.* 2007; 190(4): 485–496.
84. Johnston MV, Smith RO. Single subject designs: current methodologies and future directions. *OTJR: Occupation, Particip Health.* 2010; 30(1): 4–10.