

# Single-Case Design, Analysis, and Quality Assessment for Intervention Research

Michele A. Lobo, PT, PhD, Mariola Moeyaert, PhD, Andrea Baraldi Cunha, PT, PhD,  
and Iryna Babik, PhD

**Background and Purpose:** The purpose of this article is to describe single-case studies and contrast them with case studies and randomized clinical trials. We highlight current research designs, analysis techniques, and quality appraisal tools relevant for single-case rehabilitation research.

**Summary of Key Points:** Single-case studies can provide a viable alternative to large group studies such as randomized clinical trials. Single-case studies involve repeated measures and manipulation of an independent variable. They can be designed to have strong internal validity for assessing causal relationships between interventions and outcomes, as well as external validity for generalizability of results, particularly when the study designs incorporate replication, randomization, and multiple participants. Single-case studies should not be confused with case studies/series (ie, case reports), which are reports of clinical management of a patient or a small series of patients.

**Recommendations for Clinical Practice:** When rigorously designed, single-case studies can be particularly useful experimental designs in a variety of situations, such as when research resources are limited, studied conditions have low incidences, or when examining effects of novel or expensive interventions. Readers will be directed to examples from the published literature in which these techniques have been discussed, evaluated for quality, and implemented.

**Key words:** *n-of-1 studies, quality assessment, research design, single-subject research*

(JNPT 2017;41: 187–197)

Biomechanics & Movement Science Program, Department of Physical Therapy, University of Delaware, Newark, Delaware (M.A.L., A.B.C., I.B.); and Division of Educational Psychology & Methodology, State University of New York at Albany, Albany, New York (M.M.).

This research was supported by the National Institute of Health, Eunice Kennedy Shriver National Institute of Child Health & Human Development (1R21HD076092-01A1, Lobo PI), and the Delaware Economic Development Office (Grant #109).

Some of the information in this article was presented at the IV Step Meeting in Columbus, Ohio, June 2016.

The authors declare no conflict of interest.

**Correspondence:** Michele A. Lobo, PT, PhD, Biomechanics & Movement Science Program, Department of Physical Therapy, University of Delaware, Newark, DE 19713 (malobo@udel.edu).

Copyright © 2017 Academy of Neurologic Physical Therapy, APTA.

ISSN: 1557-0576/17/4103-0187

DOI: 10.1097/NPT.0000000000000187

## INTRODUCTION

In this special interest article we present current tools and techniques relevant for single-case rehabilitation research. Single-case (SC) studies have been identified by a variety of names, including “n of 1 studies” and “single-subject” studies. The term “single-case study” is preferred over the previously mentioned terms because previous terms suggest these studies include only 1 participant. In fact, as discussed later, for purposes of replication and improved generalizability, the strongest SC studies commonly include more than 1 participant.

A SC study should not be confused with a “case study/series” (also called “case report”). In a typical case study/series, a single patient or small series of patients is involved, but there is not a purposeful manipulation of an independent variable, nor are there necessarily repeated measures. Most case studies/series are reported in a narrative way, whereas results of SC studies are presented numerically or graphically.<sup>1,2</sup> This article defines SC studies, contrasts them with randomized clinical trials, discusses how they can be used to scientifically test hypotheses, and highlights current research designs, analysis techniques, and quality appraisal tools that may be useful for rehabilitation researchers.

In SC studies, measurements of outcome (dependent variables) are recorded repeatedly for individual participants across time and varying levels of an intervention (independent variables).<sup>1–5</sup> These varying levels of intervention are referred to as “phases,” with 1 phase serving as a baseline or comparison, so each participant serves as his/her own control.<sup>2</sup> In contrast to case studies and case series in which participants are observed across time without experimental manipulation of the independent variable, SC studies employ systematic manipulation of the independent variable to allow for hypothesis testing.<sup>1,6</sup> As a result, SC studies allow for rigorous experimental evaluation of intervention effects and provide a strong basis for establishing causal inferences. Advances in design and analysis techniques for SC studies observed in recent decades have made SC studies increasingly popular in educational and psychological research. Yet, the authors believe SC studies have been undervalued in rehabilitation research, where randomized clinical trials (RCTs) are typically recommended as the optimal research design to answer questions related to interventions.<sup>7</sup> In reality, there are advantages and disadvantages to both SC studies and RCTs that should be carefully considered to select the best design to answer

individual research questions. Although there are a variety of other research designs that could be utilized in rehabilitation research, only SC studies and RCTs are discussed here because SC studies are the focus of this article and RCTs are the most highly recommended design for intervention studies.<sup>7</sup>

When designed and conducted properly, RCTs offer strong evidence that changes in outcomes may be related to provision of an intervention. However, RCTs require monetary, time, and personnel resources that many researchers, especially those in clinical settings, may not have available.<sup>8</sup> RCTs also require access to large numbers of consenting participants who meet strict inclusion and exclusion criteria that can limit variability of the sample and generalizability of results.<sup>9</sup> The requirement for large participant numbers may make RCTs difficult to perform in many settings, such as rural and suburban settings, and for many populations, such as those with diagnoses marked by lower prevalence.<sup>8</sup> To rely exclusively on RCTs has the potential to result in bodies of research that are skewed to address the needs of some individuals whereas neglecting the needs of others. RCTs aim to include a large number of participants and to use random group assignment to create study groups that are similar to one another in terms of all potential confounding variables, but it is challenging to identify all confounding variables. Finally, the results of RCTs are typically presented in terms of group means and standard deviations that may not represent true performance of any one participant.<sup>10</sup> This can present as a challenge for clinicians aiming to translate and implement these group findings at the level of the individual.

SC studies can provide a scientifically rigorous alternative to RCTs for experimentally determining the effectiveness of interventions.<sup>1,2</sup> SC studies can assess a variety of research questions, settings, cases, independent variables, and outcomes.<sup>11</sup> There are many benefits to SC studies that make them appealing for intervention research. SC studies may require fewer resources than RCTs and can be performed in settings and with populations that do not allow for large numbers of participants.<sup>1,2</sup> In SC studies, each participant serves as his/her own comparison, thus controlling for many confounding variables that can impact outcome in rehabilitation research, such as gender, age, socioeconomic level, cognition, home environment, and concurrent interventions.<sup>2,11</sup> Results can be analyzed and presented to determine whether interventions resulted in changes at the level of the individual, the level at which rehabilitation professionals intervene.<sup>2,12</sup> When properly designed and executed, SC studies can demonstrate strong internal validity to determine the likelihood of a causal relationship between the intervention and outcomes and external validity to generalize the findings to broader settings and populations.<sup>2,12,13</sup>

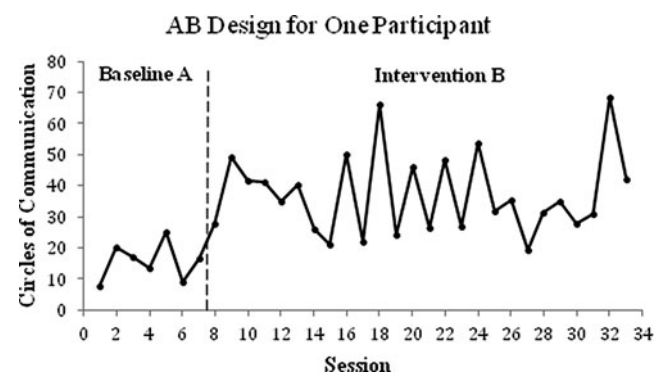
## SINGLE-CASE RESEARCH DESIGNS FOR INTERVENTION RESEARCH

There are a variety of SC designs that can be used to study the effectiveness of interventions. Here we discuss (1) AB designs, (2) reversal designs, (3) multiple baseline designs, and (4) alternating treatment designs, as well as ways repli-

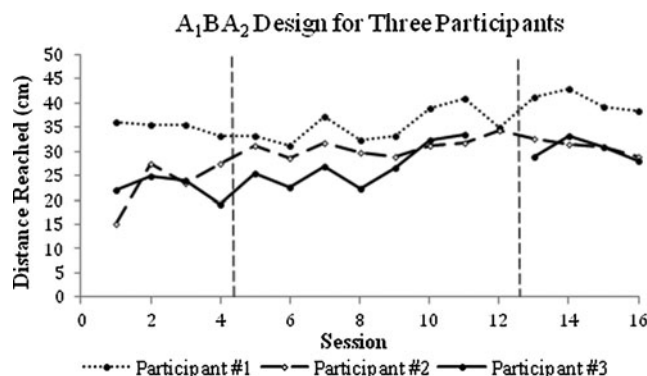
cation and randomization techniques can be used to improve internal validity of all of these designs.<sup>1-3,12-14</sup>

The simplest of these designs is the *AB design*<sup>15</sup> (Figure 1). This design involves repeated measurement of outcome variables throughout a baseline control/comparison phase (A) and then throughout an intervention phase (B). When possible, it is recommended that a stable level and/or rate of change in performance be observed within the baseline phase before transitioning into the intervention phase.<sup>2</sup> As with all SC designs, it is also recommended that there be a minimum of 5 data points in each phase.<sup>1,2</sup> There is no randomization or replication of the baseline or intervention phases in the basic AB design.<sup>2</sup> Therefore, AB designs have problems with internal validity and generalizability of results.<sup>12</sup> They are weak in establishing causality because changes in outcome variables could be related to a variety of other factors, including maturation, experience, learning, and practice effects.<sup>2,12</sup> Sample data from a single-case AB study performed to assess the impact of Floor Play intervention on social interaction and communication skills for a child with autism<sup>15</sup> are shown in Figure 1.

If an intervention does not have carryover effects, it is recommended to use a *reversal design*.<sup>2</sup> For example, a reversal  $A_1BA_2$  design<sup>16</sup> (Figure 2) includes alternation of the baseline and intervention phases, whereas a reversal  $A_1B_1A_2B_2$  design<sup>17</sup> (Figure 3) consists of alternation of 2 baseline ( $A_1, A_2$ ) and 2 intervention ( $B_1, B_2$ ) phases. Incorporating at least 4 phases in the reversal design (ie,  $A_1B_1A_2B_2$  or  $A_1B_1A_2B_2A_3B_3 \dots$ ) allows for a stronger determination of a causal relationship between the intervention and outcome



**Figure 1.** An example of results from a single-case AB study conducted on 1 participant with autism; 2 weeks of observation (baseline phase A) was followed by 7 weeks of Floor Time Play (intervention phase B). The outcome measure Circles of Communications (reciprocal communication with 2 participants responding to each other verbally or nonverbally) served as a behavioral indicator of the child's social interaction and communication skills (higher scores indicating better performance). A statistically significant improvement in Circles of Communication was found during the intervention phase as compared with the baseline. Note that although a stable baseline is recommended for single-case studies, it is not always possible to satisfy this requirement, as shown in Figures 1 to 4. Data were extracted from Dionne and Martini<sup>15</sup> utilizing Rohatgi's WebPlotDigitizer software.<sup>78</sup>

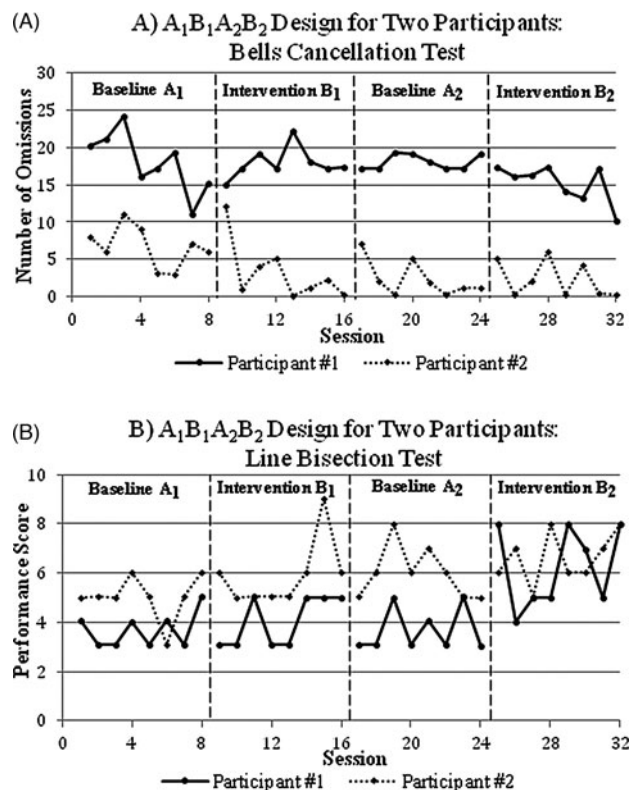


**Figure 2.** An example of results from a single-case  $A_1BA_2$  study conducted on 8 participants with stable multiple sclerosis (data on 3 participants were used for this example). Four weeks of observation (baseline phase  $A_1$ ) was followed by 8 weeks of core stability training (intervention phase B), and then another 4 weeks of observation (baseline phase  $A_2$ ). Forward functional reach test (the maximal distance the participant can reach forward or lateral beyond arm's length, maintaining a fixed base of support in the standing position; higher scores indicating better performance) significantly improved during intervention for participants 1 and 3 without further improvement observed following withdrawal of the intervention (during baseline phase  $A_2$ ). Data were extracted from Freeman et al<sup>16</sup> utilizing Rohatgi's WebPlotDigitizer software.<sup>78</sup>

variables because the relationship can be demonstrated across at least 3 different points in time—change in outcome from  $A_1$  to  $B_1$ , from  $B_1$  to  $A_2$ , and from  $A_2$  to  $B_2$ .<sup>18</sup> Before using this design, however, researchers must determine that it is safe and ethical to withdraw the intervention, especially in cases where the intervention is effective and necessary.<sup>12</sup>

A recent study used an ABA reversal SC study to determine the effectiveness of core stability training in 8 participants with multiple sclerosis.<sup>16</sup> During the first 4 weekly data collections, the researchers ensured a stable baseline, which was followed by 8 weekly intervention data points, and concluded with 4 weekly withdrawal data points. Intervention significantly improved participants' walking and reaching performance (Figure 2).<sup>16</sup> This  $A_1BA_2$  design could have been strengthened by the addition of a second intervention phase for replication ( $A_1B_1A_2B_2$ ). For instance, a single-case  $A_1B_1A_2B_2$  withdrawal design aimed to assess the efficacy of rehabilitation using visuo-spatio-motor cueing for 2 participants with severe unilateral neglect after a severe right hemisphere stroke.<sup>17</sup> Each phase included 8 data points. Statistically significant intervention-related improvement was observed, suggesting that visuo-spatio-motor cueing might be promising for treating individuals with very severe neglect (Figure 3).<sup>17</sup>

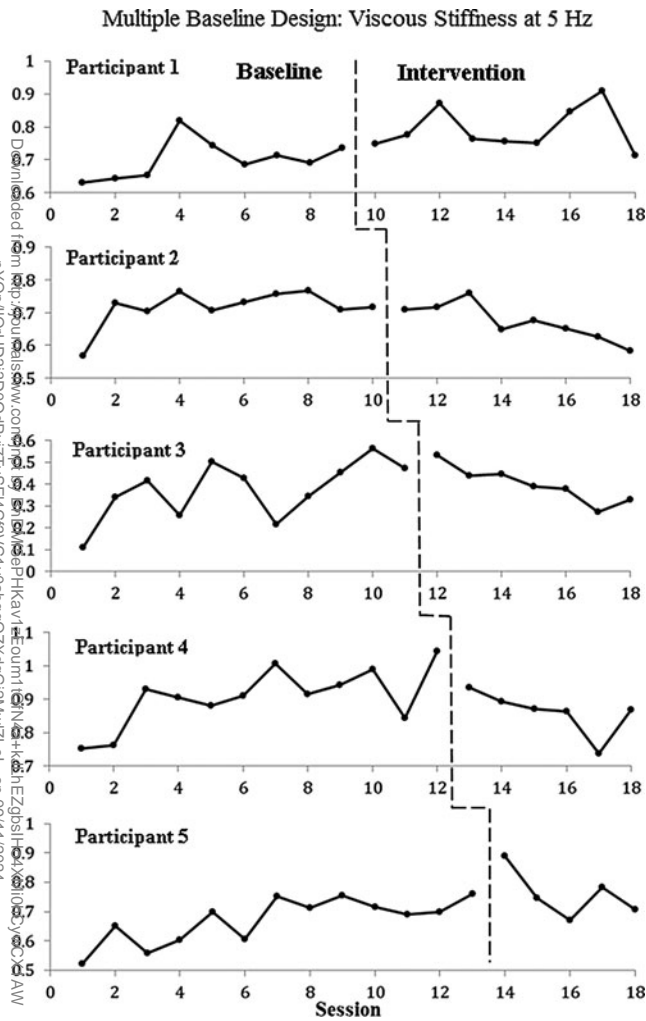
The reversal design can also incorporate a cross-over design where each participant experiences more than 1 type of intervention. For instance, a  $B_1C_1B_2C_2$  design could be used to study the effects of 2 different interventions (B and C) on outcome measures. Challenges with including more than 1 intervention involve potential carryover effects from earlier interventions and order effects that may impact the measured



**Figure 3.** An example of results from a single-case  $A_1B_1A_2B_2$  study conducted on 2 participants with severe unilateral neglect after a right hemisphere stroke. Two weeks of conventional treatment (baseline phases  $A_1$ ,  $A_2$ ) alternated with 2 weeks of visuo-spatio-motor cueing (intervention phases  $B_1$ ,  $B_2$ ). Performance was assessed in 2 tests of lateral neglect, the Bells Cancellation Test (Figure A; lower scores indicating better performance) and the Line Bisection Test (Figure B; higher scores indicating better performance). There was a statistically significant intervention-related improvement in participants' performance on the Line Bisection Test, but not on the Bells Test. Data were extracted from Samuel et al<sup>17</sup> utilizing Rohatgi's WebPlotDigitizer software.<sup>78</sup>

effectiveness of the interventions.<sup>2,12</sup> Including multiple participants and randomizing the order of intervention phase presentations are tools to help control for these types of effects.<sup>19</sup>

When an intervention permanently changes an individual's ability, a return-to-baseline performance is not feasible and reversal designs are not appropriate. *Multiple baseline designs (MBDs)* are useful in these situations (Figure 4).<sup>20</sup> Multiple baseline designs feature staggered introduction of the intervention across time: each participant is randomly assigned to 1 of at least 3 experimental conditions characterized by the length of the baseline phase.<sup>21</sup> These studies involve more than 1 participant, thus functioning as SC studies with replication across participants. Staggered introduction of the intervention allows for separation of intervention effects from those of maturation, experience, learning, and practice. For example, a multiple baseline SC study was used to investigate the effect of an antispasticity baclofen medication on stiffness

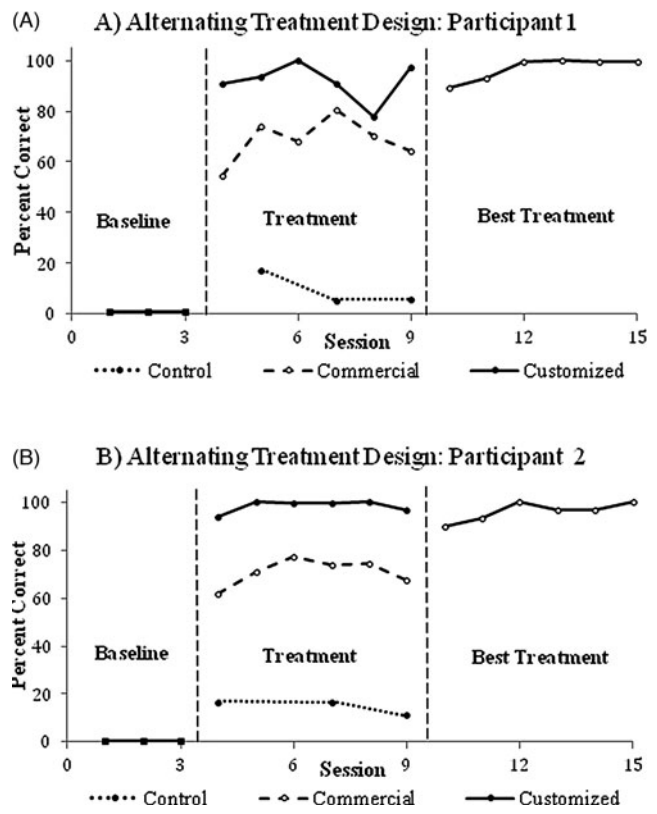


**Figure 4.** An example of results from a single-case multiple baseline study conducted on 5 participants with spasticity because of traumatic spinal cord injury. Total duration of data collection was 9 weeks. The first participant was switched from placebo treatment (baseline) to baclofen treatment (intervention) after 5 data collection sessions, whereas each consecutive participant was switched to baclofen intervention at the subsequent sessions through the ninth session. There was no statistically significant effect of baclofen on viscous stiffness at the ankle joint. Data were extracted from Hinderer et al<sup>20</sup> utilizing Rohatgi's WebPlotDigitizer software.<sup>78</sup>

in 5 adult males with spinal cord injury.<sup>20</sup> The subjects were randomly assigned to receive 5 to 9 baseline data points with a placebo treatment before the initiation of the intervention phase with the medication. Both participants and assessors were blind to the experimental condition. The results suggested that baclofen might not be a universal treatment choice for all individuals with spasticity resulting from a traumatic spinal cord injury (Figure 4).<sup>20</sup>

The impact of 2 or more interventions can also be assessed via *alternating treatment designs (ATDs)*. In ATDs, after establishing the baseline, the experimenter exposes subjects to different intervention conditions administered in close

proximity for equal intervals (Figure 5).<sup>22</sup> ATDs are prone to “carryover effects” when the effects of 1 intervention influence the observed outcomes of another intervention.<sup>1</sup> As a result, such designs introduce unique challenges when attempting to determine the effects of any 1 intervention and have been less commonly utilized in rehabilitation. An ATD was used to monitor disruptive behaviors in the school setting throughout a baseline followed by an alternating treatment phase with randomized presentation of a control condition or an exercise condition.<sup>23</sup> Results showed that 30 minutes of moderate to intense physical activity decreased behavioral disruptions through 90 minutes after the intervention.<sup>23</sup> An ATD was also used to compare the effects of commercially available and custom-made video prompts on the performance of multistep cooking tasks in 4 participants with autism.<sup>22</sup> Results showed that participants independently performed more steps with the custom-made video prompts (Figure 5).<sup>22</sup>



**Figure 5.** An example of results from a single-case alternating treatment study conducted on 4 participants with autism (data on 2 participants were used for this example). After the observation phase (baseline), effects of commercially available and custom-made video prompts on the performance of multistep cooking tasks were identified (treatment phase), after which only the best treatment was used (best treatment phase). Custom-made video prompts were most effective for improving participants' performance of multistep cooking tasks. Data were extracted from Mechling et al<sup>22</sup> utilizing Rohatgi's WebPlotDigitizer software.<sup>78</sup>

Regardless of the SC study design, *replication and randomization* should be incorporated when possible to improve internal and external validity.<sup>11</sup> The reversal design is an example of replication across study phases. The minimum number of phase replications needed to meet quality standards is 3 ( $A_1B_1A_2B_2$ ), but having 4 or more replications is highly recommended ( $A_1B_1A_2B_2A_3 \dots$ ).<sup>11,14</sup> In cases when interventions aim to produce lasting changes in participants' abilities, replication of findings may be demonstrated by replicating intervention effects across multiple participants (as in multiple-participant AB designs), or across multiple settings, tasks, or service providers. When the results of an intervention are replicated across multiple reversals, participants, and/or contexts, there is an increased likelihood that a causal relationship exists between the intervention and the outcome.<sup>2,12</sup>

Randomization should be incorporated in SC studies to improve internal validity and the ability to assess for causal relationships among interventions and outcomes.<sup>11</sup> In contrast to traditional group designs, SC studies often do not have multiple participants or units that can be randomly assigned to different intervention conditions. Instead, in *randomized phase-order designs*, the sequence of phases is randomized. Simple or block randomization is possible. For example, with simple randomization for an  $A_1B_1A_2B_2$  design, the A and B conditions are treated as separate units and are randomly assigned to be administered for each of the predefined data collection points. As a result, any combination of A-B sequences is possible without restrictions on the number of times each condition is administered or regard for repetitions of conditions (eg,  $A_1B_1B_2A_2B_3B_4B_5A_3B_6A_4A_5A_6$ ). With block randomization for an  $A_1B_1A_2B_2$  design, 2 conditions (eg, A and B) would be blocked into a single unit (AB or BA), randomization of which to different periods would ensure that each condition appears in the resulting sequence more than 2 times (eg,  $A_1B_1B_2A_2A_3B_3A_4B_4$ ). Note that AB and reversal designs require that the baseline (A) always precedes the first intervention (B), which should be accounted for in the randomization scheme.<sup>2,11</sup>

In *randomized phase start-point designs*, the lengths of the A and B phases can be randomized.<sup>2,11,24–26</sup> For example, for an AB design, researchers could specify the number of time points at which outcome data will be collected (eg, 20), define the minimum number of data points desired in each phase (eg, 4 for A, 3 for B), and then randomize the initiation of the intervention so that it occurs anywhere between the remaining time points (points 5 and 17 in the current example).<sup>27,28</sup> For multiple baseline designs, a dual-randomization or “regulated randomization” procedure has been recommended.<sup>29</sup> If multiple baseline randomization depends solely on chance, it could be the case that all units are assigned to begin intervention at points not really separated in time.<sup>30</sup> Such randomly selected initiation of the intervention would result in the drastic reduction of the discriminant and internal validity of the study.<sup>29</sup> To eliminate this issue, investigators should first specify appropriate intervals between the start points for different units, then randomly select from those intervals, and finally randomly assign each unit to a start point.<sup>29</sup>

## SINGLE-CASE ANALYSIS TECHNIQUES FOR INTERVENTION RESEARCH

The *What Works Clearinghouse* (WWC) single-case design technical documentation provides an excellent overview of appropriate SC study analysis techniques to evaluate the effectiveness of intervention effects.<sup>1,18</sup> First, visual analyses are recommended to determine whether there is a functional relationship between the intervention and the outcome. Second, if evidence for a functional effect is present, the visual analysis is supplemented with quantitative analysis methods evaluating the magnitude of the intervention effect. Third, effect sizes are combined across cases to estimate overall average intervention effects, which contribute to evidence-based practice, theory, and future applications.<sup>2,18</sup>

### Visual Analysis

Traditionally, SC study data are presented graphically. When more than 1 participant engages in a study, a spaghetti plot showing all of their data in the same figure can be helpful for visualization. Visual analysis of graphed data has been the traditional method for evaluating treatment effects in SC research.<sup>1,12,31,32</sup> The visual analysis involves evaluating level, trend, and stability of the data within each phase (ie, within-phase data examination) followed by examination of the immediacy of effect, consistency of data patterns, and overlap of data between baseline and intervention phases (ie, between-phase comparisons). When the changes (and/or variability) in level are in the desired direction, are immediate, readily discernible, and maintained over time, it is concluded that the changes in behavior across phases result from the implemented treatment and are indicative of improvement.<sup>33</sup> Three demonstrations of an intervention effect are necessary for establishing a functional relationship.<sup>1</sup>

### Within-Phase Examination

Level, trend, and stability of the data within each phase are evaluated. Mean and/or median can be used to report the level, and trend can be evaluated by determining whether the data points are monotonically increasing or decreasing. Within-phase stability can be evaluated by calculating the percentage of data points within 15% of the phase median (or mean). The *stability criterion* is satisfied if about 85% (80%–90%) of the data in a phase fall within a 15% range of the median (or average) of all data points for that phase.<sup>34</sup>

### Between-Phase Examination

Immediacy of effect, consistency of data patterns, and overlap of data between baseline and intervention phases are evaluated next. For this, several nonoverlap indices have been proposed that all quantify the proportion of measurements in the intervention phase not overlapping with the baseline measurements.<sup>35</sup> Nonoverlap statistics are typically scaled as percent from 0 to 100, or as a proportion from 0 to 1. Here, we briefly discuss the *nonoverlap of all pairs* (NAP),<sup>36</sup> the *extended celeration line* (ECL), the *improvement rate difference* (IRD),<sup>37</sup> and the *TauU*, and the *TauU-adjusted*,  $TauU_{adj}$ ,<sup>35</sup> as these are the most recent and complete techniques. We also examine the *percentage of nonoverlapping data* (PND)<sup>38</sup> and the *two standard deviations band method*, as these are frequently

used techniques. In addition, we include the *percentage of nonoverlapping corrected data (PNCD)*—an index applying to the PND after controlling for baseline trend.<sup>39</sup>

### Nonoverlap of All Pairs

Each baseline observation can be paired with each intervention phase observation to make  $n$  pairs (ie,  $N = n_A \times n_B$ ). Count the number of overlapping pairs,  $n_0$ , counting all ties as 0.5. Then define the percent of the pairs that show no overlap. Alternatively, one can count the number of positive (P), negative (N), and tied (T) pairs<sup>2,36</sup>:

$$NAP = \frac{n - n_0}{n} \times 100 = \frac{P + 0.5T}{P + T + N} \times 100$$

### Extended Celeration Line

ECL or split-middle line allows control for a positive phase A trend. Nonoverlap is defined as the proportion of phase B ( $n_B$ ) data that are above the median trend plotted from phase A data ( $n_B$ ), but then extended into phase B:

$$ECL = \frac{n_{B \text{ Above Median trend A}}}{n_B} \times 100$$

As a consequence, this method depends on a straight line and makes an assumption of linearity in the baseline.<sup>2,12</sup>

### Improvement Rate Difference

This analysis is conceptualized as the difference in improvement rates (IR) between baseline ( $IR_B$ ) and intervention phases ( $IR_T$ ).<sup>38</sup> The IR for each phase is defined as the number of “improved data points” divided by the total data points in that phase. Improvement rate difference, commonly employed in medical group research under the name of “risk reduction” or “risk difference,” attempts to provide an intuitive interpretation for nonoverlap and to make use of an established, respected effect size,  $IR_B - IR_T$ , or the difference between 2 proportions.<sup>37</sup>

### TauU and TauU<sub>adj</sub>

Each baseline observation can be paired with each intervention phase observation to make  $n$  pairs (ie,  $n = n_A \times n_B$ ). Count the number of positive (P), negative (N), and tied (T) pairs, and use the following formula:  $TauU = \frac{P - N}{P + N + T}$ .

The  $TauU_{adj}$  is an adjustment of  $TauU$  for monotonic trend in baseline. Each baseline observation can be paired with each intervention phase observation to make  $n$  pairs (ie,  $n = n_A \times n_B$ ). Each baseline observation can be paired with all later baseline observations ( $n_A \times (n_A - 1)/2$ ).<sup>2,35</sup> Then the baseline trend can be computed:  $TauU_{adj} = \frac{P - N - S_{trend}}{P + N + T}$ ;  $S_{trend} = P_A - N_A$ .

Online calculators might assist researchers in obtaining the TauU and TauU adjusted coefficients (<http://www.singlecaseresearch.org/calculators/tau-u>).

### Percentage of Nonoverlapping Data

If anticipating an increase in the outcome, locate the highest data point in the baseline phase and then calculate the percent of the intervention phase data points that exceed it. If anticipating a decrease in the outcome, find the lowest

data point in the baseline phase and then calculate the percent of the treatment phase data points that are below it:  $PND = \frac{n_{B \text{ Overlap A}}}{n_B} \times 100$ . A PND less than 50 would mark no observed effect, PND = 50 to 70 signifies a questionable effect, and PND more than 70 suggests the intervention was effective.<sup>40</sup> The percentage of nonoverlapping corrected was proposed in 2009 as an extension of the PND.<sup>39</sup> Before applying the PND, a data correction procedure is applied, eliminating preexisting baseline trend.<sup>38</sup>

### Two Standard Deviation Band Method

When the *stability criterion* described earlier is met within phases, it is possible to apply the 2-standard deviation band method.<sup>12,41</sup> First, the mean of the data for a specific condition is calculated and represented with a solid line. In the next step, the standard deviation of the same data is computed, and 2 dashed lines are represented: one located 2 standard deviations above the mean and the other 2 standard deviations below. For normally distributed data, few points (<5%) are expected to be outside the 2-standard deviation bands if there is no change in the outcome score because of the intervention. However, this method is not considered a formal statistical procedure, as the data cannot typically be assumed to be normal, continuous, or independent.<sup>41</sup>

### Statistical Analysis

If the visual analysis indicates a functional relationship (ie, 3 demonstrations of the effectiveness of the intervention effect), it is recommended to proceed with the quantitative analyses, reflecting the magnitude of the intervention effect. First, effect sizes are calculated for each participant (individual-level analysis). Moreover, if the research interest lies in the generalizability of the effect size across participants, effect sizes can be combined across cases to achieve an overall average effect size estimate (across-case effect size).

Note that quantitative analysis methods are still being developed in the domain of SC research<sup>1</sup> and statistical challenges of producing an acceptable measure of treatment effect remain.<sup>14,42,43</sup> Therefore, the WWC standards strongly recommend conducting sensitivity analysis and reporting multiple effect size estimators. If consistency across different effect size estimators is identified, there is stronger evidence for the effectiveness of the treatment.<sup>1,18</sup>

### Individual-Level Effect Size Analysis

The most common effect sizes recommended for SC analysis are (1) standardized mean difference *Cohen's d*; (2) standardized mean difference with correction for small sample sizes *Hedges' g*; and (3) the *regression-based approach*, which has the most potential and is strongly recommended by the WWC standards.<sup>1,44,45</sup> *Cohen's d* can be calculated using following formula:  $d = \frac{\bar{X}_A - \bar{X}_B}{s_p}$ , with  $\bar{X}_A$  being the baseline mean,  $\bar{X}_B$  being the treatment mean, and  $s_p$  indicating the pooled within-case standard deviation. *Hedges' g* is an extension of *Cohen's d*, recommended in the context of SC studies, as it corrects for small sample sizes. The piecewise regression-based approach does not reflect only the immediate

intervention effect but also the intervention effect across time:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 D_i + \beta_3 T_i D_i + e_i \\ + \rho e_{i-1} \text{ with } e_i \sim N(0, \sigma_e^2),$$

$i$  stands for the measurement occasion ( $i = 0, 1, \dots, I$ ),  $\rho$  indicates the autocorrelation parameter. If  $\rho$  is positive, the errors closer in time are more similar; if  $\rho$  is negative, the errors closer in time are more different, and if  $\rho$  equals 0, there is no correlation between the errors. The dependent variable is regressed on a time indicator,  $T$ , which is centered around the first observation of the intervention phase,  $D$ , a dummy variable for the intervention phase, and an interaction term of these variables. The equation shows that the expected score,  $\hat{Y}_i$ , equals  $\beta_0 + \beta_1 T_i$  in the baseline phase, and  $(\beta_0 + \beta_2) + (\beta_1 + \beta_3)T_i$  in the intervention phase.  $\beta_0$ , therefore, indicates the expected baseline level at the start of the intervention phase (when  $T = 0$ ), whereas  $\beta_1$  marks the linear time trend in the baseline scores. The coefficient  $\beta_2$  can then be interpreted as an immediate effect of the intervention on the outcome, whereas  $\beta_3$  signifies the effect of the intervention across time. The  $e_i$ 's are residuals assumed to be normally distributed around a mean of 0 with a variance of  $\sigma_e^2$ . The assumption of independence of errors is usually not met in the context of SC studies because repeated measures are obtained within a person. As a consequence, it can be the case that the residuals are autocorrelated, meaning that errors closer in time are more related to each other compared with errors further away in time.<sup>46–48</sup> As a consequence, a lag-1 autocorrelation is appropriate (taking into account the correlation between 2 consecutive errors:  $e_i$  and  $e_{i-1}$ ; for more details see Verbeke and Molenberghs.<sup>49</sup>

### Across-Case Effect Sizes

Two-level modeling to estimate the intervention effects across cases can be used to evaluate across-case effect sizes.<sup>44,45,50</sup> Multilevel modeling is recommended by the WWC standards because it takes the hierarchical nature of SC studies into account: measurements are nested within cases and cases, in turn, are nested within studies. By conducting a multilevel analysis, important research questions can be addressed (which cannot be answered by single-level analysis of SC study data), such as (1) What is the magnitude of the average treatment effect across cases? (2) What is the magnitude and direction of the case-specific intervention effect? (3) How much does the treatment effect vary within cases and across cases? (4) Does a case and/or study-level predictor influence the treatment's effect? The 2-level model has been validated in previous research using extensive simulation studies.<sup>45,46,51</sup> The 2-level model appears to have sufficient power ( $>0.80$ ) to detect large treatment effects in at least 6 participants with 6 measurements.<sup>21</sup>

Furthermore, to estimate the across-case effect sizes, the *HPS* (Hedges, Pustejovsky, and Shadish), or single-case educational design (*SCEDD*)-specific mean difference, index can be calculated.<sup>52</sup> This is a standardized mean difference index specifically designed for *SCEDD* data, with the aim of making it comparable to Cohen's  $d$  of group-comparison designs. The

standard deviation takes into account both within-participant and between-participant variability, and is typically used to get an across-case estimator for a standardized change in level. The advantage of using the *HPS* across-case effect size estimator is that it is directly comparable with Cohen's  $d$  for group comparison research, thus enabling the use of Cohen's (1988) benchmarks.<sup>53</sup>

Valuable recommendations on SC data analyses have recently been provided.<sup>54,55</sup> They suggest that a specific SC study data analytic technique can be chosen on the basis of (1) the study aims and the desired quantification (eg, overall quantification, between-phase quantifications, and randomization), (2) the data characteristics as assessed by visual inspection and the assumptions one is willing to make about the data, and (3) the knowledge and computational resources.<sup>54,55</sup> Table 1 lists recommended readings and some commonly used resources related to the design and analysis of single-case studies.

### QUALITY APPRAISAL TOOLS FOR SINGLE-CASE DESIGN RESEARCH

Quality appraisal tools are important to guide researchers in designing strong experiments and conducting high-quality systematic reviews of the literature. Unfortunately, quality assessment tools for SC studies are relatively novel, ratings across tools demonstrate variability, and there is currently no “gold standard” tool.<sup>56</sup> Table 2 lists important SC study quality appraisal criteria compiled from the most common scales; when planning studies or reviewing the literature, we recommend readers to consider these criteria. Table 3 lists some commonly used SC quality assessment and reporting tools and references to resources where the tools can be located.

When an established tool is required for systematic review, we recommend use of the WWC tool because it has well-defined criteria and is developed and supported by leading experts in the SC research field in association with the Institute of Education Sciences.<sup>18</sup> The WWC documentation provides clear standards and procedures to evaluate the quality of SC research; it assesses the internal validity of SC studies, classifying them as “meeting standards,” “meeting standards with reservations,” or “not meeting standards.”<sup>1,18</sup> Only studies classified in the first 2 categories are recommended for further visual analysis. Also, WWC evaluates the evidence of effect, classifying studies into “strong evidence of a causal relation,” “moderate evidence of a causal relation,” or “no evidence of a causal relation.” Effect size should only be calculated for studies providing strong or moderate evidence of a causal relation.

The *Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016* is another useful SC research tool developed recently to improve the quality of single-case designs.<sup>57</sup> *SCRIBE* consists of a 26-item checklist that researchers need to address while reporting the results of SC studies. This practical checklist allows for critical evaluation of SC studies during study planning, manuscript preparation, and review.

**Table 1. Recommended Readings and Resources Related to the Design and Analysis of Single-Case Studies****General Readings on Single-Case Research Design and Analysis**

- Barlow DH, Nock MK, Hersen M. *Single Case Experimental Designs: Strategies for Studying Behavior Change*. 3rd ed. Needham Heights, MA: Allyn & Bacon; 2008.<sup>3</sup>
- Kazdin AE. *Single-Case Research Designs: Methods for Clinical and Applied Settings*. New York, NY: Oxford University Press; 2010.<sup>4</sup>
- Kratochwill TR, Levin JR. *Single-Case Research Design and Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992.<sup>13</sup>
- Kratochwill TR, Levin JR. eds. *Single-Case Intervention Research: Methodological and Statistical Advances*. Washington, DC: American Psychological Association; 2014.<sup>2</sup>
- Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. Philadelphia, PA: F. A. Davis Company; 2015.<sup>12</sup>

**Design****Reversal design**

- Feeney TJ, Ylvisaker M. Context-sensitive cognitive-behavioral supports for young children with TBI. *J Posit Behav Interv*. 2008;10(2):115-128.<sup>59</sup>
- Lin CY, Chang YM. Increase in physical activities in kindergarten children with cerebral palsy by employing MaKey-MaKey-based task systems. *Res Dev Disabil*. 2014;35:1963-1969.<sup>60</sup>
- Samuel C, Louis-Dreyfus A, Kaschel R, et al. Rehabilitation of very severe unilateral neglect by visuo-spatio-motor cueing: two single case studies. *Neuropsychol Rehabil*. 2000;10(4):385-399.<sup>17</sup>

**Multiple baseline design**

- Hinderer SR, Lehmann JF, Price R, White O, deLateur BJ, Deitz J. Spasticity in spinal cord injured persons: quantitative effects of baclofen and placebo treatments. *Am J Phys Med Rehab*. 1990;69(6):311-317.<sup>20</sup>
- Lane-Brown A, Tate R. Evaluation of an intervention for apathy after traumatic brain injury: a multiple baseline, single-case experimental design. *J Head Trauma Rehab*. 2010;25(6):459-469.<sup>61</sup>

**Alternating treatment design**

- Folino A, Ducharme JM, Greenwald N. Temporal effects of antecedent exercise on students' disruptive behaviors: an exploratory study. *J School Psychol*. 2014;52(5):447-462.<sup>23</sup>
- Mechling LC, Atres KM, Foster AL, Bryant KJ. Comparing the effects of commercially available and custom-made video prompting for teaching cooking skills to high school students with autism. *Rem Spec Ed*. 2013;34(6):371-383.<sup>22</sup>

**Randomization**

- Kratochwill TR, Levin JR. Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psychol Methods*. 2010;15(2):124-144.<sup>11</sup>

**Analysis****Visual analysis**

- Lieberman LJ, Dunn JM, van der Mars H, McCubbin J. Peer tutors' effects on activity levels of deaf students in inclusive elementary physical education. *Adapt Phys Act Q*. 2000;17(1):20-39.<sup>62</sup>
- Lundblom EEG, Woods JJ. Working in the classroom: improving idiom comprehension through classwide peer tutoring. *Commun Dis Q*. 2012;33(4):202-219.<sup>63</sup>

**Percentage of nonoverlapping data**

- Banda DR, Hart SL, Liu-Gitz L. Impact of training peers and children with autism on social skills during center time activities in inclusive classrooms. *Res Autism Spect Dis*. 2010;4(4):619-625.<sup>64</sup>
- Oddo M, Barnett DW, Hawkins RO, Musti-Roa S. Reciprocal peer tutoring and repeated reading: Increasing practicality using student groups. *Psychol Schools*. 2010;47(8):842-858.<sup>65</sup>

**Nonoverlap of all pairs**

- Parker RI, Vannest K. An improved effect size for single-case research: Nonoverlap of all pairs. *Behav Ther*. 2009;40:357-367.<sup>36</sup>
- Peterson-Brown S, Karich AC, Symons FJ. Examining estimates of effect using non-overlap of all pairs in multiple baseline studies of academic intervention. *J Behav Educ*. 2012;21(3):203-216.<sup>66</sup>

**Improvement rate difference**

- Chen M, Hyppa-Martin JK, Reichle JE, Symons FJ. Comparing single case design overlap-based effect size metrics from studies examining speech generating device interventions. *Am J Intell Dev Disabil*. 2016;121(3):169-193.<sup>67</sup>
- Derakhshandeh F, Nikmaram M, Hosseiniabad HH, et al. Speech characteristics after articulation therapy in children with cleft palate and velopharyngeal dysfunction: a single case experimental design. *Int J Pediatr Otorhinolaryngol*. 2016;86:104-113.<sup>68</sup>

**Tau-U/Piecewise regression**

- Klingbeil D, Moeyaert M, Archerm C, Chimnoza TM, Zwolski SA. Examining the efficacy of peer-mediated incremental rehearsal. *School Psychol Rev*. In press.<sup>69</sup>
- Saddler K, Saddler B, Moeyaert M, Ellis-Robinson T. Effects of a summarizing strategy on written summaries of children with emotional and behavioral disorders. *Rem Spec Educ*. 2017;38(2).<sup>70</sup>

**Hierarchical Linear Modeling**

- Ingersoll B, Wainer A. Initial efficacy of project ImPACT: a parent-mediated social communication intervention for young children with ASD. *J Autism Dev Disord*. 2013;43(12):2943-2952.<sup>71</sup>
- Wade CA, Ortiz C, Gorman BS. Two-session group parent training for bedtime noncompliance in head start preschoolers. *Child Fam Behav Ther*. 2007;29(3):23-55.<sup>72</sup>



**Table 2. Summary of Important Single-Case Study Quality Appraisal Criteria**

Criteria	Requirements
1. Design <sup>1</sup>	The design is appropriate for evaluating the intervention
2. Method details <sup>56</sup>	Participants' characteristics, selection method, and testing setting specifics are adequately detailed to allow future replication
3. Independent variable <sup>1,14,18,56</sup>	The independent variable (ie, the intervention) is thoroughly described to allow replication; fidelity of the intervention is thoroughly documented; the independent variable is systematically manipulated under the control of the experimenter
4. Dependent variable <sup>1,14,18</sup>	Each dependent/outcome variable is quantifiable. <sup>57</sup> Each outcome variable is measured systematically and repeatedly across time to ensure the acceptable 0.80-0.90 interassessor percent agreement (or $\geq 0.60$ Cohen's kappa) on at least 20% of sessions <sup>73</sup>
5. Internal validity <sup>1,14,18</sup>	The study includes at least 3 attempts to demonstrate an intervention effect at 3 different points in time or with 3 different phase replications. <sup>73</sup> Design-specific recommendations: (1) for reversal designs, a study should have $\geq 4$ phases with $\geq 5$ points, (2) for alternating intervention designs, a study should have $\geq 5$ points per condition with $\leq 2$ points per phase, and (3) for multiple baseline designs, a study should have $\geq 6$ phases with $\geq 5$ points to meet the What Works Clearinghouse standards without reservations. <sup>18</sup> Assessors are independent and blind to experimental conditions
6. External validity <sup>14</sup>	Experimental effects should be replicated across participants, settings, tasks, and/or service providers
7. Face validity <sup>1,18,56</sup>	The outcome measure should be clearly operationally defined, have a direct unambiguous interpretation, and measure a construct is designed to measure
8. Social validity <sup>14,56</sup>	Both the outcome variable and the magnitude of change in outcome because of an intervention should be socially important; the intervention should be practical and cost effective
9. Sample attrition <sup>1,18</sup>	The sample attrition should be low and unsystematic, because loss of data in SC designs due to overall or differential attrition can produce biased estimates of the intervention's effectiveness if that loss is systematically related to the experimental conditions
10. Randomization <sup>1,18</sup>	If randomization is used, the experimenter should make sure that (1) equivalence is established at the baseline and (2) the group membership is determined through a random process

**Table 3. Quality Assessment and Reporting Tools Related to Single-Case Studies**

Quality Assessment and Reporting Tools	
What Works Clearinghouse Standards (WWC)	Kratochwill TR, Hitchcock J, Horner RH, et al. Institute of Education Sciences: What Works Clearinghouse: Procedures and standards handbook. <a href="http://files.eric.ed.gov/fulltext/ED510743.pdf">http://files.eric.ed.gov/fulltext/ED510743.pdf</a> . Published 2010. Accessed November 20, 2016. <sup>1</sup>
Quality indicators from Horner et al	Horner RH, Carr EG, Halle J, McGee G, Odom S, Wolery M. The use of single-subject research to identify evidence-based practice in special education. <i>Except Children</i> . 2005;71(2):165-179. <sup>14</sup>
Evaluative method	Reichow B, Volkmar F, Cicchetti D. Development of the evaluative method for evaluating and determining evidence-based practices in autism. <i>J Autism Dev Disord</i> . 2008;38(7):1311-1319. <sup>74</sup>
Certainty framework	Simeonsson R, Bailey D. Evaluating programme impact: levels of certainty. In: Mitchell D, Brown R, eds. <i>Early Intervention Studies for Young Children With Special Needs</i> . London, England: Chapman & Hall; 1991:280-296. <sup>75</sup>
Evidence in Augmentative and Alternative Communication Scales (EVIDAAC)	Schlosser RW, Sigafoos J, Belfiore P. EVIDAAC comparative single-subject experimental design scale (CSSEDARS). <a href="http://www.evidaac.com/ratings/CSSEDARS.pdf">http://www.evidaac.com/ratings/CSSEDARS.pdf</a> . Published 2009. Accessed November 20, 2016. <sup>76</sup>
Single-Case Experimental Design (SCED)	Tate RL, McDonald S, Perdices M, Togher L, Schulz R, Savage S. Rating the methodological quality of single-subject designs and n-of-1 trials: Introducing the Single-Case Experimental Design (SCED) Scale. <i>Neuropsychol Rehabil</i> . 2008;18(4):385-401. <sup>6</sup>
Logan et al scales	Logan LR, Hickman RR, Harris SR, Heriza CB. Single-subject research design: Recommendations for levels of evidence and quality rating. <i>Dev Med Child Neurol</i> . 2008;50:99-103. <sup>77</sup>
Single-Case Reporting Guideline In Behavioural Interventions (SCRIBE)	Tate RL, Perdices M, Rosenkoetter U, et al. The Single-Case Reporting guideline In Behavioural interventions (SCRIBE) 2016 statement. <i>J School Psychol</i> . 2016;56:133-142. <sup>57</sup>
Online Resources and Tools	
Theory, examples, and tools related to multilevel data analysis	Van den Noortgate W, Ferron J, Beretvas SN, Moeyaert M. Multilevel synthesis of single-case experimental data. Katholieke Universiteit Leuven web site. <a href="http://ppw.kuleuven.be/home/onderzoek/multilevel-synthesis-of-single-case-experimental-data/">http://ppw.kuleuven.be/home/onderzoek/multilevel-synthesis-of-single-case-experimental-data/</a> .
Tools for computing between-cases standardized mean difference ( <i>d</i> -statistic)	Pustejovsky JE. <i>scdhlrm</i> : a web-based calculator for between-case standardized mean differences (Version 0.2) [Web application]. <a href="https://jepusto.shinyapps.io/scdhlrm">https://jepusto.shinyapps.io/scdhlrm</a> .
Tools for computing NAP, IRD, Tau, and other statistics	Vannest KJ, Parker RI, Gonen O. Single case research: web based calculators for SCR analysis (Version 1.0) [Web-based application]. College Station, TX: Texas A&M University. Published 2011. Accessed November 20, 2016. <a href="http://www.singlecaseresearch.org/">http://www.singlecaseresearch.org/</a> .
Tools for obtaining graphical representations, means, trend lines, PND	Wright J. Intervention central. Accessed November 20, 2016. <a href="http://www.interventioncentral.org/teacher-resources/graph-maker-free-online">http://www.interventioncentral.org/teacher-resources/graph-maker-free-online</a> .
Access to free Simulation Modeling Analysis (SMA) Software	Borckardt JJ. SMA Simulation modeling analysis: time series analysis program for short time series data streams. Published 2006. <a href="http://www.clinicalresearcher.org/software.htm">http://www.clinicalresearcher.org/software.htm</a> .

Abbreviations: IRD, improvement rate difference; NAP, nonoverlap of all pairs; PND, percentage of nonoverlapping data.

## SUMMARY

Single-case studies can be designed and analyzed in a rigorous manner that allows researchers strength in assessing causal relationships among interventions and outcomes, and in generalizing their results.<sup>2,12</sup> These studies can be strengthened via incorporating replication of findings across multiple study phases, participants, settings, or contexts, and by using randomization of conditions or phase lengths.<sup>11</sup> There are a variety of tools that can allow researchers to objectively analyze findings from SC studies.<sup>56</sup> Although a variety of quality assessment tools exist for SC studies, they can be difficult to cate and utilize without experience, and different tools can provide variable results. The WWC quality assessment tool is recommended for those aiming to systematically review SC studies.<sup>1,18</sup>

SC studies, like all types of study designs, have a variety of limitations. First, it can be challenging to collect at least 5 data points in a given study phase. This may be especially true when traveling for data collection is difficult for participants, or during the baseline phase when delaying intervention may not be safe or ethical. Power in SC studies is related to the number of data points gathered for each participant, so it is important to avoid having a limited number of data points.<sup>12,58</sup> Second, SC studies are not always designed in a rigorous manner and, thus, may have poor internal validity. This limitation can be overcome by addressing key characteristics that strengthen SC designs (Table 2).<sup>1,14,18</sup> Third, SC studies may have poor generalizability. This limitation can be overcome by including a greater number of participants, or units. Fourth, SC studies may require consultation from expert methodologists and statisticians to ensure proper study design and data analysis, especially to manage issues like autocorrelation and variability of data.<sup>2</sup> Fifth, although it is recommended to achieve a stable level and rate of performance throughout the baseline, human performance is quite variable and can make this requirement challenging. Finally, the most important validity threat to SC studies is maturation. This challenge must be considered during the design process to strengthen SC studies.<sup>1,2,12,58</sup>

SC studies can be particularly useful for rehabilitation research. They allow researchers to closely track and report change at the level of the individual. They may require fewer resources and, thus, can allow for high-quality experimental research, even in clinical settings. Furthermore, they provide a tool for assessing causal relationships in populations and settings where large numbers of participants are not accessible. For all of these reasons, SC studies can serve as an effective method for assessing the impact of interventions.

## REFERENCES

- Kratochwill TR, Hitchcock J, Horner RH, et al. Single case designs technical documentation. *What Works Clearinghouse: Procedures and Standards Handbook*. <http://files.eric.ed.gov/fulltext/ED510743.pdf>. Published 2010.
- Kratochwill TR, Levin JR, eds. *Single-Case Intervention Research: Methodological and Statistical Advances*. Washington, DC: American Psychological Association; 2014.
- Barlow DH, Nock MK, Hersen M. *Single Case Experimental Designs: Strategies for Studying Behavior Change*. 3rd ed. Boston, MA: Allyn & Bacon; 2008.
- Kazdin AE. *Single-Case Research Designs: Methods for Clinical and Applied Settings*. 2nd ed. New York, NY: Oxford University Press; 2010.
- Onghena P. Single-case designs. In: Howell BED, ed. *Encyclopedia of Statistics in Behavioral Science*. Vol 4. Chichester, England: Wiley; 2005:1850-1854.
- Tate RL, McDonald S, Perdices M, Togher L, Schultz R, Savage S. Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the Single-Case Experimental Design (SCED) Scale. *Neuropsychol Rehabil*. 2008;18(4):385-401.
- Jewell DV. *Guide to Evidence-Based PT Practice*. 3rd ed. Burlington, MA: Jones & Bartlett Learning; 2014.
- Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. *Am J Prev Med*. 2007;33(2):155-161.
- Cartwright N, Munro E. The limitations of randomized controlled trials in predicting effectiveness. *J Eval Clin Pract*. 2010;16(2):260-266.
- Lefkowitz W, Jefferson TC. Medicine at the limits of evidence: the fundamental limitation of the randomized clinical trial and the end of equipoise. *J Perinatol*. 2014;34(4):249-251.
- Kratochwill TR, Levin JR. Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psychol Methods*. 2010;15(2):124-144.
- Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. Philadelphia, PA: F. A. Davis Company; 2015.
- Kratochwill TR, Levin JR. *Single-Case Research Design and Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992.
- Horner RH, Carr EG, Halle J, McGee G, Odom S, Wolery M. The use of single-subject research to identify evidence-based practice in special education. *Except Children*. 2005;71(2):165-179.
- Dionne M, Martini R. Floor Time Play with a child with autism: a single subject study. *Can J Occup Ther Ther*. 2011;78(3):196-203.
- Freeman JA, Gear M, Pauli A, et al. The effect of core stability training on balance and mobility in ambulant individuals with multiple sclerosis: a multi-centre series of single case studies. *Mult Scler*. 2010;16(11):1377-1384.
- Samuel C, Louis-Dreyfus A, Kaschel R, et al. Rehabilitation of very severe unilateral neglect by visuo-spatio-motor cueing: Two single case studies. *Neuropsychol Rehabil*. 2000;10(4):385-399.
- What Works Clearinghouse. *What Works Clearinghouse: Procedures and Standards Handbook. Version 3.0:1-91*. Washington, DC: Institute of Education Sciences.
- Wellek S, Blettner M. On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2012;109(15):276-281.
- Hinderer SR, Lehmann JF, Price R, White O, deLateur BJ, Deitz J. Spasticity in spinal cord injured persons: quantitative effects of baclofen and placebo treatments. *Am J Phys Med Rehab*. 1990;69(6):311-317.
- Ferron JM, Moeyaert M, Van den Noortgate NW, Beretvas SN. Estimating causal effects from multiple-baseline studies: implications for design and analysis. *Psychol Methods*. 2014;19(4):493-510.
- Mechling LC, Ayres KM, Foster AL, Bryant KJ. Comparing the effects of commercially available and custom-made video prompting for teaching cooking skills to high school students with autism. *Rem Spec Educ*. 2013;34(6):371-383.
- Folino A, Ducharme JM, Greenwald N. Temporal effects of antecedent exercise on students' disruptive behaviors: an exploratory study. *J School Psychol*. 2014;52(5):447-462.
- Edgington ES, Onghena P. *Randomization Tests*. Boca Raton, FL: Chapman & Hall/CRC; 2007.
- Onghena P, Edgington ES. Customization of pain treatments: single-case design and analysis. *Clin J Pain*. 2005;21:56-68.
- Todman JB, Dugard P. *Single-Case and Small-n Experimental Designs: A Practical Guide to Randomization Tests*. Mahwah, NJ: Erlbaum; 2001.
- Edgington ES. Randomization tests for one-subject operant experiments. *J Psychol*. 1975;90(1):57-68.
- Edgington ES. Nonparametric tests for single-case experiments. *Single-Case Research Design and Analysis*. Hillsdale, NJ: Erlbaum; 1992.
- Koehler MJ, Levin JR. Regulated Randomization: a potentially sharper analytical tool for the multiple-baseline design. *Psychol Methods*. 1998;3(2):206-217.
- Marascuilo LA, Busk PL. Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behav Assess*. 1988;10:1-28.

31. Ferron J, Jones PK. Tests for the visual analysis of response-guided multiple-baseline data. *J Exp Educ*. 2006;75(1):66-81.
32. Horner RH, Swaminathan H, Sugai G, Smolkowski K. Considerations for the systematic analysis and use of single-case research. *Educ Treat Children*. 2012;35(2):269-290.
33. Busse RT, Kratochwill TR, Elliott SN. Meta-analysis for single-case consultation outcomes: applications to research and practice. *J School Psychol*. 1995;33(4):269-285.
34. Neuman SB, McCormick S. *Single-Subject Experimental Research: Applications for Literacy*. Newark, DE: International Reading Association; 1995.
35. 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.
36. Parker RI, Vannest K. An improved effect size for single-case research: nonoverlap of all pairs. *Behav Ther*. 2009;40(4):357-367.
37. Parker RI, Vannest KJ, Brown L. The "improvement rate difference" for single-case research. *Except Child*. 2009;75(2):135-150.
38. Scruggs TE, Mastropieri MA. PND at 25: past, present, and future trends in summarizing single subject research. *Rem Spec Educ*. 2013;34:9-19.
39. Manolo R, Solanas A. Percentage of nonoverlapping corrected data. *Behav Res Meth*. 2009;41:1262-1271.
40. Scruggs TE, Mastropieri MA. Summarizing single-subject research: issues and applications. *Behav Modif*. 1998;22:221-242.
41. Callahan CD, Barisa MT. Statistical process control and rehabilitation outcome: the single-subject design reconsidered. *Rehabil Psychol*. 2005;50(1):24-33.
42. Beretvas SN, Chung H. A review of meta-analyses of single-subject experimental designs: methodological issues and practice. *Evid Based Commun Assess Interv*. 2008;2(3):129-141.
43. Shadish WR, Rindskopf DM. Methods for evidence-based practice: quantitative synthesis of single-subject designs. *New Dir Eval*. 2007;113:95-109.
44. Van den Noortgate DNW, Onghena P. Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behav Res Meth Instrum Comput*. 2003;35(1):1-10.
45. Van den Noortgate W, Onghena P. Combining single-case experimental data using hierarchical linear models. *Sch Psychol Q*. 2003;18(3):325-346.
46. Ferron JM, Bell BA, Hess MR, Rendina-Gobioff G, Hibbard ST. Making treatment effect inferences from multiple-baseline data: the utility of multilevel modeling approaches. *Behav Res Meth*. 2009;41(2):372-384.
47. Huitema BE, McKean JW. Reduced bias autocorrelation estimation: three Jackknife methods. *Educ Psychol Meas*. 1994;54(3):654-665.
48. McKnight SD, McKean JW, Huitema BE. A double bootstrap method to analyze linear models with autoregressive error terms. *Psychol Methods*. 2000;5(1):87-101.
49. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer; 2000.
50. Van den Noortgate W, Onghena P. A multilevel meta-analysis of single-subject experimental design studies. *Evid Based Commun Assess Interv*. 2008;2(3):142-151.
51. Ferron JM, Farmer JL, Owens CM. Estimating individual treatment effects from multiple-baseline data: a Monte Carlo study of multilevel-modeling approaches. *Behav Res Meth*. 2010;42(4):930-943.
52. Hedges LV, Pustejovsky JE, Shadish WR. A standardized mean difference effect size for single case designs. *Res Synth Methods*. 2012;3(3):224-239.
53. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
54. Manolov R, Moeyaert M. How can single-case data be analyzed? Software resources, tutorial, and reflections on analysis. *Behav Modif*. 2017;41(2):179-228.
55. Manolov R, Moeyaert M. Recommendations for choosing single-case data analytical techniques. *Behav Ther*. 2017;48(1):97-114.
56. Wendt O, Miller B. Quality appraisal of single-subject experimental designs: an overview and comparison of different appraisal tools. *Educ Treat Children*. 2012;35(2):235-286.
57. Tate RL, Perdices M, Rosenkoetter U, et al. The Single-Case Reporting guideline In BEhavioural interventions (SCRIBE) 2016 statement. *J School Psychol*. 2016;56:133-142.
58. Gast DL, Ledford JR. *Single Case Research Methodology: Applications in Special Education and Behavioral Sciences*. New York, NY: Routledge; 2014.
59. Feeney TJ, Ylvisaker M. Context-sensitive cognitive-behavioral supports for young children with TBI. *J Posit Behav Interv*. 2008;10(2):115-128.
60. Lin C-Y, Chang Y-M. Increase in physical activities in kindergarten children with cerebral palsy by employing MaKey-MaKey-based task systems. *Res Dev Disabil*. 2014;35:1963-1969.
61. Lane-Brown A, Tate R. Evaluation of an intervention for apathy after traumatic brain injury: a multiple-baseline, single-case experimental design. *J Head Trauma Rehab*. 2010;25(6):459-469.
62. Lieberman LJ, Dunn JM, van der Mars H, McCubbin J. Peer tutors' effects on activity levels of deaf students in inclusive elementary physical education. *Adapt Phys Act Q*. 2000;17(1):20-39.
63. Lundblom EEG, Woods JJ. Working in the classroom: improving idiom comprehension through classwide peer tutoring. *Commun Disord Q*. 2012;33(4):202-219.
64. Banda DR, Hart SL, Liu-Gitz L. Impact of training peers and children with autism on social skills during center time activities in inclusive classrooms. *Res Autism Spect Dis*. 2010;4(4):619-625.
65. Oddo M, Barnett DW, Hawkins RO, Musti-Rao S. Reciprocal peer tutoring and repeated reading: Increasing practicality using student groups. *Psychol Schools*. 2010;47(8):842-858.
66. Peterson-Brown S, Karich AC, Symons FJ. Examining estimates of effect using non-overlap of all pairs in multiple baseline studies of academic intervention. *J Behav Educ*. 2012;21(3):203-216.
67. Chen M, Hyppa-Martin JK, Reichle JE, Symons FJ. Comparing single case design overlap-based effect size metrics from studies examining speech generating device interventions. *Am J Intellect Dev Disabil*. 2016;121(3):169-193.
68. Derakhshandeh F, Nikmaram M, Hosseiniabad HH, et al. Speech characteristics after articulation therapy in children with cleft palate and velopharyngeal dysfunction: a single case experimental design. *Int J Pediatr Otorhinolaryngol*. 2016;86:104-113.
69. Klingbeil D, Moeyaert M, Archerm C, Chimnoza TM, Zwolski SA. Examining the efficacy of peer-mediated incremental rehearsal. *School Psychol Rev*. 2017;46:122-140.
70. Asaro-Saddler K, Saddler B, Moeyaert M, Ellis-Robinson T. Effects of a summarizing strategy on written summaries of children with emotional and behavioral disorders. *Rem Spec Educ*. 2017; doi: 0.1177/0741932516669051.
71. Ingersoll B, Wainer A. Initial efficacy of project ImPACT: a parent-mediated social communication intervention for young children with ASD. *J Autism Dev Disord*. 2013;43(12):2943-2952.
72. Wade CA, Ortiz C, Gorman BS. Two-session group parent training for bedtime noncompliance in head start preschoolers. *Child Fam Behav Ther*. 2007;29(3):23-55.
73. Hartman DP, Barrios BA, Wood DD. Principles of behavioral observation. In: Haynes SN, Hieby EM, eds. *Comprehensive Handbook of Psychological Assessment. Behavioral Assessment*. Vol 3. New York, NY: Wiley; 2004.
74. Reichow B, Volkmar F, Cicchetti D. Development of the evaluative method for evaluating and determining evidence-based practices in autism. *J Autism Dev Disord*. 2008;38(7):1311-1319.
75. Simeonsson R, Bailey D. Evaluating programme impact: levels of certainty. In: IDMRB, ed. *Early Intervention Studies for Young Children With Special Needs*. London, England: Chapman & Hall; 1991:280-296.
76. Schlosser RW, Sigafoos J, Belfiore P. EVIDAAC Comparative Single-Subject Experimental Design Scale (CSSEDARS). Published 2009. Accessed November 20, 2016 from <http://www.evidaac.com/ratings/CSSEDARS.pdf>.
77. Logan LR, Hickman RR, Harris SR, Heriza CB. Single-subject research design: recommendations for levels of evidence and quality rating. *Dev Med Child Neurol*. 2008;50:99-103.
78. Rohatgi A. WebPlotDigitizer User Manual Version 3.4. Published 2014. Accessed from <http://arohatgi.info/WebPlotDigitizer/userManual.pdf>