

## Small *N* designs for rehabilitation research

Scott D. Barnett, PhD;<sup>1\*</sup> Allen W. Heinemann, PhD;<sup>2</sup> Alexander Libin, PhD;<sup>3</sup> Arthur C. Houts, PhD;<sup>4</sup> Julie Gassaway, MS, RN;<sup>5</sup> Sunil Sen-Gupta, PhD, MPH;<sup>6</sup> Aaron Resch, MS;<sup>7</sup> Daniel F. Brossart, PhD<sup>7</sup>

<sup>1</sup>Center of Excellence: Maximizing Rehabilitation Outcomes, James A. Haley Veterans Hospital, Tampa, FL;

<sup>2</sup>Department of Physical Medicine and Rehabilitation, Feinberg School of Medicine, Northwestern University, Chicago, IL; <sup>3</sup>Georgetown University Medical Center, Department of Rehabilitation Medicine, Washington, DC;

<sup>4</sup>University of Memphis, Professor Emeritus, Memphis, TN; <sup>5</sup>Institute for Clinical Outcomes Research, Salt Lake City, UT; <sup>6</sup>Department of Veterans Affairs, Rehabilitation Research and Development Service, Washington, DC;

<sup>7</sup>Department of Educational Psychology, Texas A&M University, College Station, TX

**Abstract**—Rehabilitation research presents unique and challenging problems to investigators during both the design and analysis periods. Statistical issues regarding sample size requirements for an adequately powered study may be in direct conflict with realistic recruitment and subject retention goals. Issues of underpowered studies, sample size requirements, and recruitment goals plague rehabilitation research. Randomized clinical trials (RCTs) are typically narrow in scope and thus lack generalizability to everyday, yet specific, clinical problems; they are also costly and time-consuming and require large numbers of participants for randomization to have optimal, desired effects. Further, the RCT design may not be applicable to assistive technologies and environmental modifications—vital components of disability and rehabilitation research—nor is it appropriate in situations in which theoretical models of change are lacking or premature. Single-case designs are better suited for studies in which understanding and changing patient behavior and functional status are primary goals and the targeted sample sizes are less than 30 and frequently less than 10. Theoretical, methodological, and clinical reasons for using experimental and quasi-experimental single-case designs are presented. Recommendations for designing and conducting single-case studies that contribute to the evidence base are also discussed.

**Key words:** case reports, case series, experimental design, randomized clinical trial, rehabilitation, research design, single-case study, single-subject design, small *N*, statistics.

## INTRODUCTION

Rehabilitation research implements studies in which the long-term goals are to improve health and promote wellness for persons with physical disabilities. Rehabilitation presents a growing area of research within society as a whole and, more specifically, within the Veterans Health Administration (VHA) and Department of Defense. The recent Operation Iraqi Freedom/Operation Enduring Freedom conflicts have resulted in an unprecedented number of wounded warriors presenting for rehabilitation because of traumatic brain injury (TBI), blast injuries, amputation, and other conditions, which frequently include polytrauma characterized by lung, bowel, and inner ear injuries;

**Abbreviations:** ARIMA = autoregressive integrated moving average, AUC = area under the curve, NAP = nonoverlap of all pairs, RCT = randomized clinical trial, ROC = receiver operator curve, SCED = Single Case Experimental Design (scale), TBI = traumatic brain injury, VA = Department of Veterans Affairs, VHA = Veterans Health Administration.

\*Address all correspondence to Scott D. Barnett, PhD; Center of Excellence: Maximizing Rehabilitation Outcomes, James A. Haley Veterans Hospital, 8900 Grand Oak Cir (118M), Tampa, FL 33612; 813-558-3926; fax: 813-558-7616. Email: [Scott.Barnett2@va.gov](mailto:Scott.Barnett2@va.gov)

<http://dx.doi.org/10.1682/JRRD.2010.12.0242>

traumatic-limb or partial-limb amputation; soft tissue trauma from fragments and other missiles; and posttraumatic stress injuries [1–3]. As a result, the VHA has targeted rehabilitation research as a primary focus of the overall research portfolio.

The randomized clinical trial (RCT) is the gold standard of research designs, providing the best evidence of effect [4]. The RCT is regarded as the most rigorous design because of the prospective nature, randomization of subjects to independent study arms, and blinding process. Ideally, the randomization process balances potential confounding factors equally across study groups, and blinding reduces potential bias by blocking investigators and subjects from the hypothesis under investigation.

However, RCTs are generally narrow in scope and thus lack generalizability [5–7]; they are costly and time-consuming. The RCT design may not be applicable to assistive technologies and environmental modifications—vital components of disability and rehabilitation research. In many clinical scenarios, a meaningful control group experience is difficult—if not impossible—to design or implement. For example, many interventions in the rehabilitation setting are highly individualized (e.g., modifying assistive equipment to individual needs or abilities) and a control-group comparison is unreasonable [8]. RCTs are typically contingent on participants consenting to the randomization process, which raises concerns about the degree to which consenting participants are representative of the larger number of those who are unable or unwilling to consent. Rehabilitation research often involves specific behavioral and performance outcomes among persons who have low-incidence conditions or who have multiple and complex co-occurring conditions (e.g., polytrauma with behavioral disturbance among returning veterans). The effectiveness of randomization depends on large samples, representative of the population of concern, to distribute unmeasured factors that might otherwise influence results.

Issues of underpowered studies, sample size requirements, and recruitment goals often plague rehabilitation research. Statistical issues regarding sample size requirements for an adequately powered RCT may be in direct conflict with realistic recruitment and subject retention goals. There is simply no margin for error given the number of available subjects with infrequent or co-occurring conditions. Rehabilitation researchers are hard-pressed to balance scientific rigor with clinical feasibility. Consequently, the narrow scope and stringent requirements of

the RCT may be theoretically premature, clinically time-consuming, and of questionable generalizability for many research problems encountered in clinical rehabilitation.

In January 2010, the Department of Veteran Affairs (VA), Rehabilitation Research and Development Service convened a State-of-the-Art conference in Miami, Florida, to discuss current and future seminal issues pertinent to rehabilitation research both within and without VA. In this article, we summarize the strategy discussion about situations typically encountered by rehabilitation researchers in which small samples sizes are an issue. Quasi-experimental and experimental small *N* designs are ideal methods for clinical research in which understanding and changing maladaptive patterns in a patient's behavior and functional status are primary goals [9–10]. In this article, we summarize strategies that rehabilitation researchers may consider for studying issues in which small sample sizes are expected.

## WHY USE SMALL *N* RESEARCH DESIGNS?

Many reasons exist for conducting single-case research. Studies of costly treatment regimens, such as pharmacologic studies, may not have adequate funding to develop a large subject pool, and investigators may have to choose between increasing the number of subjects and rigorous testing of enrolled subjects. For some pathologies or disease conditions, such as TBI, pediatric oncology, gait and balance disorders, and cardiothoracic surgery, recruiting large numbers of subjects may not be feasible.

Single-case designs are intentionally used in scenarios in which compelling theoretical and clinical reasons exist to examine variable amounts, intensities, and types of interventions to achieve an outcome or resolve a behavioral problem [11]. In some situations, single-case designs may be the best choice; they may be among the most elegant and sophisticated experimental designs for use [12–13]. For these reasons, single-case designs have an impressive legacy in the study of many individualized interventions for persons with disabilities [14–16]. Moreover, an historical overview of the rehabilitation literature reveals that many current empirically supported practices first began with evidence obtained in single-case designs (e.g., behavioral methods for chronic pain rehabilitation, supported employment, and biofeedback) [11]. At times, a control condition may not be ethically appropriate because people cannot be randomized in a

treatment condition and treatment cannot be withheld. Consequently, a study may be ethically confined to quasi-experimental designs in which each person must serve as his or her own control and treatment is not withheld.

Single-case designs vary considerably in quality and rigor. The present article is chiefly concerned with experimental and quasi-experimental designs, as these have the greatest potential among the available small *N* designs to contribute to the evidence base. However, considering the variation in quality and rigor that exists in small *N* designs is instructive.

### SINGLE-CASE DESIGNS AND CONTINUUM OF CONFIDENCE

Kazdin describes a “continuum of confidence,” which conveys the extent to which one can be assured that any change is due to intervention or treatment effects (p. 258) [17]. At the lowest level of confidence is the anecdotal case study, also often referred to as an uncontrolled case study, which is a study of a single client, dyad, or group in which observations are made under uncontrolled and unsystematic conditions. These designs lack internal validity, but they permit the study of rare or low-incidence phenomena. They may generate new ideas and hypotheses. They may also play a role in the development of new interventions or therapies.

No single definition of “case study” is used across multiple disciplines, but case studies typically explore the mechanisms of a particular disease or condition, while focusing on a detailed observation of an individual person. Case studies may employ systematic observation by a clinician or a researcher, but the design does not manipulate an independent variable (e.g., a treatment) in an *a priori* manner. In clinical research, a “case” is defined as an individual affected by a disease, illness, or disability, which can be characterized in terms of the outcome of interest. Individual case studies emerge from clinical observation methods as one of the primary tools used by a clinician to understand the nature of illnesses.

Case studies may be helpful in understanding the unique problems of an individual patient [18]. Traditionally, clinicians use this design in absence of external resources. Unfortunately, this method is susceptible to an array of contextual and experimenter biases and alternative explanations of the observations that cannot be easily dismissed. Case studies (and their extension, case

reports) have a long history in medical research, particularly in the study of an infrequent or novel occurrence to describe the symptoms, signs, diagnosis, treatment, and follow-up of an individual patient or event. Case studies serve to present discoveries of new diseases and unexpected effects, either adverse or beneficial [19–22]. Vandenbroucke notes that “. . . case reports and series have a high sensitivity for detecting novelty and therefore remain one of the cornerstones of medical progress; they provide many new ideas in medicine” [23]. Further, recent pharmacologic agents involved in high-profile lawsuits from resulting adverse reactions may have been identified through case reports [24–25].

### SMALL *N* DESIGNS WITH HIGHEST LEVELS OF CONFIDENCE

In contrast, true experimental designs have the highest level of confidence. Multiple single-case designs exist that are true experimental designs. Methodologists generally accept that true experiments are those designs in which randomization (i.e., random assignment) plays a central role. In single-case research, randomization may play a role, but true experiments are those designs in which the researcher is able to control assessment occasions and the administration and termination of the treatment or intervention within the constraints of the design. The fundamental design requirement of single-case experimentation is repeated observations over time. These designs usually begin with a baseline phase in which data are collected before any treatment or intervention occurs. This phase is used to determine the stability of the variable(s) thought to be affected by the treatment. If the baseline phase is too short, then one does not have sufficient confidence that a good estimate of stability was obtained prior to treatment. Although it is often difficult to collect baseline data, an adequate baseline is essential. The baseline phase is followed by a treatment phase. Many variations exist, but a standard example would be the ABAB design in which baseline precedes a treatment phase—which is then repeated. Including the second set of AB phases increases one’s ability to rule out contradictory conclusions and thus accept that the treatment was the cause of the change. In the third phase (the second A phase), one expects the behavior under study to return to the initial baseline levels seen in the first phase. In the final phase, the intervention is again implemented, which allows one to test whether performance again improved with the

initiation of treatment. The basic logic of the ABAB design consists of making comparisons and/or predictions about performance under different conditions.

Several limitations of the ABAB design should be noted. This design requires a return to baseline performance in the second A phase. In many instances, such a return to baseline levels is not possible or would be unethical. For instance, if the intervention was designed to improve a skill set, typically one would not expect a return to baseline even if treatment was withdrawn, in contrast to withdrawing a certain drug, for example. Typically, any intervention that can be viewed as having a learning component would not be expected to return to baseline. In other cases, ceasing treatment when the client is expected to return to the baseline level is the same as expecting the client to get worse. In many cases, this would cause harm and be deemed unethical. A strong alternative (in terms of internal validity) to the ABAB design that does not require a return to baseline is the multiple-baseline design, which will be discussed shortly.

Falling below true experiments on the confidence continuum are quasi-experiments. Quasi-experiments usually do not have random assignment, but it is important to recognize that they can be among “the most effective and powerful” (p. 171) [12] nonrandomized experimental designs [13]. Furthermore, single-case quasi-experimental designs are especially useful in the study of low-incidence problems or conditions in which accessing a large representative sample for randomization of treatment would be virtually impossible. In addition, they are useful in the study of clinically complex cases. When these studies include multiple individuals, they can provide a strong basis for drawing valid inferences about a treatment effect. Such designs may address major threats to internal validity such as history, maturation, testing, and instrumentation.

One essential feature of quasi-experimental designs is that they require continuous assessment or data collection over time. Data collection occurs in all phases of the study. A second essential feature is that intervention effects are replicated over time within the same subject (as in an ABAB design) or, in the case of a multiple baseline with an AB design, intervention effects are replicated across subjects. Through the thoughtful and creative implementation of these two essential features, one may address “threats to validity, demonstrate causal relations, and build a knowledge base” (p. 385) [17].

Multiple-baseline designs are worth highlighting. A major strength of these designs is that they provide the

opportunity to replicate results. In multiple-baseline designs, inferences about treatment effects are based on examination of performance across several different baselines. These baselines may be measuring different specific behaviors from a single person, they may consist of multiple baselines across several individuals, or they may involve multiple baselines across situations, settings, or time. Multiple-baseline designs may also vary in terms of the number of baselines and the manner in which interventions are applied to these baselines. The key requirement for demonstrating unambiguous effects in a multiple-baseline design is that each baseline, whether involving a particular behavior, person, or situation, changes only after the intervention is introduced and not before.

Kazdin notes that multiple-baseline designs have a number of advantages that make them well-suited for applied settings [17]. He notes that these designs do not depend on withdrawing treatment to show that the intervention has in fact occurred. This characteristic makes multiple-baseline designs preferable to ABAB designs and their variations. The logic of the multiple-baseline design is the logic of replication to demonstrate that the observed change in a phase happens reliably at the point of intervention and also across different phase lengths. Three baselines are considered the minimum for a multiple-baseline design, but the more baselines included, the easier it is to attribute the intervention as the cause of change rather than extraneous threats to validity.

## USE OF STATISTICS IN ANALYSIS OF SINGLE-CASE RESEARCH

Historically, single-subject research designs relied on visual analysis of graphed data from the baseline and treatment conditions for each subject. However, contemporary advances have demonstrated that numerous statistical techniques can be used to analyze data from single-subject research designs [26–27]. As in many areas of statistics, no single best method exists for analyzing such data. For example, in a simple AB design, investigators may have questions about the overall change in level between the two phases and/or they may have questions about the rate of change. Each statistical method has its own set of assumptions and degrees to which violating those assumptions is problematic.

All of the available statistical techniques for single-subject research designs are somewhat influenced by autocorrelation (even visual analysis is affected by autocorrelation [28–29]). Autocorrelation is a problem because it violates the assumption of independence that most methods require. Most studies have shown that at least one-third of all published data contains positive autocorrelation to a problematic degree ( $\geq 0.20$ ) [30–33].

The recommended method for controlling autocorrelation within a single-subject research design is backcasting with an autoregressive integrated moving average model (ARIMA) (AR1 [1,0,0] first order autoregressive) model [34–36]. With this method, the auto-aggressive component is modeled and removed and the “prewhitened” data are then used for subsequent analyses. It is important to note that traditionally ARIMA modeling has been recommended for no less than 35 to 40 data points [34,37–38]. Yet in the single-subject research design, one is not trying to predict performance into the future and one is not interested in model fitting, so this limitation does not apply [e.g.,11,26]. Since using ARIMA can be unwieldy, investigators may wish to use methods that are integrated into regression software [39].

Several techniques appear to perform fairly well with single-subject research designs. One regression technique developed by Allison and colleagues removes trend from the baseline phase, which in some cases may be overly corrective [40–41]. This regression method can evaluate phase differences, and it may include a trend component. The downside to this method is that it is a parametric technique and it has all the assumptions of ordinary least squares method regression. This technique should probably be replaced with a robust regression method [e.g., 42] when a fair amount of variability is present in phase A. Brossart et al. found that at least one outlier was present in 61 percent of the data sets they examined and four or more outliers were present in more than 21 percent of the data sets [42]. In many data series examined, standard nonrobust methods were unable to detect outliers.

Brossart et al. suggest that this method may be less than ideal in terms of the effect sizes produced. The range appears to be small and to lack the coverage one typically expects from an effect size. For example, Brossart et al. reported that graphs judged to portray ineffective interventions had an average effect size of 0.36 in a regression model that examined change in mean shift between phases [26]. Graphs portraying somewhat effective interventions produced an average effect size of 0.52. Graphs

that were rated as very effective produced an average effect size of 0.67 for the same regression model. The difference between a moderately effective intervention and a very effective intervention may not be well represented with this regression model. More research is needed on this issue.

Promising nonparametric methods also exist for analyzing single-case data, for example, logistic regression [11]; nonoverlap of all pairs (NAP) [43]; and Tau-U, which combines nonoverlap of data and trend [39]. The advantages of the logistic method are that it does not assume a linear relationship between the independent and dependent variables and it does not require normally distributed variables or equal variance per cell (for example an AB design may be seen as a  $2 \times 2$  table). With this method, one may analyze all data series from a multiple-baseline study. This gives one an overall effect size regarding the treatment across all individuals. One may also analyze each subject separately to determine treatment effectiveness for each client or patient. The dependent variable is a phase variable, typically zeros for the baseline and ones for the treatment phase. The dependent variables are a participant variable and the corresponding treatment scores (for the overall analysis of multiple baselines) or just the participants' scores when each is analyzed separately. Thus, logistic regression predicts the phase (baseline or treatment) to which each score belongs based on its size. The output from a logistic regression usually includes a  $2 \times 2$  agreement table, and when analyzed using chi-square, one may calculate Pearson's phi; in which the effect size may be interpreted roughly as “prediction accuracy beyond chance” (p. 7) [11].

NAP evaluates the overlap in data between the baseline phase and the treatment phase. It may be calculated by hand but is also easily obtained as the area under the curve (AUC) from a receiver operator curve (ROC) analysis. NAP (or AUC) may be defined as “the probability that a score drawn at random from a treatment phase will exceed (overlap) that of a score drawn at random from a baseline phase” or as “the percent of nonoverlapping data between baseline and treatment phases” (p. 359) [43]. NAP scores range from 0.5 to 1.0 but may be transformed to have a range from 0 to 1.0 to represent deterioration in behavior or treatment [44]. We should note that other overlap indices exist and that the performance of these indices with single-case data remains a topic of research.

Tau-U assesses nonoverlap between phases, similar to the idea behind a ROC analysis. It consists of four indices;

three include nonoverlap and a trend component. The Tau-U index reported in the **Table** is similar to NAP and answers the question, “What is the improvement in nonoverlapping data between Phase A and B?” [39]. Alternatively, it may be interpreted as “percent of nonoverlap between phases,” or as “percent of data showing improvement between phases” (p. 291) [39]. The other three variations of Tau-U answer similar but different questions: (1) What is the improvement trend during phase B?, (2) What is the overall client improvement in phase A versus phase B with phase B trend?, and (3) What is the overall client improvement, controlling for baseline trend? [39].

### ILLUSTRATIVE EXAMPLE

The **Table** contains the results of two illustrative data sets. **Figure 1** portrays a treatment effect, and **Figure 2** does not. Several points are relevant upon examining these graphs and the **Table**. First, one should note that the effect sizes produced by each technique are different. The rough recommendations Cohen offered of what constitutes a large, medium, or small effect size do not apply to the effect sizes generated in single-case research [45]. Because guidelines do not exist at this point, researchers are encouraged to report effect sizes with confidence intervals, conduct visual analysis, and when possible, include the single-case graphs in published articles.

Second, one should note that using **Figure 1** data, the simple mean shift regression produced a smaller  $R^2$  than when baseline trend was corrected using Allison and colleagues’ regression model [40–41]. Not every data set will need to have baseline trend corrected. Investigators will need to determine whether such correction is necessary. If correction is desired, Allison and colleagues’ regression method is one option, but Tau-U also has several variations that correct for trend. Strategies for dealing with trend

in single-case research continue to be developed and evaluated.

### GUIDELINES FOR EVALUATING SMALL $N$ DESIGNS

While no universally accepted formal guidelines exist for the evaluation of small  $N$  designs, several efforts at guideline formulation have been attempted. In 2005, Horner et al. suggested guidelines for the evaluation of single-subject experimental designs [46]. Their “quality indicators” address the following areas: (1) the description of participants and settings, (2) the dependent variable, (3) the independent variable, (4) the baseline, (5) experimental control and internal validity, (6) external validity, and (7) social validity (p. 174) [46]. In addition, they proposed five standards to determine whether the results meet criteria to be called evidence based. These criteria are met when “(a) the practice is operationally defined; (b) the context in which the practice is to be used is defined; (c) the practice is implemented with fidelity; (d) results from single-subject research document the practice to be functionally related to change in dependent measures; and (e) the experimental effects are replicated across a sufficient number of studies, researchers, and participants to allow confidence in the findings.” (p. 175) [46]. Tate et al. proposed an 11-point scale (only 10 items are scored), the Single Case Experimental Design (SCED) scale, to measure the methodological quality of single-subject designs [47]. The items were designed to address specific weaknesses of single-subject designs and their threats to various components of validity. Scores range from 0 to 10, with a score of 10 reflecting the highest methodological quality rating. The SCED ratings are scored based on criteria ranging from a detailed operational definition of the target behavior to the requirement of at least three

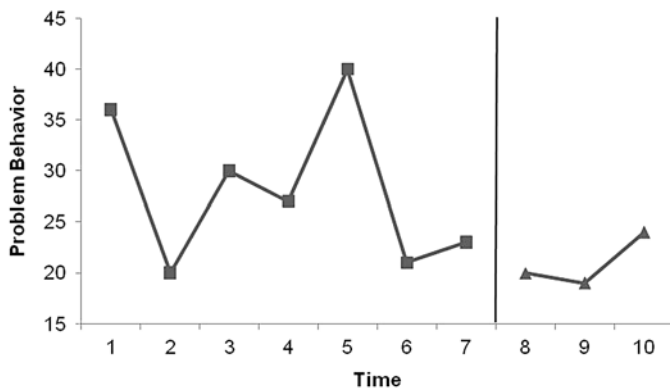
**Table.**

Results of example graphs analyzed using different statistical methods.

Example Data	Logistic Regression		Simple Mean Shift Regression ( $R^2$ )	Mean Plus Trend Regression* ( $R^2$ )	Tau-U <sub>nonoverlap</sub>	NAP
	% Correctly Classified	$\Phi$				
With treatment effect ( <b>Figure 1</b> )	80.0	0.52	0.23	0.75	0.67	0.83
With no treatment effect ( <b>Figure 2</b> )	54.5	0.04	0.02	0.01	0	0.50

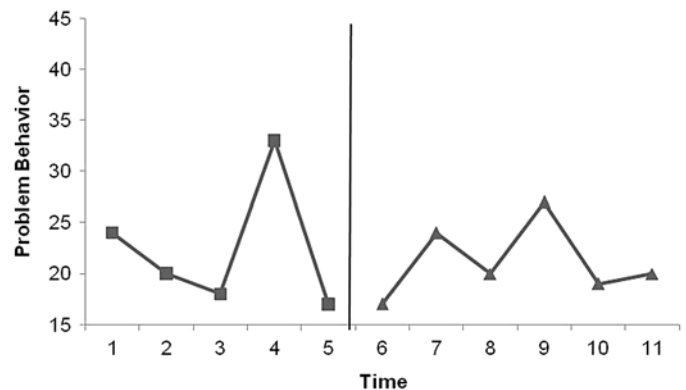
\*Allison DB, Gorman BS. Calculating effect sizes for meta-analysis: The case of the single case. Behav Res Ther. 1993;31(6):621–31.

NAP = nonoverlap of all pairs, Tau-U<sub>nonoverlap</sub> = assesses only overlap between phase A and phase B, no trend components included.



**Figure 1.**

Example of single-case data with treatment effect (reduction in problematic behavior).



**Figure 2.**

Example of single-case data with no treatment effect (no reduction in problematic behavior).

observations in the baseline phase. The SCED is an initial effort to evaluate single-case studies based on a minimum set of criteria that reflect sound single-case methodology, with the hope that its use will improve the design, oversight, and reporting of single-case research. It should be noted that a high score does not reflect the value of the study nor does it guarantee appropriate statistical analysis and/or conclusions were made [48].

Other recommendations have been narrower in focus. For example, Beeson and Robey evaluated single-case experimental designs of treatments for aphasia [49]. Their evaluation focused on the issue of effect sizes. To calculate effect sizes, they recommend using an ABA design. In their presentation, the second A phase was not expected to be a return to baseline. Instead of being referred to as a baseline, it could be called a maintenance phase. They state that an AB design only provides information about the slope in phases A and B, and therefore, such designs are not capable of providing adequate information for calculating effect sizes. Depending upon the statistical technique used, however, calculating an effect size is not difficult with an AB design. The real issue is determining what the researcher or practitioner wants the effect size to represent. The situation may warrant an effect size for a given duration of treatment, or if the treatment phase is long enough, it may be that the treatment effect stabilized in the latter part of that phase. In either case, an effect size from an AB design may be sufficient. If the investigator wants to examine the effects of treatment after it is withdrawn, then an effect size based on an ABA design is appropriate. Researchers and practi-

tioners should give careful thought to what they want the effect size to represent. In an ABA maintenance design, it would be possible to report two effect sizes, one comparing the baseline (phase A) to the treatment phase (phase B) and the other comparing the baseline (phase A) to the posttreatment phase (second A phase). Parker and Brosart discuss other variations in terms of phase contrasts for single-case designs [27].

In terms of conducting single-case research, the field has moved from conducting visual analysis only to the use of both visual and statistical analysis. No longer is visual analysis of single-case data graphed over time sufficient. All single-case data should be analyzed using both visual and statistical methods. The change in practice stems in part from the consistent finding that visual judgments have low to moderate interrater reliability [26], even among expert raters [50–53]. Efforts to increase judge reliability include modified graphs to help judges [50,54–56], visual analysis training for judges [57], or provision of contextual information regarding the data presented graphically [26], but none of these methods have resulted in higher judge reliability. Even when true parameters are known using simulated data, judges have low accuracy rates and their intrajudge consistency is also low (except when professionals rate graphs with no or small effects [58]). Another reason for the inclusion of statistical methods to supplement visual analysis is the need to report effect sizes. Many journals and funding agencies require them. Effect sizes also allow studies to be compared and meta-analyses to be conducted.

Statistical analysis can do two things, generally, very well. Statistics provide effect sizes that can quantify the amount of change for individual contrasts within a design (or an omnibus effect size for a larger design). They also inform one as to whether the amount of change could occur by chance alone (significance tests). Visual analysis plays many important and complementary roles in single-case research. For example, it allows one to determine whether the effects within a design are consistent across all or most contrasts. It allows one to identify which phase contrasts within a design can be legitimately combined for a design-wide effect size. It allows one to consider many different attributes in complex graphs, and it allows one to identify patterns that support or invalidate conclusion validity from the design. Thus, the goal is to implement a strong design with statistical and visual analysis when conducting single-case studies. All are required to draw proper, informed conclusions about intervention effectiveness.

## CONCLUSIONS

In 1894, Windelband formulated the epistemological dichotomy that divides human sciences into two categories: nomothetic science that concentrates on studying general laws and idiographic science that focuses on studying uniqueness of a single event or a person [59–60]. Despite the enduring efforts of epistemologists, psychologists, clinical researchers, and practitioners to bridge the nomothetic-idiographic divide in sociobehavioral and medical sciences [17,61–63], this dichotomous way of clinical thinking and scientific reasoning continues to be prominent today. A critical assumption underlies all levels of clinical evidence examination, from a single-case study report to a large scale RCT—predictions based on studying relations among group-level variables and a clinical change observed in a single patient are mutually exclusive entities in the decision-making process [64]. This assumption is also reflected in the compelling evidence paradox that clinicians derive from their meticulous field experience, while nomothetic-oriented researchers draw their conclusions from large sets derived from laboratory data [65]. This review demonstrates the complementary nature of large and small  $N$  studies and the ways in which investigators can use them to advance our understanding of processes and treatments that can inform improvements in patient care and outcomes.

Given the small  $N$  designs common in disability and rehabilitation research, we assert that small  $N$  studies (including pilot studies) should be conducted more often, as they are a valuable part of the evidence base [66]. Thus, administrators, grant reviewers, and journal editors should be educated on the value of small  $N$  studies, especially in the field of rehabilitation and specifically during the initial phases of research development and the study of complex behavioral issues associated with co-occurring conditions for which little reasonable recourse is available for control group experiences [66]. Rather than devaluing grant and manuscript submissions that present studies with small  $N$ s, reviewers should consider the value small  $N$  studies have in establishing initial evidence and raising questions for additional research to be conducted during later phases. We also recommend that funding administrators set aside monies specifically for small  $N$  studies. This funding will provide an opportunity for clinicians to become involved in research and advance the evidence base. Researchers must ensure that pilot studies are conducted in a systematic and thorough manner.

## ACKNOWLEDGMENTS

### Author Contributions

*Study concept and design:* S. D. Barnett, A. W. Heinemann, A. Libin, A. C. Houts, J. Gassaway, S. Sen-Gupta, A. Resch, D. F. Brossart.

*Drafting of manuscript:* S. D. Barnett, A. W. Heinemann, A. Libin, A. C. Houts, J. Gassaway, S. Sen-Gupta, A. Resch, D. F. Brossart.

*Critical revision of manuscript for important intellectual content:* S. D. Barnett, A. W. Heinemann, A. Libin, A. C. Houts, J. Gassaway, S. Sen-Gupta, A. Resch, D. F. Brossart.

*Study supervision:* S. D. Barnett.

**Financial Disclosures:** The authors have declared that no competing interests exist.

**Funding/Support:** This material was unfunded at the time of manuscript publication.

## REFERENCES

1. Belanger HG, Scott SG, Scholten J, Curtiss G, Vanderploeg RD. Utility of mechanism-of-injury-based assessment and treatment: Blast Injury Program case illustration. *J Rehabil Res Dev.* 2005;42(4):403–12. [PMID:16320137](https://pubmed.ncbi.nlm.nih.gov/16320137/)  
<http://dx.doi.org/10.1682/JRRD.2004.08.0095>
2. DePalma RG, Burris DG, Champion HR, Hodgson MJ. Blast injuries. *N Engl J Med.* 2005;352(13):1335–42. [PMID:15800229](https://pubmed.ncbi.nlm.nih.gov/15800229/)  
<http://dx.doi.org/10.1056/NEJMra042083>



3. Stuhmiller JH, Phillips YY, Richmond DR. The physics and mechanisms of primary blast injury. In: Bellamy RF, Zajtcuk R, editors. *Conventional warfare: Ballistic, blast, and burn injuries*. Washington (DC): Office of the Surgeon General at TMM Publications; 1991. p. 241–70.
4. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. *BMJ*. 1996;312(7023):71–72. [PMID:8555924](#)  
<http://dx.doi.org/10.1136/bmj.312.7023.71>
5. Kelley JM, Kaptchuk TJ. Group analysis versus individual response: The inferential limits of randomized controlled trials. *Contemp Clin Trials*. 2010;31(5):423–28. [PMID:20624483](#)  
<http://dx.doi.org/10.1016/j.cct.2010.07.003>
6. Stel VS, Jager KJ, Zoccali C, Wanner C, Dekker FW. The randomized clinical trial: An unbeatable standard in clinical research? *Kidney Int*. 2007;72(5):539–42. [PMID:17597704](#)  
<http://dx.doi.org/10.1038/sj.ki.5002354>
7. Henschel AD, Rothenberger LG, Boos J. Randomized clinical trials in children—Ethical and methodological issues. *Curr Pharm Des*. 2010;16(22):2407–15. [PMID:20513232](#)  
<http://dx.doi.org/10.2174/138161210791959854>
8. Marcel MP. When the best is the enemy of the good: The nature of research evidence used in systematic reviews and guidelines. Austin (TX): SEDL; 2009.
9. Gorman BS, Allison DB. Statistical alternatives for single-case designs. In: Franklin RD, Allison DB, Gorman BS, editors. *Design and analysis of single-case research*. Mahwah (NJ): Lawrence Erlbaum Associates; 1996. p. 159–214.
10. Saville BK, Buskist W. Traditional idiographic approaches: Small-N research designs. In: Davis SF, editor. *Handbook of research methods in experimental psychology*. Malden (MA): Blackwell; 2003. p. 66–82. <http://dx.doi.org/10.1002/9780470756973.ch4>
11. Brossart DF, Meythaler JM, Parker RI, McNamara J, Elliott TR. Advanced regression methods for single-case designs: Studying propranolol in the treatment for agitation associated with traumatic brain injury. *Rehabil Psychol*. 2008;53(3):357–69. <http://dx.doi.org/10.1037/a0012973>
12. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental design for generalized causal inference*. Boston (MA): Houghton-Mifflin; 2001.
13. Shadish WR, Rindskopf DM, Hedges LV. The state of the science in the meta-analysis of single-case experimental designs. *Evid Based Commun Assess Interv*. 2008;2(3):188–96. <http://dx.doi.org/10.1080/17489530802581603>
14. Schlosser RW. The role of single-subject experimental designs in evidence-based practice times. *Focus*. 2009;22:1–8.
15. Schlosser RW, Sigafoos J. Navigating evidence-based information sources in augmentative and alternative communication. *Augment Altern Commun*. 2009;25(4):225–35. [PMID:19903133](#)  
<http://dx.doi.org/10.3109/07434610903360649>
16. Schlosser RW, Sigafoos J. Augmentative and alternative communication interventions for persons with developmental disabilities: Narrative review of comparative single-subject experimental studies. *Res Dev Disabil*. 2006;27(1):1–29. [PMID:16360073](#)  
<http://dx.doi.org/10.1016/j.ridd.2004.04.004>
17. Kazdin AE. *Single-case research designs: Methods for clinical and applied settings*. 2nd ed. New York (NY): Oxford University Press; 2011.
18. Franklin RD, Allison DB, Gorman BS. *Design and analysis of single-case research*. Mahwah (NJ): Lawrence Earlbaum Associates; 1997.
19. Hamel J, Dufour S, Fortin D. *Case study methods*. Newbury Park (CA): Sage; 1993.
20. Stake RE. *The art of case study research*. Thousand Oaks (CA): Sage; 1995.
21. Yin RK. *Case study research: Design and methods*. 4th ed. Thousand Oaks (CA): Sage; 2009.
22. Baxter P, Jack S. Qualitative case study methodology: Study design and implementation for novice researchers. *Qual Rep*. 2008;13(4):544–59.
23. Vandembroucke JP. In defense of case reports and case series. *Ann Intern Med*. 2001;134(4):330–34. [PMID:11182844](#)
24. Arnaiz JA, Carné X, Riba N, Codina C, Ribas J, Trilla A. The use of evidence in pharmacovigilance. Case reports as the reference source for drug withdrawals. *Eur J Clin Pharmacol*. 2001;57(1):89–91. [PMID:11372600](#)  
<http://dx.doi.org/10.1007/s002280100265>
25. Loke YK, Price D, Derry S, Aronson JK. Case reports of suspected adverse drug reactions—Systematic literature survey of follow-up. *BMJ*. 2006;332(7537):335–39. [PMID:16421149](#)  
<http://dx.doi.org/10.1136/bmj.38701.399942.63>
26. Brossart DF, Parker RI, Olson EA, Mahadevan L. The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behav Modif*. 2006;30(5):531–63. [PMID:16894229](#)  
<http://dx.doi.org/10.1177/0145445503261167>
27. Parker RI, Brossart DF. Phase contrasts for multiphase single case intervention designs. *Sch Psychol Q*. 2006;21(1):46–61. <http://dx.doi.org/10.1521/sepq.2006.21.1.46>
28. Jones RR, Weinrott MR, Vaught RS. Effects of serial dependency on the agreement between visual and statistical inference. *J Appl Behav Anal*. 1978;11(2):277–83. [PMID:16795592](#)  
<http://dx.doi.org/10.1901/jaba.1978.11-277>
29. Matyas TA, Greenwood KM. Visual analysis of single-case time series: Effects of variability, serial dependence, and magnitude of intervention effects. *J Appl Behav Anal*.

- 1990;23(3):341–51. [PMID:16795732](#)  
<http://dx.doi.org/10.1901/jaba.1990.23-341>
30. Matyas TA, Greenwood KM. Serial dependency in single-case time series. In: Franklin RD, Allison DB, Gorman BS, editors. *Design and analysis of single-case research*. Mahwah (NJ): Lawrence Erlbaum Associates; 1996. p. 215–43.
  31. Parker RI, Cryer J, Byrns G. Controlling baseline trend in single-case research. *Sch Psychol Q*. 2006;21(4):418–44. <http://dx.doi.org/10.1037/h0084131>
  32. Sharpley CF, Alavosius MP. Autocorrelation in behavioral data: An alternative perspective. *Behav Assess*. 1988; 10(3):243–51.
  33. Suen HK, Ary D. *Analyzing quantitative behavioral observation data*. Hillsdale (NJ): Lawrence Erlbaum Associates; 1989.
  34. Box GE, Jenkins GM. *Time series analysis: Forecasting and control*. 2nd ed. San Francisco (CA): Holden-Day; 1976.
  35. Glass GV, Willson VL, Gottman JM. *Design and analysis of time-series experiments*. Boulder (CO): Colorado Associated University Press; 1975.
  36. Jones RR, Vaught RS, Weinrott M. Time-series analysis in operant research. *J Appl Behav Anal*. 1977;10(1):151–66. [PMID:16795544](#)  
<http://dx.doi.org/10.1901/jaba.1977.10-151>
  37. Gottman JM, Glass GV. Analysis of interrupted time-series experiments. In: Kratochwill TR, editor. *Single subject research: Strategies for evaluating change*. New York (NY): Academic Press; 1978. p. 197–235.
  38. Horne GP, Yang MC, Ware WB. Time series analysis for single-subject designs. *Psychol Bull*. 1982;91(1):178–89. <http://dx.doi.org/10.1037/0033-2909.91.1.178>
  39. Parker RI, Vannest KJ, Davis JL, Sauber SB. Combining nonoverlap and trend for single-case research: Tau-U. *Behav Ther*. 2011;42(2):284–99. [PMID:21496513](#)  
<http://dx.doi.org/10.1016/j.beth.2010.08.006>
  40. Allison DB, Gorman BS. Calculating effect sizes for meta-analysis: The case of the single case. *Behav Res Ther*. 1993;31(6):621–31. [PMID:7880208](#)  
[http://dx.doi.org/10.1016/0005-7967\(93\)90115-B](http://dx.doi.org/10.1016/0005-7967(93)90115-B)
  41. Faith MS, Allison DB, Gorman BS. Meta-analysis of single-case research. In: Franklin RD, Allison DB, Gorman BS, editors. *Design and analysis of single-case research*. Mahwah (NJ): Lawrence Erlbaum Associates; 1996. p. 245–77.
  42. Brossart DF, Parker RI, Castillo LG. Robust regression for single-case data analysis: How can it help? *Behav Res Methods*. 2011;43(3):710–19. [PMID:21437750](#)  
<http://dx.doi.org/10.3758/s13428-011-0079-7>
  43. Parker RI, Vannest KJ. An improved effect size for single-case research: Nonoverlap of all pairs. *Behav Ther*. 2009; 40(4):357–67. [PMID:19892081](#)  
<http://dx.doi.org/10.1016/j.beth.2008.10.006>
  44. Huberty CJ, Lowman LL. Group overlap as a basis for effect size. *Educ Psychol Meas*. 2000;60(4):543–63.
  45. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988.
  46. 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 children. *Except Child*. 2005; 71(2):165–79.
  47. 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. [PMID:18576270](#)  
<http://dx.doi.org/10.1080/09602010802009201>
  48. Perdices M, Tate RL. Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognized and undervalued? *Neuropsychol Rehabil*. 2009;19(6):904–27. [PMID:19657974](#)  
<http://dx.doi.org/10.1080/09602010903040691>
  49. Beeson PM, Robey RR. Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychol Rev*. 2006;16(4):161–69. [PMID:17151940](#)  
<http://dx.doi.org/10.1007/s11065-006-9013-7>
  50. DeProspero A, Cohen S. Inconsistent visual analyses of intrasubject data. *J Appl Behav Anal*. 1979;12(4):573–79. [PMID:16795617](#)  
<http://dx.doi.org/10.1901/jaba.1979.12-573>
  51. Harbst KB, Ottenbacher KJ, Harris SR. Interrater reliability of therapists' judgements of graphed data. *Phys Ther*. 1991;71(2):107–15. [PMID:1989006](#)
  52. Ottenbacher KJ. When is a picture worth a thousand *p* values? A comparison of visual and quantitative methods to analyze single subject data. *J Spec Educ*. 1990;23(4):436–49. <http://dx.doi.org/10.1177/002246699002300407>
  53. Park H, Marascuilo L, Gaylord-Ross R. Visual inspection and statistical analysis of single-case designs. *J Experimental Ed*. 1990;58:311–20.
  54. Greenspan P, Fisch GS. Visual inspection of data: A statistical analysis of behavior. *Proceedings of the Annual Meeting of the American Statistical Association*; 1992; Alexandria, VA.
  55. Hojem MA, Ottenbacher KJ. Empirical investigation of visual-inspection versus trend-line analysis of single-subject data. *Phys Ther*. 1988;68(6):983–88. [PMID:3375323](#)
  56. Skiba R, Deno S, Marston D, Casey A. Influence of trend estimation and subject familiarity on practitioners judgments of intervention effectiveness. *J Spec Educ*. 1989; 22(4):433–46. <http://dx.doi.org/10.1177/002246698902200405>

57. Wampold BE, Furlong MJ. The heuristics of visual inference. *Behav Assess*. 1981;3:79–92.
58. Ximenes VM, Manolov R, Solanas A, Quera V. Factors affecting visual inference in single-case designs. *Span J Psychol*. 2009;12(2):823–32. [PMID:19899683](#)
59. Marshall G. “Windelband, Wilhelm.” A dictionary of sociology [Internet]. *Encyclopedia.com*; 1998 [cited 2010 Dec 6]. Available from: <http://www.encyclopedia.com>
60. Lamiell JT. *Beyond individual and group differences: Human individuality, scientific psychology, and William Stern’s critical personalism*. Thousand Oaks (CA): Sage; 2003.
61. Iwakabe S, Gazzola N. From single-case studies to practice-based knowledge: Aggregating and synthesizing case studies. *Psychother Res*. 2009;19(4–5):601–11. [PMID:19579088](#)  
<http://dx.doi.org/10.1080/10503300802688494>
62. Schiller R, Tellegen A, Evens J. An idiographic and nomothetic study of personality description. In: Spielberger CD, Butcher JN, editors. *Advances in personality assessment*. Vol. 10. Hillsdale (NJ): Lawrence Erlbaum Associates; 1995.
63. Allport GW. The functional autonomy of motives. *Am J Psychol*. 1937;50(1/4):141–56. <http://dx.doi.org/10.2307/1416626>
64. Saville BK, Buskist W. Traditional idiographic approaches: Small-N research designs. In: Davis SF, editor. *Handbook of research methods in experimental psychology*. Malden (MA): Blackwell; 2003. p. 66–82. <http://dx.doi.org/10.1002/9780470756973.ch4>
65. Bellak L, Chassan JB. An approach to the evaluation of drug effects during psychotherapy: A double-blind study of a single case. *J Nerv Ment Dis*. 1964;139:20–30. [PMID:14202813](#)  
<http://dx.doi.org/10.1097/00005053-196407000-00003>
66. Tucker JA, Reed GM. Evidentiary pluralism as a strategy for research and evidence-based practice in rehabilitation psychology. *Rehabil Psychol*. 2008;53(3):279–93. [PMID:19649150](#)  
<http://dx.doi.org/10.1037/a0012963>

Submitted for publication December 27, 2010. Accepted in revised form July 11, 2011.

This article and any supplementary material should be cited as follows: Barnett SD, Heinemann AW, Libin A, Houts AC, Gassaway J, Sen-Gupta S, Resch A, Brossart DF. Small *N* designs for rehabilitation research. *J Rehabil Res Dev*. 2012;49(1):175–86.

<http://dx.doi.org/10.1682/JRRD.2010.12.0242>

ResearcherID: Scott D. Barnett, PhD: A-3226-2012



