Published in final edited form as:

Arch Phys Med Rehabil. 2012 August ; 93(8 Suppl): S111–S116. doi:10.1016/j.apmr.2011.12.017.

Small Sample Research Designs for Evidence-based Rehabilitation: Issues and Methods

James E. Graham, PhD, DC¹, Amol M. Karmarkar, PhD, OTR¹, and Kenneth J. Ottenbacher, PhD, OTR¹

¹Division of Rehabilitation Sciences, University of Texas Medical Branch

Abstract

Conventional research methods, including randomized controlled trials, are powerful techniques for determining the efficacy of interventions. These designs, however, have practical limitations when applied to many rehabilitation settings and research questions. Alternative methods are available that can supplement findings from traditional research designs and improve our ability to evaluate the effectiveness of treatments for individual patients. The focus on individual patients is an important element of evidenced-based rehabilitation. This paper examines one such alternate approach: small-N research designs. Small-N designs usually focus on ten or fewer participants whose behavior (outcomes) are measured repeatedly and compared over time. The advantages and limitations of various small-N designs are described and illustrated using three examples from the rehabilitation literature. The challenges and opportunities of applying small-N designs to enhance evidence-based rehabilitation are discussed.

Keywords

Evaluation studies as topic; Evidence-based practice; Rehabilitation; Research design

Introduction

RCTs are the most widely recommended research approach to evaluate treatment efficacy in biomedical research.^{1,2} Efficacy studies are designed to demonstrate causal relationships between treatment and outcomes under strictly controlled conditions. RCTs are also the most highly valued source of research evidence selecting treatments for individual patients based on evidence-based practice guidelines such as those established by the Center for Evidence-based Medicine (<http://www.cebm.net/>) and Cochrane Collaboration (<http://www.cochrane.org/>).

Some of the same qualities that make RCTs the gold-standard for efficacy research may limit their application in assessing the effectiveness of a given intervention for an individual patient. Effectiveness studies are designed to examine the effects of an intervention with typical patients in everyday situations wherein an investigator cannot control all the extraneous factors. RCTs tend to have strict inclusion and exclusion criteria and typically report average treatment effects obtained from statistical comparisons of group-level (aggregate) data from experimental and control groups. The article by Horn and colleagues in this issue, as well as earlier reviews by Grimmer et al.,³ and Kravitz et al.² provide more

Corresponding Author: K. Ottenbacher, Division of Rehabilitation Sciences, UTMB, 301 University Blvd., Galveston, TX 77555-1137. kottenba@utmb.edu, phone: (409) 747-1637, fax: (409) 747-1638.

Conflicts of Interest: The authors report no conflicts of interest.

details on why it is sometimes difficult to extrapolate findings from RCTs to everyday clinical practice.

An alternative research approach, broadly classified as small-N designs, involves serial observations of single persons or small groups before, during, and after an intervention period. This patient-level focus can facilitate EBP in two broad ways. First, it enables researchers to provide clinicians with practical information for making decisions to improve the care of individual patients. Second, it provides a potential avenue for including evaluation and research design in clinical practice and building the foundation for evidence-based rehabilitation at the level of the individual patient in actual treatment settings.⁴ Guyatt and colleagues⁵ have suggested a revision in the hierarchy of best evidence to include N of 1 randomized trials as the highest or “best” level of evidence for individual decision making. The purposes of this article are to: 1) provide an overview and select examples of small-N designs, 2) describe analysis and interpretation options for these designs, and 3) discuss the opportunities and challenges for small-N designs in rehabilitation research.

Small-N Designs: Overview and Examples

Small-N designs are not new. This approach is well-established in education and the behavioral sciences and is increasingly present in the clinical literature.^{6–8} The textbook by Bloom et al.⁴ is an excellent resource for using small-N designs to evaluate and inform clinical practice. This text addresses many of the issues (e.g., statistical versus visual analysis) that are not adequately covered in this article due to space limitations.

Several different terms are used to describe small-N methodology, including single-subject, N of 1 trials, single-case, and single-system designs, among others. Regardless of the terminology, the design framework is essentially the same: 1) studying a single person or small group of persons over time, 2) repeated measurement of the outcome, and 3) the sequential application and withdrawal of (or variation in) the intervention.

Unlike traditional parallel group RCTs where the independent variable (intervention) is standardized, small-N designs may involve a dynamic intervention that is manipulated across phases of the design⁹ (see examples below). The manner in which the intervention is applied distinguishes different categories of small-N designs. While the science of small-N research continues to evolve,¹⁰ there are three fundamental design frameworks frequently described in the small-N literature: basic and experimental withdrawal designs, multiple baseline designs, and changing intensity or alternating treatments designs.^{4,11} It is important to understand that the small-N designs described below vary in their ability to draw causal inferences linking the treatment (independent variable) and outcome (dependent variable), in the same manner that traditional designs based on group comparisons also vary in their ability to control threats to internal validity.

Basic (AB) and Experimental Withdrawal (ABA...) Designs

The basic design in this classification involves an AB structure, wherein “A” represents the baseline (non-treatment) phase and “B” refers to the intervention phase. The outcome measure is recorded repeatedly over both phases. This model is appealing and easy for clinicians to understand because it is closely related to what happens in day-to-day clinical practice.⁷ However, it is important to differentiate the AB design from a case study, which is principally a detailed after-the-fact descriptive summary without systematic manipulation of the independent variable; that is, in a case study there is no established baseline and transition to an intervention phase.

The key strengths of the AB design are its applicability to almost any clinical setting or problem, as well as its simplicity in evaluating whether changes occurred in the outcome following the transition from baseline to intervention. The primary weakness of the AB design is its inability to differentiate coincidence from causality with only a single transition between the baseline (non-treatment) and treatment phase.⁴ Because the information regarding patient performance is recorded repeatedly over time, AB designs can be used to document patterns of clinical change, but they cannot unequivocally demonstrate a causal connection between the change in the outcome variable and a specific treatment.

There are several experimental withdrawal designs within this framework (e.g., ABA, ABAB, BAB, etc.) that increase the ability to study effects associated with establishing a causal inference. Horner and colleagues⁹ suggest that at least three transitions are necessary to begin to establish causality and/or internal validity. For example, in the ABAB design, the targeted outcome measure is recorded repeatedly over all phases, the intervention (B) applied for a period, then withdrawn (second A phase) and then reintroduced (second B phase). The premise is to observe changes in the patterns of the outcome(s) with both the initiation and withdrawal of the intervention. Substantial improvements during the intervention phase(s) and subsequent reversal to baseline levels following withdrawal of the intervention increase confidence in the validity of the inference that a causal relationship exists.¹²

While more powerful than the basic AB design, these experimental withdrawal designs may have ethical limitations (e.g., removing a seemingly successful intervention); or practical (design) issues (e.g., interventions that produce carryover effects, or outcomes that are stable behaviors that cannot be expected to return to pre-intervention levels).⁴ An experimental withdrawal design may be an efficient approach to examining the effects of an adaptive device or assistive technology, but these designs are less useful in studies of rehabilitation interventions leading to enduring gains in performance⁶, such as increased strength.

Figure 1 shows an AB design with follow-up (for one of five participants) in a study on the effects of mass practice using constraint-induced movement therapy on lower-extremity function in patients with stroke.¹³ Performance on the *Timed Up and Go Test* was observed repeatedly during six sessions of the two-week baseline phase (A), six sessions of the two-week intervention phase (B), and two follow-up sessions (FU) at three and six months. The baseline phase demonstrates some variability in performances across the six sessions, but no discernable change (or slope) in the data points. The dashed horizontal line indicates the value that is two standard deviations below the mean of the baseline phase. Assumptions of normality notwithstanding, for data that show little to no trend over the baseline phase, the two standard deviation bands are considered thresholds for statistical significance for $\alpha = .05$.¹⁴ A gradual, but consistent slope (decrease in time needed to perform the test) can be seen across the two-week intervention phase. In addition, the faster performances in the Timed-up-and-go-Test appear to be retained at both 3- and 6-month follow-up. Note, the third phase was intentionally designed to be a follow-up rather than a baseline (A) withdrawal phase, because the strength and functional gains achieved through the intervention are lasting effects that would not be expected to return to baseline values immediately after terminating the training program.

Multiple Baseline Designs

Multiple baseline designs are viable alternatives to the experimental withdrawal designs in situations where either removing the intervention is inappropriate (unethical) or the treatment effects are long-lasting. Multiple baseline designs can target changes across three areas: 1) multiple baseline across *subjects* (the most common form of this design), 2) multiple baseline across *behaviors (outcomes)*, and 3) multiple baseline across *settings*. The

structure for these designs is a direct extension of the basic AB design, wherein two or more target AB pairs are created with the timing of the B phases systematically staggered across the different targets. In other words, baseline observations are made across all targets (persons, outcomes, or settings). A strength of this design is that the intervention is introduced at different times across each of the (multiple) baselines. The assumption is that the pattern of performance during the baseline will remain similar across phases (A, B, etc.) if the intervention has no impact on the target outcome. If the pattern of performance changes each time the intervention is introduced across persons, settings or related outcomes,^{4,15} this is viewed as evidence of a potential causal connection between the intervention and the outcome.

A strength of the multiple baseline approach is that it includes a built-in series of replications involving the treatment of interest. Replications are important in establishing generalizability in both small-N research and large group-comparisons studies (without random selection of participants). The role of replication in establishing generalizability is described in more detail in the Discussion. The sequential introduction of the treatment also helps to control some threats to internal validity that are present in the basic AB design described above, such as a coincidental change in the outcome that occurs at the same time the intervention is introduced.^{4,15}

There are two potential challenges in conducting and interpreting multiple baseline design studies.¹⁶ The first, which applies primarily to the multiple baseline across subjects version, is related to the time and resources required to concurrently identify, recruit and measure several participants. A potential solution to this limitation is to use a modified version of the design referred to as the non-concurrent multiple baseline design. This adaptation requires the investigator to determine possible baseline length and randomly assign these different baseline periods to persons as they are identified and become available to be studied. The second challenge, which applies to the multiple baseline design, arises when the various outcome measures are positively correlated. If changes in one outcome cause changes in another, then it is not possible to assess the cause-effect relationship between the intervention and the second outcome. This situation requires a revision of the outcome measures or the selection of another research design; e.g., ABAB.¹⁶

Figure 2 shows a multiple baseline design for one person. The figure is adapted from Carey and Matyas' study¹⁷ on direct and transfer effects of stimulus-specific training on joint proprioception in five patients with stroke. Two outcome measures (patients' perception of wrist position in the flexion-extension and ulnar-radial deviation planes) were recorded across 3 consecutive phases of 10 sessions each. In phase 1 (baseline), both outcomes were monitored with no intervention. Phase 2 (intervention 1) included flexion-extension stimuli provided every 2–3 days. In phase 3 (intervention 2), ulnar-radial deviation stimuli were added along with a maintenance program for the flexion-extension stimuli. Lastly, follow-up assessments were conducted 12–14 weeks following phase 3. The figure shows considerable variability overall and substantial error in proprioceptive awareness for both planes of movement during the baseline phase. Stimulus-specific training led to immediate reductions in both the variability and magnitude of errors, yet there appeared to be no transfer of flexion-extension training to accuracy of ulnar-radial perception. The results of the follow-up assessments suggest that the improvements in proprioceptive awareness were long-lasting.

Changing Intensity and Alternating Treatments Designs

The broad, flexible framework of the changing intensity and alternating treatments designs is perhaps most reflective of everyday rehabilitation clinical practice. RCTs aside, rarely does a patient receive a single intervention for a fixed duration regardless of their interim

progress. Rather, rehabilitation typically entails the use of several concurrent interventions (procedures, devices, activities, etc.) and the parameters of these interventions (type or intensity or duration, etc.) are adjusted as a patient's needs and abilities change.

As with the other small-N designs described above, the names essentially reflect the design parameters. Changing intensity designs can be illustrated with the following formula: $AB^1B^2B^3$, where the Bs with superscripts indicate successive intervention phases in which stepwise changes in intervention parameters are implemented.¹⁵ This design is not effective for establishing causality but can provide useful information in demonstrating change in patient performance over time. Additional baseline phases (withdrawal designs) can be incorporated into this design to increase internal validity, but this is neither necessary nor realistic in everyday clinical practice. Another caveat with respect to this design is that it is best suited for outcomes that can be gradually improved over time.⁴

Alternating treatments designs can be illustrated with the following formulas: ABC, ABAC, ABACA, or any other feasible combination of baseline/standard care (A) and intervention phases (B = intervention 1, C = intervention 2, etc.). Interpreting the results of alternating treatment designs must be done with caution. For example, order effects may influence the magnitude of the responses and confound interpretation. Also, it is important to understand that only adjacent phases can be compared, because there may be unmeasured factors or interactions among treatments that occur during the intervening phase. Patterns or relationships observed for non-adjacent phases can, however, facilitate the development of hypotheses that can be tested in future studies.⁴

Figure 3 shows an alternating treatment design¹⁸ comparing two treatments to improve gait in a patient with hemiparetic stroke. The treatments included two different types of ankle-foot orthoses (AFO) designed to reduce foot-drop. The baseline phase consisted of barefoot walking recorded during five sessions over seven days. The intervention phase spanned 1 month during which 12 sessions each of 3 randomly-ordered conditions were observed: barefoot, traditional ankle-foot orthosis (AFO), and tone-inhibiting dynamic AFO. Four additional barefoot-condition assessments were conducted over six days following the intervention phase. The sequence of baseline and alternating treatments is presented in the figure along with a graph showing the responses collapsed for each intervention.¹⁸ The barefoot condition demonstrated consistent increases in step length over all three phases. This might be interpreted as a learning or practice affect. The AFO interventions showed both 'main effects' for time and 'group effects' by condition. Specifically, stride length increased over time in all three conditions and the tone-inhibiting dynamic AFO yielded longer steps than the traditional AFO, which produced longer step lengths than walking barefoot.

Analysis and Interpretation in Small-N Designs

Conventional analysis and interpretation in small-N research is based on visual inspection of graphed and tabulated data within and across phases. Visual analysis includes both quantitative methods and judgments made by the investigators regarding patterns in the data. Quantitative indicators consist of the levels (e.g., means, percentages), trends (slopes of the lines of best fit), and variability (or dispersion) of data within each phase. Judgments include assessments regarding the consistency of data patterns within each phase, immediacy of change following phase transitions, degree of overlap from one phase to the next, and the extent of change, using visual inspection of graphs and/or number displays.⁹ Visual inspection has several advantages: 1) it is intuitive and economical, 2) it provides ongoing information regarding changes in the pattern of performance, and 3) it is focused on patient-

level treatments and responses.¹⁶ The disadvantage of visual inspection is the lack of standardized criteria and the potential for disagreement between or bias of raters.^{4,19}

In small-N designs the individual participant is the unit of analysis with each person serving as his or her own control. Performances during successive phases (e.g., baseline and intervention, or intervention and follow-up) are compared.⁹ The use of a control phase rather than a control group shifts the emphasis from between groups variance to within person variance.²⁰ The fact that repeated measurements are gathered from the same person increases the likelihood of serial dependency or auto-correlation within the data.⁴ This potential for dependency, along with other aspects of small-N research (e.g. small number of observations, non-normal distributions, unequal variances, etc.), violate basic assumptions for common statistical techniques such as t-tests and analysis of variance (ANOVA).

Statistical analysis options in small-N designs range from simple (e.g., celeration line, C-statistic, running medians, two standard deviation bands, etc.)^{14,16} to complex (e.g., hierarchical linear modeling).^{21,22} Detailed descriptions of these techniques are beyond the scope of this paper, but comprehensive analytic approaches for small-N designs are available in the literature.^{4,23–28} It is important to note that effect size has gained popularity as it overcomes many of the limitations of p-values in small-N research. Manolov and colleagues^{29,30} provide examples and describe the strengths and limitations of several effect size calculations including the common standardized mean difference approach, regression-based approaches, and visual-based approaches.

Opportunities and Challenges for Small-N Designs in Rehabilitation

Healthcare reform, a growing, aging, and diversifying population, and technological and scientific advances are all transforming the healthcare system and the way rehabilitation services are (and will be) delivered. One important characteristic of evidence-based practice is the focus on research information relevant to the individual patient. Sackett and colleagues³¹ originally described evidence-based medicine as the “*conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients*.”^{p. 71} The emerging field of personalized medicine promotes this viewpoint as well. Small-N research designs provide information directly relevant to the individual participants being studied. Lillie et al.⁸ stated recently that “*The ultimate goal of an n-of-1 trial is to determine the optimal or best intervention for an individual patient using objective data-driven criteria*.”^{p.61} Of course, it is important to acknowledge that the use of small-N research methods can only directly determine the ‘best’ treatments for persons that are involved in a specific study. Fields such as clinical and counseling psychology have proposed the ‘scientist-practitioner’ model in which systematic evaluation and accountability methods involving aspects of small-N research are incorporated into daily practice.^{4,27,30} The scientist-practitioner model attempts to integrate components of small-N designs directly into clinical practice. This involves a number of logistic and ethical issues (e.g., human subject protection) that are beyond the scope of this article, but discussed elsewhere.^{4,27,30,34}

An advantage of small-N design is that they allow investigators (and clinicians) to potentially identify characteristics relevant to individual patient performance. If an experimental group of 50 patients does statistically better than a control group of 50 patients, the difference could be due to a small number of persons in the treatment group showing large changes while the majority of individuals show little or no change. Individual variation may be masked by the group average. In small-N designs, each participant is assessed repeatedly and comparisons within the person are made over time allowing patterns of performance to be linked to individuals with specific characteristics. Once a clinically

significant difference appears within a small-N design, the practitioner can identify the patient variables and other relevant factors present when the result was obtained. In the small-N approach, not only are factors such as sex, age, diagnosis, level of disability, and education kept constant in the same participant over time, but so are all significant life experiences that occur before the intervention begins. This degree of individual control is only possible in large-N group comparison designs when the participants are measured repeatedly and followed over long periods. Doing this is very difficult within the context of an RCT.

The efficacy of a given treatment in a traditional RCT is usually assessed only after the treatment is completed (post-test). The typical pre-test, post-test RCT design precludes the continuous assessment and analysis of the patient's performance during treatment. Small-N designs usually allow researchers to observe within-person variability and relate environmental or physical characteristics to patient performance. Repeated observations permit a systematic analysis of the course of treatment and may suggest useful modifications as the study progresses. The researcher using a small-N design can take into account and analyze the impact of day-to-day contingencies and related events that affect patient behaviors and performances.

There are several factors that have limited the growth of small-N clinical research. Not the least of these factors is the dominance of the RCT research design hierarchy, including the publication bias of journal editors and predisposition of funding agencies.³² Another factor involves the general lack of research skills to conduct small-N studies and appropriately interpret and apply the findings among rehabilitation researchers and practitioners.³³

The small-N research approach includes a wide variety of designs, similar to the diversity in larger-N group comparison designs. We have introduced three categories of small-N designs and provided examples. There are other forms of small-N designs and many variations within each category of design. Identifying which design is the best fit for a particular research question or context depends on many factors. In general, small-N designs are practical complements to larger N trials. They can be useful in the early developmental phase of research as well as in refining the application of research findings to individual patients. Specific situations in which small-N designs may be particularly useful include 1) low-prevalence conditions or combinations of (comorbid) conditions wherein it is not feasible to recruit patients for a sufficiently-powered between-groups analysis; 2) conditions and/or interventions with the potential for substantial variability in either the responsiveness or magnitudes of responses; for example, in the early phase of a research program when there is little information about which outcomes will be sensitive to a new treatment; 3) when the sequence or duration of treatments is dependent upon the achievement of certain milestones or outcome thresholds; or 4) in unique settings or situations that may limit the feasibility of group comparison clinical trials, for instance in the early stages of establishing protocols for expensive or invasive interventions.

Small-N research can be challenging and these designs are not universally applicable to all clinical questions or patients. All designs – both small-N and large-N (group comparison) -- have limitations. Some of the design-specific caveats relevant to small-N methods were described in the sections above. Perhaps the most commonly identified limitation associated with small-N research is the perceived inability to generalize the study findings – the issue of external validity.^{6,15}

Since randomly selecting a representative sample from a large target population is not a component of small-N designs, replication is the alternative strategy used to establish the generalizability of small-N research findings. Barlow and Hersen³⁴ describe three strategies

for establishing generalizability in small-N research. The first form of generalizability involves the accumulation of a number of direct replications of the specific treatment effect on one well-defined outcome measure within a defined clinical setting. In this form of replication, participants are matched as closely as possible on subject characteristics. The aim is to establish, as clearly as possible, that a given intervention can have an effect on a certain kind of patient within a specific setting. If a series of direct replication small-N studies produces consistently positive results, then the replication process moves to the next level.

The second level of replication involves the systematic replication of the treatment across various participants, settings, clinicians, or a combination of these. Systematic replication helps to establish the generality of the findings over a wider range of situations than does direct replication. The final strategy of replication identified by Barlow and Hersen³⁴ is clinical replication. Clinical replication involves establishing the generality of related components of the intervention. These might include issues such as intensity or duration or combining multiple components of the intervention and testing them across various patients and settings.

Using replication to establish generalizability aims to develop a scientific consensus regarding the effectiveness (or lack of effectiveness) of an intervention across different persons, clinicians, and settings. It cannot be achieved in the context of a single study, but must be developed over time and involves multiple studies, often conducted by different investigators.

Meta-analysis of small-N studies has become increasingly popular as a means of integrating the findings of multiple studies and contributing to the generalizability of small-N studies.³⁵ The reader is referred to Barlow and Hersen³⁴ and Bloom et al.⁴ for more information on the issues of replication and generalizability in small-N research. Also, see the article by Johnston and others in this special issue for background on synthesizing information from several small-N trials and integrating this information with data from RCTs.

Not all rehabilitation clinicians will participate in traditional large-N group comparison experimental research designed to test hypotheses or refine theory. However, every clinician does have a responsibility to document the services they provide and changes in patient performance related to the person's individual rehabilitation goals.

Grimmer and colleagues³ contend that clinicians should be a driving force in developing more appropriate and relevant evidence for rehabilitation practice. Clinically-derived evidence can serve as the catalyst for investigators to design high-quality research that is relevant to clinical practice (see article by Whyte and others in this issue). Small-N designs represent one approach that is available to practitioners and that may allow them to contribute to the advancement of rehabilitation science and practice. While RCTs remain the gold standard for establishing treatment efficacy, it is important to recognize that small-N research can supplement and/or refine the findings from large parallel group trials when making treatment decisions for an individual patient.^{2,32} We are optimistic that the relationship between these approaches will continue to evolve into a partnership that provides the evidence necessary to validate clinical practice and improve services provided to persons with disabilities and their families.

Acknowledgments

The study was supported in part by the National Institutes of Health: R24-HD065702 (J. Graham, A. Karmarkar, K. Ottenbacher).

Abbreviations

AFO	ankle foot orthosis
EBP	evidence-based practice
RCT	randomized controlled trial

References

1. Law, MC.; MacDermid, J. Evidence-based Rehabilitation: A Guide to Practice, 2nd Ed. Thorofare, NJ: Slack; 2008.
2. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q.* 2004; 82(4):661–687. [PubMed: 15595946]
3. Grimmer K, Bialocerkowski A, Kumar S, Milanese S. Implementing evidence in clinical practice: the 'therapies' dilemma. *Physiotherapy.* 2004; 90(4):189–194.
4. Bloom, M.; Fischer, J.; Orme, JG. Evaluating Practice: Guidelines for the Accountable Professional. 6th Edition ed.. Boston: Allyn and Bacon; 2009.
5. Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, Wilson MC, Richardson WS. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. *JAMA.* 2000; 284(10):1290–1296. [PubMed: 10979117]
6. Backman CL, Harris SR, Chisholm JA, Monette AD. Single-subject research in rehabilitation: a review of studies using AB, withdrawal, multiple baseline, and alternating treatments designs. *Arch Phys Med Rehabil.* 1997; 78(10):1145–1153. [PubMed: 9339167]
7. Johnston MV, Smith RO. Single subject designs: current methodologies and future directions. *OTJR: Occupation, Participation & Health.* 2010; 30(1):4–10.
8. Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personalized Medicine.* 2011; 8(2):161–173. [PubMed: 21695041]
9. 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 Child.* 2005; 71(2):165–179.
10. McDougall D, Smith G, Black R, Rumrill P. Recent innovations in small-N designs for rehabilitation research: An extension of Cowan, Hennessey, Vierstra, and Rumrill. *Journal of Vocational Rehabilitation.* 2005; 23(3):197–205.
11. Ottenbacher, KJ. Evaluating Clinical Change: Strategies for Occupational and Physical Therapists. Baltimore, MD: Williams & Wilkins; 1986.
12. Gonnella C. Single-Subject Experimental Paradigm As A Clinical Decision Tool. *Phys Ther.* 1989; 69(7):601–609. [PubMed: 2662229]
13. Marklund I, Klassbo M. Effects of lower limb intensive mass practice in poststroke patients: single-subject experimental design with long-term follow-up. *Clin Rehabil.* 2006; 20(7):568–576. [PubMed: 16894799]
14. Backman CL, Harris SR. Case studies, single-subject research, and N of 1 randomized trials: comparisons and contrasts. *Am J Phys Med Rehabil.* 1999; 78(2):170–176. [PubMed: 10088595]
15. Cowan RJ, Hennessey ML, Vierstra CV, Rumrill PD. Small-N designs in rehabilitation research. *Journal of Vocational Rehabilitation.* 2004; 20(3):203–211.
16. Zhan S, Ottenbacher KJ. Single subject research designs for disability research. *Disabil Rehabil.* 2001; 23(1):1–8. [PubMed: 11213316]
17. Carey LM, Matyas TA. Training of somatosensory discrimination after stroke: facilitation of stimulus generalization. *Am J Phys Med Rehabil.* 2005; 84(6):428–442. [PubMed: 15905657]
18. Diamond MF, Ottenbacher KJ. Effect of a tone-inhibiting dynamic ankle-foot orthosis on stride characteristics of an adult with hemiparesis. *Phys Ther.* 1990; 70(7):423–430. [PubMed: 2192374]
19. Harbst KB, Ottenbacher KJ, Harris SR. Interrater reliability of therapists' judgements of graphed data. *Phys Ther.* 1991; 71(2):107–115. [PubMed: 1989006]

20. Small Clinical Trials: Issues and Challenges. Washington, DC: Institute of Medicine, National Academy Press; 2001. Committee on Strategies for Small-Number-Participant Clinical Research Trials.
21. Ridenour TA, Hall DL, Bost JE. A small sample randomized clinical trial methodology using N-of-1 designs and mixed model analysis. *American Journal of Drug & Alcohol Abuse*. 2009; 35(4): 260–266. [PubMed: 20180679]
22. Van den Noortgate W, Onghena P. Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers*. 2003; 35(1):1–10.
23. Edington ES. Randomized single-subject experimental designs. *Behav Res Ther*. 1996; 34(7):567–574. [PubMed: 8826764]
24. Glass, GV.; Willson, VL.; Gottman, JM. Design and Analysis of Time-series Experiments. Boulder: Colorado Associated University Press; 1975.
25. Harrop JW, Velicer WF. A Comparison of Alternative Approaches to the Analysis of Interrupted Time-Series. *Multivariate Behav Res*. 1985; 20(1):27–44.
26. Hayes SC. Single case experimental design and empirical clinical practice. *J Consult Clin Psych*. 1981; 49(2):193–211.
27. Kazdin AE. Single-case research designs in clinical child-psychiatry. *J Am Acad Child Adolesc Psychiatry*. 1983; 22(5):423–432.
28. Ottenbacher KJ. Analysis of data in idiographic research. Issues and methods. *Am J Phys Med Rehabil*. 1992; 71(4):202–208. [PubMed: 1642819]
29. Manolov R, Solanas A. Comparing $N = 1$ effect size indices in presence of autocorrelation. *Behav Modif*. 2008; 32(6):860–875. [PubMed: 18497425]
30. Manolov R, Solanas A, Leiva D. Comparing "visual" effect size indices for single-case designs. *Methodology: Eur J Res Methods Behav Soc Sci*. 2010; 6(2):49–58.
31. 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. [PubMed: 8555924]
32. Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 Trials in the Medical Literature: A Systematic Review. *Med Care*. 2011; 49(8):761–768. [PubMed: 21478771]
33. Koch LC, Schultz JC, Kontosh LG, Conyers LM. Rehabilitation research in the 21st century: concerns and potential solutions. *Rehabil Educ*. 2006; 20(1):7–19.
34. Barlow, DH.; Hersen, M. Single Case Experimental Designs: Strategies for Studying Behavior Change. New York: Pergamon Press; 1984.
35. Lipsey, MW.; Wilson, DB. Practical Meta-Analysis. Thousand Oaks, Calif.: Sage Publications; 2001.

Figure 1.

AB design with follow-up for one person. Y-axis = time to complete the Timed Up and Go test. X-axis = sequence of observations across six baseline (A), six intervention (B), and two follow-up (FU) sessions. The dashed horizontal line indicates the value that is two standard deviations below the mean of baseline phase scores.

Adapted from Marklund and Klasbo.¹³

Figure 2.

Multiple baseline design across outcomes with follow-up for one person. Y-axis = magnitude of error in perceived joint position. X-axis = sequence of observations across consecutive phases. Vertical lines indicate the transitions from baseline to stimulus-specific intervention, to follow-up phases, respectively.

Adapted from Carey and Matyas.¹⁷

	B	C	C	B	A	B	B	C	
A	A	B	A	C	B	A	C	A	A
	C	A	B	A	C	C	A	B	

A = Barefoot, B = Traditional AFO, C = Dynamic AFO

Figure 3.

Alternating treatment design with follow-up for one person. Upper chart: Y-axis = step length in meters. X-axis = sequence of observations across consecutive phases. Solid vertical lines indicate phase transitions: baseline, intervention, and follow-up. The lower chart illustrates potential patterns of randomly-assigned intervention sequences applied during each session. Note: The lower chart does not show all 12 alternating intervention sequences.

Adapted from Diamond and Ottenbacher.¹⁸