

Harvard Data Science Review • Special Issue 3: Personalized (N-of-1) Trials

The Family of Single-Case Experimental Designs

Leonard H. Epstein¹ Jesse Dallery²

¹Jacobs School of Medicine and Biomedical Sciences, Division of Behavioral Medicine, Department of Pediatrics, University at Buffalo, Buffalo, New York, United States of America,

²Department of Psychology, University of Florida, Gainesville, Florida, United States of America

The MIT Press

Published on: Sep 08, 2022

DOI: <https://doi.org/10.1162/99608f92.ff9300a8>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

Single-case experimental designs (SCEDs) represent a family of research designs that use experimental methods to study the effects of treatments on outcomes. The fundamental unit of analysis is the single case—which can be an individual, clinic, or community—ideally with replications of effects within and/or between cases. These designs are flexible and cost-effective and can be used for treatment development, translational research, personalized interventions, and the study of rare diseases and disorders. This article provides a broad overview of the family of single-case experimental designs with corresponding examples, including reversal designs, multiple baseline designs, combined multiple baseline/reversal designs, and integration of single-case designs to identify optimal treatments for individuals into larger randomized controlled trials (RCTs). Personalized N-of-1 trials can be considered a subcategory of SCEDs that overlaps with reversal designs. Relevant issues for each type of design—including comparisons of treatments, design issues such as randomization and blinding, standards for designs, and statistical approaches to complement visual inspection of single-case experimental designs—are also discussed.

Keywords: single-case experimental designs, reversal designs, multiple baseline designs, personalized medicine

1. Introduction

Single-case experimental designs (SCEDs) represent a family of experimental designs to examine the relationship between one or more treatments or levels of treatment and changes in biological or behavioral outcomes. These designs originated in early experimental psychology research (Boring, 1929; Ebbinghaus, 1913; Pavlov, 1927), and were later expanded and formalized in the fields of basic and applied behavior analysis (Morgan & Morgan, 2001; Sidman, 1960). SCEDs have been extended to a number of fields, including medicine (Lillie et al., 2011; Schork, 2015), public health (Biglan et al., 2000; Duan et al., 2013), education (Horner et al., 2005), counseling psychology (Lundervold & Belwood, 2000), clinical psychology (Vlaeyen et al., 2020), health behavior (McDonald et al., 2017), and neuroscience (Soto, 2020).

SCEDs provide a framework to determine whether changes in a target behavior(s) or symptom are in fact a function of the intervention. The fundamentals of an SCED involve repeated measurement, replication of conditions (e.g., baseline and intervention conditions), and the analysis of effects with respect to each individual serving as his or her own control. This process can be useful for identifying the optimal treatment for an individual (Dallery & Raiff, 2014; Davidson et al., 2021), treating rare diseases (Abrahamyan et al., 2016), and implementing early phase translational research (Czajkowski et al., 2015). SCEDs can be referred to as ‘personalized (N-of-1) trials’ when used this way, but they also have broad applicability to a range of scientific questions. Results from SCEDs can be aggregated using meta-analytic techniques to establish

generalizable methods and treatment guidelines (Shadish, 2014; Vannest et al., 2018). Figure 1 presents the main family of SCEDs, and shows how personalized (N-of-1) trials fit into these designs (Vohra et al., 2016). The figure also distinguishes between experimental and nonexperimental single-case designs. In the current article, we provide an overview of SCEDs and thus a context for the articles in this special issue focused on personalized (N-of-1) trials. Our focus is to provide the fundamentals of these designs, and more detailed treatments of data analysis (Moeyaert & Fingerhut, 2022; Schork, 2022) conduct and reporting standards (Kravitz & Duan, 2022; Porcino & Vohra, 2022), and other methodological considerations are provided in this special issue. Our hope is that this article will inspire a diverse array of students, engineers, scientists, and practitioners to further explore the utility, rigor, and flexibility of these designs.

Single-case Experimental Designs	Reversal (ABA, ABAB, ABACA, etc.)	Multiple Baseline
	N-of-1 trials	Changing Criterion
	Alternating treatment	Hybrid (e.g., reversal + multiple baseline)
Non-Experimental Designs	Bi-phase designs (AB)	1-phase (e.g., B phase only)
	Pre-post designs	Case descriptions

Figure 1. The main family of single-case experimental designs and nonexperimental designs. A = Baseline, B and C refer to different treatments.

The most common approach to evaluating the effectiveness of interventions on outcomes is using randomized controlled trials (RCTs). RCTs provide an idea of the *average* effect of an intervention on outcomes. People do not all change at the same rate or in the same way, however; variability in both how people change and the effect of the intervention is inevitable (Fisher et al., 2018; Normand, 2016; Roustit et al., 2018). These sources of variability are conflated in a typical RCT, leading to heterogeneity of treatment effects (HTE). Research on HTE has shown variability in outcomes in RCTs, and in some studies very few people actually exhibit the benefits of that treatment (Williams, 2010). One approach in RCTs is to assess moderators of treatment response to identify individual differences that may predict response to a treatment. This approach may not limit variability in response, and substantial reduction in variability of treatment for subgroups in comparison to the group as a whole is far from assured. Even if variability is reduced, the average effect for that subgroup may not be representative of individual members of the subgroup.

SCEDs can identify the optimal treatment for an individual person rather than the average person in a group (Dallery & Raiff, 2014; Davidson et al., 2021; Hekler et al., 2020). SCEDs are multiphase experimental designs in which a great deal of data is collected on a single person, said person serves as his or her own control (Kazdin, 2011, 2021), and the order of presentation of conditions can be randomized to enhance experimental control. That is, a person's outcomes in one phase are compared to outcomes in another phase. In

a typical study, replications are achieved within and/or across several individuals; this allows for strong inferences about causation between behavior and the treatment (or levels thereof). Achieving replications is synonymous with achieving experimental control.

We provide an overview of three experimental designs that can be adapted for personalized medicine: reversal, multiple baseline, and combined reversal and multiple baseline designs, and we discuss how SCEDs can be integrated into RCTs. These designs focus on demonstrating experimental control of the relationship between treatment and outcome. Several general principles common to all of the designs are noteworthy (Lobo et al., 2017). First, in many studies, treatment effects are compared with control conditions with a no-intervention baseline as the initial condition. To reduce threats to internal validity of the study, the order of assignment of interventions can be randomized (Kratochwill & Levin, 2010) and, when possible, the intervention and data collection can be blinded. The demonstration of experimental control across conditions or people needs to be replicated several times (three replications is the minimum) to ensure confidence of the relationship between treatment and outcome (Kratochwill et al., 2010; Kratochwill & Levin, 2015). Demonstrating stability of data within a phase or, otherwise, no trend in the direction of treatment effects prior to starting treatment is particularly important. Stability refers to the degree of variability in the data path over time (e.g., data points must fall within a 15% range of the median for a condition). Thus, phase length needs to be flexible for the sake of determining stability and trend within a phase, but a minimum of 5 data points per phase has been recommended (Kratochwill et al., 2013). The focus of the intervention's effects is on clinically rather than statistically significant effects with the target effect prespecified and considered in interpretation of the relevance of the effect for clinical practice (Epstein et al., 2021). In addition, multiple dependent outcomes can be simultaneously measured (Epstein et al., 2021). SCEDs can be used to test whether a variable mediates the effect of a treatment on symptoms or behavior (Miočević et al., 2020; Riley & Gaynor, 2014). Visual inspection of graphical data is typically used to determine treatment effects, and statistical methods are commonly used to assist in interpretation of graphical data (Epstein et al., 2021). Furthermore, a growing number of statistical approaches can summarize treatment effects and provide effect sizes (Kazdin, 2021; Moeyaert & Fingerhut, this issue; Pustejovsky, 2019; Shadish et al., 2014). Data across many SCED trials can be aggregated to assess the generality of the treatment effects to help address for whom and under what conditions an intervention is effective (Branch & Pennypacker, 2013; Shadish, 2014; Van den Noortgate & Onghena, 2003).

2. Reversal Designs

A reversal design collects behavioral or biological outcome data in at least two phases: a baseline or no-treatment phase (labeled as 'A') and the experimental or treatment phase (labeled as 'B'). The design is called a reversal design because there must be reversals or replications of phases for each individual; for example, in an ABA design, the baseline phase is replicated (Kazdin, 2011). Ideally, three replications of treatment effects are used to demonstrate experimental control (Kratochwill et al., 2010; Kratochwill & Levin, 1992). Figure 2

shows hypothetical results from an A1B1A2B2 design. The graph shows three replications of treatment effects (A1 versus B1, B1 versus A2, A2 versus B2) across four participants. Each phase was carried out until stability was evident from visual inspection of the data as well as absence of trends in the direction of the desired effect. The replication across participants increases the confidence in the effectiveness of the intervention. Extension of this design is possible by comparing multiple interventions, as well. The order of the treatments should be randomized, especially when the goal is to combine SCEDs across participants.

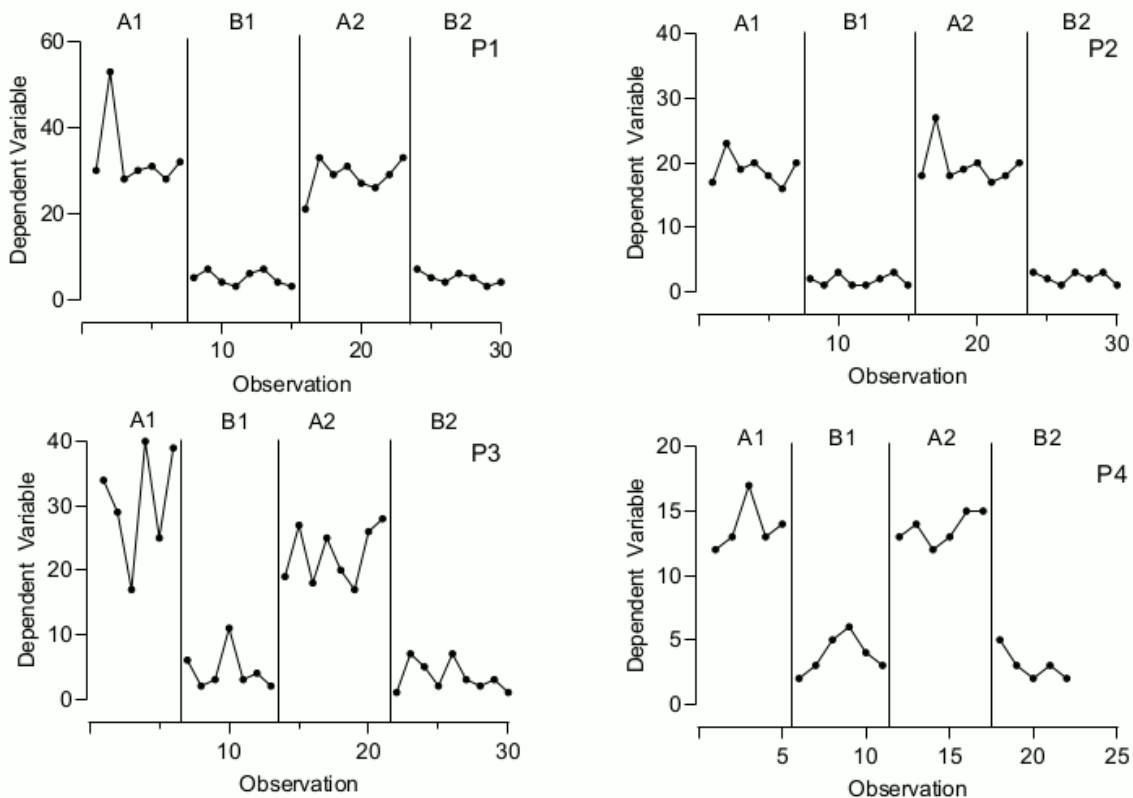


Figure 2. Example of a reversal design showing replications within and between subjects. A1 = First Baseline, B1 First Treatment, A2 = Return to Baseline, B2 = Return to Treatment. P1–P4 represent different hypothetical participants.

Reversal designs can be more dynamic and compare several treatments. A common approach in personalized medicine would be to compare two or more doses of or different components of the same treatment (Ward-Horner & Sturme, 2010). For example, two drug doses could be compared using an A1B1C1B2C2 design, where A represents placebo and B and C represent the different drug doses (Guyatt et al., 1990). In the case of drug studies, the drug/placebo administration can be double blinded. A more complex design could be A1B1A2C1A3C2A4B2, which would yield multiple replications of the comparison between drug and placebo. Based on the kinetics of the drug and the need for a washout period, the design could also be A1B1C1B2C2.

This would provide three demonstration of treatment effects: B1 to C1, C1 to B2, and B2 to C2. Other permutations could be planned strategically to identify the optimal dose for each individual.

Advantages of SCED reversal designs are their ability to experimentally show that a particular treatment was functionally related to a particular change in an outcome variable *for that person*. This is the core principle of personalized medicine: an optimal treatment for an individual can be identified (Dallery & Raiff, 2014; Davidson et al., 2021; Guyatt et al., 1990; Hekler et al., 2020; Lillie et al., 2011). These designs can work well for studying the effect of interventions on rare diseases in which collecting enough participants with similar characteristics for an RCT would be unlikely. An additional strength is the opportunity for the clinical researcher who also delivers clinical care to translate basic science findings or new findings from RCTs to their patients, who can potentially benefit (Dallery & Raiff, 2014; Hayes, 1981). Research suggests that the trickle-down of new developments and hypotheses to their support in RCTs can take more than 15 years; many important advancements in the medical and behavior sciences are likely not to be implemented rapidly enough (Riley et al., 2013). The ability to test new intervention developments using scientific principles could speed up their translation into practice.

Limitations to SCED designs, however, are worth noting. Firstly, in line with the expectation that the outcome returns to baseline levels, reversals may require removal of the treatment. If the effect is not quickly reversible, then the designs are not relevant. A washout period may be placed in-between phases if the effect is not immediately reversible; for example, a drug washout period could be planned based on the half-life of drug. Secondly, the intervention should have a relatively immediate effect on the outcome. If many weeks to months are needed for some interventions to show effects, a reversal design may not be optimal unless the investigator is willing to plan a lengthy study. Thirdly, the design depends on comparing stable data over conditions. If achieving stability due to uncontrolled sources of biological or environmental variation is not possible, a reversal design may not be appropriate to evaluate a treatment, though it may be useful to identify the sources of variability (Sidman, 1960). Finally, for a reversal to a baseline, a no-treatment phase may be inappropriate in investigating treatment effects for a very ill patient.

3. Multiple Baseline Designs

An alternative to a reversal design is the multiple baseline design, which does not require reversal of conditions to establish experimental control. There are three types of multiple baseline designs: multiple baseline across people, behaviors, and settings. The most popular is the multiple baseline across people, in which baselines are established for three or more people for the same outcome (Cushing et al., 2011; Meredith et al., 2011). Treatment is implemented after different durations of baseline across individuals. The order of treatment implementation across people can be randomized (Wen et al., 2019). Figure 3 shows an example across three individuals. In this hypothetical example, baseline data for each person are relatively stable and not decreasing, and reductions in the dependent variable are only observed after introduction of the intervention. Inclusion of one control person, who remains in baseline throughout the study and provides a control for extended

monitoring, is also possible. Another variation is to collect baseline data intermittently in a ‘probe’ design, which can minimize burden associated with simultaneous and repeated measurement of outcomes (Byiers et al., 2012; Horner & Baer, 1978). If the outcomes do not change during baseline conditions and the changes only occur across participants after the treatment has been implemented—and this sequence is replicated across several people—change in the outcome may be safely attributed to the treatment. The length of the baselines still must be long enough to show stability and no trend toward improvement until the treatment is implemented.

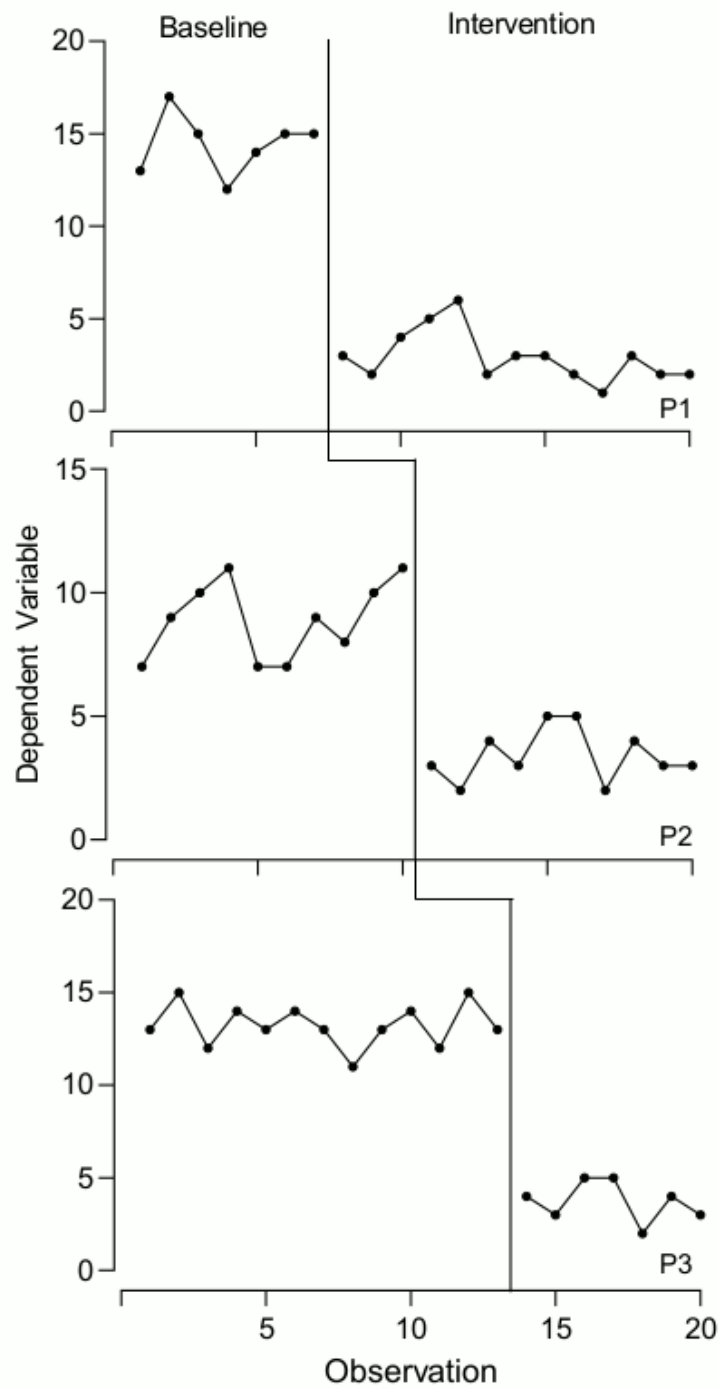


Figure 3. Example of a multiple baseline design showing replications between subjects. P1–P3 represent different hypothetical participants.

The two other multiple baseline designs focus on individual people: the multiple baseline across settings and the multiple baseline across behaviors (Boles et al., 2008; Lane-Brown & Tate, 2010). An example of a

multiple baseline across settings would be a dietary intervention implemented across meals. An intervention that targets a reduction in consumption of high-glycemic index foods, or foods with added sugar across meals, could be developed with the order of meals randomized. For example, someone may be randomized to reduce sugar-added or high-glycemic index foods for breakfast without any implementation at lunch or dinner. Implementation of the diet at lunch and then dinner would occur after different durations of baselines in these settings. An example of multiple baseline across behaviors might be to use feedback to develop a comprehensive exercise program that involves stretching, aerobic exercise, and resistance training. Feedback could target improvement in one of these randomly selected behaviors, implemented in a staggered manner.

The main limitation to a multiple baseline design is that some people (or behaviors) may be kept in baseline or control conditions for extended periods before treatment is implemented. Of course, failure to receive an effective treatment is common in RCTs for people who are randomized to control conditions, but unlike control groups in RCTs, all participants eventually receive treatment.

Finally, while the emphasis in personalized medicine is the identification of an optimal treatment plan for an individual person, situations in which multiple baselines across people prove relevant for precision medicine may arise. For example, identification of a small group of people with common characteristics—perhaps with a rare disease and for which a multiple-baseline-across-people design could be used to test an intervention more effectively than a series of personalized designs—is possible. In a similar vein, differential response to a common treatment in a multiple-baseline-across-people design can help to identify individual differences that can compromise the response to a treatment.

4. Integrating Multiple Baseline and Reversal Designs

While reversal designs can be used to compare effects of interventions, multiple baseline designs provide experimental control for testing one intervention but do not compare different interventions. One way to take advantage of the strengths of both designs is to combine them. For example, the effects of a first treatment could be studied using a multiple-baseline format and, after experimental control has been established, return to baseline prior to the commencement of a different treatment, which may be introduced in a different order. These comparisons can be made for several different interventions with the combination of both designs to demonstrate experimental control and compare effects of the interventions.

Figure 4 shows a hypothetical example of a combined approach to identify the best drug to decrease blood pressure. Baseline blood pressures are established for three people under placebo conditions before new drug X is introduced across participants in a staggered fashion to establish relative changes in blood pressure. All return to placebo after blood pressures reach stability, drug Y is introduced in a staggered sequence, participants are returned to placebo, and the most effective intervention for each individual (drug X or Y) is reintroduced to replicate the most important result: the most effective medication. This across-subjects design establishes experimental control for two different new drug interventions across three people while also

establishing experimental control for five comparisons within subjects (placebo—drug X, drug Y—placebo, placebo—drug Y, drug Y—placebo, placebo—more effective drug). Though this combined design strengthens confidence beyond either reversal or multiple baseline designs, in many situations, experimental control demonstrated using a reversal design is sufficient.

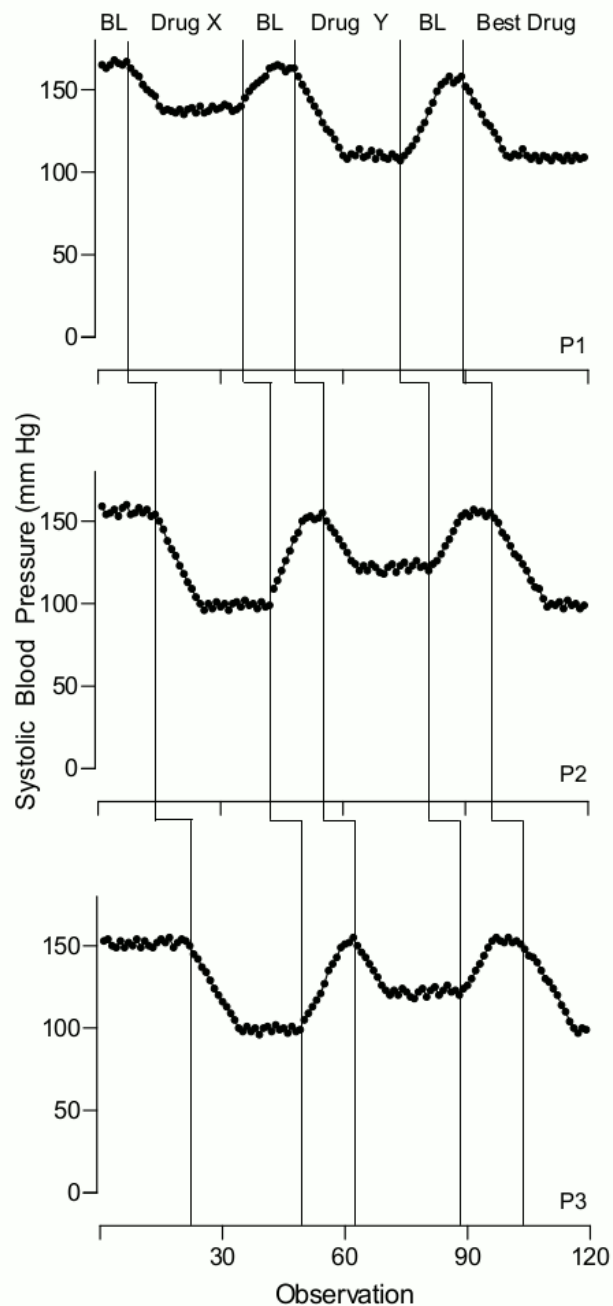


Figure 4. Example of a combined reversal and multiple baseline to determine the best drug to lower blood pressure. BL = Baseline. Drug X and Drug Y represent hypothetical drugs to lower blood pressure, and Best Drug represents a reversal to the most effective drug as identified for each hypothetical participant, labeled P1–P3.

5. Other Varieties of Single-Case Experimental Designs

Other less commonly used designs within the family of SCEDs may be useful for personalized medicine. One of the most relevant may be the alternating treatment design (Barlow & Hayes, 1979; Manolov et al., 2021), in which people are exposed to baseline and one or more treatments for very brief periods without the concern about stability before changing conditions. While the treatment period may be short, many more replications of treatments—and ineffective treatments—can be identified quickly. This type of design may be relevant for drugs that have rapid effects with a short half-life and behavioral interventions that have rapid effects (Coyle & Robertson, 1998)—for example, the effects of biofeedback on heart rate (Weems, 1998). Another design is the changing criterion design, in which experimental control is demonstrated when the outcome meets certain preselected criteria that can be systematically increased or decreased over time (Hartmann & Hall, 1976). The design is especially useful when learning a new skill or when outcomes change slowly over time (Singh & Leung, 1988)—for example, gradually increasing the range of foods chosen in a previously highly selective eater (Russo et al., 2019).

6. Integrating Single-Case Experimental Designs Into Randomized Controlled Trials

SCEDs can be integrated into RCTs to compare the efficacy of treatments chosen for someone based on SCEDs versus a standardized or usual care treatment (Epstein et al., 2021; Schork & Goetz, 2017). Such innovative designs may capture the best in SCEDs and randomized controlled designs. Kravitz et al. (2018) used an RCT in which one group ($n = 108$) experienced a series of reversal AB conditions, or a personalized (N-of-1) trial. The specific conditions were chosen for each patient from among eight categories of treatments to reduce chronic musculoskeletal pain (e.g., acetaminophen, any nonsteroidal anti-inflammatory drug, acetaminophen/oxycodone, tramadol). The other group ($n = 107$) received usual care. The study also incorporated mobile technology to record pain-related data daily (see Dallery et al., 2013, for a discussion of technology and SCEDs). The results suggested that the N-of-1 approach was feasible and acceptable, but it did not yield statistically significant superior results in pain measures compared to the usual care group. However, as noted by Vohra and Punja (2019), the results do not indicate a flaw in the methodological approach: finding that two treatments do not differ in superiority is a finding worth knowing.

Another example of a situation where an integrated approach may be useful is selecting a diet for weight control. Many diets for weight control that vary in their macronutrient intake—such as low carb, higher fat versus low fat, and higher carb—have their proponents and favorable biological mechanisms. However, direct comparisons of these diets basically show that they achieve similar weight control with large variability in outcome. Thus, while the average person on a low-fat diet does about the same as the average person on a low-carb diet, some people on the low-carb diet do very well, while some fail. Some of the people who fail on the low-fat diet would undoubtedly do well on the low-carb diet, and some who fail on the low-fat diet would do well on the low-carb diet. Further, some would fail on both diets due to general problems in adherence.

Personalized medicine suggests that diets should be individualized to achieve the best results. SCEDs would be one way to show ‘proof of concept’ that a particular diet is better than a standard healthy diet. First, people would be randomized to experimental (including SCEDs) or control (not basing diet on SCEDs). Subject selection criteria would proceed as in any RCT. For the first 3 months, people in the experimental group would engage in individual reversal designs in which 2-week intervals of low-carb and low-fat diets would be interspersed with their usual eating, and weight loss, diet adherence, food preferences, and the reinforcing value of foods in the diet would be measured to assess biological, behavioral, and subjective changes.

Participants in the control group would experience a similar exposure to the different types of diets, but the diet to which they are assigned would be randomly chosen rather than chosen using SCED methods. In this way, they would have similar exposure to diets during the first 3 months of the study, but this experience would not impact group assignment. As with any RCT, the study would proceed with regular measures (e.g., 6, 12, 24 months) and the hypothesis that those assigned to a diet that results in better initial weight loss, and that they like and are motivated to continue, would do better than those receiving a randomly selected diet. The study could also be designed with three groups: a single-case design experimental group similar to the approach in the hypothetical study above and two control groups, one low-fat and one low-carb.

An alternative design would be to have everyone experience SCEDs for the first 3 months and then be randomized to either the optimal treatment identified during the first 3 months or an intervention randomly chosen among the interventions to be studied. This design has the advantage of randomization being after 3 months of study so that dropouts and nonadherers within the first 3 months would not be randomized in an intent-to-treat format.

The goal of either hypothesized study, or any study that attempts to incorporate SCEDs into RCTs, is that matching participants to treatments will provide superior results in comparison to providing the same treatment to everyone in a group. Two hypotheses can be generated in these types of designs: first, that the mean changes will differ between groups, and second, that the variability will differ between groups with less variability in outcome for people who have treatment selected after a single-case trial than people who have a treatment randomly selected. A reduction in variability plus mean differences in outcome should increase the effect size for people treated using individualized designs, increase power, and allow for a smaller sample size to ensure confidence about the differences observed between groups.

7. Limitations of Single-Case Experimental Designs

Single-case experimental designs have their common limitations. If a measure changes with repeated testing without intervention, it may not be useful for an SCED unless steps can be taken to mitigate such reactivity, such as more unobtrusive monitoring (Kazdin, 2021). Given that the effects of interventions are evaluated over time, systematic environmental changes or maturation could influence the relationship between a treatment and outcome and thereby obscure the effect of a treatment. However, the design logic of reversal and multiple

baseline designs largely control for such influences. Since SCEDs rely on repeated measures and a detailed study of the relationship between treatment and outcome, studies that use dependent measures that cannot be sampled frequently are not candidates for SCEDs. Likewise, the failure to identify a temporal relationship between the introduction of treatment and initiation of change in the outcome can make attribution of changes to the intervention challenging. A confounding variable's association with introduction or removal of the intervention, which may cause inappropriate decisions about the effects of the intervention, is always possible. Dropout or uncontrolled events that occur to individuals can introduce confounding variables to the SCED. These problems are not unique to SCEDs and also occur with RCTs.

8. Single-Case Experimental Designs in Early Stage Translational Research

The emphasis of a research program may be on translating basic science findings to clinical interventions. The goal may be to collect early phase translational research as a step toward a fully powered RCT—(Epstein et al., 2021). The fact that a large amount of basic science does not get translated into clinical interventions is well known (Butler, 2008; Seyhan, 2019); this served in part as the stimulus for the National Institutes of Health (NIH) to develop a network of clinical and translational science institutes in medical schools and universities throughout the United States. A common approach to early phase translational research is to implement a small, underpowered RCT to secure a ‘signal’ of a treatment effect and an effect size. This is a problematic approach to pilot research, and it is not advocated by the NIH as an approach to early phase translational research (National Center for Complementary and Integrative Health, 2020). The number of participants needed for a fully powered RCT may be substantially different from the number projected from a small-sample RCT. These small, underpowered, early phase translational studies may provide too large an estimate of an effect size, leading to an underpowered RCT. Likewise, a small-sample RCT can lead to a small effect size that can, in turn, lead to a failure to implement a potentially effective intervention (Kraemer et al., 2006). Therefore, SCEDs—especially reversal and multiple baseline designs—are evidently ideally suited to early phase translational research. This use complements the utility of SCEDs for identifying the optimal treatment for an individual or small group of individuals.

9. Conclusion

Single-case experimental designs provide flexible, rigorous, and cost-effective approaches that can be used in personalized medicine to identify the optimal treatment for an individual patient. SCEDs represent a broad array of designs, and personalized (N-of-1) designs are a prominent example, particularly in medicine. These designs can be incorporated into RCTs, and they can be integrated using meta-analysis techniques. SCEDs should become a standard part of the toolbox for clinical researchers to improve clinical care for their patients, and they can lead to the next generation of interventions that show maximal effects for individual cases as well as for early phase translational research to clinical practice.

Acknowledgments

We thank Lesleigh Stinson and Andrea Villegas for preparing the figures.

Disclosure Statement

Preparation of this special issue was supported by grants R01LM012836 from the National Library of Medicine of the National Institutes of Health and P30AG063786 from the National Institute on Aging of the National Institutes of Health. Funding to authors of this article was supported by grants U01 HL131552 from the National Heart, Lung, and Blood Institute, UH3 DK109543 from the National Institute of Diabetes, Digestive and Kidney Diseases, and RO1HD080292 and RO1HD088131 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication. The views expressed in this paper are those of the authors and do not represent the views of the National Institutes of Health, the U.S. Department of Health and Human Services, or any other government entity.

References

- Abrahamyan, L., Feldman, B. M., Tomlinson, G., Faughnan, M. E., Johnson, S. R., Diamond, I. R., & Gupta, S. (2016). Alternative designs for clinical trials in rare diseases. *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics*, 172(4), 313–331. <https://doi.org/10.1002/ajmg.c.31533>
- Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis*, 12(2), 199–210. <https://doi.org/10.1901/jaba.1979.12-199>
- Biglan, A., Ary, D., & Wagenaar, A. C. (2000). The value of interrupted time-series experiments for community intervention research. *Prevention Science*, 1(1), 31–49. <https://doi.org/10.1023/a:1010024016308>
- Boles, R. E., Roberts, M. C., & Vernberg, E. M. (2008). Treating non-retentive encopresis with rewarded scheduled toilet visits. *Behavior Analysis in Practice*, 1(2), 68–72. <https://doi.org/10.1007/bf03391730>
- Boring, E. G. (1929). *A history of experimental psychology*. Appleton-Century-Crofts.
- Branch, M. N., & Pennypacker, H. S. (2013). Generality and generalization of research findings. In G. J. Madden, W. V. Dube, T. D. Hackenberg, G. P. Hanley, & K. A. Lattal (Eds.), *APA handbook of behavior analysis, Vol. 1. Methods and principles* (pp. 151–175). American Psychological Association. <https://doi.org/10.1037/13937-007>

- Butler D. (2008). Translational research: Crossing the valley of death. *Nature*, 453(7197), 840–842. <https://doi.org/10.1038/453840a>
- Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology*, 21(4), 397–414. [https://doi.org/10.1044/1058-0360\(2012/11-0036\)](https://doi.org/10.1044/1058-0360(2012/11-0036))
- Coyle, J. A., & Robertson, V. J. (1998). Comparison of two passive mobilizing techniques following Colles' fracture: A multi-element design. *Manual Therapy*, 3(1), 34–41. <https://doi.org/10.1054/math.1998.0314>
- Cushing, C. C., Jensen, C. D., & Steele, R. G. (2011). An evaluation of a personal electronic device to enhance self-monitoring adherence in a pediatric weight management program using a multiple baseline design. *Journal of Pediatric Psychology*, 36(3), 301–307. <https://doi.org/10.1093/jpepsy/jsq074>
- Czajkowski, S. M., Powell, L. H., Adler, N., Naar-king, S., Reynolds, K. D., Hunter, C. M., Laraia, B., Olster, D. H., Perna, F. M., Peterson, J. C., Epel, E., Boyington, J. E., Charlson, M. E., Related, O., Czajkowski, S. M., Powell, L. H., Adler, N., Reynolds, K. D., Hunter, C. M., ... Boyington, J. E. (2015). From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. *Health Psychology*, 34(10), 971–982. <https://doi.org/10.1037/hea0000161>
- Dallery, J., Cassidy, R. N., & Raiff, B. R. (2013). Single-case experimental designs to evaluate novel technology-based health interventions. *Journal of Medical Internet Research*, 15(2), Article e22. <https://doi.org/10.2196/jmir.2227>
- Dallery, J., & Raiff, B. R. (2014). Optimizing behavioral health interventions with single-case designs: From development to dissemination. *Translational Behavioral Medicine*, 4(3), 290–303. <https://doi.org/10.1007/s13142-014-0258-z>
- Davidson, K. W., Silverstein, M., Cheung, K., Paluch, R. A., & Epstein, L. H. (2021). Experimental designs to optimize treatments for individuals. *JAMA Pediatrics*, 175(4), 404–409. <https://doi.org/10.1001/jamapediatrics.2020.5801>
- Duan, N., Kravitz, R. L., & Schmid, C. H. (2013). Single-patient (n-of-1) trials: A pragmatic clinical decision methodology for patient-centered comparative effectiveness research. *Journal of Clinical Epidemiology*, 66(8 Suppl), S21–S28. <https://doi.org/10.1016/j.jclinepi.2013.04.006>
- Ebbinghaus, H. (1913). *Memory; A contribution to experimental psychology*. Teachers College, Columbia University.
- Epstein, L. H., Bickel, W. K., Czajkowski, S. M., Paluch, R. A., Moeyaert, M., & Davidson, K. W. (2021). Single case designs for early phase behavioral translational research in health psychology. *Health Psychology*,

40(12), 858–874. <https://doi.org/10.1037/hea0001055>

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, 115(27), E6106–E6115. <https://doi.org/10.1073/pnas.1711978115>

Guyatt, G. H., Heyting, A., Jaeschke, R., Keller, J., Adachi, J. D., & Roberts, R. S. (1990). N of 1 randomized trials for investigating new drugs. *Controlled Clinical Trials*, 11(2), 88–100. [https://doi.org/10.1016/0197-2456\(90\)90003-k](https://doi.org/10.1016/0197-2456(90)90003-k)

Hartmann, D., & Hall, R. V. (1976). The changing criterion design. *Journal of Applied Behavior Analysis*, 9(4), 527–532. <https://doi.org/10.1901/jaba.1976.9-527>

Hayes, S. C. (1981). Single case experimental design and empirical clinical practice. *Journal of Consulting and Clinical Psychology*, 49(2), 193–211. <https://doi.org/10.1037/0022-006X.49.2.193>

Hekler, E., Tiro, J. A., Hunter, C. M., & Nebeker, C. (2020). Precision health: The role of the social and behavioral sciences in advancing the vision. *Annals of Behavioral Medicine*, 54(11), 805–826. <https://doi.org/10.1093/abm/kaaa018>

Horner, R. D., & Baer, D. M. (1978). Multiple-probe technique: A variation on the multiple baseline1. *Journal of Applied Behavior Analysis*, 11(1), 189–196. <https://doi.org/10.1901/jaba.1978.11-189>

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2), 165–179. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2004-22378-004&site=ehost-live>

Kazdin, A. E. (2011). Single-case research designs: Methods for clinical and applied settings. In *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). Oxford University Press. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2010-18971-000&site=ehost-live>

Kazdin, A. E. (2021). Single-case experimental designs: Characteristics, changes, and challenges. *Journal of the Experimental Analysis of Behavior*, 115(1), 56–85. <https://doi.org/10.1002/jeab.638>

Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavage, J. A. (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. In *Archives of General Psychiatry*, 63(5), 484–489. <https://doi.org/10.1001/archpsyc.63.5.484>

Kratochwill, T R, Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. What Works Clearinghouse.

Kratochwill, T. R. & Levin, J. R. (1992). *Single-case research design and analysis: New directions for psychology and education*. Lawrence Erlbaum.

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*, 15(2), 124–144.

<https://doi.org/10.1037/a0017736>

Kratochwill, T. R., & Levin, J. R. (2015). *Single-case research design and analysis: New directions for psychology and education*. Routledge. <https://doi.org/10.4324/9781315725994>

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, 34(1), 26–38.

<https://doi.org/10.1177/0741932512452794>

Kravitz, R., & Duan, N. (2022). Conduct and implementation of personalized trials in research and practice. *Harvard Data Science Review*, (Special Issue 3). <https://doi.org/10.1162/99608f92.901255e7>

Kravitz, R. L., Schmid, C. H., Marois, M., Wilsey, B., Ward, D., Hays, R. D., Duan, N., Wang, Y., MacDonald, S., Jerant, A., Servadio, J. L., Haddad, D., & Sim, I. (2018). Effect of mobile device-supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: A randomized clinical trial. *JAMA Internal Medicine*, 178(10), 1368–1377. <https://doi.org/10.1001/jamainternmed.2018.3981>

Lane-Brown, A., & Tate, R. (2010). Evaluation of an intervention for apathy after traumatic brain injury: A multiple-baseline, single-case experimental design. *Journal of Head Trauma Rehabilitation*, 25(6), 459–469.

<https://doi.org/10.1097/HTR.0b013e3181d98e1d>

Lillie, E. O., Patay, B., Diamant, J., Issell, B., Topol, E. J., & Schork, N. J. (2011). The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personalized Medicine*, 8(2), 161–173.

<https://doi.org/10.2217/pme.11.7>

Lobo, M. A., Moeyaert, M., Cunha, A. B., & Babik, I. (2017). Single-case design, analysis, and quality assessment for intervention research. *Journal of Neurologic Physical Therapy*, 41(3), 187–197.

<https://doi.org/10.1097/NPT.0000000000000187>

Lundervold, D. A., & Belwood, M. F. (2000). The best kept secret in counseling: Single-case (N = 1) experimental designs. *Journal of Counseling and Development*, 78(1), 92–102. <https://doi.org/10.1002/j.1556-6676.2000.tb02565.x>

Manolov, R., Tanious, R., & Onghena, P. (2021). Quantitative techniques and graphical representations for interpreting results from alternating treatment design. *Perspectives on Behavior Science*. Advance online publication. <https://doi.org/10.1007/s40614-021-00289-9>

McDonald, S., Quinn, F., Vieira, R., O'Brien, N., White, M., Johnston, D. W., & Sniehotta, F. F. (2017). The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: A systematic literature overview. *Health Psychology Review*, 11(4), 307–323.

<https://doi.org/10.1080/17437199.2017.1316672>

Meredith, S. E., Grabinski, M. J., & Dallery, J. (2011). Internet-based group contingency management to promote abstinence from cigarette smoking: A feasibility study. *Drug and Alcohol Dependence*, 118(1), 23–30.

<https://doi.org/10.1016/j.drugalcdep.2011.02.012>

Miočević, M., Klaassen, F., Geuke, G., Moeyaert, M., & Maric, M. (2020). Using Bayesian methods to test mediators of intervention outcomes in single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, 14(1–2), 52–68. <https://doi.org/10.1080/17489539.2020.1732029>

Moeyaert, M., & Fingerhut, J. (2022). Quantitative synthesis of personalized trials studies: Meta-analysis of aggregated data versus individual patient data. *Harvard Data Science Review*, (Special Issue 3).

<https://doi.org/10.1162/99608f92.3574f1dc>

Morgan, D. L., & Morgan, R. K. (2001). Single-participant research design: Bringing science to managed care. *American Psychologist*, 56(2), 119–127. <https://doi.org/10.1037/0003-066X.56.2.119>

National Center for Complementary and Integrative Health. (2020, May 18). *Pilot studies: Common uses and misuses*. National Institutes of Health. <https://www.nccih.nih.gov/grants/pilot-studies-common-uses-and-misuses>

Normand, M. P. (2016). Less is more: Psychologists can learn more by studying fewer people. In *Frontiers in Psychology*, 7(94). <https://doi.org/10.3389/fpsyg.2016.00934>

Pavlov, I. P. (1927). *Conditioned reflexes*. Clarendon Press.

Porcino, A., & Vohra, S. (2022). N-of-1 trials, their reporting guidelines, and the advancement of open science principles. *Harvard Data Science Review*, (Special Issue 3). <https://doi.org/10.1162/99608f92.a65a257a>

Pustejovsky, J. E. (2019). Procedural sensitivities of effect sizes for single-case designs with directly observed behavioral outcome measures. *Psychological Methods*, 24(2), 217–235. <https://doi.org/10.1037/met0000179>

Riley, A. R., & Gaynor, S. T. (2014). Identifying mechanisms of change: Utilizing single-participant methodology to better understand behavior therapy for child depression. *Behavior Modification*, 38(5), 636–664. <https://doi.org/10.1177/0145445514530756>

Riley, W. T., Glasgow, R. E., Etheredge, L., & Abernethy, A. P. (2013). Rapid, responsive, relevant (R3) research: A call for a rapid learning health research enterprise. *Clinical and Translational Medicine*, 2(1),

Article e10. <https://doi.org/10.1186/2001-1326-2-10>

Roustit, M., Gai, J., Gaget, O., Khouri, C., Mouhib, M., Lotito, A., Blaise, S., Seinturier, C., Subtil, F., Paris, A., Cracowski, C., Imbert, B., Carpentier, P., Vohra, S., & Cracowski, J. L. (2018). On-demand sildenafil as a treatment for raynaud phenomenon: A series of *n*-of-1 trials. *Annals of Internal Medicine*, 169(10), 694–703. <https://doi.org/10.7326/M18-0517>

Russo, S. R., Croner, J., Smith, S., Chirinos, M., & Weiss, M. J. (2019). A further refinement of procedures addressing food selectivity. *Behavioral Interventions*, 34(4), 495–503. <https://doi.org/10.1002/bin.1686>

Schork N. J. (2015). Personalized medicine: Time for one-person trials. *Nature*, 520(7549), 609–611. <https://doi.org/10.1038/520609a>

Schork, N. (2022). Accommodating serial correlation and sequential design elements in personalized studies and aggregated personalized studies. *Harvard Data Science Review*, (Special Issue 3). <https://doi.org/10.1162/99608f92.f1eef6f4>

Schork, N. J., & Goetz, L. H. (2017). Single-subject studies in translational nutrition research. *Annual Review of Nutrition*, 37, 395–422. <https://doi.org/10.1146/annurev-nutr-071816-064717>

Seyhan, A. A. (2019). Lost in translation: The valley of death across preclinical and clinical divide – Identification of problems and overcoming obstacles. *Translational Medicine Communications*, 4(1), Article 18. <https://doi.org/10.1186/s41231-019-0050-7>

Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology*, 52(2), 109–122. <https://doi.org/10.1016/j.jsp.2013.11.009>

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology*, 52(2), 123–147. <https://doi.org/10.1016/j.jsp.2013.11.005>

Sidman, M. (1960). *Tactics of scientific research*. Basic Books.

Singh, N. N., & Leung, J. P. (1988). Smoking cessation through cigarette-fading, self-recording, and contracting: Treatment, maintenance and long-term followup. *Addictive Behaviors*, 13(1), 101–105. [https://doi.org/10.1016/0306-4603\(88\)90033-0](https://doi.org/10.1016/0306-4603(88)90033-0)

Soto, P. L. (2020). Single-case experimental designs for behavioral neuroscience. *Journal of the Experimental Analysis of Behavior*, 114(3), 447–467. <https://doi.org/10.1002/jeab.633>

Van den Noortgate, W., & Onghena, P. (2003). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers*, 35(1), 1–10.

<https://doi.org/10.3758/bf03195492>

Vannest, K. J., Peltier, C., & Haas, A. (2018). Results reporting in single case experiments and single case meta-analysis. *Research in Developmental Disabilities*, 79, 10–18. <https://doi.org/10.1016/j.ridd.2018.04.029>

Vlaeyen, J. W. S., Wicksell, R. K., Simons, L. E., Gentili, C., De, T. K., Tate, R. L., Vohra, S., Punja, S., Linton, S. J., Sniehotta, F. F., & Onghena, P. (2020). From boulder to stockholm in 70 years: Single case experimental designs in clinical research. *Psychological Record*, 70(4), 659–670.
<https://doi.org/10.1007/s40732-020-00402-5>

Vohra, S., Punja, S. (2019). A case for n-of-1 trials. *JAMA Internal Medicine*, 179(3), 452.
<https://doi.org/10.1001/jamainternmed.2018.7166>

Vohra, S., Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Tate, R., Nikles, J., Zucker, D. R., Kravitz, R., Guyatt, G., Altman, D. G., Moher, D., & CENT Group (2016). CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *Journal of Clinical Epidemiology*, 76, 9–17.
<https://doi.org/10.1016/j.jclinepi.2015.05.004>

Ward-Horner, J., & Sturmey, P. (2010). Component analyses using single-subject experimental designs: A review. *Journal of Applied Behavior Analysis*, 43(4), 685–704. <https://doi.org/10.1901/jaba.2010.43-685>

Weems, C. F. (1998). The evaluation of heart rate biofeedback using a multi-element design. *Journal of Behavior Therapy and Experimental Psychiatry*, 29(2), 157–162. [https://doi.org/10.1016/S0005-7916\(98\)00005-6](https://doi.org/10.1016/S0005-7916(98)00005-6)

Wen, X., Eiden, R. D., Justicia-Linde, F. E., Wang, Y., Higgins, S. T., Thor, N., Haghdel, A., Peters, A. R., & Epstein, L. H. (2019). A multicomponent behavioral intervention for smoking cessation during pregnancy: A nonconcurrent multiple-baseline design. *Translational Behavioral Medicine*, 9(2), 308–318.
<https://doi.org/10.1093/tbm/iby027>

Williams, B. A. (2010). Perils of evidence-based medicine. *Perspectives in Biology and Medicine*, 53(1), 106–120. <https://doi.org/10.1353/pbm.0.0132>

©2022 Leonard H. Epstein and Jesse Dallery. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.