

Single-Case Intervention Research in the Health Sciences: Randomization + Replication = Respectability (Almost)

Joel R. Levin^{1,*}, Thomas R. Kratochwill²

¹University of Arizona, USA

²University of Wisconsin – Madison, USA

*Correspondence should be addressed to Joel R. Levin, jrlevin@arizona.edu

Received date: November 19, 2025, **Accepted date:** January 21, 2026

Citation: Levin JR, Kratochwill TR. Single-Case Intervention Research in the Health Sciences: Randomization + Replication = Respectability (Almost). Arch Pharmacol Ther. 2026;8(1):1–5.

Copyright: © 2026 Levin JR, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Elaboration and extensions are provided of an earlier article on the application of time-series-based single-case intervention designs (SCIDs) and analyses for researchers in the health sciences. The research potential, as well as the strengths and weaknesses of such designs, are detailed. In this Comment we expand on the argument that the scientific credibility of SCIDs can be considerably enhanced through a researcher's incorporation of various randomization and replication procedures, attention to unwanted operational effects, along with the adoption of appropriate data-analysis methods.

Keywords: Single-case intervention research, Randomization, Replication, Operational issues, Internal validity, External validity, Construct validity, Researcher and patient bias, Statistical conclusion validity, Scientific credibility

Abbreviations: SCID: Single-Case Intervention Design; SCD: Single-Case Design; RCT: Randomized Controlled Trial

Commentary

A few years ago we presented a class of intervention designs that is relatively unfamiliar to health-sciences researchers [1], namely, single-case intervention designs (SCIDs) out of the applied behavior analysis tradition (see, for example, the *Journal of Applied Behavior Analysis*). The main thrust of that article was to introduce health-sciences researchers to the virtually unlimited possibilities for conducting intervention studies with only scarce resources (specifically, with small numbers of participants, restricted amounts of time, and financial constraints). In that article, the rationale underlying SCIDs was presented, along with the features and procedural characteristics of basic single-case intervention designs. The uninitiated health-science researcher is referred to that article for an introductory presentation, or “primer”, on the logic, design, conduct, analysis, and associated inferences associated with SCIDs. In the present Comment we expand on the argument that the “scientific credibility” of SCID research can be considerably enhanced through the incorporation

of various randomization and replication procedures, along with attention to concerning operational issues. SCIDs, not to be confused with *N*-of-1 trials designs in the medical-research literature (e.g., [2]), are time-series designs that are characterized by a small number of participants and a large number of measurement/observation occasions [3,4]. Briefly: (1) *N*-of-1 trials designs are not as methodologically rigorous with respect to controlling unwanted (or experimentally “confounding”) sources of variation – and therefore not as scientifically credible – as are many SCIDs. In addition: (2) with the vast number of SCID design variations available, there are more design-and-analysis alternatives to address directly a researcher's questions and purposes, as well as to allow for the inclusion of one or more participants; and (3) SCIDs are associated with compendia of desired Standards and Guidelines that researchers must adhere to for their study to be deemed “acceptable” by the scientific community (e.g., [5]). For additional differences between the two methodological approaches, see [1].

The Role of Randomization

As with conventional large-scale randomized controlled trials (RCTs) intervention research, in SCIDs various forms of randomization can turn an observational study into a scientifically credible experimental study [1]. Two common types of randomization in two-treatment (e.g., drug vs. placebo or Drug 1 vs. Drug 2) conventional intervention research involve those that control for or remove unwanted sources of treatment contamination variation. In single-case one-sample alternating treatment or crossover designs a researcher randomly determines, on a participant-by-participant basis, the order in which the two treatments are administered. In two independent-samples designs, a researcher randomly assigns participants to the two different treatments. In addition, these two forms of randomization (phase-order randomization and between-intervention case randomization) can be combined in single-case two independent-samples replicated B (Treatment 1 phase) vs. C (Treatment 2 phase) or crossover designs. Each of these randomization forms serves to prevent a researcher from falsely claiming to have produced an intervention effect either when there is none or when two treatments are essentially equivalent in their efficacy.

As a hypothetical pharmacology example of the latter two independent-samples application, consider the recent speculation that a simple COVID vaccination might prove to be a potential cancer antidote and a likely more economical alternative to a *bona fide* cancer vaccine [6]. One could design a preliminary SCID study in which a handful of participants are randomly assigned to one of two conditions, a cancer vaccine condition and a COVID vaccine condition. After a series of several cancer-indicating baseline blood tests taken over a period of weeks or months in the two experimental conditions, participants are then administered their respective vaccines. Continuing intervention-series blood tests would be taken repeatedly over several weeks or months to track those cancer indicators and to compare them in the two experimental conditions.

As another related example, suppose that a pharmacology researcher wishes to conduct a preliminary investigation of whether a steroid pill such as dexamethasone taken along with a cancer-treatment vaccine is a more effective regimen than the cancer-treatment vaccine alone. Again, in a two independent-samples SCID, patients would be randomly assigned to either an A (baseline phase), B (cancer drug phase) condition or an A (baseline phase), B+C (cancer drug + steroid phase) condition, with cancer-relevant outcomes repeatedly compared over time. For a recent actual health-sciences SCID example, see [7].

Health-sciences researchers can implement two additional forms of randomization to increase the validity of their inferences concerning intervention efficacy: One form is

within-intervention case randomization, where for certain designs (e.g., multiple-baseline designs and multiple-probe designs) the order within the design structure in which participants are treated and tested is randomly determined. A second form of randomization is intervention start-point randomization, where the specific time period or session in which, for each participant, a series of “baseline” (or no-treatment) observations is systematically transitioned to a randomly determined series of intervention (or treatment) observations. This latter form of randomization, which contrasts with a traditional “response-guided” intervention process, serves to remove all potential researcher bias associated with subjective decisions about when to transition from administering a baseline series to administering an intervention series (see, for example, [8–11]).

The Role of Replication

Replication is the second hallmark component of the scientific credibility equation in SCID research. Because SCID research is typically based on small sample sizes, generalization to other participants and contexts may be limited. Within a single-study, replication of procedures and intervention effects can be provided within a single participant (e.g., in ABAB (i.e., two alternating baseline and intervention phases) and alternating treatment designs, (i.e., two-or-more repeated alternating conditions phases) across participants in one-sample multiple-participant designs, and in two independent-samples designs. Replication across studies can be provided through implementation of the same or similar design and procedures with new participants, thereby strengthening the generalizability of a study’s findings.

Attention to Operational Issues

In SCID research, various forms of randomization are generally *necessary* to preserve the integrity of the design and statistical analysis (“internal validity” and “statistical conclusion validity” foci), whereas within- and across-study replication is necessary to permit generalization of the results to other contexts and participant populations (an “external validity” focus) – see, for example, [12]. However, neither randomization nor replication *per se* are *sufficient* to allow for confidence in the study’s operations and conclusions. For design-and-analysis integrity and results generalization to be validly claimed, researchers must pay additional attention to, and have evidence to rule out, specific “operational issues” that could have arisen during the study’s conduct. As was originally noted by [13]:

After the data have been collected, the researcher can examine the *operational issues* of the study *as actually conducted* (including, for example, attention to unwanted baseline trend, outlying observations, missing data, participant attrition, etc.). Operational issues encompass unplanned-for problems and concerns that could occur

during the conduct of the SCID, which, if unattended to, serve to undermine the study's credibility. These issues are akin to the host of confounding procedural, investigator, and participant "effects" that have been enumerated in the conventional group intervention-research literature and that must be taken into account to preserve the study's internal validity (e.g., [12,14,15]).

Researcher-administered and analyzed surveys, questionnaires, checklists, etc., along with patient-produced mental and physical health protocols are particularly susceptible to unintended operational effects – or worse, biased outcomes. Even a carefully conducted SCID experiment does not rule out the possibility that various types of researcher and patient bias can become internal validity threats. At a minimum, patients should not be aware of the specific test responses that are expected or desirable, thereby negating the "demand characteristics" associated with the intervention. If possible, testers/data collectors should not be aware of (i.e., they should be "blind" to) whether the data being collected are part of the baseline series or the intervention series, and in two-treatment studies, they should be "blind" to the patient's experimental condition.

Additional Considerations

To remove the parenthetical "Almost" from this article's title, at least three additional critical actions are necessary. First, the just-mentioned operational issues concern must be attended to and satisfactorily accounted for. Second, adherence to established SCID Standards and Guidelines must be carefully followed (namely, those provided by [5,13,16,17], and others). For example, an adequate number of baseline and intervention outcome measures must be collected for each participant to constitute an acceptable SCID study. Third, the outcome measures taken must possess a high degree of "construct validity" (i.e., that the outcome measures accurately reflect the participants' underlying traits, behaviors, or processes of research interest). With these actions stated and successfully achieved, a strong case can be made that Randomization + Replication = Respectability, where Respectability equates to scientific credibility.

Data-Analysis Comments

In addition, the analysis of SCID data is a specialized topic that should not go unmentioned. Unfortunately, data-analysis options for these designs are not well established in the health sciences research community. As was noted in the opening of this Commentary, SCID data are of a special kind in that they are collected multiple times over the course of a time-series experiment. This feature of data collection requires understanding that multiply measured SCID outcomes within a single participant are invariably "autocorrelated", rather than

"independent" as they are in the singly measured within-participant data of conventional randomized controlled trials studies. The autocorrelated nature of the data in turn means that conventional statistical tests (e.g., traditional *t* and chi-squared tests, analyses of variance, simple regression analyses, etc.) are statistically inappropriate and invalid when applied to the types of experiments outlined here and in our earlier article (see, for example, [18]). Specifically, when conventional statistical procedures are applied to the analysis of the typically autocorrelated data of SCIDs, the resulting statistical conclusions are too liberal, in the sense of their declaring treatment effects "statistically significant" when in fact they are not or declaring effect sizes larger than they actually are. Our point here is that application of conventional statistical tests will often lead to misleading or erroneous conclusions (namely, about the magnitude of statistically based *p*-values and effect sizes, along with decisions about the test's statistical significance), which compromise the "statistical conclusion validity" of the experiment.

What are some appropriate data-analysis options for SCID studies? Although it is beyond the scope of this Commentary to provide the specifics, what can be said here is that different forms of data analysis include visual analysis (e.g., [19,20]), effect-size estimation (e.g., [21,22]), directed acyclic graphs (e.g., [23]), descriptive statistics (numerous studies), randomization tests (e.g., [24]), multilevel modeling (e.g., [25,26]), Bayesian analysis [27,28], and artificial neural networks (e.g., [29]). These various single-case data-analysis methods have different purposes and possess different strengths and weaknesses. The statistical power of different valid quantitative procedures to detect various effect types over time (e.g., strong, weak, immediate, delayed, abrupt, gradual, nonlinear, deteriorating) should also be considered when a researcher is deciding on which statistical test to adopt [1]. Readers are referred to textbooks that feature different procedures related to the analysis and interpretation of SCID data (e.g., [3,4]).

Conclusion

We conclude this Commentary by providing a sample of recent relevant single-case intervention articles [30–37]. Although such techniques might initially seem imposing and challenging to implement in applied research settings, they are now commonly adopted in numerous cognitive, behavioral, mental health and other health-related domains, the latter of which are illustrated through empirical studies in the accompanying References (e.g., [7,8,28,31]). Specifically, as can be appreciated from an inspection of the content in those References, dozens of scientifically credible SCID studies have been conducted with children experiencing cognitive or behavioral difficulties and who are taking various medications. Preliminary SCID investigations of different drug efficacies, dosage levels, or combinations could well

lead to more definitive large-scale RCTs. Researchers in the pharmacological-sciences field are well-advised to take note of SCID methodology, with the goal of implementing the tactics therein to expand the design, conduct, and analysis of their own intervention-research arsenals.

References

1. Levin JR, Kratochwill TR. Randomized Single-Case Intervention Designs and Analyses for Health Sciences Researchers: A Versatile Clinical Trials Companion. *Ther Innov Regul Sci*. 2021 Jul;55(4):755–64.
2. Mirza RD, Punja S, Vohra S, Guyatt G. The history and development of N-of-1 trials. *J R Soc Med*. 2017 Aug;110(8):330–40.
3. Kazdin AE. *Single-case research designs: Methods for clinical and applied settings*. 3rd ed. New York: Oxford University Press; 2021.
4. Kratochwill TR, Levin JR. *Single-case intervention research: Methodological and statistical advances*. Washington, DC: American Psychological Association; 2014.
5. Kratochwill TR, Hitchcock JH, Horner RH, Levin JR, Odom SL, Rindskopf DM, et al. Single-case intervention research design standards. *Remedial and Special Education*. 2013 Jan;34(1):26–38.
6. Jarvis L. What if COVID vaccine could save cancer patients? *Arizona Daily Star*, 2025, Nov. 8, p. A10.
7. Hanniffy J, Kelly ME. Using Behavioural Skills Training with Healthcare Staff to Promote Greater Independence for People Living with Dementia: A Randomised Single-Case Experimental Design. *Behav Sci (Basel)*. 2025 Jun 26;15(7):870.
8. Borst M, Moeyaert M, van Rood Y. The effect of eye movement desensitization and reprocessing on fibromyalgia: A multiple-baseline experimental case study across ten participants. *Neuropsychol Rehabil*. 2024 Dec;34(10):1422–54.
9. Ferron J, Rohrer LL, Levin JR. Randomization Procedures for Changing Criterion Designs. *Behav Modif*. 2023 Nov;47(6):1320–44.
10. Levin JR, Kratochwill TR, Ferron JM. Randomization procedures in single-case intervention research contexts: (Some of) "the rest of the story". *J Exp Anal Behav*. 2019 Nov;112(3):334–48.
11. Onghena P. Randomization tests for extensions and variations of ABAB single-case experimental designs: A rejoinder. *Behavioral assessment*. 1992 Jan 1;14(2):153–71.
12. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin; 2002.
13. Kratochwill TR, Horner RH, Levin JR, Machalicek W, Ferron J, Johnson A. Single-case design standards: An update and proposed upgrades. *J Sch Psychol*. 2021 Dec;89:91–105.
14. Rosenthal R. *Experimenter effects in behavioral research*. New York, NY: Appleton-Century-Crofts; 1966.
15. Weber SJ, Cook TD. Subject effects in laboratory research: an examination of subject roles, demand characteristics, and valid inference. *Psychological bulletin*. 1972 Apr;77(4):273–95.
16. Kratochwill TR, Horner RH, Levin JR, Machalicek W, Ferron J, Johnson A. Single-case intervention research design standards: Additional proposed upgrades and future directions. *J Sch Psychol*. 2023 Apr;97:192–216.
17. Kratochwill TR, Horner RH, Levin JR, Machalicek W, Ferron J, Johnson A. Single-case design standards: An update and proposed upgrades. *Journal of School Psychology*. 2021 Dec 1;89:91–105.
18. Levin JR. Randomized classroom trials on trial. In: Phye GD, Robinson DH, Levin JR, editors. *Empirical methods for evaluating educational interventions*. San Diego (CA): Academic Press; 2005. p. 3–27.
19. Wolfe K, McCammon MN, LeJeune LM, Check AR, Slocum TA. A review of visual analysis reporting procedures in the functional communication training literature. *Sch Psychol*. 2024 Nov;39(6):548–56.
20. Tanious R, Manolov R. *Visual Analysis of Single-Case Experimental Designs Data: Beyond Time-Series Graphs*. *Single Case in the Social Sciences*. 2025 Feb 26;2(1):43–64.
21. Ferron JM, Kirby MS, Lipien L. Fine-grained effect sizes. *Sch Psychol*. 2024 Nov;39(6):613–24.
22. Parker RI, Vannest KJ. Non-overlap analysis for single-case research. In: Kratochwill TR, Levin JR, editors. *Single-case intervention research: Methodological and statistical advances*. Washington (DC): American Psychological Association; 2014. p. 127–51.
23. Hall GJ, Putzeys S, Kratochwill TR, Levin JR. Discovering Internal Validity Threats and Operational Concerns in Single-Case Experimental Designs Through Directed Acyclic Graphs. *Educational Psychology Review*. 2024 Dec;36(4):128.
24. Ferron JM, Levin JR. Single-case permutation and randomization statistical tests: Present status, promising new developments. In: Kratochwill TR, Levin JR, editors. *Single-case intervention research: Methodological and statistical advances*. Washington (DC): American Psychological Association; 2014. p. 153–83.
25. Manolov R, Moeyaert M. Multilevel Model Selection Applied to Single-Case Experimental Design Data. *Journal of Behavioral Education*. 2025 Aug 5:1–6.
26. Shadish WR, Kyse EN, Rindskopf DM. Analyzing data from single-case designs using multilevel models: new applications and some agenda items for future research. *Psychol Methods*. 2013 Sep;18(3):385–405.
27. Grekov P, Pustejovsky JE. A gentle introduction to Bayesian posterior predictive checking for single-case researchers. *Journal of Behavioral Education*. 2026 Jan 17:1–57.
28. Xue Y, Xue Y, Moeyaert M, McMahan D. The effectiveness of augmented reality-based interventions for individuals with

- autism and/or intellectual and developmental disabilities: A Bayesian three-level meta-analysis of single-case experimental design data. *British Journal of Educational Technology*. 2025 Aug 29;55–78.
29. Lanovaz MJ, Bailey JD. Tutorial: Artificial neural networks to analyze single-case experimental designs. *Psychological Methods*. 2024 Feb;29(1):202–18.
30. Falkenström F, Fjällström R, Bengtsson D. The quasi-experimental multiple baseline panel design: A suitable design for psychotherapy outcome research in clinical practice. *Journal of Consulting and Clinical Psychology*. 2025 Oct 30;93(12):777–88.
31. Krasny-Pacini A, Chabran E, Evans J, Clauss F, Sarda MA, Isner-Horobeti ME, et al. A proposed regulatory and ethical framework for the application of single-case experimental design methodology in rehabilitation research and clinical practice. *Neuropsychological Rehabilitation*. 2025 Jun ;35(10):2055–87.
32. Kratochwill TR, Levin JR. Randomization in single-case design experiments: Addressing threats to internal validity. *School psychology*. 2025 Feb 20;40(6):754–68.
33. Kratochwill TR, Levin JR, Morin KL, Lindström ER. Examining and enhancing the methodological quality of nonconcurrent multiple-baseline designs. *Perspectives on Behavior Science*. 2022 Sep;45(3):651–60.
34. Levin JR, Ferron JM, Gafurov BS. Novel randomization tests for two-sample multiple-baseline designs. *Journal of Education for Students Placed at Risk (JESPAR)*. 2022 Oct 2;27(4):353–66.
35. Levin JR, Kratochwill TR, Ferron JM. Randomization procedures in single-case intervention research contexts: (Some of) “the rest of the story”. *Journal of the Experimental Analysis of Behavior*. 2019 Nov;112(3):334–48.
36. Lloveras LA, Tate SA, Vollmer TR, Gravina NE, Dallery J. The Compound Multiple-Baseline Design. *Perspectives on Behavior Science*. 2025 Mar;48(1):133–44.
37. Morin KL, Lindström ER, Kratochwill TR, Levin JR, Blasko A, Weir A, et al. Nonconcurrent multiple-baseline and multiple-probe designs in special education: A systematic review of current practice and future directions. *Exceptional Children*. 2024 Jan;90(2):126–47.