

Single-Case Intervention Designs in Health-Sciences Research: How Do We Randomize These? Let Us Count the Ways

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: August 14, 2026, **Accepted date:** September 25, 2026

Citation: Levin JR, Kratochwill TR. Single-Case Intervention Designs in Health-Sciences Research: How Do We Randomize These? Let Us Count the Ways. Arch Pharmacol Ther. 2026;8(1):53–60.

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

The present Commentary extends the discussion in an earlier Commentary by Levin & Kratochwill (2026) on the promise of randomized single-case intervention designs for researchers in the health sciences. Four different randomization schemes for various single-case intervention designs are introduced, accompanied by actual single-case intervention studies illustrating that such randomization has begun to appear in the scientific research literature.

Keywords: Single-case intervention research, Randomization schemes, Internal validity, External validity, Scientific credibility, Research examples

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

Commentary

Recently health-sciences intervention researchers were introduced to a class of small-sample intervention-research designs that have been popularized in the educational and behavioral science literatures [1–3]. Such methods, known as single-case intervention designs (SCIDs) are part of the time-series research family that permit interventionists to conduct scientifically credible research based on only a few experimental “units” comprised of individual participants, small groups, or other aggregates (see, for example, [4]; SCIDs have their origins in experimental and applied behavioral analysis science [5] and are now adopted extensively in educational and social science research [6–8]). Although most SCID research applications are with small numbers of participants, moderately sized sample studies are possible, as this class of designs is not defined strictly by its sample size.

The major characteristic of this methodology is the repeated assessment of client or patient outcomes, typically with intervention replication in the design but also, as is highlighted throughout this Commentary, with the design-

enhancing option of incorporating various randomization schemes. With SCIDs, data (outcome measures) are typically collected in multiple sessions in each “phase” of the design and are spread over days, weeks, or months. These designs and corresponding analyses can be implemented as stand-alone preliminary/pilot and replication studies or as an embedded companion to large-sample randomized controlled trials studies (RCTs). Quantitative analysis methods associated with such studies afford researchers statistical significance testing with p-values and confidence intervals, effect-size estimators, and meta-analysis procedures [9]. Single-case designs and analyses also enable researchers to examine such issues as whether their intervention-produced effects over time are temporary or permanent, immediate or delayed, abrupt or gradual, reversible or irreversible, and linear or curvilinear, as well as to accommodate additional variables and statistical analyses that address other research questions.

In our earlier Commentary, we introduced SCIDs, explained their logic, and discussed the advantages, cautions, and data-analysis options associated with SCIDs relative to other small-sample intervention designs that are more familiar to health-

sciences researchers. As was also indicated in the Commentary, SCIDs differ both conceptually and methodologically from early N-of-1 studies that commonly appear in the medical research literature (e.g., [10,11]). Most of the N-of-1 studies were of the “multiple crossover trials” variety that involved some form of randomization (e.g., a randomized crossover between treatment and control conditions). More recently, an expanded range of “N-of-1” designs has been suggested beyond these earlier studies [12]. Nevertheless, few of what we call SCIDs can be found in the medical research literature, and studies that have adopted the randomization schemes discussed in this Commentary are limited.

SCIDs include a variety of rigorous, systematic design structures with a priori decision application criteria and built-in replication features both within and across the experimental units. Replication in the design structure elevates the study’s causal inference status. When coupled with various randomization schemes (four of which will be presented here), SCIDs constitute experimental intervention studies with high levels of both “internal and external validity” [13]. Intervention studies with high internal validity permit researchers to rule out potential factors that are confounded with the intervention per se, thereby strengthening causal inferences about the connection between the intervention and the experimental outcome(s). Intervention studies with high external validity are those associated with various forms of participant sampling and results replication that allow researchers to generalize their study’s findings to other participant populations and situational contexts (e.g., [6]). Although committed researchers may strive to attain high marks in satisfying both validity types, more often the two are in a yin-and-yang struggle with one another [14].

Single-Case Intervention Designs

There are several different SCID types, including, among others, ABAB (reversal or withdrawal) designs, alternating intervention designs, changing criterion designs, and multiple-baseline designs [4,5]. It is beyond the scope of this Commentary to discuss the nuances of each design type, but we do provide example applications of each design based on the type of randomization that can be applied. The pictorial inclusion of a much-implemented SCID, the multiple-baseline design, presented here in **Figure 1**, provides a more concrete and readily grasped representation of that design and its sought-after experimental outcomes.

As a single-case intervention design option, the multiple-baseline design possesses methodologically strong characteristics. The design has three variations consisting of the manner in which the intervention is implemented and replicated: across participants (e.g., a drug trial across patients), across settings (e.g., hospital, clinic, home), or across behaviors

(e.g., multiple phobias). We note that the replication “across participants” variation (with at least four patients preferred) is the strongest of the three design variations. A pictorial representation of a hypothetical multiple-baseline design across participants is displayed in Figure 1, a graph presented previously by [1]. There it may be seen that the design includes four cases, or units, depicted in four levels, or tiers, with a built-in “stagger” between adjacent units. In a time-series fashion, measurements or outcomes are obtained for each participant at multiple time periods throughout the study (29 time periods for Participants 1–3 and 32 for Participant 4 in this example). The vertical dashes in each participant’s outcome series represent the point at which the intervention has been introduced, thereby transitioning the participant’s time series from an A (baseline or control) phase to a B (intervention or experimental) phase. Given that a reasonably stable pattern of baseline responding (i.e., a flat series with low variability) has been established for each unit, if the intervention is effective, then the pattern of a unit’s series should change (with respect to mean level, slope, and/or variability) coincident with or shortly following the introduction of the intervention. The specific logic of the multiple-baseline design is that the change should occur sequentially in a downward “stairstep fashion” at the point of intervention only for the intervened participant and not for any other participant, thereby ruling out other potentially contaminating, contributing factors to the change.

In **Figure 1** it may be seen through an informal “visual analysis” [15] that such a baseline pattern with a sequential, intervention-point change does occur in each participant’s outcomes (specifically an anticipated increase in each participant’s mean level), thereby providing replicated evidence for the success of the intervention. Degrees of success are measured by the magnitude of the change and its consistency across the study’s N units. With concurrent multiple-baseline designs, the replication component is based on generalizing across units that are treated and tested at the same chronological time periods, whereas with nonconcurrent designs the generalization is based on treating and testing units at different time periods. Although the latter design is viewed less favorably than the former by most SCID researchers, it has been argued [16,17] that the latter is useful for ruling out external validity confounding factors such as history and maturation [13,18].

Four Design Randomization Schemes and Actual SCID Research Examples

As was recently discussed elsewhere [16], one or more of four different design randomization schemes can be incorporated into various SCIDs to enhance their scientific credibility. These four schemes will now be summarized, along with examples of published studies in which they have been implemented.

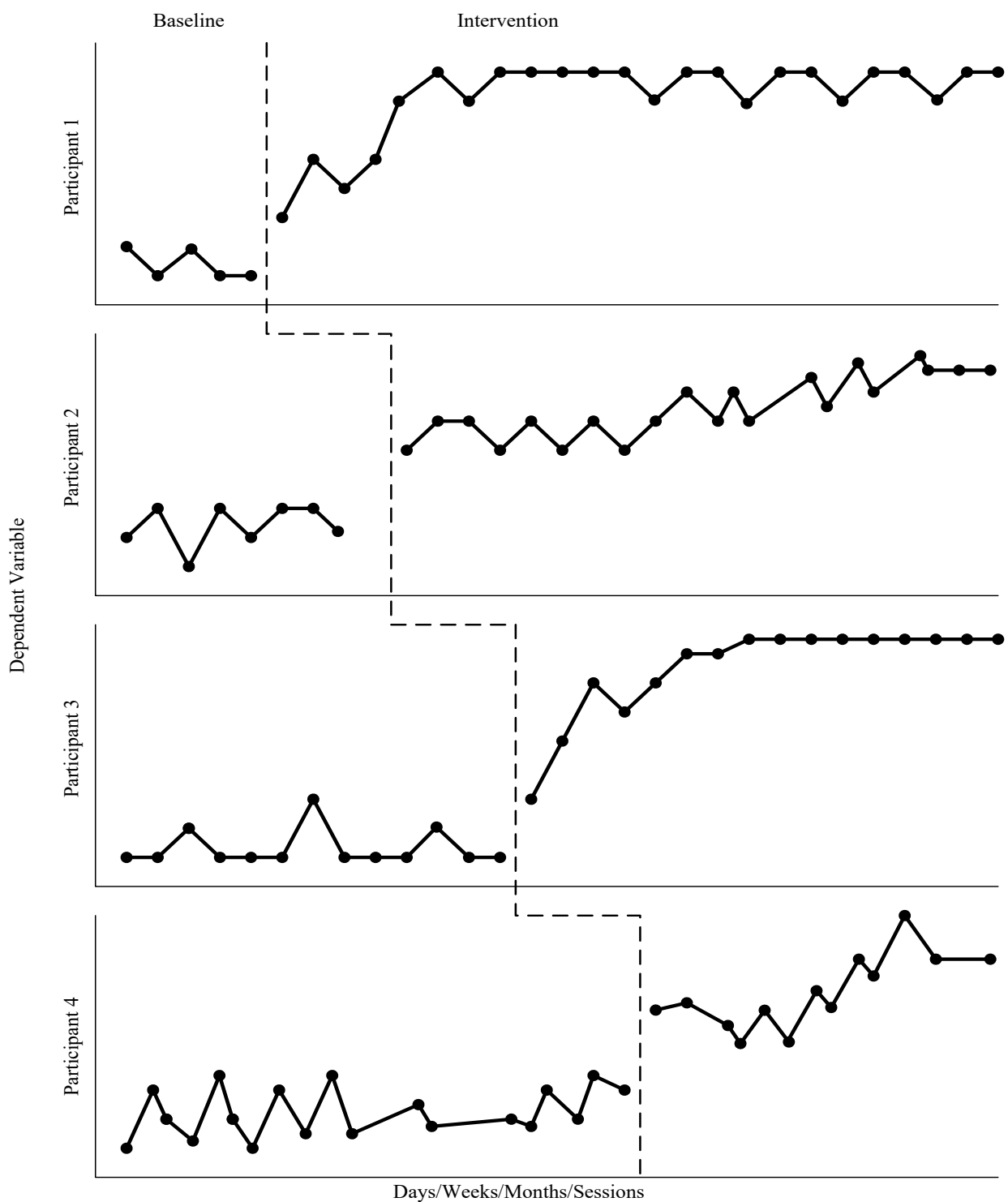


Figure 1. Graphical representation of the outcomes from a hypothetical four-participant multiple-baseline design (From [39]). Given that each participant's staggered outcome increase is coincident with the transition from the baseline (control) series to the intervention (experimental) series, the scientific credibility of the study's findings (that is, the plausibility of a causal relationship between the intervention and the experimental outcomes) is strengthened.

Because of the paucity of randomized SCIDs that have appeared in the health-sciences research literature, all but one of the research examples are taken from published studies in the educational and behavioral sciences domains, but which have direct implications and applications in the health-sciences domain as well. (For a literature review documenting the burgeoning adoption of SCIDs in the instructional, behavioral, and general and special education fields [19,20]).

Extensions of the implications of SCID applications in the educational and behavioral sciences domains to the pharmacological science domain are straightforward: Researchers conducting, say, small-sample drug efficacy intervention studies could administer multiple cycles of treatment followed by assessment (e.g., blood-test measures, patient symptoms, questionnaire well-being self-reports) in which the four current randomization schemes can be adopted, including the “blinding” of patients in drug vs. placebo and comparative drug treatment studies. For drug safety studies, multiple cycles of outcome measures could include patient physiological reactions, illnesses, or mortality.

The three single-case intervention designs (of many possible) on which we focus are the randomized multiple-baseline design, the randomized two independent samples design, and the randomized phase-ordered crossover intervention design. In each of these examples, statistical randomization tests that match the study’s design structure [21,22] were selected for data analysis, although any number of alternative single-case statistical procedures could have been adopted as well [23–25].

Multiple-baseline designs

Despite high marks for its scientific integrity and appeal, the traditional multiple-baseline design described earlier can be substantially improved by the addition of two different randomization schemes to it, each of which will be discussed in turn. With one type of design randomization, referred to as within-intervention case randomization, units are randomly assigned to the N tiers of the design, thereby eliminating the researcher’s intended or unintended biases in assigning participants to tiers based on perceived “intervention need” or “outcome favorability” criteria.

Educational intervention example 1

In a study by Rodgers et al. [26], 10 first- to third-grade below-average readers were randomly assigned to the tiers of a multiple-baseline design [27] to examine the effectiveness of an enhanced reading-program intervention. In the study, all students transitioned from a standard reading program (A phase) consisting of between 7 and 25 outcome measurements across the 10 tiers to an enhanced reading program (B phase) consisting of between 11 and 29 outcome observations, with

a 2-observation stagger between adjacent multiple-baseline tiers. A visual analysis indicated that the enhanced reading-program intervention increased the children’s performance, from modestly to substantially, for 9 of the 10 children on the primary reading measure of interest, along with a statistically significant randomization-test outcome across the 10 children ($p=0.03$) and an effect moderate in magnitude.

Comment: The staggered design controls for temporal simultaneity effects. The within-intervention case randomization controls for researcher bias effects, and the associated statistical analysis takes into account that randomization.

Conclusion: There is scientifically credible evidence for a positive effect of the enhanced reading intervention program on third-grade below-average readers’ reading performance. (In a pharmacology research context, one need only substitute “placebo” for “standard reading program” and “drug” for “enhanced reading program” in the A and B phases, respectively, along with a series of blood-test results for each patient.)

With a second design randomization scheme, referred to as intervention start-point randomization, experimenter bias is again eliminated with respect to the researcher deciding when it would be a “good time” for the intervention phase to commence. Rather, a predetermined interval of “acceptable” intervention start points is decided upon by the researcher or research team and a specific intervention start point is randomly selected (with or without replacement) for each unit from within that interval.

Two behavioral intervention examples

In the context of two actual behavioral intervention research studies, we now describe two different methods for randomly selecting an intervention start-point for each unit from an established intervention start-point interval.

Behavioral intervention example 1

Bice-Urbach et al. [28] sought to determine whether the addition of telecommunication could provide support to expert consultants’ delivery of evidence-based behavioral interventions to 6 behavior-disordered kindergarten through 5th grade students in a teacher-student dyad format. Within a 17–32 outcome observation multiple-baseline design across dyads, each dyad was assigned an intervention start time randomly selected from two potential adjacent start times, starting with Session 8 or 9 for the dyad in Tier 1 and ending with Session 23 or 24 for the dyad in Tier 6. This “dual intervention start-point” approach [29] resulted in a 2- and 4-observation stagger between adjacent dyads and is known to be statistically more powerful than a multiple-baseline

design with only a single intervention start point, as in the [26] study – see [30,31]. For the primary outcome measure, a visual analysis revealed that there was a sizable decrease in the level of “disruptive” behavior for all six of the participating students, which was substantiated across students by a randomization test ($p < 0.001$).

Comment: The staggered design controls for temporal simultaneity effects. The within-intervention case randomization controls for researcher bias effects. The dual intervention start-point randomization also controls for researcher bias effects and serves to increase the power of the associated statistical randomization test. That statistical test takes into account both randomization schemes.

Conclusion: There is scientifically credible evidence for a positive effect of the telecommunication intervention on students’ “disruptive” behavior. (As in the preceding section, in a pharmacology research context one need only substitute “placebo” for “no telecommunication” and “drug” for “telecommunication” in the A and B phases, respectively, along with a series of blood-test results for each patient.)

Behavioral intervention example 2

Lively et al. [32] recruited 3 male students from a juvenile justice facility to study to see whether a self-monitoring worksheet could improve the students’ on-task behavior. Through [33]’s “restricted” sampling-without-replacement adaptation of [34]’s procedure, a multiple-baseline design with 27 outcome observations per student was “constructed” by randomly sampling three nonoverlapping intervention start times from within a 15-session interval. This procedure resulted in intervention start times – following at least 5 required non-self-monitoring (baseline) sessions and at least 5 required self-monitoring (intervention) sessions – for Sessions 6, 11, and 17 in the three tiers, with the three students randomly assigned to tiers. A visual analysis of the outcome data indicated that all three students exhibited a dramatic improvement in their on-task behavior following the administration of the self-monitoring worksheet, with virtually no overlap in each student’s baseline and intervention series. The results of a randomization test applied to the A- to B-phase change in on-task behavior across students produced a p -value of 0.002.

Comment: The staggered design controls for temporal simultaneity effects. The within-intervention case randomization controls for researcher bias effects. The intervention start-point randomization also controls for researcher bias effects and serves to increase the power of the associated statistical randomization test. That statistical test takes into account both randomization schemes.

Conclusion: There is scientifically credible evidence for a dramatic positive effect of the self-monitoring worksheet

intervention on troubled male students’ on-task behavior. (As in the preceding section, in a pharmacology research context one need only substitute “placebo” for “no self-monitoring worksheet” and “drug” for “self-monitoring worksheet” in the A and B phases, respectively, along with a series of blood-test results for each patient.)

Two independent samples designs

As with conventional two independent samples experimental or RCT designs (e.g., drug vs. placebo conditions), in the present context a third type of design randomization, between-intervention case randomization, consists of units being randomly assigned to two (or more) different experimental conditions. The random assignment process is required to turn an observational or quasi-experimental study into a scientifically credible or true experimental study [13].

Educational intervention research example 2

Hwang et al. [35] implemented a multi-component mnemonic (memory-enhancing) strategy to improve children’s learning of the centuries and decades in which various inventions were introduced to the public. The mnemonic strategy had proven to be effective in large-scale two-sample experimental research and the authors wished to determine whether the essential findings of that research could be replicated in a comparable single-case two-sample experiment. The participating children were 8 underperforming students from low socioeconomic backgrounds, with 4 students each randomly assigned to either a Century condition (with mnemonic line drawings that connected the inventions to their centuries) or a Century + Decade mnemonic condition (with mnemonic line drawings that connected the inventions to their centuries and their decades). In 7 experimental sessions wherein the dates of 8 different inventions were presented, students were also randomly assigned an intervention start point beginning in either Session 4 or Session 5. Through a two-sample randomization test of students’ immediate memory for the dates it was found that, from the baseline phase (where nonmnemonic line drawings of the inventions and their dates were presented) to the intervention phase, as was anticipated: (i) across the two experimental conditions, students greatly improved their learning of the inventions’ centuries ($p = 0.008$), with no between-phase overlap for 7 of the 8 students; and (ii) there was no statistical difference in century memory for students in the two conditions ($p > 0.50$). However, for decade memory, although there was a statistical increase across all students $p = 0.01$, which could be attributed to the clear superiority of Century + Decade students (no between-phase overlap for all 4 students) over Century students (complete between-phase overlap for all 4 students), $p < 0.001$.

Comment: A two independent samples SCID design with random assignment of the students to the two

experimental conditions represents between-intervention case randomization, thereby controlling for pre-experimental conditions differences in participant composition, as in conventional RCT studies. The intervention start-point randomization controls for researcher bias effects and serves to increase the power of the associated statistical randomization test. That statistical test takes into account both randomization schemes.

Conclusion: There is scientifically credible evidence of a highly positive effect on memory for the decades of historical inventions by low SES underperforming children in the mnemonic Century + Decade condition relative to the memory children in the nonmnemonic Century condition. These results serve as a replication of results previously obtained in the conventional large-sample intervention research literature [36]. (Similar to what has been stated previously, in a pharmacology research context one need only substitute “no drug” or “placebo” for “nonmnemonic line drawings” in the A baseline phase and “Drug 1” and “Drug 2” for “mnemonic Century line drawings” and “mnemonic Century + Decade line drawings” for the B intervention phase, respectively, coupled with a series of blood-test results for each patient.)

Phase-ordered designs

For the most common SCIDs that encompass the fourth type of design randomization, namely, phase-order randomization, the experimental phases are randomly ordered in some fashion [37,38]. Such SCIDs include: (i) ABAB (reversal or withdrawal) designs for effects that are “reversible”; that is, the effect is expected not to be present or to disappear during the A (baseline or placebo) phases and to be present or to reappear during the B (intervention or active ingredient) phases; (ii) alternating treatment designs based on two different interventions, B and C, where different intervention effects are assumed to go back and forth repeatedly between the B (e.g., Drug 1) and C (e.g., Drug 2) phases and where the two interventions typically are systematically alternated (rather than randomly ordered); and (iii) single-case two-phase crossover designs where, as with their conventional large-sample counterparts, the order in which the B and C intervention phases are presented to each unit can be randomly determined. An actual published randomized-phase crossover-design study follows. For information about how an ABAB design or an alternating AB design can plausibly and acceptably commence with a B (intervention) phase, see [39].

In a pharmacological research context, an example of an ABAB design application for assessing the efficacy of a drug is this: In the initial A (no drug) phase, repeated measurements over several weeks of, say, the patient’s cholesterol blood-test level results are taken. In the initial B (cholesterol-lowering drug) phase, repeated measurements of the patient’s blood-

test results continue to be taken, with the expectation that if the drug is effective, a graph of the patient’s cholesterol level results will reverse and exhibit a decrease. In the second A (no drug) phase, the drug is “withdrawn” for several weeks to see whether a graph of the blood-test results documents another reversal and instead would exhibit an increase in the patient’s cholesterol level. Finally, in the second B (cholesterol-lowering drug) phase, the patient resumes taking the drug for several weeks, along with the blood tests. If the drug is effective, a graph of the blood-test results should again reverse and exhibit a cholesterol-level decrease. A design with outcomes such as these would provide convincing evidence that the drug was effective in reducing the patient’s cholesterol level.

Health-sciences intervention research example

de Jong et al. [40] conducted a single-case crossover design study with patients experiencing pain from post-traumatic neck injuries. They compared two intervention protocols for reducing pain and stress in a sample of 8 individuals who experienced neck pain caused by whiplash more than 12 weeks after a motor vehicle accident. One intervention was the standard “operant graded activity program” (GA) and the other was “graded exposure *in vivo*” (GEXP), a program “that has catastrophic interpretations and associated pain-related fear as its primary target”. Following a two-week baseline period (Phase A in SCID terms), the patients were randomly assigned to one of two intervention orders (phase-order randomization), consisting of 20 one-hour sessions of GA (B) over a 10-week period followed by 12 one-hour sessions of GEXP (C) over a 6-week period (i.e., BC), or the reverse order (CB). A variety of repeated short- and long-term pain and stress measures were taken throughout the duration of the study. Specialized permutation and randomization tests were conducted to compare the GA and GEXP programs on each of the individual patients’ outcome measures through [41]’s SCRT software. The resulting individual p-values were then aggregated to produce an overall across-patients p-value for each measure [21], although no specific p-values were reported. The findings overwhelmingly supported the GEXP program over the GA program for reducing patients’ post-traumatic neck pain and associated fears. However, as [40] caution: “Because the experimental design did not include washout periods between the different treatment components, it is likely that carry-over effects occurred. Indeed, when GA followed GEXP, the improvements remained stable...”

Comment: As in conventional large-sample crossover-design research, randomizing the two intervention orders rules out testing, adaptation, “maturation”, fatigue, etc. as “plausible rival hypotheses” [13] for the superiority of one intervention program over the other – but see the authors’ “carry-over effects” caution. Statistical analyses were conducted that took into account the single-case, time-series nature of the study.

Conclusion: There is scientifically credible evidence of an overwhelmingly positive effect on post-traumatic neck pain-and-fear reduction by patients in the “graded exposure in vivo” program relative to patients in the standard “operant graded activity program”. (Similar to what has been discussed before, in a pharmacology research context one need only insert Drug 1 and Drug 2 for the two crossover-design interventions, along with a series of blood-test results and washout sessions for each patient.) Also see [42] for an instructional strategy study conducted in a crossover design format that incorporates crossover-point randomization (intervention start-point randomization) for each participant, along with composite randomization tests that assess the differential effects of the B and C interventions.

Closing Comments

In sum, single-case intervention designs and analyses can provide health-sciences researchers with a vast array of powerful, versatile new tools for use either in their small-sample preliminary/pilot studies and replication attempts or as companion studies embedded in large-sample investigations. It is hoped that intervention researchers in the health sciences will follow the lead of intervention researchers in the educational and social sciences – and more recently, the recommendations of [12] – with respect to conducting studies based on small samples with SCID methodology. By embracing and adopting the high-quality features that randomized SCIDs have to offer, health-sciences researchers will likely elevate the scientific credibility of their intervention research endeavors. Such approaches will also both extend the scientific literature to complement large-sample RCT studies and promote evidence-based practice in the health sciences.

Acknowledgements

The invaluable assistance of AI Meigin with selected technical aspects of this article is greatly appreciated.

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. Levin JR, Kratochwill TR. Single-Case Intervention Research in the Health Sciences: Randomization + Replication = Respectability (Almost). *Arch Pharmacol Ther.* 2026;8(1):1–5
3. Tanious R, Onghena P. Randomized Single-Case Experimental Designs in Healthcare Research: What, Why, and How? *Healthcare (Basel).* 2019 Nov 13;7(4):143.
4. Horner RH, Odom SL. Constructing single-case research designs: Logic and options. In: Kratochwill TR, Levin JR, editors. *Single-case intervention research: Methodological and statistical advances.* Washington, DC: American Psychological Association; 2014. p. 27–51.
5. Kazdin AE. Single-case experimental designs: Characteristics, changes, and challenges. *J Exp Anal Behav.* 2021 Jan;115(1):56–85.
6. Barlow DH, Nock MK, Hersen M. *Single case experimental designs: Strategies for studying behavior change.* 3rd ed. New York: Pearson; 2009.
7. Kazdin AE. *Single-case research designs: Methods for clinical and applied settings.* 3rd ed. New York: Oxford University Press; 2021.
8. Ledford JR, Gast DL, editors. *Single case research methodology: Applications in special education and behavioral sciences.* 4th ed. New York: Routledge; 2024.
9. Kratochwill TR, Levin JR. *Single-case intervention research: Methodological and statistical advances.* American Psychological Association; 2014.
10. Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. *Medical care.* 2011 Aug 1;49(8):761–8.
11. 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.
12. De Carvalho MCL, de Matos Dourado Simões M, Adler UC, Junior ABV, de Sousa CNS, et al. N-of-1 trials in clinical research: Methodological foundations, statistical approaches and implementation challenges. *Br J Clin Pharmacol.* 2026 Mar;92(3):809–21.
13. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin; 2002.
14. Kazdin AE. Drawing causal inferences from randomized controlled trials in psychotherapy research. *Psychother Res.* 2023 Nov;33(8):991–1003.
15. Kratochwill TR, Levin JR, Horner RH, Swoboda CM. Visual analysis of single-case intervention research: Conceptual and methodological issues. In: Kratochwill TR, Levin JR, editors. *Single-case intervention research: Methodological and statistical advances.* Washington, DC: American Psychological Association; 2014. p. 91–125.
16. Kratochwill TR, Levin JR. Randomization in single-case design experiments: Addressing threats to internal validity. *Sch Psychol.* 2025 Nov;40(6):754–68.
17. Slocum TA, Pinkelman SE, Joslyn PR, Nichols B. Threats to Internal Validity in Multiple-Baseline Design Variations. *Perspect Behav Sci.* 2022 Jan 27;45(3):619–638.
18. Kratochwill TR, Levin JR, Morin KL, Lindström ER. Examining and enhancing the methodological quality of nonconcurrent multiple-baseline designs. *Perspect Behav Sci.* 2022;45:651–60.
19. Kittelman A, Mowery A, Conley KM, Quinn S, Bao W, Campbell AR, et al. A systematic literature review of randomization

-
- procedures within single-case research designs. *J Sch Psychol*. 2025 Jun;110:101458.
20. Tanious R, Onghena P. A systematic review of applied single-case research published between 2016 and 2018: Study designs, randomization, data aspects, and data analysis. *Behav Res Methods*. 2021 Aug;53(4):1371–84.
 21. Gafurov BS, Levin JR. ExPRT (Excel Package of Randomization Tests): Statistical Analyses of Single-Case Intervention Data [Software]. Version 4.4.1. 2026 [Cited 2026 October 02]. Available from: <http://ex-prt.weebly.com>
 22. Onghena P, Edgington ES. Customization of pain treatments: single-case design and analysis. *Clin J Pain*. 2005 Jan-Feb;21(1):56–68.
 23. Kratochwill TR, Ferron J, Levin JR. Reporting guidelines for single-case experimental design research: Recommendations for data analysis. *J Sch Psychol*. 2026 Apr;115:101528.
 24. Manolov R, Moeyaert M. Recommendations for Choosing Single-Case Data Analytical Techniques. *Behav Ther*. 2017 Jan;48(1):97–114.
 25. Manolov R, Rochat L. Analyzing data in single-case experimental designs: objectives and available software options. *J Behav Cogn Ther*. 2024 Dec 1;34(4):100511.
 26. Rodgers E, D'Agostino JV, Levin JR, Rasinski T. Pairing phrase-cued text with readers theatre: Effects on reading prosody and automaticity. *J Res Read*. 2025 May;48(2):153–74.
 27. Wampold BE, Worsham NL. Randomization tests for multiple-baseline designs. *Behav Assess*. 1986;8(2):135–43.
 28. Bice-Urbach BJ, Kratochwill TR. Teleconsultation: The use of technology to improve evidence-based practices in rural communities. *J Sch Psychol*. 2016 Jun;56:27–43.
 29. Koehler MJ, Levin JR. Regulated randomization: A potentially sharper analytical tool for the multiple-baseline design. *Psychol Methods*. 1998 Jun;3(2):206–17.
 30. Ferron J, Sentovich C. Statistical power of randomization tests used with multiple-baseline designs. *The Journal of Experimental Education*. 2002 Jan 1;70(2):165–78.
 31. Levin JR, Ferron JM. Different randomized multiple-baseline models for different situations: A practical guide for single-case intervention researchers. *J Sch Psychol*. 2021 Jun;86:169–77.
 32. Lively R, Myers D, Levin JR. Using self-monitoring to support student behavior in a juvenile justice facility. *J. Corr. Educ.*. 2019;70:36–52.
 33. Levin JR, Ferron JM, Gafurov BS. Comparison of randomization-test procedures for single-case multiple-baseline designs. *Dev Neurorehabil*. 2018 Jul;21(5):290–311.
 34. Marascuilo LA, Busk PL. Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behav Assess*. 1988;10(1):1–28.
 35. Hwang Y, Levin JR. Demonstration of single-case two-sample randomization-test procedures: Underperforming children's implementation of an effective memory strategy. *J Educ Stud Placed Risk (JESPAR)*. 2023 Jul 3;28(3):294–305.
 36. Hwang Y, Levin JR. Examination of middle-school students' independent use of a complex mnemonic system. *The Journal of Experimental Education*. 2002 Jan 1;71(1):25–38.
 37. 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.
 38. Levin JR, Ferron JM, Kratochwill TR. Nonparametric statistical tests for single-case systematic and randomized ABAB...AB and alternating treatment intervention designs: new developments, new directions. *J Sch Psychol*. 2012 Oct;50(5):599–624.
 39. Kratochwill TR, Levin JR. Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psychol Methods*. 2010 Jun;15(2):124–44.
 40. de Jong JR, Vangronsveld K, Peters ML, Goossens ME, Onghena P, Bulté I, et al. Reduction of pain-related fear and disability in post-traumatic neck pain: a replicated single-case experimental study of exposure in vivo. *J Pain*. 2008 Dec;9(12):1123–34.
 41. Onghena P, Van Damme G. SCRT 1.1: Single-case randomization tests. *Behav Res Methods Instrum Comput*. 1994;26:369.
 42. Hwang Y, Levin JR, Johnson EW. Pictorial mnemonic-strategy interventions for children with special needs: Illustration of a multiply randomized single-case crossover design. *Dev Neurorehabil*. 2018 May;21(4):223–37.
-