



Randomized Single-Case Intervention Designs and Analyses for Health Sciences Researchers: A Versatile Clinical Trials Companion

Joel R. Levin, PhD¹ · Thomas R. Kratochwill, PhD²

Received: 1 October 2020 / Accepted: 7 March 2021 / Published online: 1 April 2021
© The Drug Information Association, Inc 2021

Abstract

Aim An expository note introduces health sciences researchers to randomized single-case intervention designs, an adaptation of interrupted time-series methodology, and the staple of a scientifically credible small-sample research paradigm.

Methods Detailed examples illustrating two different randomized single-case procedures are presented to highlight the techniques' advantages relative to small-sample nonparametric procedures that are commonly applied in the medical and health sciences fields.

Results Numerous positive outcomes, based on both statistical simulation studies and actual intervention research investigations, support the applicability and value of these procedures.

Conclusion Randomized single-case intervention designs are recommended for consideration by health sciences researchers.

Introduction and Background

From our field of interest and expertise in psychological intervention research, we introduce a valuable design-and-analysis methodology that may be of consequence to health sciences intervention researchers. In 2010, a lengthy article appeared in *Psychological Methods* on the value of conducting randomized experiments in an area of quantitative experimental research known as single-case intervention designs (SCIDs; [1]). In medical research contexts, these experiments have traditionally fallen under the rubric of “N-of-1 Trials” (e.g., [2]) and both these and SCIDs are a class of interrupted time-series design born out of the statistics and econometrics literatures (see, for example, [3]).

Early systematically controlled interrupted time-series designs were being implemented in other domains, but primarily in certain branches of psychology such as behavioral and clinical (e.g., the *Journal of Applied Behavior Analysis*)—for examples, and discussion, see [4]. Such designs are characterized by the inclusion of only a few participants or

other entities (“cases”), two or more “phases” (e.g., baseline and intervention phases), and multiple outcome observations per case (see Fig. 1 for a hypothetical “multiple-baseline” design), and which should not be confused with traditional “clinical case reports” or “case studies” based on a specific individual’s ongoing records and protocols. Of interest from a historical perspective, medical case studies date back to research directed at combating the 1918 flu pandemic [5]. British medical scientists, in search of the “active transmission ingredients” of the flu virus, conducted a series of several controlled (to the best of their abilities) investigations, in which the virus was introduced through different means and in different combinations to individual ferrets and white mice (pp. 73–78).

Basic Differences Between N-of-1 Trials Designs and Randomized SCIDs and Analyses

In the medical literature, N-of-1 trials methodology has been featured in two major applications: one in clinical practice and the other in clinical science research [6]. Their use in evaluation of practice in the social and behavioral science intervention research literatures has been suggested for many years [7–9]. However, the failure to implement this methodology in practice was recognized decades ago and is based on a number of considerations, including the multiple facets

✉ Joel R. Levin
jrlevin@u.arizona.edu

¹ Emeritus of Educational Psychology, University of Arizona, 3850 E. Placita de Peri, Tucson, AZ 85718, USA

² Emeritus of Educational Psychology, University of Wisconsin-Madison, S 7572 Oak Road, Merrimac, WI 53561, USA

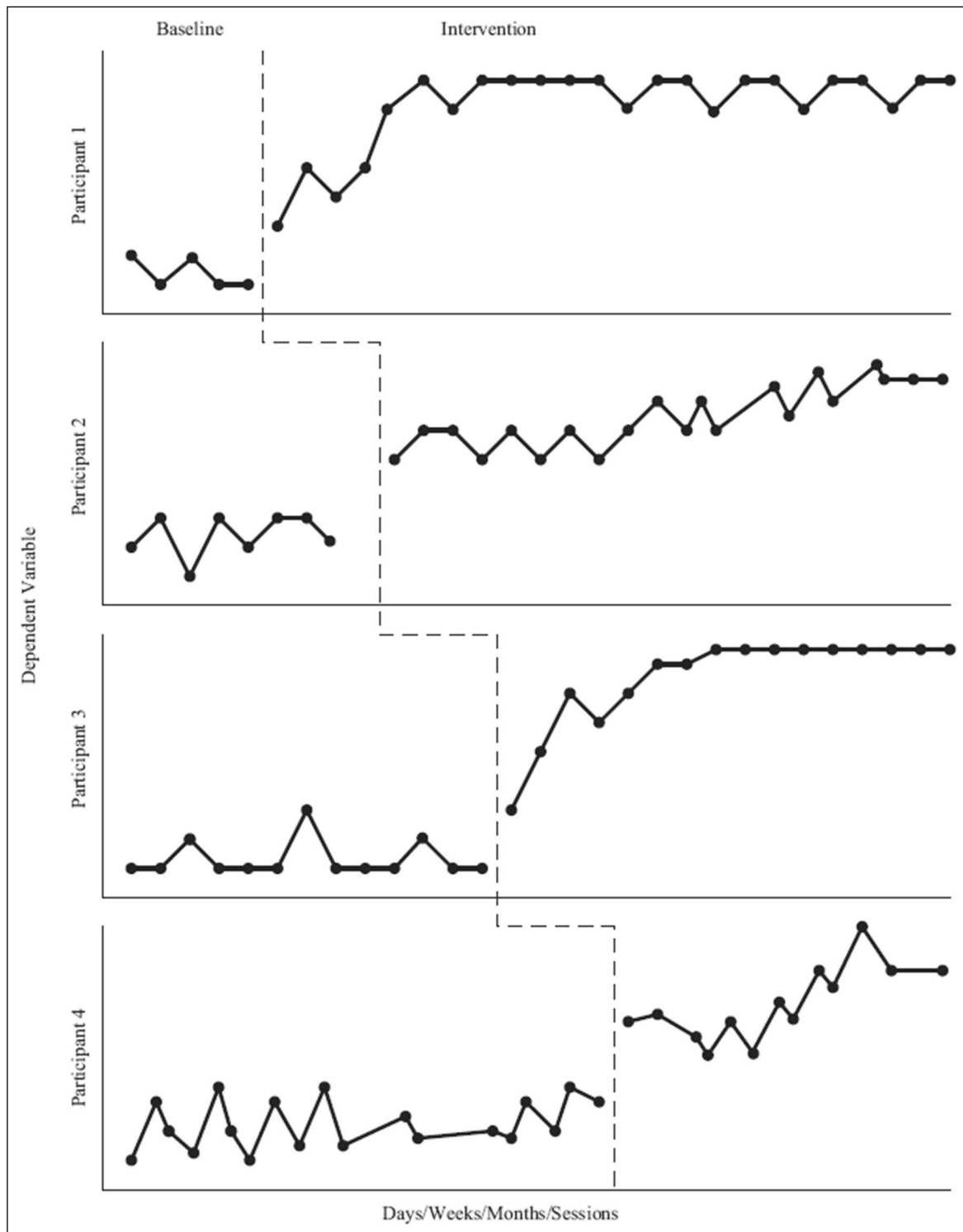


Figure 1 Schematic of a One-Condition Multiple-Baseline Design Based on Four Participants and 29 Outcome Observations. The Dashed Vertical Lines Represent the Participants' Staggered Intervention Start Points (From [1])

of *N*-of-1 designs, challenges in measurement, and overall resource and time commitments required [4]. It is evident that there is a major lack of cross-fertilization in citation

of the medical and the social science intervention research literatures. In particular, and with few exceptions, such as the international work of Tate et al. [10, 11], there is a decisive

failure of behavioral science researchers to cite and adopt advances in medical researchers' *N*-of-1 research design methodology, measurement, and data analyses – and, as was alluded to above, vice versa with respect to the implementation of single-case intervention designs (SCIDs) in medical intervention research.

For an informative primer on *N*-of-1 trials procedures and analyses in the medical intervention research field, see [12]. There are a number of similarities between *N*-of-1 trials structures and procedures and those of randomized SCIDs, but there are important differences as well. Based on our review of the *N*-of-1 trials literature, let us indicate the apparent differences that exist between *N*-of-1 trials designs and randomized SCIDs. First, *N*-of-1 trials procedures were devised primarily as one-case replicated crossover designs: “*N*-of-1 trials in clinical medicine are multiple crossover trials [akin to alternating treatment designs in the SCID literature], usually randomized and often blinded, conducted in a single patient” [13, p. 1]. In contrast, randomized SCIDs include more varied and versatile design structures and are often implemented with multiple cases [13]. Second, novel randomization schemes can be incorporated into SCIDs, which generally have not been considered in *N*-of-1 trials studies. Third, the systematic alternation of A and B phases in *N*-of-1 trials studies is not conducive to enhance the precision of the statistical analyses that are typically conducted, whereas randomized SCID features work in concert with statistically valid randomization test analyses and various effect size measures. Fourth, unlike certain types of invalid statistical tests that are often conducted in *N*-of-1 trials studies (e.g., conventional parametric and nonparametric *t*-tests—see [12] and [14]), in randomized SCIDs attention is always paid to the outcome observations' autocorrelation, which is taken into account in the randomization test procedures that are applied. For a given case's series, the autocorrelation coefficient reflects the degree to which each outcome observation is correlated with each successive outcome observation, as has been recognized by some medical intervention researchers (e.g., [15]). Fifth, the *N*-of-1 trial models that respect the autocorrelated nature of the data are generally quite complex (e.g., time-series analyses, hierarchical linear modeling), whereas the SCID randomized test procedures are straightforward and easy for an intervention researcher to implement and interpret. Sixth, SCID randomization tests serve as a valuable complement to visual analyses of the data, which have traditionally been the primary assessment tool for SCID researchers in the social and behavioral sciences.

When a SCID researcher adopts novel forms of randomization and accompanying data analyses, along with stringent experimental controls, it is possible to produce experimental results that rival the scientific credibility of those produced in “gold standard” conventional large-scale randomized

clinical trials. We have illustrated how SCID researchers are able to conduct a wide variety of randomized experiments with very small samples, an approach that can be more expeditious, more economical, more feasible, and often more illuminating than the clinical trials findings associated with experimental psychological and medical intervention research [10, 11]. Importantly, SCID studies based on participants randomly selected from specified populations, as well as replicated findings of individual SCID investigations, permit limited generalizations of SCID conclusions. The randomized SCIDs to be presented here are “fixed” in the sense that all design- and procedure-related decisions are made pre-experimentally, in advance of data collection. Other more adaptive-like (e.g., [16, 17]) SCIDs that incorporate a randomization component include probabilistically grounded “response-guided” designs, and they are discussed elsewhere [18, pp. 160–172]. Additional consideration of adaptive designs is included later in this article.

Methods

Four Types of Randomization to Improve an SCID Study's Internal Validity and Statistical Power

In SCIDs, four distinct types of randomization can be incorporated by the researcher [18, 19], either individually or in combination, to increase the “internal” and “statistical conclusion” validities of the study [20].

Within-Case Intervention-Order Randomization

This form of randomization is implemented when each case is to receive both A (Baseline or Placebo) and B (Intervention) phases or B (Intervention 1) and C (Intervention 2) phases, in two- or multiple-phase crossover designs and in single-case “alternating treatment” designs. Most intervention researchers routinely adopt this form of randomization so that the order in which the two phases are administered is not confounded with the targeted intervention as a result of adaptation, practice, fatigue, and the like. As will be illustrated here, in SCID designs, within-case randomization also serves to increase the statistical power of the randomization tests that are conducted to analyze the data.

Between-Case Intervention Randomization

This form of randomization is implemented when some cases receive only one of the preceding two intervention options and other cases receive the other option, as is characteristic in conventional two independent samples and matched-pairs “group” designs. Randomly assigning interventions to cases in SCIDs counteracts any biases that

could arise from the initial inequality of the two intervention options and is required for a valid inferential statistical test to be conducted.

Case Randomization

In SCID multiple-baseline studies (Fig. 1, and to be described later), randomly assigning the N cases to the N different staggered intervention start positions within the design guards against biases associated with preferential case placement within the study. Such randomization is also required for a valid inferential statistical to be performed.

Intervention Start-Point Randomization

This form of randomization can be implemented in all SCIDs and serves to eliminate bias arising from a researcher's decision about when the "preferred" time is to transition from one phase to the next for each case. With this randomization strategy, the specific process of which is described later and the statistical power of the associated inferential randomization test that is conducted are greatly improved.

Hypothetical Example of a Two-Condition SCID Intervention Study

We now provide a hypothetical example, contrasting a small-sample design commonly applied by health sciences researchers with a few alternative randomized SCIDs that we are promoting here. Suppose that a new cholesterol-lowering drug has been developed, which is thought to be superior in most respects to a current alternative drug. As a precursor, or as a companion, to a large-sample randomized clinical trial, a much smaller-scale experiment is to be conducted with an available sample of 8 participants.

A Common Small-Sample Two-Condition Design for Health Sciences Researchers

For the present example, suppose that from the eight participants, four are randomly assigned to each of the two experimental conditions, new drug treatment and a placebo (or a current alternative). The data are collected and analyzed in a double-blind fashion. The data consist of pre- and post-intervention lipid-panel outcome measures, along with other ancillary related and side-effect variables of interest (e.g., blood pressure, pulmonary function). For illustrative purposes, say that there is one pre-intervention measure and one post-intervention measure for each outcome variable of interest. After the new drug or placebo intervention has been administered, the pre-to-post-average difference (mean change) data associated with each lipid-panel variable are

statistically analyzed by means of a two-sample nonparametric randomization t -test or by its rank-test analog, such as Wilcoxon, Mann–Whitney, or Kruskal–Wallis test (e.g., [21–23]). It should be noted that for each analyzed outcome measure, even if the single most extreme difference is possible between the two conditions emerged (viz., where the outcomes of all four new drug participants are more positive than the outcomes of all four placebo participants), the smallest statistical significance probability (p value) would be $2/70 = 0.028$ for a two-tailed test and 0.014 for a one-tailed test. For any other less extreme difference, the result would not be statistically significant even at the $\alpha = 0.05$ level for a two-tailed test. Unless the new cholesterol-lowering drug is far superior to the placebo, it can be shown for this particular example based on a total of 8 participants and a common two-sample nonparametric randomization test, the likelihood of statistically documenting the new drug's effectiveness (i.e., the test's statistical power) is quite low.

A Randomized Two-Condition SCID Design

A fundamental characteristic of SCIDs is that for each participant, there is (1) an A (baseline or control/placebo) phase, analogous to a traditional pretest, but consisting of multiple observations of a given outcome variable, as well as (2) a B (intervention or experimental) phase, analogous to a traditional posttest, and consisting of multiple observations of the same outcome variable [24] – see Fig. 1. The inclusion of multiple A- and B-phase outcome observations improves the stability/reliability of the outcome variable and with it, the statistical power of the accompanying statistical test (e.g., [19]). For the present two-condition example, let us suppose that there are 12 outcome observations for all participants.

A randomized SCID is constructed through direct extensions of the common small-sample two-condition intervention study that was just described, beginning with the assumption that participants are again randomly assigned to the two experimental conditions B (Intervention 1) and C (Intervention 2 or Placebo). As was indicated earlier, each of these to-be-described extensions serves to improve the methodological rigor (the internal validity of the research design) and the statistical power (the statistical conclusion validity) of the data analysis, and with them the scientific validity, or "credibility" [25], of the research.

Extension 1: Random Assignment of Participants to Staggered Tier Positions

Likely the most popular, as well as the most methodologically sound, SCID is the multiple-baseline design, wherein participants are systematically assigned to time-staggered intervention start-point positions, or "tiers," of

the design, as is represented in Fig. 1 for a one-condition design. For our present two-condition example based on 12 observations per participant, it might be decided that the intervention will be administered just before the 4th outcome observation for one participant in each condition (A-B and A-C), just before the 6th outcome observation for another participant in each condition, just before the 8th for another, and just before the 10th for another. With that process, at least 3 of the outcome observations will be associated with each participant's A phase and at least 3 of them will be associated with each participant's B or C phase. When, in addition, the participants are *randomly* assigned to those tier positions, viable statistical randomization test possibilities become available [26–28]. Convincing evidence of an intervention effect is present when expected changes in each participant's outcome measure is coincident with the point of, or just following, the introduction of the intervention (i.e., at each participant's staggered intervention start-point position). In Fig. 1, a case-by-case "horizontal analysis" [29] indicates that the start of each case's increase in level is coincident with the introduction of the case's intervention start point (i.e., Session 4 for Participant 1, Session 6 for Participant 2, Session 8 for Participant, 3, and Session 10 for Participant 4). At the same time, a session-by-session "vertical analysis" [29] reveals that as each participant begins its increase in level, each of the lower-tier participants does not, remaining at the same baseline level. Analogously, in a two-sample context (as in the present example), the same "coincident" criterion would apply to between-condition differences.

In the context of the present two-condition example, with four participants randomly assigned to the tiers in both the new drug (B) and the current alternative (C) conditions, the most extreme difference between the two conditions' mean outcomes would be associated with a two-tailed p value of 0.00005 and with a one-tailed p value of 0.000025. Less extreme differences between the two conditions' mean outcomes would also yield statistically significant results. Of more meaningful "significance," the statistical power of the resulting SCID randomization test conducted to document the new drug's greater effectiveness is now increased, relative to the previously described health sciences researcher's common two-sample nonparametric randomization test. Even with only three participants randomly assigned to each experimental condition, the most extreme between-conditions mean difference would yield two- and one-tailed p values of 0.0027 and 0.0014, respectively, for the SCID randomization test, which can be shown to be more powerful than the common two-sample nonparametric randomization test with 4 participants per condition.

Extension 2: Random Assignment of an Intervention Start Point to Each Participant

The most ingenious randomized SCID extension of the common two-sample statistical procedure, initially proposed by [30] and adapted by [31], is that each participant's intervention start point should be randomly sampled from an "acceptable" interval of potential intervention start points. So, let us say for the present example based on 12 outcome observations per participant, it had been decided that the new drug (B) or the current alternative (C) could be administered to each participant anywhere from just before the 5th outcome observation to just before the 9th outcome observation, resulting in 5 potential intervention start points for each participant. That decision would ensure that each participant would furnish at least 4 A-phase observations and 4 B- or C-phase observations in the design and for the statistical analysis. Additional planned observations can be included at the discretion of the researcher and depending on the intervention under consideration. As with Extension 1, adopting this design and analysis would result in considerably more statistical power relative to a common small-sample two-condition nonparametric randomization test.

Combination of Extensions 1 and 2

Combining the SCID design and-analysis tactics of Extensions 1 and 2 is also possible [32, 33]. For our present example based on eight participants, suppose that we could include 15 outcome observations per participant. With that, we could randomly assign 4 participants in each condition to the 4 staggered positions within a multiple-baseline design (Extension 1). The first participant in each condition could then be randomly assigned an intervention start point that is just before either Observation 3 or 4; the second participant, just before either Observation 6 or 7; the third, just before either Observation 9 or 10; and the fourth, just before either Observation 12 or 13 (Extension 2). Implementing this combined design and analysis further increases the power of the associated statistical randomization tests, which would clearly be of benefit to health sciences intervention researchers.

An Additional Valuable SCID for Health Sciences Intervention Researchers

In addition to the two-sample procedures just discussed, two-period crossover designs from the conventional large-sample literature have been developed for SCID researchers [34] and are most commonly implemented in the N -of-1 trials designs of the health sciences field [2]. In a nutshell,

with two different intervention conditions (or an intervention and a placebo condition), B and C, cases are initially randomly assigned in equal numbers to either B or C for a series of outcome observations to form a balanced-order (or completely counterbalanced) design. Then, following a predetermined long enough “washout” interval included to control for plausible drug or other intervention carryover effects, at some randomly determined acceptable crossover point for each case based on a required minimum number of B and C outcome observations, Condition B cases are crossed over to Condition C, and vice versa for Condition C cases, for a continuing series of outcome observations (as is diagrammed in Fig. 2, and which includes a randomized intervention start point for each case).

As with the conventional large-sample crossover design, the SCID crossover design analog enables a researcher to determine whether there was a *general* intervention effect (i.e., a change from the B phase to the C phase averaged across the two experimental conditions), as well as the more critical *differential* intervention effect favoring one experimental condition over the other (e.g., Condition C > Condition B). Importantly, as was the case for the previously discussed two-sample SCIDs, randomized single-case crossover designs can be constructed to have respectable statistical power for detecting the differences between intervention conditions.

Statistical Power Advantages Associated with Randomization Test Procedures

We now provide empirical evidence, derived from large-scale Monte Carlo simulation studies, documenting certain power benefits associated with randomization statistical tests under various experimental conditions.

	Week														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Case 1	B	B	B	B	B	B	B	C	C	C	C	C	C	C	C
Case 2	C	C	C	C	C	C	B	B	B	B	B	B	B	B	B
Case 3	C	C	C	C	C	C	C	C	C	B	B	B	B	B	B
Case 4	B	B	B	B	B	C	C	C	C	C	C	C	C	C	C

Figure 2 Hypothetical Single-Case Randomized Crossover Design with Two Interventions, B and C. Half of the Cases Are Randomly Selected to Receive a B-C Order of Intervention Administration and Half to Receive a C-B Order. With 15 Sessions and a Minimum of 5 Sessions Required for Each Intervention, Each Case Receives a Crossover Start Point Randomly Selected Between Week 6 and Week 10 Inclusive

First, we consider a multiple-baseline design based on four participants, 30 total observations, an autocorrelation of 0.30, and a one-tailed Type I error probability of 0.05 [35]. With a statistical randomization test that assumes only random assignment of participants to a single staggered intervention start point, the power to detect a Cohen *d* effect size of 1.0 is 0.53; and for a randomization test in which each case has been assigned to one of two “acceptable” randomly determined staggered intervention start points, the power is 0.63. The respective powers for an effect size of 1.5 are 0.82 and 91.

In another multiple-baseline simulation study [33] based on four participants, 19 total observations, an autocorrelation of 0.30, a one-tailed Type I error probability of 0.05, and either one fixed intervention staggered start point or one of two randomly assigned staggered intervention start points for each case, the respective powers to detect a 1.0 effect size were 0.41 and 0.51, and the respective powers to detect a 1.5 effect size were 0.66 and 0.81. The powers for a randomization test based three potential intervention start points were virtually identical to those based on two.

As a final example, consider a simple two-phase A-B (baseline–intervention 1)/A-C (baseline–placebo or intervention 2) design or a B-C (two intervention) crossover design. A researcher decides to assign one of 5 randomly preselected acceptable crossover points to each case and either ignores (Situation 1) or accounts for (Situation 2) the crossover-point randomization process in the associated statistical analysis. Unlike the example in Fig. 2, the researcher is not restricting the assignment of B and C phases in a completely counterbalanced design (i.e., where exactly two cases are administered the interventions in a B-C order and two in a C-B order). In one simulation study [34]), based on 4 cases, a total of 15 outcome observations, and a crossover point randomly selected from among the 5 crossover points for each case (see Fig. 2), along with a series autocorrelation of 0.30, and a one-tailed Type I error probability of 0.05, under Situation 1 powers are 0.37 and 0.61 to detect effect sizes of 1.0 and 1.5, respectively, as compared to 0.80 and 0.98 under Situation 2. Even when the design is restricted so that exactly two cases apiece are randomly assigned to the B-C and C-B orders in a completely counterbalanced design, under Situation 2 the lower respective powers of 0.73 and 0.93 still surpass by far the above Situation 1 powers. As a pertinent aside, it is worth noting that because there is no need for extended washout intervals in the earlier discussed two independent samples A (baseline)-B (intervention 1) vs. A (baseline)-C (placebo or intervention 2) SCIDs, such designs might provide a viable option for health science intervention researchers—albeit typically with a loss in statistical power relative to within-case SCIDs, such as two-period crossover designs and alternating treatment designs.

Issues and Concerns

A wide variety of randomized SCIDs and associated analyses have been developed that could be considered by health science intervention researchers, including randomized ABAB reversal designs, alternating treatment designs, and changing criterion designs, among others (see, for example, [36–39]). A number of issues and concerns raised by reviewers of an earlier version of this manuscript will now be addressed.

Number and Spacing of Outcome Observations

Readers might be concerned that the various SCID examples presented here require that an equal number of evenly spaced and time-coordinated observations must be obtained for all cases in the design, but that concern should be allayed. In all SCIDs in common use in the behavioral sciences, standards have been established for a required minimal number of outcome observations per A or B phase (e.g., five) to represent an acceptable design [40]. Moreover, the observations across cases do not have to be coordinated in real time nor collected continuously at evenly spaced time points as, for example, with “nonconcurrent” multiple-baseline designs. In addition, with a multiple-probe design, only a few select observations need be collected periodically throughout the intervention-phase interval. Finally, there are statistical randomization tests that are specifically tailored to allow for different numbers of outcome observations collected for each case [41].

Attention to Participant Characteristics and Changes in Them

In SCID studies, participants may be selected based on certain characteristics (i.e., disease status, health condition) at the onset of the experimental trial. During baseline and/or the intervention conditions, these characteristics may change, thereby not matching the original pre-specified inclusion/exclusion criteria. For example, a patient diagnosed with ADHD might begin a trial with a traditional diagnosis but as the trial evolves, (s)he might develop additional medical/psychiatric conditions. Of course, such circumstances could occur in traditional RCTs in which certain patients in either the experimental or placebo/control condition display a change in status as the experiment proceeds. Nevertheless, the SCID researcher has available options in such evolving circumstances and these options may actually provide benefits to the investigation.

First, the researcher can document the change in participant status and proceed with the trial, noting that the status has changed and the results (positive or negative) could be impacted by the changed status. This information is useful

to establish the generalizability of the findings, especially in cases where comorbidity is likely to occur with certain medical conditions. Second, the researcher has the option to continue the trial with the original participant while at the same time recruiting additional participants with the pre-established status characteristics likely to remain in the study throughout the duration. This option allows the researcher to compare outcomes across participants who are exposed to the same intervention in the successive trials. Third, and a major advantage of the repeated measurement characteristic of SCIDs is that although such ongoing assessment will reveal the point at which the participant’s status has changed, it might not be reflective of any substantial variation in outcome performance of the participant’s response to the intervention. Thus, a change in participant status might suggest that the baseline or intervention data were not affected by the change in status as reflected by no change in level, variance, or slope in the data series. Again, such a finding yields scientific information on the generalizability of the intervention and ultimately such SCID studies will provide insights into the scope of the intervention across diverse participant characteristics and can guide subsequent RCTs across phases. A recurring theme here is that scientifically informative advances gleaned from SCIDs are best obtained through multiple replication efforts (see [42] for an overview of replication types).

Adaptive Intervention Designs

SCID trials may be especially useful as adaptive designs, or adaptive platform trials, which can be established with protocols that are similar to applications with conventional large-sample studies. Essentially, and paralleling an adaptive design in which multiple interventions need to be tested [43], SCID researchers can introduce multiple interventions to the same participant over time. Traditional adaptive designs used in medical research share some similarity with our earlier referenced response-guided SCIDs of behavioral and social science research, but with a critical difference. The formal introduction of an adaptive design strategy in a SCID context should include an a priori algorithm that outlines how the study will progress with the to-be-implemented methodological-component and participant-response variations.

As an example, the SCID researcher might introduce multiple drugs or varying dosage levels of a drug in a within-series ABABACACADAD design. In such a design configuration, a randomized start point could also be used for each drug variation, given appropriate consideration of carryover effects, washout effects, etc. In accord with best practices for adaptive designs, planned modifications can be outlined prior to the trial, such as drug type, dosage level, randomized intervention start points, and patient response, among other

variables. The researcher must be mindful, however, that such designs carry with them the potential problem of order effects, which, if confounded with the intervention, would need to be ruled out as an outcome contributor—for instance, by conducting a supplementary SCID in which the order of drug administration is appropriately randomized or counterbalanced across patients (as was discussed earlier). Additional examples of these “rapidly changing designs” can be found in [42].

Comprehensive Documentation

In SCIDs, investigators can take into account the estimand framework, wherein the medical researcher’s objective is to follow the addendum to the ICH E9 guideline to document an alignment among several important features of a clinical trial study, including an *N*-of-1 trial. These features include specification of the research objectives, details of the study process, data analyses (visual and statistical), and how the results will be interpreted [44]. Increasingly, these components of the estimand framework are being integrated into appraisal guidelines for SCIDs in the behavioral and social sciences, for both the conduct of the study and in literature reviews and meta-analyses (e.g., [10, 11, 45]).

Within- and Between-Case Outcome Measure Variability

In SCIDs, the researcher should take into account two types of case variability with respect to the outcome measure. Intra-subject variability is typically examined in within-subject designs, where the researcher has repeated measurement of the outcome measure(s) and examines the case’s response to the intervention as it is replicated across the series (e.g., ABABAB type designs). Such designs are labeled “intra-subject replication designs” but are only one class of design option in SCID research. In this class of designs, variability can be assessed through visual and statistical analysis of the data and may reveal several possible reasons for the participant’s variability in the data (e.g., unreliability of the outcome measures, variation in intervention integrity, among several possibilities—see [13] for a discussion). With their reliance on repeated measurements, SCIDs are especially helpful in identifying sources of participant variability.

Inter-subject variability typically refers to variance across cases and although it is addressed differently in data analysis in large-*N* RCTs and SCID research, any SCID experiment with multiple participants (such as the multiple-baseline design across cases) must take into account this source of variability. We have already addressed this issue in a previous section with illustrations of the statistical power of randomization tests in multiple-case SCID studies. In addition to the use of randomization tests, inter-subject variability can

be assessed through various effect size measures in multiple-case experiments and in meta-analyses of SCID studies.

We hasten to add that SCIDs would have different applications and interpretations in “treatment” medical research than in “prevention” medical research. As in the examples discussed here, with treatment-related research, a series of “undesirable” or “abnormal” state baseline observations would be collected on one randomly constituted participant group, followed by a medical treatment, which in turn would be followed by a series of intervention observations. A second randomly constituted participant group would have baseline observations collected, followed by either no treatment or a placebo, followed by a series of intervention observations. A statistical comparison of the two participant groups’ outcome observations would afford efficacy evidence of the medical treatment in question. In contrast, with preventive-related research, a series of baseline observations would be collected on participants who are not experiencing symptoms of the medical condition in question. The preventive measure (e.g., a vaccination; physical safety information) would then be implemented in one randomly constituted group and no preventive measure or a placebo would be implemented in another randomly constituted group. Again, a series of intervention series observations would then be collected, followed by a statistical comparison of the two groups’ pre- and post-intervention series outcomes to assess the efficacy of the preventive measure that was administered. With any of the randomized SCIDs and associated statistical analyses considered here, a researcher can examine A-phase to B-phase changes in mean (level), trend (slope), and variability for either expected immediate or delayed intervention effects [41, 46]. Effect size indices can also be calculated and are recommended to complement and elucidate the various randomization test results of interest. Freely available for statistical analyses of a multitude of randomized SCID variations is [47] *ExPRT (Excel Package of Randomization Tests)* randomization test software.

Conclusions

We strongly encourage health sciences researchers to incorporate various randomized SCID schemes and structures into their research, including timely research that could be targeted at the Covid-19 pandemic. There are times when it could be opportune for an intervention researcher to (1) conduct an initial small-scale randomized SCID, (2) examine the data in search of aspects of the intervention that “worked” either well or poorly (across the entire design, or selectively on a case-by-case basis), and then (3) redesign and conduct a thought-to-be-improved version of the study. With randomized SCIDs, although the number of outcome observations required of each case may seem excessive, the

number of cases required to produce a scientifically credible, statistically sound, study is a fraction of what is needed for a comparably valid conventional RCT. The same is undoubtedly true with respect to cost considerations associated with the two types of intervention research. In short, we recommend that randomized SCIDs be considered not as a replacement for, but as a versatile precursor or a companion to, conventional multiple-participant large-sample randomized experimental trials that are aimed at accelerating public health research breakthroughs. In many respects, such designs can be viewed analogously to Phase 1 clinical trials medical research investigations.

Authors Contribution

Both authors have contributed to the writing of, and have agreed with the content of the manuscript, and both have given explicit consent to submit the manuscript to *Therapeutic Innovation & Regulatory Science*. There are no other responsible parties associated with the work reported in the manuscript.

Funding

There was no funding to support the preparation of this article.

Declarations

Conflict of interest

The authors have no conflicts of interest to declare.

Ethical Approval

No ethical issues are connected to the preparation and reporting of this article. The authors are grateful to two anonymous reviewers for their helpful editorial suggestions.

References

- Kratochwill TR, Levin JR. Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psych Meth*. 2010;15:122–44.
- Kravitz RL, Duan N, eds, and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). *Design and Implementation of N-of-1 Trials: A User's Guide*. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014. www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm.
- Box GEP, Jenkins GM. *Time series analysis: forecasting and control*. San Francisco: Holden-Day; 1970.
- Kratochwill TR, Piersel WC. Time-series research: contributions to empirical clinical practice. *Behav Assess*. 1983;5:165–76.
- Kolata G. *Flu: The story of the great influenza pandemic of 1918 and the search for the virus that caused it*. New York: Farrar, Straus, & Giroux; 1999.
- Mirza RD, Punja S, Vohra S, Guyatt G. The history and development of N-of-1 trials. *J R Soc Med*. 2017;110(8):330–40. <https://doi.org/10.1177/0141076817721131>. PMID:28776473;PMCID:PMC5546658.
- Barlow DH, Hayes SC, Nelson RO. *The scientist practitioner: research and accountability in clinical and educational settings*. New York: Pergamon Press; 1984.
- Bloom M, Fischer J, Orme JG. *Evaluating practice: guidelines for the accountable professional*. Boston: Pearson; 2009.
- Hayes SC, Barlow DH, Nelson-Gray RO. *The scientist practitioner: research and accountability in the age of managed care*. Boston: Allyn and Bacon; 1999.
- Tate R, et al. The single-case reporting guideline in behavioural interventions (SCRIBE) 2015 statement. *Arch Sci Psych*. 2016;2016(4):10–31.
- Tate R, et al. The single-case reporting guideline in behavioural interventions (scribe) 2015 statement: explanation and elaboration. *Arch Sci Psych*. 2016;2016(4):1–9.
- Kravitz RL, Duan N, Vohra S, Li J, The DEcIDE methods center N-of-1 guidance panel. Introduction to N-of-1 trials: indications and barriers. In Kravitz RL, Duan N (Eds.) and the DEcIDE methods center N-of-1 guidance panel (Duan N, Eslick I, Gabler NB, Kaplan HC). In Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). *Design and Implementation of N-of-1 Trials: A User's Guide*, Chapter 1. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality 2014 Feb. www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm.
- Kazdin AE. *Single-case research designs: methods for clinical and applied settings*. 3rd ed. Oxford: Oxford University Press; 2021.
- Schmid CH, Duan N, The DEcIDE methods center N-of-1 guidance panel. Statistical design and analytic considerations for N-of-1 trials. In Kravitz RL, Duan N (Eds.) and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). *Design and Implementation of N-of-1 Trials: A User's Guide*, Chapter 4. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality 2014 Feb.
- Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Per Med*. 2011;8(2):161–73. <https://doi.org/10.2217/pme.11.7>. PMID:21695041;PMCID:PMC3118090.
- Lin J, Bunn V. Comparison of multi-arm multi-stage design and adaptive randomization in platform clinical trials. *Contemp Clin Trials*. 2017;54:48–59.
- Li Q, Lin J, Lin Y. Adaptive design implementation in confirmatory trials: Methods, practical considerations and case studies. *Contemp Clin Trials*. in press.
- 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.
- 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;112:334–48.
- Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin; 2002.
- Marascuilo LA, McSweeney M. *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA: Brooks/Cole; 1977.
- Craig AR, Fisher WW. Randomization tests as alternative analysis methods for behavior analytic data. *J Exp Anal Behav*. 2019;111:309–28.

23. Edgington ES, Onghena P. Randomization tests. 4th ed. Boca Raton: Chapman & Hall, Boca Raton, FL; 2007.
24. Horner R, Odom SL (2014) 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.
25. Levin JR. Crafting educational intervention research that's both credible and creditable. *Educ Psych Rev*. 1994;6:231–43.
26. Levin JR, Ferron JM, Gafurov BS. An improved two independent-samples randomization test for single-case AB-type intervention designs: a 20-year journey. *J Mod Appl Stat Meth*. 2019;18(1):1–20. <https://doi.org/10.22237/jmasm/15566/70480>.
27. Revusky SH. Some statistical treatments compatible with individual organism methodology. *J Exp Anal of Behav*. 1967;10:319–30.
28. Wampold B, Worsham N. Randomization tests for multiple-baseline designs. *Behav Assess*. 1986;8:135–43.
29. Horner RH. Overview of single-case intervention designs. Madison, WI: Paper presented at the Single-Case Intervention Design and Analysis Institute; 2017.
30. Edgington ES. Randomization tests for one-subject operant experiments. *J Psychol*. 1975;90:57–8.
31. Marascuilo LA, Busk PL. Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behav Assess*. 1988;10:1–28.
32. Koehler MJ, Levin JR. Regulated randomization: a potentially sharper analytical tool for the multiple-baseline design. *Psych Meth*. 1998;3:206–17.
33. Levin JR, Ferron JM, Gafurov BS. Comparison of randomization-test procedures for single-case multiple-baseline designs. *Dev Neurorehab*. 2018;21:290–311.
34. Levin JR, Ferron JM, Gafurov BS. Improved randomization tests for a class of single-case intervention designs. *J Mod Appl Stat Meth*. 2014; 13(2):2–52. Retrievable from <http://digitalcommons.wayne.edu/jmasm/vol13/iss2/2>.
35. Ferron J, Sentovich C. Statistical power of randomization tests used with multiple-baseline designs. *J Exp Educ*. 2002;2002(70):165–78.
36. Ferron JM, Rohrer LL, Levin JR. Randomization procedures for changing criterion designs. *Behav Mod*. 2019. <https://doi.org/10.1177/0145445519847627>.
37. Heyvaert M, Onghena P. Randomization tests for single-case experiments: State of the art, state of the science, and state of the application. *J Context Behav Sci*. 2014;3:51–64.
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;50:599–624.
39. Onghena P. Randomization tests for extensions and variations of ABAB single-case experimental designs: a rejoinder. *Behav Assess*. 1992;14:153–71.
40. Kratochwill TR, et al. Single-case intervention research design standards. *Remed Spec Educ*. 2013;34:26–38.
41. Levin JR, Ferron JM, Gafurov BS. Additional comparisons of randomization-test procedures for single-case multiple-baseline designs: alternative effect types. *J Sch Psych*. 2017;63:13–34.
42. Barlow D, Nock M, Hersen M. Single case experimental designs: Strategies for studying behavior for change. 3rd ed. Boston: Allyn and Bacon; 2009.
43. Angus DC, et al. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nat Rev Drug Discov*. 2019;18(10):797.
44. Lawrance R, Degtyarev E, Griffiths P, Trask P, Lau H, D'Alessio D, Griebisch I, Wallenstein G, Cocks K, Rufibach K. What is an estimand & how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials? *J Patient Rep Outcomes*. 2020;4(1):68. <https://doi.org/10.1186/s41687-020-00218-5>. PMID:32833083;PMCID:PMC7445213.
45. Kratochwill TR, et al. Single-case designs technical documentation. 2010: https://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf
46. Levin JR, Ferron JM, Gafurov BS. Randomization tests of trend and variability for single-case multiple-baseline designs. *Educ Psychol Rev*. 2020. <https://doi.org/10.1007/s10648-020-09549-7>.
47. Gafurov BS, Levin JR. ExPRT (Excel Package of Randomization Tests), Version 4.2. 2021 (Jan); Retrievable from <http://ex-prt.weebly.com>.

Joel R. Levin PhD