

## Single-Case Experimental Designs for the Evaluation of Treatments for Self-Injurious and Suicidal Behaviors

SHIREEN L. RIZVI, PhD, AND MATTHEW K. NOCK, PhD

Single-case experimental designs (SCEDs) provide a time- and cost-effective alternative to randomized clinical trials and offer significant advantages in terms of internal and external validity. A brief history and primer on SCEDs is provided, specifically for use in suicide intervention research. Various SCED methodologies, such as AB, ABAB, multiple baseline, and changing criterion designs, are described. Advantages and disadvantages of their use specifically for intervention research for suicidal and self-injurious behaviors are detailed.

There are many potential obstacles to conducting research on treatments for suicidal and self-injurious behavior. One such obstacle is the relatively low base rate of the behavior, which has implications for the ability to carry out an adequately powered study with meaningful results (Cohen, 1986). For example, if a researcher developed a new treatment for suicidal behavior and wanted to compare it to an existing treatment in a randomized clinical trial (RCT), in order to have adequate statistical power ( $\geq .80$ ) to detect a moderate effect size (i.e.,  $d = .50$ ) for the treatment, a total of 100 participants would need to be recruited (Cohen, 1988). Even in large urban settings this enrollment would be difficult to achieve in a timely manner. Treatment developers in more rural areas are at an even greater disadvantage. In addition, the cost of conducting such research is staggeringly high and often requires large-scale fed-

eral grants which are difficult to obtain, especially among junior researchers, graduate students, and clinicians.

It is also important to note that although the group comparison approach offers many design options (see Nock, Janis, & Wedig, 2008), it introduces numerous methodological limitations that often are overlooked (see Barlow, Nock, & Hersen, 2009). Problems include the obfuscation of individual variability as well as the emphasis on between-group differences only. Thus, a substantial amount of valuable information about each individual's progress in treatment can get lost in the analysis. For these reasons, it is important to consider alternatives to RCTs so that the advancement of the field is not hindered by such concerns.

One excellent alternative is the single-case experimental design (SCED). Single-case designs are not to be confused with case studies (e.g., Freud & Breuer, 1895), with the latter referring to narrative descriptions of the events that transpired in a particular form of therapy with a particular client. Clearly case studies can provide useful information (see Kazdin, 1981), however, since they typically lack systematic assessment and do not use an experimental manipulation, they do

---

SHIREEN RIZVI is with the New School for Social Research in New York; and MATTHEW NOCK is with Harvard University.

Address correspondence to Shireen L. Rizvi, PhD, New School for Social Research, 65 Fifth Ave, Room 334, New York, NY 10003; E-mail: RizviS@newschool.edu

not allow one to test the causal effects of a treatment. In contrast, although SCEDs are not used as frequently as RCTs, they can provide the same level of experimental rigor and, if implemented properly, can possess a high degree of internal validity. Rather than requiring a large number of individuals to test research hypotheses, sometimes only one participant is needed, thus making this method an easy, efficient way to demonstrate the causal effect of an intervention. SCEDs can provide a careful examination of both between-subject *and* within-subject factors.

### A BRIEF HISTORY OF SCEDs

In the earliest days of psychological science, researchers such as Wundt, Ebbinghaus, and Pavlov used single subjects or small groups of subjects to make scientific advances in the study of perception, learning and memory, and the laws of conditioning. For instance, Ebbinghaus' ground-breaking research on learning and memory was performed over the course of 5 years using a single research subject—himself (Ebbinghaus, 1885/1913). The use of single case studies was also the primary method in the earliest days of clinical psychology. The founders of behavioral psychology used SCEDs to inform research and practice in psychology and psychopathology (Skinner, 1938; Watson, 1925).

Despite the significant advances yielded by the use of SCEDs, they have been severely underutilized over the past few decades. This trend may be due in large part to the increase in funding of large RCTs and to the development of new and sophisticated methods for collecting and analyzing large amounts of data. While the use of RCTs has clearly led to advances in the treatment of psychopathology in general, and the treatment of self-injurious and suicidal behaviors in particular (e.g., Brown et al., 2005; Linehan & Comtois et al., 2006), the incorporation of SCEDs into the current clinical research armamentarium will greatly enhance the flex-

ibility and efficiency of clinical researchers working in this area.

### A BRIEF PRIMER ON SCED METHODS

In the classic between-group comparison designs, inferences about treatment effectiveness are typically drawn by observing changes in the target behavior(s) among those receiving treatment compared to those receiving no treatment or a comparison treatment (Kazdin, 2003; Nock et al., 2008). In SCEDs, by contrast, inferences about treatment effectiveness typically are drawn by observing changes in the target behavior(s) over time within the individual(s) when treatment is present compared to when it is absent. There are many types of SCED designs that have varying degrees of complexity, including AB, ABAB or reversal, multiple baseline, and changing criterion designs. (In general, "A" indicates a baseline or no-treatment phase and "B" indicates a treatment phase.) Each of these unique designs are described below, but their common elements are highlighted here (see also Barlow et al., 2009; Nock, Michel, & Photos, 2007; Rizvi, Monroe-Devita, & Dimeff, 2007):

1. Identification of a specific target behavior. Before the study has begun, a specific behavior that can be reliably and validly measured must be identified.
2. Continuous measurement. The foundation of SCED rests on its measurement. The same reliable and valid measurement must be applied on a regular basis so that any change over time can be accurately assessed and interpreted.
3. Stability of the specific target behavior. In order for the effects of the intervention to be the most clear, it needs to be demonstrated that the target behavior changes *only* when the intervention is applied. If the target behavior is unstable and vacil-

lates widely before the intervention is applied, then it becomes increasingly difficult to demonstrate that the intervention has any effect. Kazdin (2001) describes a stable rate of behavior as one in which there is little variability as well as a lack of a trend (or slope). (This issue of stability is particularly relevant in suicide research and will be elaborated further below.)

4. A baseline period. A baseline period is necessary during which data are gathered on the target behavior before any intervention is applied. Without a baseline phase, there is no way of knowing whether the intervention had any true effect or whether it is responsible for any changes in the individual. In this sense, the baseline period has not only a descriptive function, but also a predictive function in that it is presumed to predict how frequently the behavior would continue to occur in the absence of the intervention (Barlow et al., 2009; Kazdin, 2003).
5. Systematic application of intervention. Once a baseline period has been established, the intervention must be applied in a systematic and conscientious manner. Ideally, the only difference between the A and B phase is the addition of the intervention. This discrimination allows for the most valid conclusion to be drawn.

#### *Basic AB and ABAB Designs*

These elements are used in various formats to create the different designs testing treatment effectiveness. For example, in the simplest AB design, a baseline period "A," is followed by an intervention period "B" and the effects on the target behavior of interest by the application of the intervention are assessed. Although AB designs are most likely to mimic what occurs in "real life" clinical settings (i.e., treatment occurs following a

waiting period), they are generally considered the weakest form of SCED due to the lack of experimental control over threats to internal validity. That is, there are many plausible explanations for changes in the target behavior, other than the intervention, including the mere passage of time, regression to the mean, or the occurrence of another event that was responsible for the change.

In order to increase the level of experimental control, an ABAB design can be employed. In ABAB designs, following a specified period of intervention (the first B phase), treatment is then withdrawn and the effects on the behavior are documented (the second A phase). If a treatment is responsible for the change in the behavior, then, in many instances, one would expect that behavior will regress to initial levels during the second A period. Finally, the treatment is applied again (the second B phase) with the hypothesis that the behavior will again change as a result of the treatment. Figure 1 displays hypothetical data illustrating both positive and null outcomes using an ABAB methodology to reduce suicide ideation.

Of course, there are some ethical concerns with using a withdrawal design and this concern is heightened when working with vulnerable or at-risk populations, such as suicidal individuals. If an intervention is working and the client is improving, it would be difficult to justify withdrawing the intervention in order to measure its effects. The ability to use an ABAB design depends in large part on the client, the type of intervention, and the target behavior. If elimination of self-injurious behavior is the target behavior, for example, withdrawing treatment in order to determine if self-injurious behavior reappears would be unethical. This problem makes ABAB designs less desirable for suicide intervention research, except in certain cases.

In some circumstances, for reasons outside the control of the investigator, treatment may be interrupted or stopped completely, leading to a return to baseline procedures. For instance, the client may stop the treatment temporarily or the clinician/staff may go on vacation. In such situations, one

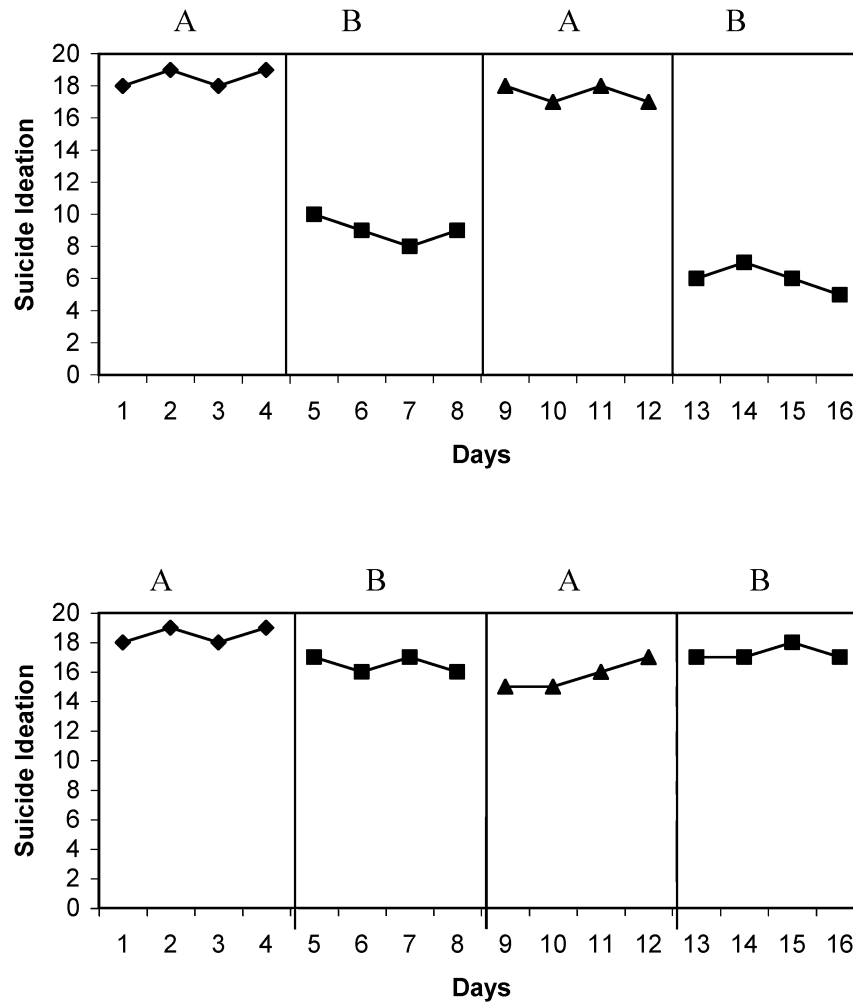


Figure 1. Hypothetical examples for ABAB data.

*Top:* Data indicating positive results for ABAB design. *Bottom:* Data indicating null results for ABAB design.

has a quasi-experimental ABA (or ABAB if treatment is resumed) design, which would allow the investigator to draw stronger inferences about treatment effectiveness than would be possible with an AB design, but it would not allow the experimenter to draw causal inferences, which are possible only with a true experimental design.

As an example of a quasi-experimental ABAB design, Wallenstein and Nock (2007) recently reported on a SCED in which they tested the effectiveness of aerobic exercise in a woman with a long history of nonsuicidal self-injury (NSSI). The woman in this case

was already receiving outpatient treatment, but continued to engage in NSSI on a fairly regular basis. During an initial baseline phase, the woman continued to attend her outpatient treatment but was asked by the investigators to systematically record information each day, including data on her NSSI urges and behaviors. After some behavioral stability was observed, the investigators asked the client to begin engaging in aerobic exercise (see Wallenstein & Nock, 2007, for details), at which time a significant decrease in NSSI was observed, as shown in Figure 2. After several weeks the client decided to stop

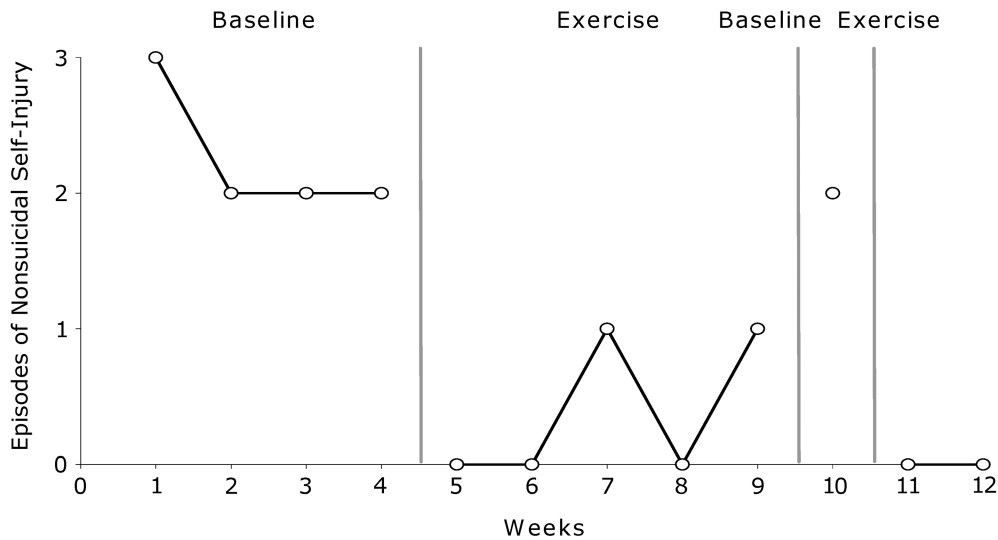


Figure 2. Example of a quasi-experimental ABAB design for non-suicidal self-injury.

From D. H. Barlow, M. K. Nock, & M. Hersen, *Single Case Experimental Designs*, 3/e. Published by Allyn and Bacon, Boston, MA. Copyright © 2009 by Pearson Education. Reprinted by permission of the publisher.

the exercise regimen (i.e., a quasi-experimental return to baseline), at which time her NSSI re-emerged at baseline levels. She then re-initiated the exercise regimen, which led to a cessation of NSSI. Studies such as this can be very useful for testing novel treatments while also maintaining ethical procedures, and thus can serve as a useful tool for clinical researchers seeking to develop more effective treatments for self-injurious thoughts and behaviors.

Another limitation of ABAB designs for use in suicide intervention research is the occurrence of carryover effects, which refers to the continued impact of the treatment phase on subsequent phases. For example, if a researcher was interested in examining the effects of cognitive restructuring on suicidal ideation, she might monitor suicide ideation for a certain period of time to establish a baseline A, followed by a period of time in which cognitive restructuring of hopeless thoughts was added, B. If she then were to “withdraw” the cognitive restructuring piece, it is not expected that suicide ideation would return to baseline levels, because the individual cannot unlearn the techniques learned during B. This is, of course, a favorable out-

come from a clinical perspective, but not experimentally, as it precludes the experimenter from establishing the causal effect of the treatment. Carryover effects can occur in both psychosocial and medication treatments, but may be more likely in the former. One potential solution is to use shorter treatment phases; however, the usefulness of this strategy depends on the nature of the intervention. In some cases, ABAB designs simply may not be the methodology of choice for psychosocial intervention research on suicidal behaviors and other design options, such as multiple baseline or changing criterion designs, may be more suitable.

#### *Multiple Baseline Designs*

Multiple baseline designs are those in which treatment is applied sequentially (i.e., in an AB fashion) across different individuals, behaviors, or settings. There typically is only one baseline (i.e., no removal of the treatment variable), and the power of these designs comes from demonstrating that change occurs when, and only when, the intervention is directed at the behavior, setting, or individual in question. A multiple baseline de-

sign *across individuals* requires taking repeated measurements of the same behavior in different individuals for varying lengths of time, followed by the intervention phase designed to affect the frequency or level of that particular behavior. In other words, the length of the baseline is different across different individuals (see Figure 3 for an example of hypothetical data for a multiple baseline design across individuals). This methodology allows for the determination of how the introduction of the specific intervention changes the baseline behavior and, since more than one individual is included, the causal effects of the intervention are more clearly delineated. For example, Rizvi and Linehan (2005) applied a multiple baseline across individuals design in the evaluation of a new treatment designed specifically to address maladaptive shame in suicidal individuals with borderline personality disorder. The five individuals had different baseline lengths, ranging from 2 to 10 weeks, after which an 8–10 session treatment was applied.

Multiple baseline designs *across behaviors* can be used to demonstrate the effects of an intervention applied to different behaviors within the same individual in a systematic manner (e.g., Nock, 2002). In the case of suicidal thoughts and behaviors, the investigator might first apply the treatment variable to the reduction of suicidal behaviors (e.g., suicide attempts, gestures, preparatory acts) if frequent in nature, and subsequently apply the treatment to the reduction of suicidal thoughts (e.g., daily frequency or intensity of such thoughts). Multiple baseline designs also can be used *across settings*, such as to demonstrate the effectiveness of decreasing suicidal thoughts or behaviors first in a therapy session, then at home, then at work or school. This type of design is especially important for demonstrating generality of treatment effects and may be particularly useful for demonstrating long-lasting effects following intensive treatment such as inpatient stays.

#### *Changing Criterion Designs*

Changing criterion designs are those in which treatment is introduced following a

brief baseline period; however, they are executed in such a way that the criterion for reinforcement of the target behavior is changed over time to require increasing levels of behavior change. As in the multiple baseline design, this design does not require withdrawal of treatment, and here a relation between treatment and behavior change is demonstrated by showing that the target behavior changes when and only when the criterion for reinforcement is changed. Applying treatment in this way should create a step-like pattern of behavior change that matches the changing criterion (e.g., Luiselli, 2000). The criterion for reinforcement typically is changed in the same direction facilitating clinical improvement over time; however, the clinical investigator can incorporate a mini-reversal to increase the strength of the argument that can be made for causation.

As an example of a changing criterion design, a clinician working with a suicidal client who is somewhat reluctant to use new skills being taught in treatment might use a changing criterion design to increase skill use. In the first phase, the client may be rewarded or reinforced for using one skill per day (see Kazdin, 2001, for a review of factors influencing the effectiveness of reinforcement programs). After the client is consistently doing so, the clinician may then change the criterion for reinforcement to two skills per day, then 3, then 5, then 7, then 10, and so on (see Figure 4 for hypothetical data demonstrating the step-like effects of the program). In this instance the dependent variable is skill use (i.e., the positive opposite of maladaptive behavior use) rather than suicidal ideation or behaviors. This designation of the dependent variable has the advantages of focusing on a higher frequency behavior and focusing treatment on the development of positive behaviors rather than on the negative aspects of a client's behavior.

#### *Data Analysis in SCED*

Once a design has been employed and data have been collected, the next phase is determining whether the intervention had a

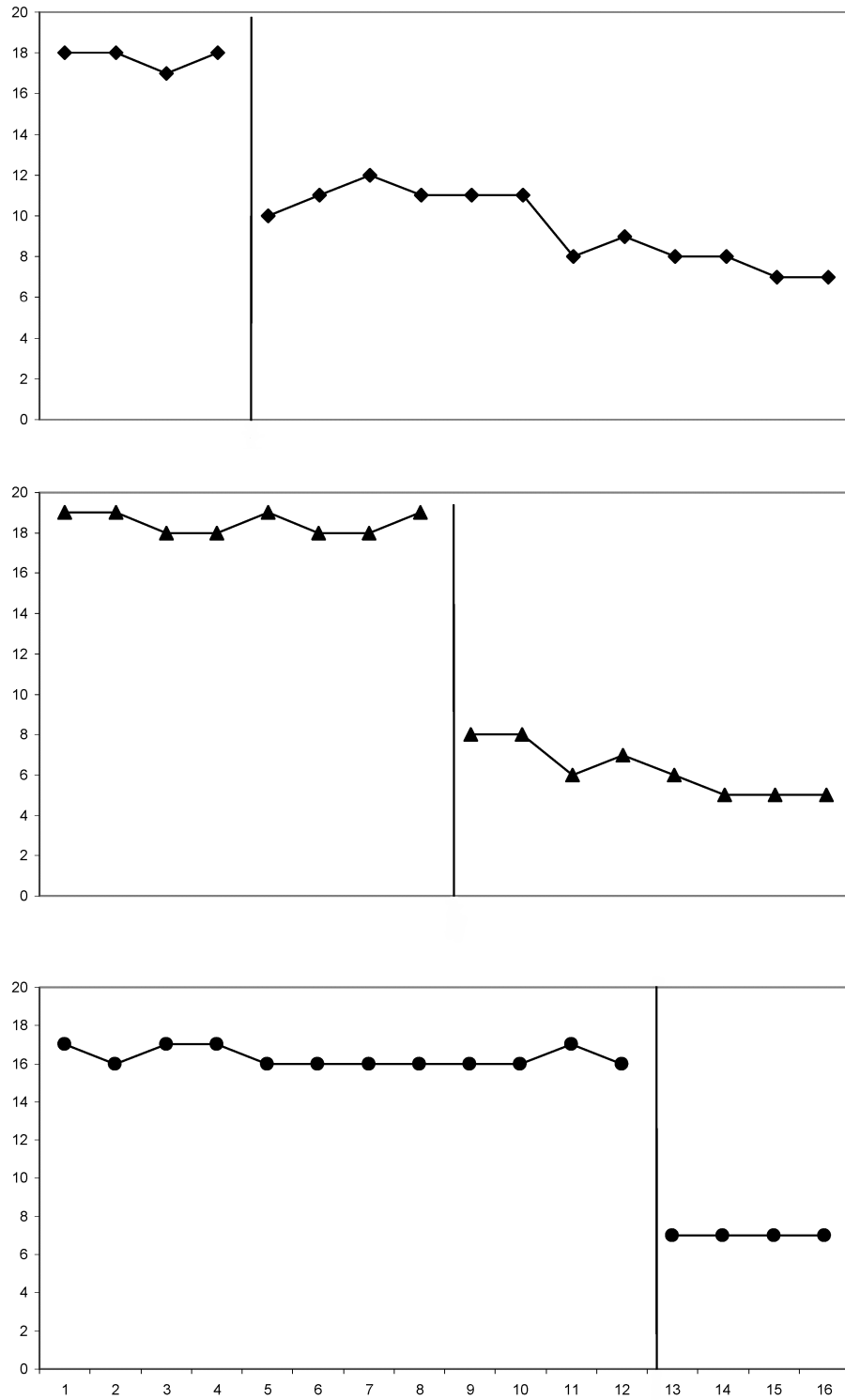


Figure 3. Hypothetical example for multiple-baseline across individuals.

Note. In this example, the baseline phase was 4, 8, and 12 days respectively for the 3 subjects.

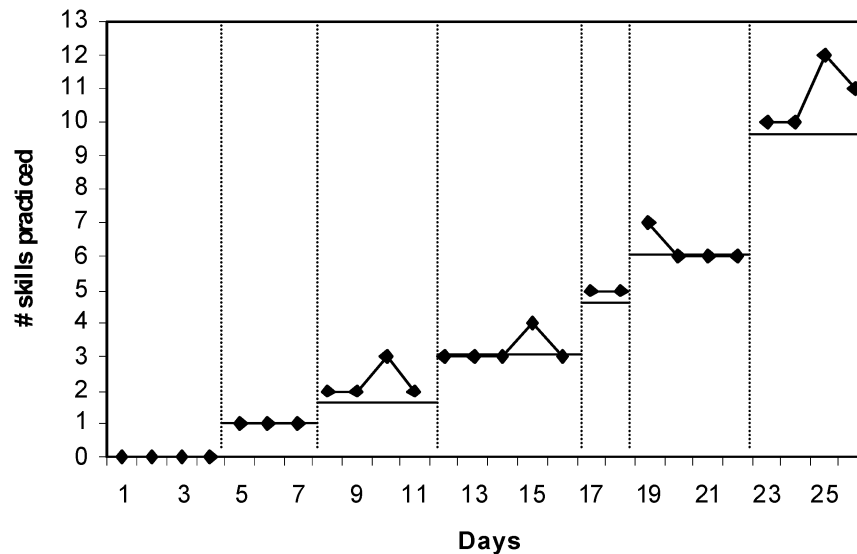


Figure 4. Hypothetical example for changing criterion designs.

*Note.* In this hypothetical example, the criterion for reinforcement was raised from the use of one skill per day for days 1–6 all the way to ten skills per day (days 23–26).

legitimate effect on the target behavior. Although standard statistical analyses are not often employed, four objective criteria are generally used to verify treatment effects (Kazdin, 2003). These criteria are (1) change in mean rate of the behavior from the baseline phase to the intervention phase, (2) change in slope from baseline (presumably horizontal) to the intervention phase (presumably decelerating for behavior one is hoping to reduce), (3) shift in level from baseline to intervention, and (4) a small latency to change from one phase to the next. These last two criteria refer to the change that occurs between the last point of the baseline phase and the first point of the intervention phase. In order to determine whether these criteria have been met, graphs are constructed that display rates of the target behavior throughout the baseline and intervention phases (see figures for examples).

Statistical analyses specifically for the examination of SCED have been developed (see Barlow et al., 2009); however, SCED researchers often rely on visual inspection of the data to determine if clinical change has occurred. How much change in slope or how

big a shift in level is significant in the eye of the beholder. Thus, it is also important to consider whether changes are *clinically* significant. Especially when suicidal behavior is the target of the intervention, it may not be that useful to demonstrate that near-lethal self-injurious behavior changes from an average of five times a day to two times a day if any episode increases the individual's risk of death. In these cases, even if the four criteria are met, it may only be abstinence of behavior that is considered a treatment success. Alternatively, when working with an individual who reports having suicidal thoughts every day of his/her life for the past 10 years, a change to thinking about suicide only a few times a week can be noteworthy and signify that the treatment has had a positive effect.

#### SCEDs IN SUICIDE INTERVENTION RESEARCH

There are many reasons why SCED methodology is well matched to intervention research for suicidal behavior. SCED methodology has numerous advantages that make



it a plausible and scientifically rigorous alternative to larger-scale group comparison designs. Importantly, when establishing the efficacy of a novel treatment, SCED focuses on change at the individual level. Rather than looking at mean differences between pretreatment and posttreatment scores, for example, SCED emphasizes the importance of examining change within the individual over time. This distinction is important for several reasons. For one, in SCEDs, there could be dozens of individual data points that are all considered in the determination of whether a treatment is effective. This comprehensive assessment is in contrast to group designs in which sometimes just two data points (pretreatment, posttreatment) are considered. The more data points that are collected, the less vulnerable one is to random occurrences that may affect scores. Furthermore, one can consider the stability of change by looking at observations over time and can dispel the concern that the findings were a fluke. Finally, in group comparison approaches, the average score obtained may not reflect the actual performance of any individual. SCEDs allow for more careful scrutiny of each person's change over time and eliminate the possibility that important shifts and trends will be lost in the analyses.

As highlighted earlier, another advantage is that far fewer participants are needed to carry out a well-executed SCED than are required in a group-comparison approach. Depending on the design, it is possible that only one participant is needed to demonstrate a causal relation between treatment and outcome variables. Since recruitment of an adequate sample size can be an onerous task for most researchers, a small sample is both time- and cost-efficient. Indeed, given this advantage, it is remarkable that SCED is not a more popular choice of methodology. SCEDs are an especially practical choice when evaluating a new research idea or testing a novel hypothesis. It would be premature to use a large-scale RCT to determine whether a treatment is effective; rather, it is important to start with one individual and document positive effects prior to imple-

menting a large between-groups experiment (Moras, Telfer, & Barlow, 1993).

Furthermore, because group comparison methods often attempt to reduce between-subject variability by obtaining homogenous samples in order to detect "true" treatment effects, they are often criticized for lacking external validity. A common complaint is that RCTs do not generalize to clinical practice because individual complexities have not been adequately sampled. Although the small sample size in SCED methodologies may also elicit comments related to external validity, a researcher is far less likely to have a laundry list of exclusion criteria when using such a design. Moreover, generality can be tested and demonstrated by systematically replicating the experiment across multiple, diverse subjects.

Of course, no methodology is perfect and it is important to also consider the disadvantages to using SCED in suicide intervention research. As mentioned earlier, one important consideration is the ethical implications of withholding treatment during a baseline period or withdrawing treatment in ABAB designs. While the ethical considerations of withholding or withdrawing treatment are always paramount in treating individuals with mental health problems, they are even more so when dealing with high-risk individuals.<sup>1</sup> Thus, it will be important to come up with flexible alternatives to the no treatment phase. One option is to have the participant(s) be engaged in treatment-as-usual (TAU) during the entire length of the baseline and intervention phases. Thus, TAU serves as the control against which the experimental intervention is compared. If a client's behavior is stable during TAU but improves with the addition of the experimental treatment, there is evidence that the new intervention is successful. Another, and perhaps stronger, solution is to use an alternating

---

1. This is not to imply that it is always unethical to withdraw treatment. Certainly, when a treatment is not working, it may be more unethical to continue in the absence of data to its effectiveness.

treatments design in which the client is always receiving some form of treatment, but the treatment variable alternates over time (see Barlow et al., 2009). Multiple baseline and changing criterion designs also circumvent the problems associated with withdrawing treatment from self-injurious or suicidal subjects.

A second disadvantage of SCED methodology in suicide research is the importance of establishing a stable baseline of the specific target behavior prior to the implementation of an intervention. Depending on the behavior, this necessity may be very difficult to achieve and may require a re-conceptualization of the target behavior. For example, suppose an intervention is developed to reduce suicide attempts in individuals with a history of suicidal behavior; that is, the primary outcome variable of interest is number of suicide attempts with the intent to die. The behavior of suicide attempts is not one that is expected to be frequent or stable. Even in an individual who repeatedly attempts suicide, a likely pattern would be an attempt one day followed by several days (weeks? months?) with no attempt. This all-or-nothing pattern of the outcome variable does not fit well with the SCED method. In this case, a more appropriate target behavior for a SCED study might be suicide ideation, as operationalized by degree of desire to die. Ideation can be measured on a daily basis using either a score on a questionnaire, such as the Scale for Suicide Ideation (SSI; Beck, Brown, & Steer, 1997); self-report ratings on a daily monitoring card or an ecological momentary assessment (EMA) study in which participants carry hand-held computers (e.g., PalmPilots) that alert them to complete questions at various points during the day (Nock & Prinstein, 2008); or a performance-based test that does not rely on self-report and thus may be more sensitive to change in self-injurious or suicidal thoughts (Nock & Banaji, 2007a, 2007b). Using ideation, as opposed to attempts, in a chronically suicidal group of individuals allows for a greater likelihood of stability during the baseline phase. Unfortunately, even this may be difficult, depending on the population. In-

dividuals with borderline personality disorder, for example, often describe fluctuating urges to harm themselves and thus achieving a stable baseline, with little variability, may be impossible (Rizvi & Linehan, 2005). One creative alternative for this problem is to demonstrate that the target behavior is consistently unstable and has no discernible pattern until an intervention is initiated, at which point a positive change in behavior occurs. Of course the key to this technique is to show that the behavior changes in the desired direction upon implementation of the intervention; it is not sufficient nor necessarily advantageous to demonstrate that a behavior simply becomes more stable. For example, demonstrating that an individual consistently engages in a high use of self-injury is not a goal of any intervention. One solution to this problem is to switch the focus of the SCED from decreasing the frequency of an infrequent undesirable behavior (e.g., suicide attempts) to increasing the frequency of the use of specific emotion regulation, self-soothing, or other skills taught in treatment. A researcher could monitor daily use of skills while also monitoring the occurrence of suicidal thoughts and behaviors over the entire course of treatment.

Some treatment developers may consider the transparency of SCEDs to be another disadvantage. Nothing is hidden from evaluators in these designs; each relapse, increase in maladaptive behavior, or lack of progress is obvious when analyzing the data with such scrutiny. However, this transparency is also one of the greatest strengths of these designs for suicide intervention research in that it requires researchers and clinicians to remain alert and vigilant to client change. When treating such high risk individuals as those contemplating suicide, it is critical that this level of attention be given. Of course, unlike in RCTs, it would be difficult to have assessors be blind in that the assessor, often the treatment provider, is aware of the treatment that the participant is receiving. There is always the risk that evaluator bias and/or demand characteristics may then influence the results, although such risk

can be diminished by using observable behavior that is verifiable as the dependent variable, which is the gold standard in SCED research.

In general, certain SCED methodologies are better suited to suicide intervention research than others. For instance, AB and ABAB designs may be the least useful, based on the disadvantages listed above, but multiple-baseline, changing criterion, and alternating treatments designs are well-suited to these behavior problems given that they do not require the removal of treatment for their evaluation.

#### SCEDs AND THE FUTURE OF SUICIDE INTERVENTION RESEARCH

There are now a number of psychosocial treatments that have empirical support in the treatment of suicidal behavior. For example, Dialectical Behavior Therapy (DBT; Linehan, 1993; Linehan et al., 1991; Linehan et al., 2006) has been shown to reduce frequency and lethality of suicidal behavior in individuals with borderline personality disorder (see Lieb, Schmahl, Linehan, & Bohus, 2004, for a review of DBT's research support). A manual-based Cognitive Therapy (CT) approach has also received support for reducing suicide attempts (Brown et al., 2005). These findings provide hope that suicidal behaviors can be successfully treated even in multiply-disordered individuals, yet there are clearly areas in which improvement is needed and SCEDs are well-suited to address these questions. For one, despite the overall success of these treatments, a number of individuals do not appear to respond even with an adequate trial. Although both CT and DBT appear to be more successful than control conditions at reducing the number of suicide attempts in some study participants, they have not been shown to be successful for everyone (e.g., Brown et al., 2005; Linehan et al., 2006). A way in which SCED methodology can be used is to try novel approaches with treatment nonresponders and carefully

attend to subtle (and not so subtle) changes in behavior as a result. Thus, rather than ending treatment by labeling an individual a "treatment failure," each individual can become a participant in a single-subject study designed to both develop better treatment protocols as well as to continue to try to help the individual improve.

The mechanisms of change for existing interventions are not yet understood (Kazdin & Nock, 2003; Nock, 2007). Although we know that certain treatments work for the majority of individuals in treatment trials, our understanding of *how* and *for whom* they work is limited. One way in which possible treatment mechanisms can be better explored is through the use of SCEDs. These methodologies can be used to evaluate components of existing treatments in a more cost-efficient manner than RCTs. For instance, specific treatment strategies can be implemented one-by-one and their effects documented in order to ascertain which have influence on the target behavior. As a hypothetical example, consider the following: DBT skills training offers four modules of skills-mindfulness, interpersonal effectiveness, emotion regulation, and distress tolerance. Although the skills, with individual therapy and other elements of comprehensive DBT, have been found to be effective at reducing suicidal behavior, it is possible that only emotion regulation skills are responsible for that change. To test this possibility, a multiple-baseline study across individuals can be employed in which the sequence of skills training varies and the effect on the dependent variable of interest is gauged. Should the multiple baseline design demonstrate that when, and only when, the emotion regulation module is administered do we see an improvement in suicidal behavior would suggest that we have identified an active ingredient or active component of treatment. This result could then be followed with a series of experiments that measure emotion regulation skills and suicidal behavior over the course of treatment to test whether changes in emotion regulation skills do in fact precede changes in suicidal behavior, which is a key requirement for

demonstrating the operation of a mechanism of change (Kazdin & Nock, 2003; Nock, 2007). Such a series of findings, which could be obtained studying only a few clients, would provide a major advance in our understanding of how treatments for suicidal behaviors actually work.

Suicidal behaviors are a serious public health concern, and a number of methodological and ethical considerations must be taken into account when conducting clinical research in these areas. With careful applica-

tion, SCEDs can be used to demonstrate the effectiveness of novel interventions and/or isolate the active ingredients of change within existing interventions while maintaining a high degree of ethical and research standards. Because of their significant advantages, SCEDs are a viable alternative to large-scale RCTs. It is our hope that SCEDs become more widely used for treatment development and evaluation efforts for suicidal behavior and nonsuicidal self-injury in the near future.

## REFERENCES

- BARLOW, D. H., NOCK, M. K., & HERSEN, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Boston, MA: Allyn and Bacon.
- BECK, A. T., BROWN, G. K., & STEER, R. A. (1997). Psychometric characteristics of the Scale for Suicide Ideation with psychiatric outpatients. *Behavior Research and Therapy, 35*, 1039–1046.
- BROWN, G. K., HAVE, T. T., HENRIQUES, G. R., XIE, S. X., HOLLANDER, J. E., & BECK, A. T. (2005). Cognitive therapy for the prevention of suicide attempts: A randomized controlled trial. *JAMA, 294*, 563–570.
- COHEN, J. (1986). Statistical approaches to suicidal risk factor analysis. *Annals of the New York Academy of Sciences, 487*, 34–41.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- EBBINGHAUS, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. A. Bussenius, Trans.). New York: Teachers College Press. (Original work published 1913)
- FREUD, S., & BREUER, J. (1895). *Studies in hysteria* (N. Luckhurst & R. Bowlby, Trans.). London: Penguin Books.
- KAZDIN, A. E. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology, 49*, 183–192.
- KAZDIN, A. E. (2001). *Behavior modification in applied settings* (6th ed.). Belmont, CA: Wadsworth.
- KAZDIN, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston, MA: Allyn and Bacon.
- KAZDIN, A. E., & NOCK, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*, 1116–1129.
- LIEB, K., SCHMAHL, C., LINEHAN, M. M., & BOHUS, M. (2004). Borderline personality disorder. *The Lancet, 364*, 453–461.
- LINEHAN, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York: Guilford.
- LINEHAN, M. M., ARMSTRONG, H. E., SUAREZ, A., ALLMON, D., & HEARD, H. L. (1991). Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Archives of General Psychiatry, 48*, 1060–1064.
- LINEHAN, M. M., COMTOIS, K. A., MURRAY, A. M., BROWN, M. Z., GALLOP, R. J., HEARD, H. L., ET AL. (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry, 63*, 757–766.
- LUISELLI, J. K. (2000). Cueing, demand fading, and positive reinforcement to establish self-feeding and oral consumption in a child with chronic food refusal. *Behavior Modification, 24*, 348–358.
- Moras, K., Telfer, L. A., & Barlow, D. H. (1993). Efficacy and specific effects data on new treatments: A case study strategy with mixed anxiety-depression. *Journal of Consulting and Clinical Psychology, 61*, 412–420.
- NOCK, M. K. (2002). A multiple-baseline evaluation of the treatment of food phobia in a young boy. *Journal of Behavior Therapy and Experimental Psychiatry, 33*, 217–225.
- NOCK, M. K. (2007). Conceptual and design essentials for evaluating mechanisms of change. *Alcoholism: Clinical and Experimental Research, 31*, 4S–12S.
- NOCK, M. K., & BANAJI, M. R. (2007a). Assessment of self-injurious thoughts using a behavioral test. *American Journal of Psychiatry, 164*, 820–823.

- NOCK, M. K., & BANAJI, M. R. (2007b). Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. *Journal of Consulting and Clinical Psychology, 75*, 707–715.
- NOCK, M. K., JANIS, I. B., & WEDIG, M. M. (2008). Research design. In A. M. Nezu & M. Nezu (Eds.), *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions* (pp. 201–218). New York: Oxford University Press.
- NOCK, M. K., MICHEL, B. D., & PHOTOS, V. (2007). Single-case research designs. In D. McKay (Ed.), *Handbook of research methods in abnormal and clinical psychology* (pp. 337–350). Thousand Oaks, CA: Sage Publications.
- NOCK, M. K., & PRINSTEIN, M. J. (2008). *Real-time occurrence of self-injurious thoughts and behaviors: Results from an electronic diary study*. Manuscript submitted for publication.
- RIZVI, S. L., & LINEHAN, M. M. (2005). The treatment of maladaptive shame in borderline personality disorder: A pilot study of “Opposite Action.” *Cognitive and Behavioral Practice, 12*, 437–447.
- RIZVI, S. L., MONROE-DEVITA, M., & DIMEFF, L. A. (2007). Evaluating your DBT program. In L. A. Dimeff & K. Koerner (Eds.), *Dialectical behavior therapy in clinical practice: Applications across disorders and settings* (pp. 326–350). New York: Guilford.
- SKINNER, B. F. (1938). *The behavior of organisms*. New York: Appleton.
- WALLENSTEIN, M. B., & NOCK, M. K. (2007). Physical exercise for the treatment of non-suicidal self-injury: Evidence from a single-case study. *American Journal of Psychiatry, 164*, 350–351.
- WATSON, J. B. (1925). *Behaviorism*. New York: Norton.

Manuscript Received: June 16, 2008

Revision Accepted: June 18, 2008