



HHS Public Access

Author manuscript

J Appl Behav Anal. Author manuscript; available in PMC 2022 February 01.

Published in final edited form as:

J Appl Behav Anal. 2020 April ; 53(2): 596–619. doi:10.1002/jaba.691.

The consecutive controlled case series: Design, data-analytics, and reporting methods supporting the study of generality

Louis P. Hagopian

Kennedy Krieger Institute and Johns Hopkins University School of Medicine

Abstract

Single-case experimental designs (SCEDs) have proven invaluable in research and practice because they are optimal for asking many experimental questions relevant to the analysis of behavior. The consecutive controlled case series (CCCS) is a type of study in which a SCED is employed in a series of consecutively encountered cases that undergo a common procedure or share a common characteristic. Additional design elements, data-analytic, and reporting methods enable researchers to ask experimental questions relevant to the study of generality of procedures and processes. The current paper discusses the CCCS methodologies, including the retrospective, prospective, and randomized CCCS. These methodologies can be applied to examine the generality of clinical procedures (including their general efficacy, the limits of their generality, and variables that may mediate generality); study the epidemiology and phenomenology of clinical problems; and compare the efficacy of 2 clinical procedures within a randomized controlled trial combining SCEDs with randomized group designs.

Keywords

consecutive controlled case series; design; generality; methodology; reliability

The predominant use of single-case experimental designs (SCEDs) in the field of applied behavior analysis can be attributed to the fact they are optimally suited for rigorously analyzing environment–behavior interactions. The SCEDs are characterized by the repeated measurement of behavior across precisely defined and controlled experimental conditions, and the replication of effects across those conditions (Barlow & Hersen, 1984; Kazdin, 2011). The essential design elements of SCEDs focus primarily on the demonstration of experimental control via the replication of effects. Reliability refers to the consistency of an effect when a condition, or entire analysis is replicated (Sidman, 1960). The extent to which observed effects within an experimental analysis are demonstrated to be reliable increases the degree of certainty that those effects are a product of the experimental manipulations, which is also referred to as the study’s internal validity. In the context of functional analysis of behavior, demonstrating that an effect is reliable enables the clinician to make definitive statements about the controlling variables of the behavior of concern. When evaluating a clinical intervention, demonstrating that a treatment effect is reliable provides justification

for its continued application, and increases accountability of providers to consumers. Thus, even when employed with a single participant, SCEDs enable researchers and clinicians to answer a variety of experimental and clinical questions relevant to the analysis of behavior.

Whereas reliability refers to the consistency and reproducibility of an effect within a given clinical or experimental context, generality refers to the extent to which that effect is reproducible beyond the context in which it was originally observed (Sidman, 1960). For an effect to have generality, it must first be demonstrated to be reliable. The extent to which a study's findings have generality is referred to as the study's external validity. The reliability and generality of experimental findings have long been characterized as being of paramount importance to the analysis of behavior in basic and applied contexts (Baer et al., 1968; Sidman, 1960). Recognizing that applied research has previously focused more on ensuring reliability over generality, there is renewed interest in examining the generality of findings associated with clinical procedures (Hanley, 2017; Tincani & Travers, 2019).

Generality can be demonstrated with a SCED in at least two ways. First, intraparticipant generality can be demonstrated by reproducing an effect within the same individual across different contexts. This demonstration is often called "generalization", when referring to a treatment having generality across contexts. Second, generality can be demonstrated by reproducing the findings across cases within the same study or in a replication study (Sidman, 1960). "Direct replication" involves repeating a study (or analysis) as it was originally conducted for the purpose of examining the reliability of the functional relations identified in the original study (or analysis), and lends support for the generality of those original findings across new participants or investigators. "Systematic replication" intentionally alters some parameter of the original study (e.g., the behavior targeted for treatment) for the purpose of examining if functional relations identified in the original study have generality that extends across other variables. Systematic replications of clinical interventions can, for example, demonstrate that an intervention shown to be efficacious for one clinical problem or population is efficacious for some circumstances but not for others. As the number of replication studies increases, it becomes possible to begin defining the conditions under which the procedure is and is not efficacious. Thus, when speaking of the generality of a clinical procedure, the reference is to the limits of its demonstrated efficacy (Johnston & Pennypacker, 2009).

DEMONSTRATING GENERALITY ACROSS A BODY OF LITERATURE

Conducting one replication study at a time contributes to the base of support for the efficacy and generality of a clinical procedure, particularly when replication studies are performed by different groups of investigators. As that base of support builds, systematic or quantitative literature reviews of those studies can be used to summarize findings on a clinical procedure and begin to help define the limits of its generality (e.g., Watkins et al., 2019; Wong et al., 2015). Literature reviews can be qualitative, quantitative, or can be conducted in accordance with a defined methodology, such as the American Psychological Association's (APA's) Divisions 12 and 16 and Empirically Supported Treatments (EST; Chambless & Hollon, 1998; Kratochwill & Stoiber, 2002). The APA's methods define criteria for evaluating the

quality and specifying the number of studies needed to characterize a treatment as having one of three different levels of support; with “empirically supported” being the highest level.

Only a handful of studies have applied APA’s EST criteria to examine the documented efficacy of behavior analytic interventions such as non-contingent reinforcement (NCR; Carr et al., 2009) and functional communication training (FCT; Kurtz et al., 2011). Both of these interventions exceeded the criteria for being classified as empirically supported. Findings from such analyses are important in documenting a sufficient level of empirical support for a procedure’s efficacy. However, this method examines empirical support by the number of methodologically rigorous studies that demonstrate efficacy, without considering how many studies exist in which the treatment failed to produce a positive outcome. Thus, a clinical procedure may meet criteria to be characterized as empirically supported if the requisite number of studies are identified even if there are a comparable number of studies documenting failures.

Meta-analysis is yet another method by which one can combine outcomes from multiple studies to examine the efficacy of a procedure, which can also contribute to some understanding of its generality. Though the use of this statistical method has been quite limited in applied behavior analysis, some examples do exist (Becraft et al., in press; Richman et al., 2015; Virués-Ortega, 2010). This method compiles datasets from individual studies, calculates an effect size for each dataset to allow a uniform outcome metric across cases, and combines them to create a super-dataset to examine outcomes across the collective. Given that this method involves statistical analysis of all datasets, the results include datasets reporting on positive and well as negative outcomes. Essentially, these types of studies ask whether the overall effect size produced by a treatment across all included datasets is statistically significant. For example, Richman et al. (2015) showed that NCR produced effect sizes that were statistically significant and associated with “a substantial decrease in problem behavior and accounted for 60% of the problem behavior variance between baseline and treatment.” (pp. 144–145). Findings are quantitative, objective, and readily consumable by the broader scientific community; and they can inform us about the limits of the generality of a procedure. Documenting that effect sizes are statistically significant across all included datasets is useful, but aggregating outcomes across cases to report the overall effect inherently limits the analysis of behavior (this is discussed in detail below) and, therefore, may limit the direct study of generality of a clinical procedure.

In summary, reviewing findings reported in published studies has great utility in summarizing a body of literature, synthesizing findings, identifying gaps in knowledge, documenting the level of empirical support of clinical procedures, and providing some information about the generality of a procedure. Whether literature reviews are conducted using a defined methodology (e.g., APA’s EST criteria), or employ quantitative or statistical analyses (e.g., meta-analysis), it is important to realize that any conclusions reached using these approaches represent an actuarial and statistical summary of results of studies in the published literature. As such, these methods are limited by the studies that are sampled from the literature. Although they typically include only studies that are methodologically rigorous, the generality of their findings of the studies sampled may be limited by their design (e.g., whether all consecutive cases are included in reporting of findings, the criteria

for selection of procedures, or assignment of participants to conditions) or publication bias, which are discussed further below.

Examining Generality within Studies

Although replication studies and reviews of the published literature can document generality and help begin to define the limits of generality of procedures across studies, it is also possible to design individual studies in a manner that produce findings that can contribute to our understanding of generality in other ways. The extent to which one can make inferences about the generality of findings of any particular study using a SCED is impacted by several design elements and data-analytic methods. These will be discussed further below, and include (a) whether or not all consecutively encountered cases that underwent the procedure or share a common characteristic are included in the reporting of results (i.e., no selective inclusion), (b) whether or not the study reports on participants enrolled in a prospective study with predefined criteria for the selection of procedures or retrospectively on clinical cases for whom procedures were selected based on a clinical process, and (c) whether or not the results of multiple cases are interpreted and reported in a manner that preserves the analysis of behavior without obscuring individual outcomes. The sample size and the characterization of participants also interact with these design elements to impact the extent to which one can examine the generality of findings (although an increase in sample size does not necessarily result in a corresponding increase in generality). It is noteworthy that all of these design elements impacting generality are peripheral to the demonstration of experimental control for studies employing SCEDs, which highlights differences in reliability and generality, and the methods used to demonstrate each.

THE CONSECUTIVE CONTROLLED CASE SERIES (CCCS)

Rationale and Applications

Throughout this paper, the term “consecutive controlled case series” (CCCS) is used to refer to a type of study in which a SCED is employed for each case in a series of consecutively encountered cases that undergo a common procedure or share a common characteristic. When this and other design, data-analytic, and reporting elements are brought together within a single study, they enable one to ask different types of experimental questions about the generality of procedures or functional relations. First, because all cases exposed to the procedure are included in reporting outcomes, one can go beyond simply asking whether a clinical procedure can produce a positive outcome (i.e., the demonstration of efficacy) to asking how often or in what proportion of cases can a clinical procedure produce a positive outcome (i.e., the evaluation of the general efficacy of a procedure). Second, because all cases exposed to a certain procedure (or who share a common characteristic) regardless of outcomes are included when reporting findings, they enable one to identify functional relations that have generality across cases; and allow one to determine their relative prevalence (the study of the phenomenology and epidemiology of a clinical problem). Third, they enable identification of the variables that may mediate the generality of clinical procedures, which can be examined in subsequent analyses (see Branch & Pennypacker, 2013 for a discussion of scientific generality).

The CCCS is not a new method or approach. Studies employing similar methods have been conducted in the field of applied behavior analysis for decades, though the designs, data-analytic methods, and terminology have varied considerably. Terms such as “consecutive case series”, “large-scale analysis”, “experimental-epidemiological analysis”, and “controlled consecutive case series” have been used (additional discussion of terminology follows below). Common to all these types of studies, however, is that they are designed to ask experimental questions related to the generality of clinical procedures or processes. A comprehensive review of the literature describing studies employing these methods is beyond the scope of the current discussion, which is aimed at describing these evolving methodologies, suggesting a common terminology, and promoting their continued development. Exemplar studies will be noted to illustrate how the various design elements listed above can be used in combination to conduct a study and analyze findings to increase opportunities to study generality. CCCS studies have several applications, including (a) to examine the general efficacy and generality of clinical procedures (assessment and treatment procedures), (b) to study the epidemiology and phenomenology of a clinical problem, and (c) to compare the general efficacy of two clinical procedures in a behavior analytic randomized controlled trial. More broadly, these methods enable researchers to ask experimental questions related to the generality of clinical procedures and processes.

Examination of the General Efficacy of Clinical Procedures.

An early example of what could be characterized as a CCCS study was described by Derby et al. (1992), which was conducted to examine the general efficacy of their brief functional assessment (FA) approach. The authors referred to their study as a “large-scale analysis” with 79 cases. The brief FA involves the use of single-session FA conditions (Iwata, Dorsey et al., 1982/1994) followed by a contingency reversal (i.e., providing reinforcement for mands) for the hypothesized function of problem behavior. This approach had been demonstrated to be effective in two earlier studies reporting on a total of 11 participants (Cooper et al., 1990; Northup et al., 1991). Derby et al. explicitly noted that all 79 cases where the procedure was applied were included in the reporting of outcomes (i.e., no cases were excluded) to enable them to evaluate the general efficacy of the procedure. Thus, the authors reported findings in terms of the percentage of cases where positive outcomes (i.e., functions of problem behavior were identified) were obtained using the brief FA procedure. They found that problem behavior was exhibited during the brief FA in only 63% of cases (i.e., problem behavior was not exhibited during the assessment in 37% of cases), and that the brief FA was effective in identifying a function for 74% of those cases where problem behavior occurred. Thus, the brief FA produced a positive outcome (an identified function) in only 47% of the total sample. Additionally, the brief FA informed the development of treatment that reduced problem behavior in 84% of those cases where a function was identified. They concluded that the brief FA has greater generality to individuals with higher rate problem behavior. By reporting on the percentage of cases where successful outcomes were achieved, rather than aggregating or averaging results, the authors’ analysis of the findings across the sample retained a behavior analytic framework.

A number of other studies which share features of what we define as a CCCS study have examined the outcomes of functional behavioral assessment procedures, including the

analog FA conducted with inpatients (Hagopian et al., 2013), with outpatients (Kurtz et al., 2013), and with students (Mueller et al., 2011); and the interview-informed synthesized contingency analysis (IISCA; Greer et al., 2019; Jessel et al., 2018; Slaton et al., 2017).¹ These methods have also been used to examine the outcomes associated with treatment procedures including FCT (Greer et al., 2016; Hagopian et al., 1998; Rooker et al., 2013) and NCR (Phillips et al., 2017). In addition to evaluating the general efficacy of assessment and treatment procedures, these methods have been used to examine the generality of procedures to specific populations, including preschool children with severe behavior (Kurtz et al., 2003), individuals with fragile X syndrome (Kurtz et al., 2015), and for specific problem behaviors (e.g., pica; Call et al., 2015).

Findings from CCCS studies can also help to build an empirical base for defining limits of the generality of the procedures. For example, Phillips et al. (2017) examined the general efficacy of NCR and defined some of its limits in what they termed a CCCS study of 27 cases (inpatients). They reported on the general efficacy of NCR across all cases and across functional classes of behavior. An 80% or greater reduction was achieved with NCR alone in 74% of applications overall; when NCR was combined with supplemental procedures, an 80% reduction was achieved in 96% of applications (general efficacy). Findings also revealed that NCR was more efficacious for socially maintained than for automatically maintained problem behavior, and it was least effective for Subtype-2 automatically maintained self-injurious behavior (SIB; i.e., a subtype characterized by relatively low differentiation of responding across the play and no-interaction conditions of the FA). Thus, the authors were able to identify a limit of NCR's generality for the sample of inpatient participants.

Examination of the Epidemiology and Phenomenology of a Clinical Problem.

—CCCS methodologies can also be applied to examine the epidemiology and phenomenology of a clinical problem. In a study on the epidemiology of the functions of SIB, Iwata, Pace et al. (1994) reported on 152 cases with SIB who underwent FA and treatment in what was termed an “experimental-epidemiological analysis”. Because every case that underwent FA and treatment was included and the sample was composed of a diverse population treated in different settings, there was a high degree of confidence that the findings would have broad generality. Consequently, findings on the relative prevalence of operant functions of SIB, and the relative efficacy of various interventions across those functions obtained with that sample, could be reported (and extrapolated to others with SIB).

In a CCCS focused on the phenomenology of automatically maintained SIB, Hagopian et al. (2015) identified distinct response patterns observed in the FA as the basis for a subtyping model for this functional class of SIB. A subsequent prospective CCCS study showed subtypes differed little with respect to human operant performance (Rooker et al., 2018), and another showed that individuals with treatment-resistant subtypes (Subtypes 2 and 3) presented with more severe injuries at the time of admission to treatment relative to those with socially maintained SIB (Rooker et al., 2019). These studies illustrate how CCCS

¹[Correction added on 1 April 2020, after first online publication: the reference citation of Jessel et al. (2018) in subsection ‘Examination of the General Efficacy of Clinical Procedures’ has been corrected to ‘Jessel’ instead of ‘Jessel’.]

methodologies can be used to describe and contribute to knowledge on the phenomenology of a clinical problem.

Comparing the Efficacy of Two Clinical Procedures in a Randomized

Controlled Trial.—A third application of CCCS methodologies is to perform a rigorous comparison of two clinical procedures. This design combines the use of SCEDs (for the evaluation of individual outcomes) with randomized group designs (to assign participants to receive one procedure or the other). In the current discussion, this type of study is termed a randomized CCCS (this design is described in detail further below). Though not common, studies combining SCEDs with randomization to groups have been conducted for decades.

Durand and Carr (1992) used such a design to compare FCT to time-out for attention-maintained behavior with respect to both their immediate effects and maintained effects across novel trainers. They randomized matched pairs of participants (six per group) to receive either FCT or time-out from reinforcement and they evaluated each treatment using a multiple baseline across participants design. FCT and time-out were equally efficacious in reducing problem behavior when first implemented, but participants who received FCT showed better maintenance of effects when treatment was applied by novel teachers. In contrast, treatment gains were lost when novel teachers implemented time-out. Studies using a combination of SCEDs and randomized group designs have been employed by other researchers to compare treatments for food selectivity (Peterson et al., 2016), instructional procedures for students (Gast & Wolery, 1988), and training methods for human service providers (Neef et al., 1991). The potential value and applications of these mixed SCED-randomized group designs is discussed in detail below.

Elements of the CCCS: Design, Data-Analytics, and Reporting Methods—A thesis of the current discussion is that studies can be designed, and their findings analyzed and reported in a manner that could lead to more definitive statements about the generality of procedures and processes. Here, five elements related to the design, data-analytic and reporting methods that could be used to construct CCCS studies are described in detail, and why they are important with respect to generality (see Table 1 for a summary of these design elements). Though a somewhat cumbersome term, CCCS is derived from “consecutive case series” (CCS), a nonexperimental method used in other fields to report on a series of consecutive cases (these are distinguished further below).

After describing the CCCS elements and analytic methods of the CCCS, a distinction will be made between the “retrospective CCCS” and “prospective CCCS”. Finally, the “randomized CCCS” design will be described, which is a type of experimental design that combines the prospective CCCS with randomized group designs. Figure 1 provides a schematic to illustrate how the retrospective and prospective CCCS studies are essentially SCEDs, whereas the randomized CCCS designs include elements of both SCEDs and randomized group designs. Articulating these elements and variations of the CCCS design is not meant to establish a standard, but merely summarize what investigators might consider when designing these types of studies to support a more systematic study of the generality of procedures and processes, suggest a common terminology, and promote their use and continued refinement.

Element 1. An SCED Is Employed with Each Case.²

The use of a SCED with each case is necessary to demonstrate the reliability of an observed effect in the context of an assessment procedure or treatment evaluation. When describing a CCCS study, it is important to document that the experimental analysis performed with each case is as methodologically rigorous as an analysis reported in a study with an n-of-1. Thus, for every case included in a CCCS (a) the independent variable should be clearly defined and controlled, (b) the dependent variable should be operationally defined and data collected with good reliability, and (c) an SCED should be employed where phase changes are initiated at appropriate times given the level, trend, and stability of data. It is important to recognize that SCEDs are a type of response-dependent adaptive experimental design in which demonstration of experimental control is, in part, tied to the participant's response.

Graphically depicting data from an analysis using a SCED is the optimal method to illustrate the design structure and response patterns, magnitude of effects across conditions, and to document the demonstration of experimental control (condition effects, reversals and replications). For larger scale studies, authors' and editors' ability to include graphs of every individual assessment and treatment analysis was quite limited in the past. However, with online publishing and the use of supplemental tables and figures, it is now possible for journals to include every relevant graphically depicted assessment and treatment analysis. In the field of genetics, massive datasets are routinely made available to reviewers and readers to enable them to confirm authors' interpretations; this is either hosted by the publishing journal or by an online data sharing platform such as Figshare (<https://figshare.com>). The potential value of sharing data also has been raised in the field of applied behavior analysis for the purpose of increasing transparency, promoting replication, and assuring research integrity (Hales et al., 2019; Tincani & Travers, 2019). These data sharing methods could be employed for the purpose of substantiating that SCEDs were used with every case included in a CCCS.

Though optimal, graphically depicting data for each case is not the only way to substantiate that a SCED was used in each case included in a CCCS. The design structure, documentation of an effect of sufficient magnitude, the replication of an effect, and response patterning can be demonstrated indirectly in at least two other ways. The authors could construct a comprehensive table that includes key information for every phase or condition of the analysis, including (a) the number of sessions in each condition or phase, (b) quantitative measures to document the magnitude of effects across every phase of a sequential analysis, or every condition of a multielement design (e.g., mean rates of responding across test and control conditions, overlapping data points across conditions, or percentage change from baseline to treatment), (c) a measure of stability (e.g., range or standard deviation), and (d) measures of trend (e.g., slope). A similar method was employed by Hagopian et al. (1998) when reporting outcomes obtained using FCT across 21 cases (see the Appendix of that article). For a large-scale analysis, such a table could be a supplemental

²Behavioral data reported in a CCCS must come from the primary source (behavioral data obtained via direct observation, or permanent product data). Reporting on data obtained from secondary sources such written records and discharge reports, including those summarizing behavioral data, would better be characterized as a "chart review" rather than a CCCS.

document, but a table summarizing the design and outcomes could be included within the manuscript itself (see Greer et al., 2016 for an example).

A third way to substantiate that a SCED was used with each case would involve combining two methods: (a) presenting a summary table reporting on the experimental design used with each individual case and the magnitude of effects across all conditions (e.g., mean rates of responding for each condition for each case; see Phillips et al., 2017), in combination with (b) the use of a structured and validated set of criteria for interpretation of outcomes. The use of such interpretative criteria, in combination with quantitative data, could provide a sufficient level of evidence that an effect was observed, as structured interpretation methods require the user to take into consideration the design, magnitude of effects, its replication, and response patterning in order to arrive at a conclusion. Ideally, a structured set of criteria, such as the procedures described by Fisher et al. (2003) or Barnard-Brak et al. (2018) could be used for a study reporting on treatment outcomes using a reversal design, and the criteria described by Roane and colleagues (Roane et al., 2013) could be used for a study reporting on functional analysis outcomes. Use of such criteria has the added benefit of allowing other researchers to replicate the methods used for interpretation of data in replication studies. Authors, reviewers, and journal editors will ultimately determine whether those efforts are needed to substantiate that a SCED was used with every case. But in light of the fact that the use of a SCED with every case is what establishes the CCCS as an experimental study and distinguishes it from nonexperimental consecutive case series (see below), there is good justification for the proposed approach. At a minimum, it would be beneficial to graphically depict outcomes of select cases to illustrate both the structure of the experimental analysis, and present exemplars of varied outcomes.

There are situations in which a methodologically rigorous SCED was initiated, but efforts to demonstrate experimental control were not successful because of treatment failure. In a reversal design, this would occur if either the initial application of treatment fails to produce an effect, or if the reapplication of treatment after a reversal to baseline fails to replicate the treatment effect. Although there would be no robust demonstration of experimental control in such cases, cases of treatment failure must be included when reporting outcomes because excluding these cases would result in an overestimate of the general efficacy of the procedure. It is also important to document the number of cases where baseline levels are not recovered after a treatment is withdrawn during a reversal phase, as can occur as a function of carryover, learning, or the influences of other variables (although these should not be reported as failures; see Phillips et al., 2017). See also Tincani and Travers (2019) for further discussion of this issue.

Element 2. All Consecutively Encountered Cases That Underwent the Procedure of Interest or Share a Common Characteristic Are Included when Reporting Outcomes.

Although peripheral to the demonstration of experimental control for studies employing SCEDs, the criteria for including and excluding cases when reporting outcomes has significant implications with respect to determining the generality of findings. If a study's purpose is to report on a specific phenomenon or to illustrate that a procedure *can be* effective, then selecting and reporting findings from only demonstrative cases would be

appropriate and consistent with the study's purpose, so long as statements about the generality of those findings are limited accordingly (see Hales et al., 2019). Including only selective cases with certain outcomes would not allow one to make accurate estimates of the procedure's general efficacy. Selective inclusion of cases based on positive outcomes would restrict the range of outcomes; limiting opportunities to examine variables that distinguish responders from nonresponders, which is critical to defining the limits of a procedure's generality. Therefore, if a study's purpose is not merely to demonstrate that a procedure *can* produce a certain outcome, but to examine *how often* or *under what conditions* the procedure produces a certain outcome (i.e., to examine the general efficacy, or to define the limits of generality, respectively), reporting findings obtained with every consecutively encountered case is imperative.

Biases favoring inclusion of cases with positive outcomes is related to, but not the same as "publication bias", which is a broader term that refers to biases in the article submission, review and editorial process that favors publication of studies reporting positive findings. This is a well-established phenomenon in the research literature across many disciplines, including human subjects research employing randomized group designs. For a comprehensive review of publication bias in studies reporting on randomized controlled trials using group designs, the reader is referred to Dwan et al. (2013). Given that publication bias is ubiquitous in basic and applied sciences, it is likely present in the literature on applied behavior analysis (though evidence for it is quite limited; Sham & Smith, 2014). The reader is referred to Tincani and Travers (2019) and Laraway et al. (2019) for further discussion of publication bias, selective data reporting, and replication research.

In light of the implications of case inclusion methods with respect to the generality of findings, it is important that authors of CCCS studies make explicit statements indicating whether every consecutive case exposed to the procedure is included in reporting the outcomes. The inclusion criteria may be focused on the application of a specific clinical procedure (e.g., assessment or treatment procedure; Derby et al., 1992) or sequence of procedures (e.g., FCT schedule thinning; Greer et al., 2016), a shared clinical target (e.g., SIB; Iwata, Pace et al., 1994), or shared clinical condition (e.g., Fragile-X syndrome; Kurtz et al., 2013). Including this information is necessary to provide a more accurate account of the obtained outcomes for the sample and informs statements about its generality to others. Beyond ensuring that the cases and outcomes are not biased, this detailed information would inform replication studies and permit additional analysis aimed at understanding sources of variation that could enable the analysis of generality.

Element 3: Criteria for Selecting Procedures and Participants Are Described.

Also peripheral to the demonstration of experimental control with SCEDs, but relevant to generality, is how participants are selected for inclusion,³ and how they are assigned to conditions. Formal prospective research studies define inclusion and exclusion criteria for enrollment in the study and prescribe how participants are assigned to conditions—and apply these criteria uniformly across participants. For example, Greer et al., (2019)

³SCRIBE guidelines suggest reporting inclusion/exclusion criteria and recruitment methods (Tate et al., 2016).

compared outcomes achieved with different functional assessment methods across 12 cases who were recruited into a formal study with defined inclusion/ exclusion criteria and assigned to undergo a predefined set of analyses (this could be referred to a prospective CCCS and is discussed in detail further below).

In contrast, a retrospective study reports on outcomes obtained in the context of prior service delivery, where participants and procedures were selected based on a clinical process which can be difficult to define and subject to biases (i.e., therapist judgement, caregiver preference) that can impact the generality of findings (this could be termed a retrospective CCCS and is discussed in detail further below). For example, if a clinician uses a preferred procedure for less challenging cases in which they anticipate it will be efficacious, but not for more challenging cases where they doubt it will be efficacious, the general efficacy of the procedure may be overestimated when findings are examined in a retrospective CCCS. Alternatively, if the therapist reserves a procedure only for cases thought to have more treatment-resistant problems, the results of a retrospective CCCS may underestimate its general effectiveness. An example of a retrospective CCCS by Rooker et al. (2013) reported on outcomes across 50 cases where FCT was used in the context of service delivery. Although all cases where FCT was used were included in reporting the outcomes, the selection of FCT over other available treatment options was the outcome of a clinical process (e.g., therapist judgment, caregiver preference). That process likely varied across cases in ways that are difficult to describe in retrospect and that were potentially subject to biases. Because the criteria for selecting FCT for some cases and not for others cannot be fully known, the extent to which one can ascertain to whom the findings may have generality is somewhat limited. Including information on how participants and procedures are selected is particularly relevant for studies in the field of applied behavior analysis because the literature is composed of studies reporting on data obtained as the product of formal prospective studies, as well as studies reporting retrospectively on data obtained in the context of clinical service delivery.

Despite the inherent limitations of retrospective studies, findings from prospective studies do not necessarily have greater generality. A prospective study with very narrow criteria for participant inclusion may produce findings that have limited generality to the broader population of people with those problems. The use of narrow and nonrepresentative inclusion criteria is a known problem with some large-scale randomized controlled trials that enroll only patients who have the single diagnosis of interest (e.g., depression), even though most individuals with that diagnosis have additional diagnoses (see Rothwell, 2005 for a discussion of this topic). Although prospective studies do not necessarily produce findings that have greater generality than retrospective studies, they do enable one to more precisely determine where and to whom the findings may have generality. In contrast, inexact and difficult-to-define clinical processes leading to the selection of one procedure over others might limit the extent to which the generality can be ascertained with retrospective studies. A prospective research study with inclusion and assignment criteria that are representative of the problem and the use of the intervention in practice would produce findings that have greater generality relative to a study with informal and judgment-based criteria. In light of how the selection of participants or assignment to conditions could impact the generality of findings, it is important for authors of CCCS studies to indicate whether the study is

retrospective or prospective and consider this when making statements about the generality of findings (this is discussed further below).

Element 4. Findings Are Examined Within and Across Participants in a Manner That Preserves the Analysis of Individual Outcomes.

Behavior–environment relations occur at the level of the individual, so any efforts to analyze outcomes across cases must retain this level of analysis. This is particularly true for the study of generality, which relies on identifying the conditions under which outcomes are similar, and when they diverge. Reporting on averages across cases not only obscures the analysis of individual outcomes, but also has the potential to lead to erroneous conclusions. The average outcome implies it was the typical outcome when in fact it may have never occurred. Compiling outcomes obtained with each individual case using a uniform outcome measure allows comparisons across cases. This could include both absolute levels of behavior, and proportional change across conditions (i.e., level of differentiation in the functional analysis, percentage reduction from baseline to treatment for behavior reduction).

Outcomes across participants are better summarized in terms of the proportion of cases where certain outcomes are achieved as opposed to averages. This type of summary provides a way to more accurately gauge the general efficacy of the procedure across the sample. For reporting on assessment findings, the proportion of cases where a certain categorical outcome was achieved could be reported. For example, the authors could report the percentage of cases where different operant functions were identified (Hagopian et al., 2013; Iwata, Pace et al., 1994), the percentage of cases where there was correspondence between assessment methods (Querim et al., 2013), or the percentage of cases where certain response patterns were observed that distinguished participants (Hagopian et al., 2015). For reporting on treatment findings, the percentage of cases where some benchmark outcome was achieved (such as an 80% or 90% reduction relative to baseline; Phillips et al., 2017; Rooker et al., 2013) is useful to characterizing how often the treatment resulted in that outcome with implications for how often others who share similar characteristics might respond as well. In studies reporting on cases where a procedure was applied more than once, findings can also be reported in terms of the number or proportion of applications where some goal was achieved (e.g., Greer et al., 2016; Hagopian et al., 2013).

When studies report on larger numbers of cases, one can use statistical analysis methods to examine effects. Given that producing socially meaningful behavior change is a fundamental goal of applied behavior analysis, statistical analysis is better used to document and quantify an effect that has been determined to have social significance, rather than used as a benchmark to determine if an effect is present. Therefore, when reporting outcomes, authors could start with summarizing individual outcomes (including graphs depicting exemplar cases), describing the proportion of cases achieving some socially meaningful outcome, followed by a summary of outcomes of statistical analyses. Adopting this minimalist approach to statistical analyses will retain a behavior analytic conceptual framework and have the added benefit of making the study more consumable by the broader scientific community.

Graphically summarizing outcomes across cases in accordance with these same principles can facilitate and support the interpretation of findings (see Branch, 2019 for a discussion). Thus, in lieu of bar graphs depicting averages across cases, plotting each individual case or application outcome as a distinct data point will depict outcomes across the collective and document their distribution (see Figure 2 for an example of this type of data depiction reprinted from Greer et al., 2016). If the outcomes vary across cases, then additional methods should be employed to identify factors associated with positive and negative outcomes (see Derby et al., 1992; Hagopian et al., 2015; Iwata, Pace et al., 1994; Phillips et al., 2017). Nonparametric statistical analyses (e.g., conditional probability analysis; Hagopian et al., 2018) could enable the formal quantification of variables that are associated with different outcomes and inform the development of hypotheses that could be tested in a subsequent experimental analysis. Using such methods when reporting outcomes provides an opportunity to directly examine variations in outcomes which is necessary to the study of generality.

Defining the limits of generality of clinical procedures for a particular clinical problem can also inform research into the phenomenological analysis of a clinical problem. Predictive behavioral markers (i.e., objective behavioral measures that predict response to treatment by way of their direct involvement in the causal or treatment-action pathways) can also be thought of as variables that are a determinant of the generality of a treatment procedure (Hagopian et al., 2018). A subsequent prospective, confirmatory, experimental analysis could provide more definitive evidence toward establishing such a functional relation. Thus, the predictive behavioral marker could be used to categorize participants who are otherwise similar to (a) those who test positive for that variable (i.e., they possess the characteristic feature thought to be mediate positive outcomes) and (b) those who test negative for that variable. Next, both groups would be subjected to the treatment that has been associated with more positive outcomes (i.e., treatment is “matched” to that predictive behavioral marker) versus a “mismatched” treatment or a nonspecific treatment. If those who test positive for the variable have a better response to the matched treatment than the mismatched treatment, and respond more favorably than those who test negative, there would be some confirmation of the functional relation. This prospective hypothesis-testing approach could be useful in identifying variables that mediate generality.

Element 5: Multiple Cases Are Included and Well Characterized.

Reporting on more participants does not necessarily result in a corresponding increase in the generality of findings (Branch & Pennypacker, 2013; Johnston & Pennypacker, 2009), just as a larger sample of behavior of an individual does not necessarily result in a corresponding increase in generality of findings for that individual. However, as the number of participants within a study increases, so does the number of direct and systematic replications within the study as participants vary across multiple dimensions. When findings are reproduced across cases, this both contributes to the demonstration of generality, and allows for the direct analysis of generality (and provides an opportunity to explore the variables that mediate it). The larger the sample, the more likely it is to be composed of sufficient numbers of both responders and nonresponders. This provides increased opportunities to observe and examine the sources of variation in outcomes and identify variables that mediate the

generality of procedures (see above, and Hagopian et al., 2018). Just as the replication of outcomes across an increasing number of direct and systematic replication studies in the body of literature provides more evidence of the generality of a procedure, so can the replication of outcomes across a greater variety of participants and problems within a study. Thus, sample size and the extent to which participants are well described interacts with other design elements to impact the generality of findings.

Among all of the design elements discussed above, the characterization of participants appears to have received the most attention in discussion on SCEDs (Kazdin, 2011; Kratochwill & Stoiber, 2002; Smith, 2012) and is among the recommended reporting requirements of Single Case Reporting Guideline In Behavioral Interventions (SCRIBE; Tate et al., 2016). It is important for authors to precisely characterize participants and the setting where the study was conducted (as well as inclusion/exclusion criteria, see Element 3 above). This information is needed to determine to whom the findings are likely to have generality, and can guide systematic replication studies aimed at examining the generality of the procedure to other populations. Because various settings have their own inclusion and exclusion criteria for clients or patients, it is also important to describe how the setting from which participants are drawn provides some information about the participants. For example, findings from studies reporting on inpatients may have greater generality to those with more severe and treatment-resistant problems, as hospitalization is necessitated (and authorized by insurance) only after outpatient treatment has failed and the behavior poses risks to self or others. Although CCCS studies allow authors to make more definitive statements about the generality of findings, it is important that those conclusions be limited based on factors such as the setting, population, and criteria for inclusion. In addition, describing participants in greater detail provides additional data that could be examined in future review or quantitative analyses of the literature. Finally, this type of information can make the study more consumable by the broader scientific community where participant variables, including diagnostic classification measures and standardized test results, are of greater interest. If authors seek broader dissemination of their findings outside the field of behavior analysis, reporting such measures could facilitate citation or publication in non-behavior analytic journals.

CCCS Variants

Retrospective CCCS.—The term “retrospective CCCS” could be used to describe a study that reports on a series of cases for which data were obtained in the course of service delivery then later were compiled and reported. An informal and noncomprehensive review of studies that have elements of the CCCS suggests that many are retrospective studies. Some have examined the general efficacy of clinical procedures (e.g., FA; Iwata, Pace et al., 1994; IISCA; Jessel et al., 2019; FCT; Rooker et al., 2013; NCR; Phillips et al., 2017), the generality of procedures to specific clinical populations (e.g., Fragile X syndrome; Kurtz et al., 2013), or the epidemiology and phenomenology of SIB (Iwata, Pace et al., 1994 and Hagopian et al., 2015, respectively). These studies summarize outcomes obtained in the course of delivering standard clinical care and the assessment or intervention selection is based on the clinician’s judgment about the needs of the client or student. Thus, these services do not meet the definition of research as defined by the U.S. Department of

Health and Human Services' Office for Human Research Protections (OHRP); however, many institutions require Institutional Review Board (IRB) oversight to ensure protection of confidentiality when reporting outcomes of clinically obtained data. Any clinical service agency (e.g., outpatient clinic, school) that utilizes SCEDs to evaluate outcomes could compile datasets over time and later summarize and report on them in a retrospective CCCS study. An IRB approval would be required when publishing clinically acquired data to ensure protection of confidentiality. This method supports an objective, data-based program evaluation process for continued improvement of clinical practices of a program, and also could contribute to the overall knowledge base when the results are published.

Despite its practical utility, an obvious and inherent scientific limitation to the retrospective CCCS, is that each case is assigned to receive a procedure based on a clinical process which may be difficult to retrospectively define, subject to biases, and probably are inconsistently applied across cases. These uncontrolled and unknown sources of variation could impact the generality of outcomes. Clinicians may select a clinical procedure for a given client based on an expectation that it would produce a good outcome and avoid a procedure suspected to prove ineffective, which can be referred to as "treatment assignment bias". This could result in findings that provide an overestimate of the procedure's general effectiveness. This limitation can be addressed by prospectively assigning participants to treatment based on predefined criteria.

Prospective CCCS.—The term "prospective CCCS" could be used to refer to a study that prospectively assigns cases to undergo a specific clinical procedure (e.g., assessment, treatment procedure), or share a characteristic according to criteria defined a priori. This process eliminates assignment bias because participants are assigned to undergo the procedure based on prospectively defined criteria rather than clinician judgment. As a result, one can more precisely define to whom the findings may have greater generality. A prospective CCCS could be conducted under at least two conditions. First, a clinical program may have established a standard recommended practice that involves applying a defined intervention to a specific clinical problem. Such clinical protocols are not uncommon as many programs strive to ensure empirically supported interventions are applied. An IRB approval would be required when publishing data obtained in the course of conducting a program evaluation/improvement project to ensure protection of confidentiality of individuals. The Derby study (1992) established the use of the Brief FA as a standard of care for their clinic, so that study could be characterized as a prospective CCCS in that everyone meeting inclusion criteria for the clinic underwent that procedure. Any clinical service agency with an established practice standard involving the use of a clinical procedure could compile and analyze data as part of a program evaluation or continuous improvement project.

Second, a prospective CCCS could be conducted by enrolling individuals as research participants and defining the procedures that are applied in accord with the study's predetermined criteria. Examples of prospective CCCS studies include those comparing various functional behavioral assessment methods (e.g., Greer et al., 2019) or those examining the phenomenology of a clinical problem (e.g., Rooker et al., 2018; 2019). Prospective CCCS studies conducted outside a program's clinical standard of care would

require IRB approval, and informed consent per the U.S. government's OHRP definition of research. If the target for behavior change is a clinically relevant target, the prospective assignment of participants to a treatment would constitute a clinical trial as defined by the National Institutes of Health (NIH; and would also require IRB approval and informed consent for participation in research). The uniform application of procedures across participants based on prospectively defined criteria for assignment of participants to conditions allows for more precise determination of to whom the findings might have greater generality relative to a retrospective study. As noted previously, however, this does not necessarily produce findings that have greater generality, because a narrowly defined set of inclusion criteria that produce a nonrepresentative sample would limit generality.

Randomized CCCS Designs.—Although conducting a CCCS study provides additional ways to examine the generality of a clinical procedure or the epidemiology or phenomenology of a clinical problem, research questions sometimes are focused on comparing the general efficacy of two clinical procedures. One type of comparison may involve that of two existing procedures known to be efficacious, but it is unclear which is superior (i.e., “clinical equipoise”). For example, although both FCT and NCR are known to be efficacious for socially maintained problem behavior, it is not clear if one has greater general efficacy, or what the limits are of each procedure's generality. Another comparison would be that of a new clinical procedure to a current standard of care procedure that is widely used. Such a comparison would allow for the examination of the general efficacy of the new procedure as well as relative efficacy compared to an established procedure. A practical benefit of both types of comparisons is that all participants receive some type of clinical procedure that could be beneficial, which may be more desirable than being assigned to a no-treatment condition (i.e., placebo, wait-list control). This is particularly true in situations where the target for treatment is a severe clinical problem, where assigning participants to no-treatment control condition could be unsafe and ethically unacceptable. For example, a current clinical trial for treatment-resistant subtypes of automatically maintained SIB (Hagopian, 2019; Identification No. [NCT03995966](#)) involves comparing a newly developed treatment (i.e., combining multiple sources of reinforcer competition targeting SIB, and other components targeting self-restraint) to NCR with competing stimuli. Given what is known about the severity of these treatment-resistant subtypes of self-injury and the risks for injury and increased restraint (Hagopian et al., 2017; Rooker et al., 2019), a no-treatment control condition would not be acceptable.

For group designs, randomization controls for threats to internal validity by eliminating bias related to participant assignment, increasing the probability that groups are equivalent and assignment to groups or conditions is free from bias. In fact, random assignment distinguishes experimental from quasi-experimental or nonexperimental group designs. Combining the control offered through randomization to groups or conditions with demonstration of functional control within cases inherent to SCEDs provides added rigor and retains the focus on the analysis of behavior at the level of the individual. Whereas the retrospective or prospective CCCS is best characterized as a type of study involving multiple SCEDs, the term “randomized CCCS” could be used to refer to a design that combines elements of SCEDs with randomized group designs. The Randomized CCCS is a type of

prospective CCCS. The SCED would necessarily be adaptive as it is when performing an analysis with a single participant, in that it would include ongoing monitoring of behavior, and consideration of response patterns to inform phase changes and treatment modifications. General criteria for making phase changes and for adding treatment components would be defined in a manner that could be both replicable and remain adaptive to individual variation. Table 2 provides a summary of similarities and differences among the CCCS variations.

Two types of randomized CCCS designs are described and illustrated in Figure 3: a) the “randomized parallel-group CCCS”, and b) the “randomized crossover CCCS”. In the randomized parallel-group CCCS, participants are randomly assigned to undergo either one of two clinical procedures (thus creating two groups). Each participant undergoes the procedure assigned, and then his/her participation ends regardless of the outcome (Top Panel). As noted above, studies that could now be characterized as using a randomized parallel-group CCCS design or something very similar to it have been reported for some time (e.g., Durand & Carr, 1992; Neef et al., 1991; Peterson et al., 2016). With the randomized crossover CCCS, participants are randomly assigned to one of two sequences (starting with procedure A then crossing over to procedure B; or vice versa). If the first assigned procedure produces a defined positive outcome, the trial is ended; but if it fails to produce the targeted outcome, participants “crossover” to undergo the other procedure (after a brief washout period to minimize carryover, and a new baseline is established; see Bottom Panel). The randomized crossover CCCS is currently being employed in a clinical trial for treatment resistant subtypes of automatically maintained SIB (Hagopian, 2019; Identification No. [NCT03995966](#)). Participants are assigned randomly to receive either NCR with competing stimuli or a novel treatment involving multiple sources of reinforcer competition and other components targeting self-restraint. If the first treatment is efficacious through generalization (based on uniform criteria), the trial is ended for that participant; if not, then the participant crosses over to the other treatment.

With both designs, randomization to treatment group or treatment sequence provides additional controls to ensure groups (or treatment sequences) are equivalent, and the use of SCEDs enable the demonstration of control within cases. Each treatment is evaluated using an SCED, which permits comparison of the two treatments (across groups or within cases) like the typical randomized group design (parallel-group, or crossover design), but it also allows for highly detailed analysis of outcomes with each individual case. The randomized crossover CCCS design would be better suited in situations where participants are limited in number (as each participant will undergo both procedures). The randomized parallel-group CCCS design would be better suited if either intervention is likely to produce unreversible and lasting effects that do not wane during the washout period, or if more participants are available. Either the randomized parallel-group or crossover CCCS would be optimal for conducting controlled behavior-analytic clinical trials. The general efficacy of each treatment would best be examined by comparing the proportion of cases achieving various benchmarks for efficacy, rather than by comparing averaged outcomes. That is, a binary system (i.e., percentage of responders vs. nonresponders based on whether an 80% reduction in problem behavior was achieved) could be used. Alternatively, one could report on the percentage of cases with strong effects (i.e., 80–100% reduction in problem

behavior), moderate effects (50–80% reduction), or weak effects (less than 50% reduction). Since outcomes have multiple dimensions (e.g., quality of life, satisfaction) those too should be reported in terms of percentage of cases rather than as averages.

These mixed SCED-randomized group designs would also allow for traditionally used statistical analyses to compare the relative efficacy of two treatments. Although the value of inferential statistics to evaluate outcomes remains debatable within the field of applied behavior analysis, grant-awarding agencies and non-behavior analytic journals typically require the use of statistical analysis. Attending to these contingencies is important for those behavioral researchers seeking research funding and wishing to disseminate findings to other scientific communities. As noted previously, statistical analysis need not be used as a standard to document that an effect is present but could be used to quantify and document an effect determined to be socially meaningful. The randomized CCCS may provide a vehicle for conducting a behavior analytic trial that meets both the standards of rigor required of applied behavior analysis and of the broader scientific community.

Distinguishing Features of the CCCS

CCCS studies are designed to examine the general efficacy of a procedure (i.e., to determine how often a procedure can be efficacious); they have the potential to identify the conditions under which it is efficacious, and thus help define the limits of its generality. To achieve these purposes, CCCS studies used SCEDs in combination with other design, data-analytic and reporting methods above and beyond the fundamental design requirements of the SCED. That is, for studies designed with the purpose of examining a functional relation or demonstrating that a clinical procedure can be efficacious, only the fundamental design elements of SCEDs are necessary (e.g., a precisely defined dependent variable, controlled manipulation of the independent variable, phase changes based on response patterning, replication of effects, etc.). Unlike the CCCS, SCED demonstration studies do not require (a) that all cases where the procedure was applied are included in the reporting of results, (b) that one specify whether conditions are prospectively assigned or whether the study is a retrospective report of data acquired in the course of service delivery, or (c) more than one participant.

The CCCS study is also distinguished from the nonexperimental “consecutive case series” used in medicine and psychology (National Cancer Institute, n.d.) and clinical replication studies (Barlow et al., 1984). These similar types of studies involve the application of a defined treatment to a series of cases presenting with similar problems. Although those methods would allow the use of a SCED for each case in the series, neither of those methods require it. Thus, an AB design can be, and often is, used (e.g., Gillespie et al., 2002; Ren et al., 2000), and outcomes may be examined using indirect measures of behavior (Peterson & Halstead, 1998). Thus, the retrospective or prospective CCCS could be considered a type of clinical replication study, with the qualifier that it is a controlled study using an SCED for each case (i.e., consecutive controlled case series). Despite overlap in some features of the CCCS with these other approaches, there are enough differences to set them apart.

The CCCS is also distinguished from the “Series N-of-1 trials” used in other fields, which involve the use of rapid reversals across series of cases (Vohra et al., 2015). These have

been used in neurorehabilitation (Perdices & Tate, 2009), psychopharmacology (Mirza et al., 2017), and across various other medical specialties and health conditions (Punja et al., 2014). Series N-of-1 trials are limited to the use of reversal designs and can also involve randomization of conditions and blinding to add additional rigor and control. Unlike the adaptive use of SCEDs in applied behavior analysis, where there is repeated measurement of behavior and ongoing analysis of response patterns (i.e., trend, variability, magnitude of effects) to guide phase changes and demonstrate experimental control, N-of-1 trials need not (a) establish a baseline, (b) use direct behavioral observation, (c) conduct ongoing analysis of response patterning to guide phase changes, or (d) include more than two data points per phase. The use of N-of-1 trials has been advocated in recent years because the focus involves evaluation of individual outcomes, and the analysis of heterogeneity across individuals is congruent with the paradigm of precision medicine (Schork, 2015; Shamseer et al., 2015; Vohra 2016). Recognition of the importance of preserving individual outcomes when examining findings across multiples cases aligns nicely with the fundamental purposes of the CCCS methodologies.

SUMMARY AND CONCLUSIONS

As noted by Johnston and Pennypaker (2009), a field's experimental questions influence its methods, and ultimately its direction and development. It follows that an over-reliance on any one experimental method can limit the development of a field. The SCEDs have proven invaluable in both the conduct of applied research and clinical practice because they are optimal for asking many important experimental questions. They can help demonstrate a functional relation between variables essential for the analysis of environment-behavior interactions. They can also isolate the operative mechanisms of behavior change procedures, and they can demonstrate that a clinical procedure can be efficacious. Documenting the generality of clinical procedures or processes (i.e., functional relations) requires that findings are reproduced across contexts, cases, and studies. The CCCS methodologies extend the use of SCEDs to enhance our ability to study generality within a single study. By conducting multiple replications across all consecutively encountered cases and employing data-analytic methods that preserve the analysis of behavior, researchers can ask additional questions pertaining to the generality of procedures or processes. These designs enable researchers to go beyond asking whether a procedure can be efficacious to asking how often the procedure can produce positive outcomes (i.e., general efficacy). If there are sufficient numbers of responders and nonresponders within a CCCS, one can potentially identify variables that are correlated with positive and negative outcomes and ask questions to identify the conditions under which the procedure is efficacious (i.e., determine the limits of generality).

The retrospective CCCS is also ideally suited to evaluate the general efficacy of clinical procedures delivered in the course of providing clinical services. Their use would support rigorous program evaluation efforts that promote the integration of research and practice and inform service improvements, and advance our knowledge about the efficacy of procedures. The prospective CCCS offers a higher degree of control as participants are assigned to conditions and, thus, could be used in the conduct of a more rigorous program evaluation project or for conducting a formal research study. These methods also enable the advancement of our understanding of a clinical problem in terms of identifying unique

functional relations that have generality among individuals with that clinical problem, and their relative prevalence (i.e., phenomenology and epidemiology). Finally, when the experimental question focuses on the general efficacy of two clinical procedures, combining SCEDs with randomized group designs (i.e., the randomized CCCS) can allow one to conduct a wholly behavior analytic randomized clinical trial. The various methods discussed here each have their purposes and should be used accordingly to answer different types of experimental questions (see Table 3).

In discussing the experimental analysis of behavior and its focus on the individual organism relative to other disciplines, Skinner (p. 21, 1966) noted that “*instead of studying a thousand rats for one hour each... the investigator is likely to study one rat for a thousand hours.*” Because behavior– environment interactions occur at the level of the individual, efforts to understand generality must retain a granular level of analysis. To this end, the use of SCEDs will remain the cornerstone of behavior analytic research. At the same time, to study generality we must extend our analyses across multiple individuals. Although the demonstration of generality of clinical procedures and process seemingly requires documenting a common outcome, the analysis of generality and its determinants begins with documenting differences in outcomes. Thus, for questions related to the analysis of generality, we may in reference to Skinner’s comment, need to study one thousand subjects for a thousand hours.

Acknowledgments

Manuscript preparation was supported by Grant 2R01HD076653-06 from the EuniceK. Shriver National Institute of Child Health and Human Development (NICHD), and U54 HD079123 from the Intellectual and Developmental Disabilities Research Centers (IDDRC). Its contents are solely the responsibility of the author and do not necessarily represent the official views of NICHD or IDDRC.

The author wishes to acknowledge Lynn Bowman, John Michael Falligant, Greg Hanley, Patricia Kurtz, and the reviewers for their assistance with this manuscript.

REFERENCES

- Baer DM, Wolf MM, & Risley TR (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91–97. 10.1901/jaba.1968.1-91. [PubMed: 16795165]
- Barlow DH, Hayes SC, & Nelson-Gray RO (1984). *The scientist practitioner: Research and accountability in clinical and educational settings* (No. 128). Pergamon Press.
- Barlow DH, & Hersen M (1984). *Single-case experimental designs: Strategies for studying behavior change* (2nd ed.). Pergamon Press.
- Barnard-Brak L, Richman DM, Little TD, & Yang Z (2018). Development of an in-vivo metric to aid visual inspection of single-case design data: Do we need to run more sessions? *Behaviour Research and Therapy*, 102, 8–15. 10.1016/j.brat.2017.12.003. [PubMed: 29277023]
- Becraft JL, Borrero JC, Shuyan S, & McKenzie AA (in press). A primer for using multilevel models to meta-analyze single case design data with AB phases. *Journal of Applied Behavior Analysis*.
- Branch MN (2019). The “reproducibility crisis:” Might the methods used frequently in behavior-analysis research help? *Perspectives on Behavior Science*, 42, 77–89. 10.1007/s40614-018-0158-5. [PubMed: 31976422]
- Branch MN, & Pennypacker HS (2013). Generality and generalization of research findings. In Madden GJ, Dube WV, Hackenberg TD, Hanley GP, & Lattal KA (Eds.), *APA handbook of behavior analysis, Vol. 1. Methods and principles* (pp. 151–175). American Psychological Association. 10.1037/13937-007.

- Call NA, Simmons CA, Mevers JEL, & Alvarez JP (2015). Clinical outcomes of behavioral treatments for pica in children with developmental disabilities. *Journal of Autism and Developmental Disorders*, 45, 2105–2114. 10.1007/s10803-015-2375-z. [PubMed: 25636679]
- Carr JE, Severtson JM, & Lepper TL (2009). Noncontingent reinforcement is an empirically supported treatment for problem behavior exhibited by individuals with developmental disabilities. *Research in Developmental Disabilities*, 30, 44–57. 10.1016/j.ridd.2008.03.002. [PubMed: 18467073]
- Chambless DL, & Hollon SD (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7–18. 10.1037/0022-006X.66.1.7. [PubMed: 9489259]
- Cooper LJ, Wacker DP, Sasso GM, Reimers TM, & Donn LK (1990). Using parents as therapists to evaluate appropriate behavior of their children: Application to a tertiary diagnostic clinic. *Journal of Applied Behavior Analysis*, 23, 285–296. 10.1901/jaba.1990.23-285. [PubMed: 2249965]
- Derby KM, Wacker DP, Sasso G, Steege M, Northup J, Cigrand K, & Asmus J (1992). Brief functional assessment techniques to evaluate aberrant behavior in an outpatient setting: A summary of 79 cases. *Journal of Applied Behavior Analysis*, 25, 713–721. 10.1901/jaba.1992.25-713. [PubMed: 1429322]
- Durand VM, & Carr EG (1992). An analysis of maintenance following functional communication training. *Journal of Applied Behavior Analysis*, 25, 777–794. 10.1901/jaba.1992.25-777. [PubMed: 1478902]
- Dwan K, Gamble C, Williamson PR, & Kirkham JJ (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. *PloS One*, 8, e66844. 10.1371/journal.pone.0066844. [PubMed: 23861749]
- Figshare. (n.d.). <https://figshare.com>
- Fisher WW, Kelley ME, & Lomas JE (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis*, 36, 387–406. 10.1901/jaba.2003.36-387. [PubMed: 14596583]
- Gast DL, & Wolery M (1988). Parallel treatments design: A nested single subject design for comparing instructional procedures. *Education and Treatment of Children*, 11, 270–285. <https://www.jstor.org/stable/42899073>.
- Gillespie K, Duffy M, Hackmann A, & Clark DM (2002). Community based cognitive therapy in the treatment of post-traumatic stress disorder following the Omagh bomb. *Behaviour Research and Therapy*, 40, 345–357. 10.1016/S0005-7967(02)00004-9. [PubMed: 12002894]
- Greer BD, Fisher WW, Saini V, Owen TM, & Jones JK (2016). Functional communication training during reinforcement schedule thinning: An analysis of 25 applications. *Journal of Applied Behavior Analysis*, 49, 105–121. 10.1002/jaba.265. [PubMed: 26482103]
- Greer BD, Mitteer DR, Briggs AM, Fisher WW, & Sodawasser AJ (2020). Comparisons of standardized and interview-informed synthesized reinforcement contingencies relative to functional analysis. *Journal of Applied Behavior Analysis*, 53(1), 82–101. 10.1002/jaba.601. [PubMed: 31281974]
- Hagopian LP (2019). A clinical trial for self-injurious behavior. <http://clinicaltrials.gov/ct2/home> (Identification No. [NCT03995966](https://doi.org/10.1186/1745-6215-9-66))
- Hagopian LP, Fisher WW, Sullivan MT, Acquisto J, & LeBlanc LA (1998). Effectiveness of functional communication training with and without extinction and punishment: A summary of 21 inpatient cases. *Journal of Applied Behavior Analysis*, 31, 211–235. 10.1901/jaba.1998.31-211. [PubMed: 9652101]
- Hagopian LP, Rooker GW, Jessel J, & DeLeon IG (2013). Initial functional analysis outcomes and modifications in pursuit of differentiation: A summary of 176 inpatient cases. *Journal of Applied Behavior Analysis*, 46, 88–100. 10.1002/jaba.25. [PubMed: 24114088]
- Hagopian LP, Rooker GW, & Yenokyan G (2018). Identifying predictive behavioral markers: A demonstration using automatically reinforced self-injurious behavior. *Journal of Applied Behavior Analysis*, 51, 443–465. 10.1002/jaba.477. [PubMed: 29781180]
- Hagopian LP, Rooker GW, & Zarccone JR (2015). Delineating subtypes of self-injurious behavior maintained by automatic reinforcement. *Journal of Applied Behavior Analysis*, 48, 523–543. 10.1002/jaba.236. [PubMed: 26223959]

- Hagopian LP, Rooker GW, Zarcone JR, Bonner AC, & Arevalo AR (2017). Further analysis of subtypes of automatically reinforced SIB: A replication and quantitative analysis of published datasets. *Journal of Applied Behavior Analysis*, 50, 48–66. 10.1002/jaba.368. [PubMed: 28032344]
- Hales AH, Wessellmann ED, & Hilgard J (2019). Improving psychological science through transparency and openness: An overview. *Perspectives on Behavior Science*, 42, 13–31. 10.1007/s40614-018-00186-8. [PubMed: 31976419]
- Hanley GP (2017). Editor's note. *Journal of Applied Behavior Analysis*, 50, 3–7. 10.1002/jaba.366. [PubMed: 28083903]
- Iwata BA, Dorsey MF, Slifer KJ, Bauman KE, & Richman GS (1994a). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis*, 27, 197–209. 10.1901/jaba.1994.27-197 (Reprinted from *Analysis and Intervention in Developmental Disabilities*, 2, 3–20, 1982). [PubMed: 8063622]
- Iwata BA, Pace GM, Dorsey MF, Zarcone JR, Vollmer TR, Smith RG, ... Willis KD (1994b). The functions of self-injurious behavior: An experimental-epidemiological analysis. *Journal of Applied Behavior Analysis*, 27, 215–240. 10.1901/jaba.1994.27-215. [PubMed: 8063623]
- Jessel J, Ingvarsson ET, Metras R, Kirk H, & Whipple R (2018). Achieving socially significant reductions in problem behavior following the interview-informed synthesized contingency analysis: A summary of 25 outpatient applications. *Journal of Applied Behavior Analysis*, 51, 130–157. 10.1002/jaba.436. [PubMed: 29319154]
- Jessel J, Metras R, Hanley GP, Jessel C, & Ingvarsson ET (2019). Evaluating the boundaries of analytic efficiency and control: A consecutive controlled case series of 26 functional analyses. *Journal of Applied Behavior Analysis*, 53(1), 25–43. 10.1002/jaba.544. [PubMed: 30734285]
- Johnston JM, & Pennypacker HS (2009). *Strategies and tactics of behavioral research* (3rd ed.). Routledge.
- Kazdin AE (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). Oxford University Press.
- Kratochwill TR, & Stoiber KC (2002). Evidence-based interventions in school psychology: Conceptual foundations of the Procedural and Coding Manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly*, 17, 341–389. 10.1521/scpq.17.4.341.20872.
- Kurtz PF, Boelter EW, Jarmolowicz DP, Chin MD, & Hagopian LP (2011). An analysis of functional communication training as an empirically supported treatment for problem behavior displayed by individuals with intellectual disabilities. *Research in Developmental Disabilities*, 32, 2935–2942. 10.1016/j.ridd.2011.05.009. [PubMed: 21696917]
- Kurtz PF, Chin MD, Huete JM, Tarbox RS, O'Connor JT, Paclawskyj TR, & Rush KS (2003). Functional analysis and treatment of self-injurious behavior in young children: A summary of 30 cases. *Journal of Applied Behavior Analysis*, 36, 205–219. 10.1901/jaba.2003.36-205. [PubMed: 12858985]
- Kurtz PF, Chin MD, Robinson AN, O'Connor JT, & Hagopian LP (2015). Functional analysis and treatment of problem behavior exhibited by children with fragile X syndrome. *Research in Developmental Disabilities*, 43, 150–166. 10.1016/j.ridd.2015.06.010. [PubMed: 26183339]
- Kurtz PF, Fodstad JC, Huete JM, & Hagopian LP (2013). Caregiver- and staff-conducted functional analysis outcomes: A summary of 52 cases. *Journal of Applied Behavior Analysis*, 46, 738–749. 10.1002/jaba.87. [PubMed: 24114788]
- Laraway S, Snyckerski S, Pradhan S, & Huitema BE (2019). An overview of scientific reproducibility: Consideration of relevant issues for behavior science/analysis. *Perspectives on Behavior Science*, 42, 33–57. 10.1007/s40614-019-00193-3. [PubMed: 31976420]
- Mirza RD, Punja S, Vohra S, & Guyatt G (2017). The history and development of N-of-1 trials. *Journal of the Royal Society of Medicine*, 110, 330–340. 10.1177/0141076817721131. [PubMed: 28776473]
- Mueller MM, Nkosi A, & Hine JF (2011). Functional analysis in public schools: A summary of 90 functional analyses. *Journal of Applied Behavior Analysis*, 44, 807–818. 10.1901/jaba.2011.44-807. [PubMed: 22219531]

- National Cancer Institute (n.d.). <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/consecutive-case-series>
- Neef NA, Trachtenberg S, Loeb J, & Sterner K (1991). Video-based training of respite care providers: An interactional analysis of presentation format. *Journal of Applied Behavior Analysis*, 24, 473–486. 10.1901/jaba.1991.24-473. [PubMed: 1836458]
- Northup J, Wacker D, Sasso G, Steege M, Cigrand K, Cook J, & DeRaad A (1991). A brief functional analysis of aggressive and alternative behavior in an outclinic setting. *Journal of Applied Behavior Analysis*, 24, 509–522. 10.1901/jaba.1991.24-509. [PubMed: 1752840]
- Perdices M, & Tate RL (2009). Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognized and undervalued? *Neuropsychological Rehabilitation*, 19, 904–927. 10.1080/09602010903040691. [PubMed: 19657974]
- Peterson AL, & Halstead TS (1998). Group cognitive behavior therapy for depression in a community setting: A clinical replication series. *Behavior Therapy*, 29, 3–18. 10.1016/S0005-7894(98)80015-4.
- Peterson KM, Piazza CC, & Volkert VM (2016). A comparison of the modified-sequential oral sensory approach to an applied-behavior analytic approach in the treatment of food selectivity in children with autism spectrum disorder. *Journal of Applied Behavior Analysis*, 49, 485–511. 10.1002/jaba.332. [PubMed: 27449267]
- Phillips CL, Iannaccone JA, Rooker GW, & Hagopian LP (2017). Noncontingent reinforcement for the treatment of severe problem behavior: An analysis of 27 consecutive applications. *Journal of Applied Behavior Analysis*, 50, 357–376. 10.1002/jaba.376. [PubMed: 28177118]
- Punja S, Vohra S, Eslick I, & Duan N (2014). An ethical framework for N-of-1 trials: Clinical care, quality improvement, or human subjects research? In Kravitz RL & Duan N (Eds.), *Design and implementation of N-of-1 trials: A user's guide* (Publication No. 13(14)-EHC122-EF) (pp. 13–22). Agency for Healthcare Research and Quality. <https://effectivehealthcare.ahrq.gov/products/n-1-trials/research-2014-3>
- Querim AC, Iwata BA, Roscoe EM, Schlichenmeyer KJ, Ortega JV, & Hurl KE (2013). Functional analysis screening for problem behavior maintained by automatic reinforcement. *Journal of Applied Behavior Analysis*, 46, 47–60. 10.1002/jaba.26. [PubMed: 24114084]
- Ren CJ, Patterson E, & Gagner M (2000). Early results of laparoscopic biliopancreatic diversion with duodenal switch: a case series of 40 consecutive patients. *Obesity Surgery*, 10, 514–523. 10.1381/096089200321593715. [PubMed: 11175958]
- Richman DM, Barnard-Brak L, Grubb L, Bosch A, & Abby L (2015). Meta-analysis of non-contingent reinforcement effects on problem behavior. *Journal of Applied Behavior Analysis*, 48, 131–152. 10.1002/jaba.189. [PubMed: 25754894]
- Roane HS, Fisher WW, Kelley ME, Mevers JL, & Bouxsein KJ (2013). Using modified visual-inspection criteria to interpret functional analysis outcomes. *Journal of Applied Behavior Analysis*, 46, 130–146. 10.1002/jaba.13. [PubMed: 24114090]
- Rothwell PM (2005). External validity of randomised controlled trials: “To whom do the results of this trial apply?”. *The Lancet*, 365, 82–93. 10.1016/S0140-6736(04)17670-8.
- Rooker GW, Hagopian LP, Becraft JL, Javed N, Fisher AB, & Finney KS (2019). Injury characteristics across functional classes of self-injurious behavior. *Journal of Applied Behavior Analysis*. E-pub ahead of print. 10.1002/jaba.664
- Rooker GW, Hausman NL, Fisher AB, Gregory MK, Lawell JL, & Hagopian LP (2018). Classification of injuries observed in functional classes of self-injurious behaviour. *Journal of Intellectual Disability Research*, 62, 1086–1096. 10.1111/jir.12535. [PubMed: 30043452]
- Rooker GW, Jessel J, Kurtz PF, & Hagopian LP (2013). Functional communication training with and without alternative reinforcement and punishment: An analysis of 58 applications. *Journal of Applied Behavior Analysis*, 46, 708–722. 10.1002/jaba.76. [PubMed: 24114463]
- Schork NJ (2015). Personalized medicine: Time for one-person trials. *Nature*, 520, 609–611. 10.1038/520609a. [PubMed: 25925459]
- Sham E, & Smith T (2014). Publication bias in studies of an applied behavior-analytic intervention: An initial analysis. *Journal of Applied Behavior Analysis*, 47, 663–678. 10.1002/jaba.146. [PubMed: 24990802]

- Shamseer L, Sampson M, Bukutu C, Schmid CH, Nikles J, Tate R, ... CENT group. (2015). CONSORT extension for reporting N-of-1 trials (CENT) 2015: Explanation and elaboration. *British Medical Journal*, 350, 1–30. 10.1136/bmj.h1793.
- Sidman M (1960). *Tactics of scientific research*. Basic Books.
- Skinner BF (1966). Operant behavior. In Honig WK (Ed.), *Operant behavior: Areas of research and application* (pp. 12–32). Appleton-Century-Crofts.
- Slaton JD, Hanley GP, & Raftery KJ (2017). Interview-informed functional analyses: A comparison of synthesized and isolated components. *Journal of Applied Behavior Analysis*, 50, 252–277. 10.1002/jaba.384. [PubMed: 28304093]
- Smith JD (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, 17, 510–550. 10.1037/a0029312. [PubMed: 22845874]
- Tate RL, Perdices M, Rosenkoetter U, Shadish W, Vohra S, Barlow DH, & Cen JW (2016). The single-case reporting guidelines in behavioural interventions (SCRIBE) 2016 statement. *Archives of Scientific Psychology*, 4, 1–9. 10.1037/arc0000026.
- Tincani M, & Travers J (2019). Replication research, publication bias, and applied behavior analysis. *Perspectives on Behavior Science*, 42, 59–75. 10.1007/s40614-019-00191-5. [PubMed: 31976421]
- Virués-Ortega J (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose–response meta-analysis of multiple outcomes. *Clinical Psychology Review*, 30, 387–399. 10.1016/j.cpr.2010.01.008. [PubMed: 20223569]
- Vohra S (2016). N-of-1 trials to enhance patient outcomes: Identifying effective therapies and reducing harms, one patient at a time. *Journal of Clinical Epidemiology*, 76, 6–8. 10.1016/j.jclinepi.2016.03.028. [PubMed: 27134139]
- Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, ... the CENT group. (2015). CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *British Medical Journal*, 350, 1–7. 10.1136/bmj.h1738.
- Watkins L, Ledbetter-Cho K, O'Reilly M, Barnard-Brak L, & Garcia-Grau P (2019). Interventions for students with autism in inclusive settings: A best-evidence synthesis and meta-analysis. *Psychological Bulletin*, 145, 490–507. 10.1037/bul0000190 [PubMed: 30869925]
- Wong C, Odom SL, Hume KA, Cox AW, Fettig A, Kucharczyk S, ... Schultz TR (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders*, 45, 1951–1966. 10.1007/s10803-014-2351-z. [PubMed: 25578338]

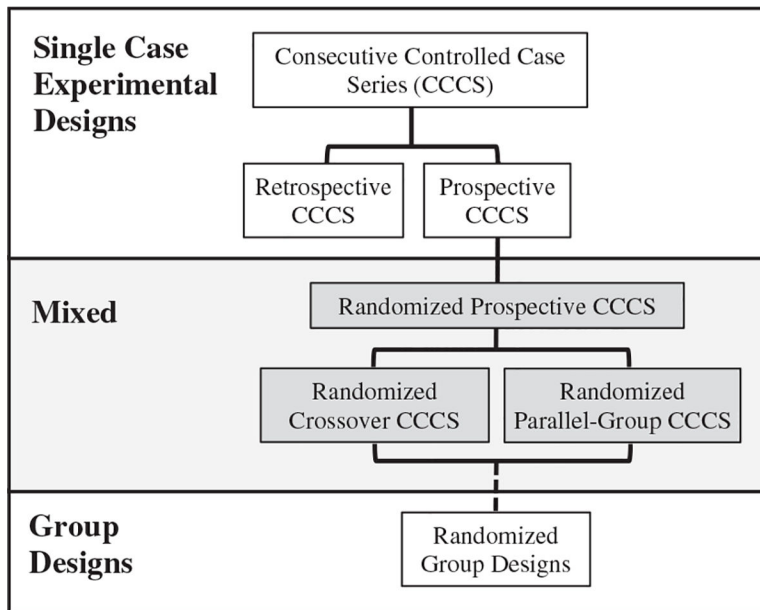


Figure 1. A Schematic of the Retrospective and Prospective CCCS Studies. *Note.* Shows how studies use Single Case Experimental Designs (SCEDs), whereas the randomized CCCS designs combine SCEDs and randomized group designs.

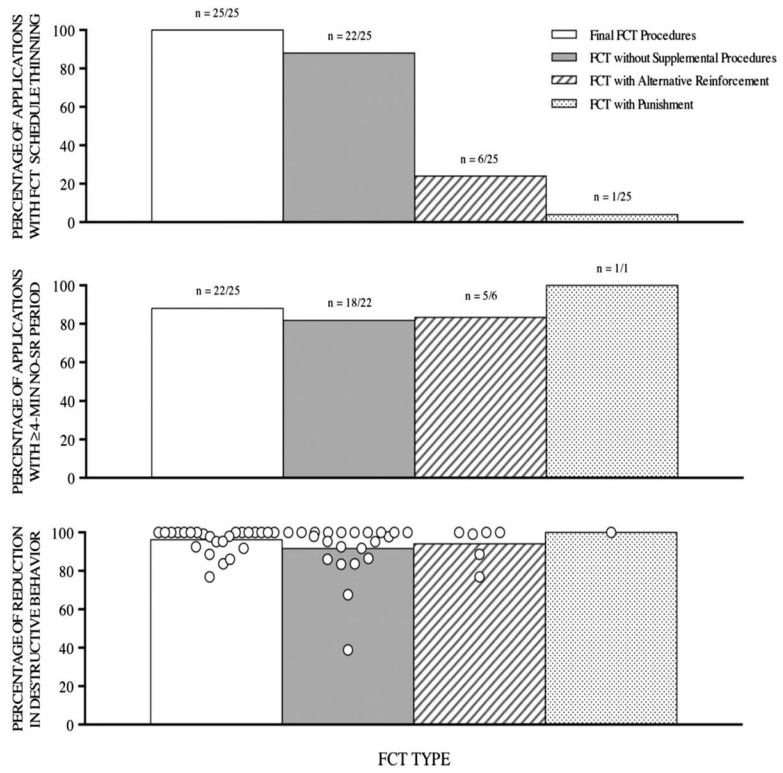


Figure 2.
Reprinted from Greer et al. (2016) Note. Figure illustrates how results are reported in terms of proportion of applications, and findings obtained with each application are depicted separately allowing analysis of the relative distribution of outcomes, as well as differences across conditions or groups.

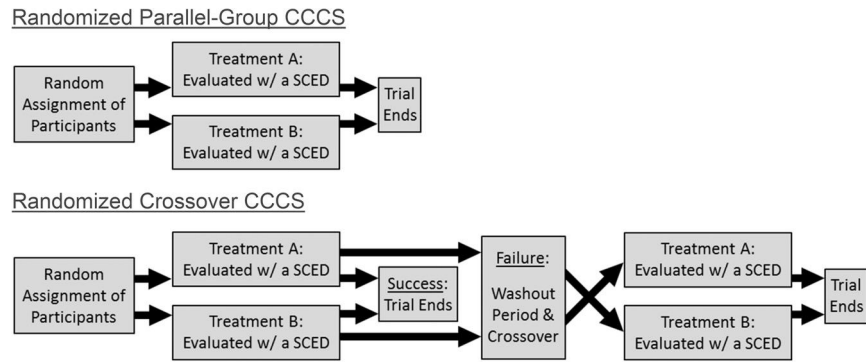


Figure 3.
Schematic Illustrations of Research Participant Progression Through the Two Types of Randomized CCCS Designs. Note. CCCS = Consecutive Controlled Case Series.

Table 1

Design Elements, Data-analytics, and Reporting Methods of CCCS Studies

Element	Explanation	Additional information
Element 1. A SCED is employed with each case.	A SCED is employed with each case where the IV is controlled, the DV operationally defined, and phase changes are initiated at appropriate times based on level, trend, and stability of responding.	Treatment failures of any type must be included when reporting outcomes (even when efforts to demonstrate an initial effect, or replicate an effect fail).
Element 2. All consecutively encountered cases that underwent the procedure of interest or share a common characteristic are included when reporting outcomes.	Reporting findings obtained with every consecutively encountered case regardless of outcomes is imperative in order to examine how often or under what conditions the procedure produces a certain outcome (i.e., general efficacy and the limits of generality, respectively).	Inclusion of all cases provides opportunities to examine variables that distinguish responders from nonresponders, which is critical to defining the limits of a procedure's generality, and identifying variables are correlated with or possibly mediate generality.
Element 3: Criteria for selecting procedures and participants are described.	The extent to which criteria for selecting participants and how they are assigned to conditions are described enables one to more precisely define where and to whom the findings may have generality.	It is important to distinguish between a prospective CCCS study (with predefined criteria) from a retrospective CCCS study (where criteria may be difficult to describe in retrospect, and were potentially subject to biases).
Element 4. Findings are examined within and across participants in a manner that preserves the analysis of individual outcomes.	While behavior-environment interactions occur at the level of the individual, the study of generality also requires us to extend our analysis of behavior across multiple individuals.	Outcomes across participants are better described in terms of the percentage of cases where certain outcomes were obtained rather than averages, which can lead to erroneous conclusions.
Element 5: Multiple cases are included and well-characterized.	Sample size and the extent to which the setting and the participants are described interacts with other design elements to impact the extent to which one can determine to whom the findings may have more or less generality.	A larger sample does not necessarily result in a corresponding increase in the generality of findings, but with a sufficient number of responders and nonresponders comes more opportunities to study generality and examine its determinants.

Note. CCCS = Consecutive Controlled Case Series.

Table 2

Similarities and Differences Among the CCCS Variants

Study Dimension	Retrospective CCCS	Prospective CCCS	Randomized CCCS; Parallel-Group & Crossover CCCS
Type of study	A retrospective study involving the analysis of clinical data (non-research)	A prospective program evaluation project; or a prospective research study	A prospective clinical trial; combines SCEDs with randomized group designs. Could be used to conduct a randomized controlled trial
Source of data	Data obtained in the course of clinical service delivery	Data obtained from a prospective protocol with pre-defined criteria for selection of procedures or participants	Same as prospective CCCS
Assignment to procedures & conditions	Based on a clinical process that includes clinician/caregiver preference and judgement	Prospective assignment of participants to the procedure based on predefined criteria	Prospective random assignment of participants to: a) either treatment exclusively (parallel-group CCCS), or b) one treatment, followed by the other treatment if needed (crossover CCCS)
Sources of experimental control	Within-subjects via single-case experimental designs (SCEDs)	Same as retrospective CCCS	Same as prospective CCCS; and across randomized groups or conditions eliminating bias and random differences
Types of comparisons	Across baseline and treatment conditions within and across participants	Same as retrospective CCCS	Same as prospective CCCS; and across groups or treatment conditions
Role of IRB	IRB oversight of protection of confidentiality related to publication; not for treatment	Same as retrospective CCCS for program evaluation; IRB oversight of all procedures for research studies	IRB oversight for all procedures

Note. CCCS = Consecutive Controlled Case Series; IRB = Institutional review board.

Table 3

Study Purpose, Exemplar Experimental Questions, and Type of Study or Experimental Design

Study Purpose/Exemplar Experimental Questions (EQ)	Type of Study or Design			
	SCED	CCCS	Rand. CCCS	Rand. Group
<u>Demonstration of Experimental Control</u> EQ: Can a functional relation be demonstrated?	✓	✓	✓	✓
<u>Demonstration of the Efficacy of a Clinical Procedure</u> EQ: Can this clinical procedure produce a positive outcome?	✓	✓	✓	✓
<u>Evaluation of the General Efficacy of a Clinical Procedure</u> EQ: In what proportion of cases does this clinical procedure produce positive outcomes?		✓	✓	✓
<u>Study of Epidemiology/Phenomenology of a Clinical Problem</u> EQ: For a clinical problem, what functional relations are present and what is their prevalence?		✓		
<u>Evaluation of the General Efficacy of Two Clinical Procedures within a Randomized Controlled Trial</u> EQ: In what proportion of cases does each clinical procedure produce positive outcomes, and what variables are associated with various outcomes?			✓	✓

Note. SCED = Single case experimental design (not involving CCCS methods).

CCCS = Consecutive controlled case series (Retrospective and Prospective).

Rand. CCCS = Randomized Prospective CCCS (Parallel-Group and Crossover design).

Rand. Group = Randomized Group design (Parallel-Group and Crossover design).