Contact

Jill Constantine jconstantine@mathematica-mpr.com
John Deke jdeke@mathematica-mpr.com

Title

Understanding What Works Clearinghouse Evidence Standards

Preferred Conference Section

Research Methods

Symposium Justification

This symposium will provide education researchers with an opportunity to learn more about new What Works Clearinghouse (WWC) standards for regression discontinuity and single case designs (RDD and SCD). We will help researchers better understand the process for setting these standards and what the implications of these standards are for their work.

The development of standards for RDD and SCD is an exciting development as it allows for a wider range of studies to be reviewed by the WWC. The potential for studies based on these designs to meet WWC evidence standards without reservations may be of particular interest, since this designation was previously reserved only for randomized controlled trials.

In the symposium, the presenters will review the different components and criteria of each standard. They will follow with specific examples of particular aspects of study designs that meet standards, meet standards with reservations, and fail to meet standards. An important aspect of the symposium will be to allow time for audience discussion and participation around challenges in developing and executing RDD and SCD designs that meet WWC evidence standards.

The authors are ideally suited to explain the standards and answer questions – Jill Constantine leads the WWC and oversaw the development of these standards. John Deke was a member of the panel that advised the development of the RD standards and he drafted the standards document. Shannon Monahan played a similar role in the development of the SCD standards.

List of authors

Jill Constantine, jconstantine@mathematica-mpr.com
John Deke, jdeke@mathematica-mpr.com (contact for RDD paper)
Shannon Monahan, SMonahan@mathematica-mpr.com (contact for SCD paper) (all from Mathematica Policy Research)
**Discussant**

Petra Todd (availability not yet confirmed)
Title: Understanding the What Works Clearinghouse Standards for Regression Discontinuity Designs

Author(s): John Deke and Jill Constantine
Regression discontinuity designs (RDDs) are considered to be one of the strongest nonexperimental designs available (Shadish, Cook, & Campbell, 2002) for the purpose of identifying the effects of an intervention. RDD can be used in situations in which assignment to a treatment group is based on a cutoff value on a continuous assignment variable (for example, when all students who score below the 20th percentile on a reading test are offered after-school tutoring.) The impact of the intervention is calculated as the regression-adjusted difference in the predicted outcomes for treatment and control group members at the cutoff value of the assignment variable.

Since Goldberger (1972a and 1972b) showed the theoretical appeal of the approach, numerous researchers have contributed to our understanding of RDD (Cook [2007] reviews this literature). In the past decade, researchers have made a renewed effort to bolster the theoretical underpinnings of RDD and advance the state of the art in estimating impacts and standard errors (Hahn, Todd, & Van der Klaauw, 2001; Imbens & Kalyanaraman, 2009; Lee & Card, 2008).

Meanwhile, the creation of the Institute for Education Sciences (IES) by the Education Sciences Reform Act of 2002 highlights the need for more rigorous education research that can identify causal, not just correlational, relationships between educational practices and student outcomes. The What Works Clearinghouse (WWC) was created to help advance IES’ mission “to provide rigorous and relevant evidence on which to ground education practice and policy and share this information broadly” (http://www.ies.ed.gov/aboutus/). A key function of the WWC is to determine evidence that is “rigorous.” Toward that end, the WWC has posted a set of standards that it uses to assess the rigor of research evidence from studies based on either experimental or matching designs (http://www.ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf). Recently the WWC has added a new standard for RDD studies (http://www.ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf).

The purpose of this paper is to explain the RDD standards and some of their implications for researchers who are designing new RDD studies. Based on our involvement in the development of these standards, we hope to help researchers better understand the standards, what they can do to design and conduct studies that are more likely to meet these standards, and how the standards might evolve in the future.
Not applicable.

**Population / Participants / Subjects:**  
*Description of the participants in the study: who, how many, key features or characteristics.*  
(May not be applicable for Methods submissions)

Not applicable.

**Intervention / Program / Practice:**  
*Description of the intervention, program or practice, including details of administration and duration.*  
(May not be applicable for Methods submissions)

Not applicable.

**Significance / Novelty of study:**  
*Description of what is missing in previous work and the contribution the study makes.*

Unlike other quasi-experimental designs (such as propensity score matching), RDD studies have the potential to be placed in the same WWC evidence category as randomized experiments. However it is one thing for a study to have the potential to reach the highest evidence category and quite another thing for a study actually to achieve that potential. The WWC RDD standards set a high bar, one that many published RDD studies do not reach.

Given our close involvement in the development of the WWC RDD standards (Jill Constantine is the project director for the WWC and John Deke drafted the standards), we are in a unique position to help other researchers better understand what they need to do in order to produce studies that will meet these evidence standards.

**Statistical, Measurement, or Econometric Model:**  
*Description of the proposed new methods or novel applications of existing methods.*

These standards apply to studies that follow a regression discontinuity design (RDD).

**Usefulness / Applicability of Method:**  
*Demonstration of the usefulness of the proposed methods using hypothetical or real data.*

RDD is widely used to estimate impacts of interventions. It is regarded as one of the most rigorous nonexperimental methods available.

**Research Design:**  
*Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).*  
(May not be applicable for Methods submissions)

Not applicable.
Data Collection and Analysis:
Description of the methods for collecting and analyzing data.
(May not be applicable for Methods submissions)

Not applicable.

Findings / Results:
Description of the main findings with specific details.
(May not be applicable for Methods submissions)

Not applicable.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

The WWC RDD standards set a high, but attainable, bar. With a clear understanding of the methodologies and reporting requirements needed to meet the standards, many more researchers should be able to conduct RDD studies that meet WWC standards. Because RDD can be employed in many settings in which experiments are not feasible, the proliferation of RDD studies that meet WWC standards will raise the level of rigor in the education literature beyond what could have been obtained with experiments alone and will therefore facilitate better decision making on the part of education policymakers and practitioners.
Appendices
Not included in page count.

Appendix A. References


Appendix B. Tables and Figures
Not included in page count.
Title: What Works Clearinghouse (WWC) Standards for Evaluating Single Case Designs (SCDs)

Author(s): Shannon Monahan, Thomas Kratochwill, and Stephen Lipscomb
Abstract Body

Limit 5 pages single spaced.

Background / Context:

The What Works Clearinghouse (WWC) seeks to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence for what works in education. The WWC was established in 2002 by the U.S. Department of Education's Institute of Education Sciences (IES). It serves as a decision-making resource by helping the education community locate and recognize credible and reliable evidence. Reviewers working for the WWC assess the quality of studies that evaluate the effectiveness of specific educational products, interventions, practices or approaches. To ensure fairness and consistency in the review of these studies, the WWC uses rigorous and transparent standards for systematically reviewing and synthesizing existing research. From its inception, the WWC reviewed experimental and quasi-experimental group comparison studies. In an effort to expand the pool of rigorous scientific evidence available for review, in 2009, the WWC assembled a panel of national experts in single-case design (SCD) and analysis to develop WWC SCD Standards.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

This presentation delineates the criteria the WWC uses to identify SCDs and the standards that the WWC uses to evaluate the rigor of SCD studies. The WWC SCD standards are bifurcated into Design and Evidence standards (see Appendix B, Figure 1). The Design standards evaluate the internal validity of the design. Reviewers trained in visual analysis then apply the Evidence standards to evaluate the effect of interventions and practices on relevant outcomes.

Significance / Novelty of study:

SCDs are adaptations of interrupted time-series designs and can provide a rigorous experimental evaluation of intervention effects (Horner & Spaulding, in press; Kazdin, 1982, in press; Kratochwill, 1978; Kratochwill & Levin, 1992; Shadish, Cook, & Campbell, 2002). As experimental designs, a central goal of SCDs is to determine whether a causal relation (i.e., functional relation) exists between the introduction of a researcher-manipulated independent variable (i.e., an intervention) and change in a dependent (i.e., outcome) variable (Horner & Spaulding, in press; Levin, O'Donnell, & Kratochwill, 2003). SCDs can provide a strong basis for establishing causal inference, and these designs are widely used in applied and clinical disciplines in psychology and education, such as school psychology and the field of special education. The WWC has recognized that single case research designs provide an alternative design approach that is especially valuable in situations that are not amenable to study through randomized trials (e.g., small numbers of participants available) and also provide the scientific rigor needed to document experimental control. The development of WWC SCD standards not only expands the pool of studies available for WWC review, it also offers guidance to the research community on the highest quality attributes to consider when designing these studies.
Statistical, Measurement, or Econometric Model:

Although the basic SCD has many variations, these designs often involve repeated, systematic measurement of a dependent variable before, during, and after the active manipulation of an independent variable (e.g., applying an intervention). Experimental control involves replication of the intervention in the experiment and this replication is addressed with one of the following methods (Horner, et al., 2005):

- Introduction and withdrawal (i.e., reversal) of the independent variable (e.g., ABAB design)
- Iterative manipulation of the independent variable across different observational phases (e.g., alternating treatments design)
- Staggered introduction of the independent variable across different points in time (e.g., multiple baseline design)

The WWC SCD standards are intended to guide WWC reviewers in identifying and evaluating the rigor of SCDs as well as characterizing effects on outcomes. The first section of the WWC SCD standards assists with identifying whether a study is a SCD. As depicted in Appendix B Figure 1, a SCD should be reviewed using the ‘Criteria for Designs that Meet Evidence Standards’, to determine those that Meet Evidence Standards, those that Meet Evidence Standards with Reservations, and those that Do Not Meet Evidence Standards.

Studies that meet evidence standards (with or without reservations) should then be reviewed using the ‘Criteria for Demonstrating Evidence of a Relation between an Independent Variable and a Dependent Variable’ (see Appendix B, Figure 1). This process results in a categorization scheme that is similar to that used for evaluating evidence credibility by inferential statistical techniques (hypothesis testing, effect-size estimation, and confidence-interval construction) in traditional group designs. This review results in a sorting of SCD studies into three groups: those that have Strong Evidence of a Causal Relation, those that have Moderate Evidence of a Causal Relation, and those that have No Evidence of a Causal Relation.

CRITERIA FOR DESIGNS THAT MEET EVIDENCE STANDARDS

In order to Meet Evidence Standards, the following design criteria must be present:

- The independent variable (i.e., the intervention) must be systematically manipulated, with the researcher determining when and how the independent variable conditions change.
- Each outcome variable must be measured systematically over time by more than one assessor, and the study needs to collect inter-assessor agreement in each phase and on
at least twenty percent of the data points in each condition (e.g., baseline, intervention) and the inter-assessor agreement must meet minimal thresholds.

- The study must include at least three attempts to demonstrate an intervention effect at three different points in time or with three different phase repetitions.*

- For a phase to qualify as an attempt to demonstrate an effect, the phase must have a minimum of three data points and preference is given to 5 data points per phase.

**CRITERIA FOR DEMONSTRATING EVIDENCE OF A RELATION BETWEEN AN INDEPENDENT VARIABLE AND AN OUTCOME VARIABLE**

Single-case researchers traditionally have relied on visual analysis of the data to determine (a) whether evidence of a relation between an independent variable and an outcome variable exists; and (b) the strength or magnitude of that relation (Hersen & Barlow, 1976; Kazdin, 1982; Kennedy, 2005; Kratochwill, 1978; Kratochwill & Levin, 1992; McReynolds & Kearns, 1983; Richards, Taylor, Ramasamy, & Richards, 1999; Tawney & Gast, 1984; White & Haring, 1980). An inferred causal relation requires that changes in the outcome measure resulted from manipulation of the independent variable. A causal relation is demonstrated if the data across all phases of the study document at least three demonstrations of an effect at a minimum of three different points in time. An effect is documented when the data pattern in one phase (e.g., an intervention phase) differs more than would be expected from the data pattern observed or extrapolated from the previous phase (e.g., a baseline phase; Horner et al., 2005).

WWC rules for conducting visual analysis involve four steps (Kratochwill et al., 2010; Parsonson & Baer, 1978). The first step is documentation of a predictable baseline pattern of data. If a convincing baseline pattern is documented, then the second step consists of examining the data within each phase of the study to assess the within-phase pattern(s). The key question is to assess whether there are sufficient data with sufficient consistency to demonstrate a predictable pattern of responding. The third step in the visual analysis process is to compare the data from each phase with the data in the adjacent (or similar) phase to assess whether manipulation of the independent variable was associated with an “effect.” An effect is demonstrated if manipulation of the independent variable is associated with predicted change in the pattern of the dependent variable. The fourth step in visual analysis is to integrate all the information from all phases of the study to determine whether there are at least three demonstrations of an effect at different points in time (i.e., documentation of a causal or functional relation) (Horner et al., in press).

For studies that meet WWC evidence standards (with and without reservations), the following rules are used to determine whether the study provides evidence of a causal relation. In order to provide *Strong Evidence*, at least two WWC reviewers certified in visual (or graphical) analysis must verify that a causal relation was documented. Specifically this is operationalized as at least three demonstrations of the intervention effect along with no non-effects by:

---

* The three demonstrations criterion is based on professional convention (Horner, Swaminathan, Sugai, & Smolkowski, under review). More demonstrations further increase confidence in experimental control (Kratochwill & Levin, in press).
• Documenting the consistency of level, trend, and variability within each phase

• Documenting the immediacy of the effect, the proportion of overlap, the consistency of the data across phases in order to demonstrate an intervention effect, and comparing the observed and projected patterns of the outcome variable

• Examining external factors and anomalies (e.g., a sudden change of level within a phase)

Overall, the WWC SCD Standards provide guidance to reviewers for evaluating the internal validity of a study and for rating the strength of the evidence presented.

Usefulness / Applicability of Method:

The goal of a SCD is usually to answer “Is this intervention more effective than the current “baseline” or “business-as-usual” condition?” SCDs are implemented when pursuing the following research objectives (Horner et al., 2005):

• Determining whether a causal relation exists between the introduction of an independent variable and a change in the dependent variable. For example, a research question might be “Does Intervention B reduce a problem behavior for this case (or these cases)?”

• Evaluating the effect of altering a component of a multi-component independent variable on a dependent variable. For example, a research question might be “Does adding Intervention C to Intervention B further reduce a problem behavior for this case (or these cases)?”

• Evaluating the relative effects of two or more independent variables (e.g., alternating treatments) on a dependent variable. For example, a research question might be “Is Intervention B or Intervention C more effective in reducing a problem behavior for this case (or these cases)?”

SCDs are especially appropriate for pursuing research questions in applied and clinical fields. This application is largely because disorders with low prevalence may be difficult to study with traditional group designs that require a large number of participants for adequate statistical power (Odom, et al., 2005). Currently, as the WWC has expanded into special education topic areas (e.g., Children Classified with an Emotional or Behavioral Disorder), the number of SCDs identified in literature reviews has increased. In some instances, the evidence base for a single intervention may consist of over 50 SCDs. The inclusion of SCDs for review gives the WWC access to this wide body of relevant research.
Conclusions:

The development of WWC SCD Standards serves two primary purposes. First, it expands the pool of rigorous research that can be reviewed by the WWC and that will ultimately be evaluated to inform educational decision making. Second, it offers guidelines to the research community at large of best practice in single case design.

A noted limitation of the current WWC SCD standards is the singular focus on visual analysis and the inability to calculate effect-size (ES). Most researchers using SCDs base their inferences on visual analysis, and there are no agreed-upon methods or standards for effect size estimation though several parametric and non-parametric quantitative methods have been proposed. Each quantitative method has flaws, and most are not comparable with those used in group-comparison studies. Shadish et al. (2008) have developed an estimator for continuous outcomes that is promising, though the distribution theory is still being derived and tested. As the field reaches greater consensus about appropriate statistical analyses and quantitative effect-size measures, new WWC standards for effect demonstration will need to be developed.
Appendices
Not included in page count.

Appendix A. References


Appendix B. Tables and Figures

FIGURE 1

PROCEDURE FOR APPLYING SCD STANDARDS: FIRST EVALUATE DESIGN, THEN IF APPLICABLE, EVALUATE EVIDENCE

Evaluate the Design

Meets Evidence Standards

Meets Evidence Standards with Reservations

Does Not Meet Evidence Standards

Conduct Visual Analysis for Each Outcome Variable

Strong Evidence

Moderate Evidence

No Evidence

Effect-Size Estimation