

What Works Clearinghouse™ **Standards for Alternative Designs**

Jonathan Jacobson, Ph.D.

jonathan.jacobson@ed.gov

Senior Research Scientist

National Center for Education Evaluation and Regional Assistance (NCEE)

Institute of Education Sciences (IES)

U.S. Department of Education (ED)

September 23, 2016

Overview

- **About the *What Works Clearinghouse*TM**
- **WWC Design Standards and Study Ratings**
- **WWC Standards for Group Designs Without Random Assignment**
- **WWC Standards for Regression Discontinuity Designs**
- **WWC Resources**

About the *What Works Clearinghouse*TM

- Begun in 2002 as an initiative of IES
- Mission is to be a *trusted source of scientific evidence for what works to improve student outcomes or other relevant education outcomes*
- Reviews *original* research on education *interventions*
 - “Intervention” means a program, product, practice, or policy
 - WWC does not review descriptive studies, qualitative studies, secondary data analyses, or research syntheses
- Current work includes 5 contracts, funded at about \$10 million/year total

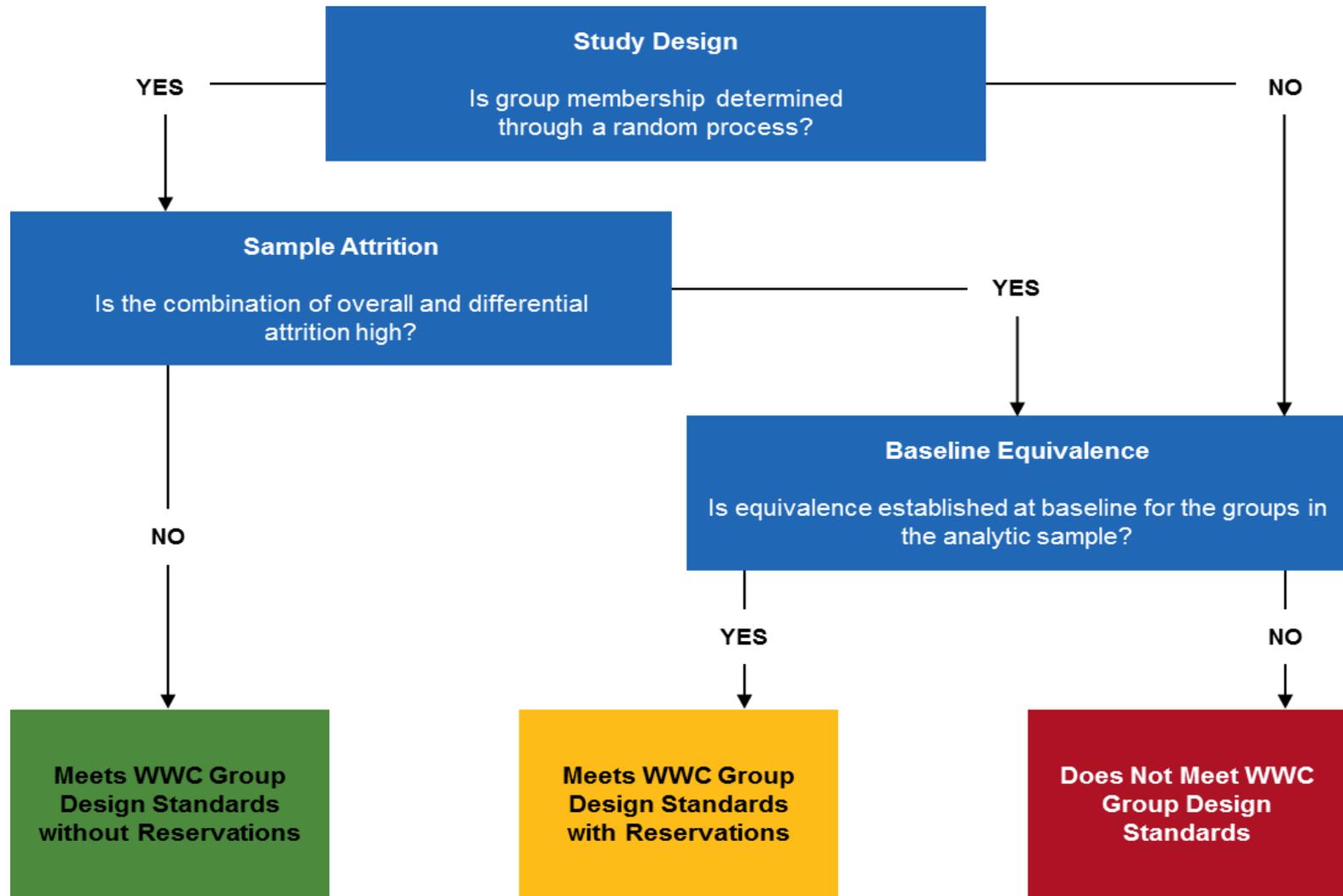
WWC Design Standards

- Developed by panels of national experts for different types of designs (*group designs, regression discontinuity designs, and single case designs*)
- Focus of standards is on the internal validity of study findings
- Applied to each study by a team of certified reviewers
- Used with a review protocol defining eligible *studies*, eligible *outcomes*, and required measures of *baseline equivalence*
- Result in a rating for each eligible study, e.g.
 - *Meets WWC Group Design Standards without Reservations*
 - *Meets WWC Group Design Standards with Reservations*
 - *Does Not Meet WWC Group Design Standards*

Notes About WWC Study Ratings

1. Ratings are of study *findings*, not interventions
2. The study as a whole receives the rating of the *highest-rated finding* reviewed by the WWC
3. Study ratings can change when *WWC standards* change
4. The *sign, size, and statistical significance* of the estimated effect are reported by the WWC but do not affect the WWC study rating
5. Current WWC procedures synthesize findings only from *group design studies* meeting WWC standards

WWC Review of a Group Design Study



Assessing Baseline Equivalence for Group Design Studies



- Calculate the difference *in the analytic sample* between the treatment and comparison groups for each baseline characteristic specified in the review protocol
- T-C differences between 0.05 and 0.25 standard deviations require statistical adjustment (such as regression or ANCOVA) when calculating impacts
- If there is a difference greater than 0.25 standard deviations for any required characteristic, then the finding does not meet WWC standards

Regression Discontinuity Designs (RDDs)

- Similar to randomized controlled trials (RCTs) in that treatment and control groups formed by design, not by unobserved self-selection
- Different from RCT in that groups are *not* formed randomly – they are formed purposefully on the basis of a continuous measure of need or merit known as a *forcing variable (or assignment variable)*
- “*Fuzzy*” RDDs (FRDDs) are RDDs in which the forcing variable is an imperfect predictor of receiving the intervention, and impacts are estimated for subjects that comply with the assigned condition

RDD Example: Intervention Has Positive Impact



Experts Advising the Development of RDD Standards

- **Panel for Development of Pilot RDD Standards (September 2011)**
 - *Thomas Cook, Northwestern University*
 - *Guido Imbens, Stanford University (formerly Harvard)*
 - *J.R. Lockwood, ETS (formerly RAND)*
 - *Jack Porter, University of Wisconsin*
 - *Jeff Smith, University of Michigan*
 - *Peter Schochet and John Deke, Mathematica Policy Research*

- **Panel for Development of Revised RDD Standards (December 2015)**
 - *Thomas Cook, Northwestern University*
 - *Sean Reardon, Stanford University*
 - *Rocio Titiunik, University of Michigan*
 - *Petra Todd, University of Pennsylvania*
 - *Wilbert Van Der Klaauw, Federal Reserve Bank of New York*
 - *Glen Waddell, University of Oregon*
 - *John Deke and Lisa Dragoset, Mathematica Policy Research*

Screening for *Eligibility* Under the Revised RDD Standards

- Treatment assignment is based on the forcing variable – units on one side of a cutoff value are in the treatment group, units on the other side are in the comparison group
- The forcing variable must be ordinal with at least 4 values above and 4 values below the cutoff value
- There must be no factor confounded with the cutoff value of the forcing variable
- The forcing variable used to calculate impacts is the same as the variable used to assign units to treatment

WWC Standards 1 and 2 Applying to RDD Studies

- **Standard 1: Integrity of the Forcing (Assignment) Variable.** Primary concern is “manipulation” of the forcing variable. *To completely satisfy the standard,* institutional, graphical, and statistical criteria must be met.
- **Standard 2: Attrition.** Primary concern is nonresponse bias. Assessed using the same attrition boundary as used by the WWC for RCTs. *To completely satisfy the standard,* overall and differential must be low and assessed at the cutoff, or within the bandwidth used for the impact analysis or selected specially for attrition.

Standard 3: *Continuity of the outcome – forcing variable relationship*

- Primary concern is phantom impacts from lack of “smoothness” in the relationship between the outcome and the forcing variable
- Three criteria:
 - A. baseline equivalence on key covariates (measured at the cut point) as assessed for group designs*
 - B. no graphical evidence of unexplained discontinuities away from the cutoff*
 - C. no statistical evidence of unexplained discontinuities away from the cutoff*
- *To completely satisfy this standard, all three criteria must be met*
- *To partially satisfy this standard, two criteria must be met (including A and either B or C)*

Standard 4: *Functional form and bandwidth*

- Primary concern is misspecification bias
- Six criteria:
 - A. An adjustment must be made for the forcing variable
 - B. Impacts would ideally be estimated within a *justified bandwidth*
 - C. (Alternatively, impacts could be estimated using all data and a “best-fit” functional form)
 - D. Evidence must be provided that impacts are robust to bandwidth/functional form (five different types of acceptable evidence)
 - E. A graphical analysis showing a scatter plot and fitted curve must be included – the plot and curve cannot be obviously inconsistent with bandwidth and functional form choices
 - F. The relationship between the forcing variable and outcome must not be constrained to be the same on both sides of the cutoff
- *To completely satisfy this standard, all criteria except C must be met*
- *To partially satisfy this standard, criteria A, B or C, and E must be met*

“Fuzzy” RDD Standards, and Study Ratings

- FRDDs must meet the same criteria required by the WWC for RCTs that estimate a complier average causal effect (CACE). In addition, *to completely satisfy the Fuzzy RDD standard*, the bandwidth must be justified for the FRDD impact or separate bandwidths must be estimated for the numerator and denominator used to calculate the FRDD impact estimate, or the smallest of the numerator and denominator bandwidths must be used
- Eligible RDDs and FRDDs can obtain one of three possible ratings:
 - *Meets WWC RDD Standards without Reservations*
 - *Meets WWC RDD Standards with Reservations*
 - *Does Not Meet WWC RDD Standards*

WWC Resources

- New website: <http://ies.ed.gov/ncee/wwc>
- WWC [Procedures and Standards Handbook](#)
- 📧 The What Works Clearinghouse [Help Desk](#)
- Follow the WWC
 -  [Facebook](#)
 -  [Twitter](#)
 -  [NewsFlash](#)