Are legislators more responsive to high quality evidence? A field experiment

Angèle Delevoye and Trevor Incerti

29 May 2019

Introduction

Research Questions

- Do policymakers give more credence to high quality research?

Research Questions

- Do policymakers give more credence to high quality research?

- Can policymakers recognize differences in research quality?

Theory

Pre-existing literature

- Stuff here

Evidence standards

- Evidence standards and descriptions already adopted in federal legislation.
    - Secondary Education Act, No Child Left Behind, Every Student Succeeds Act

Evidence standards

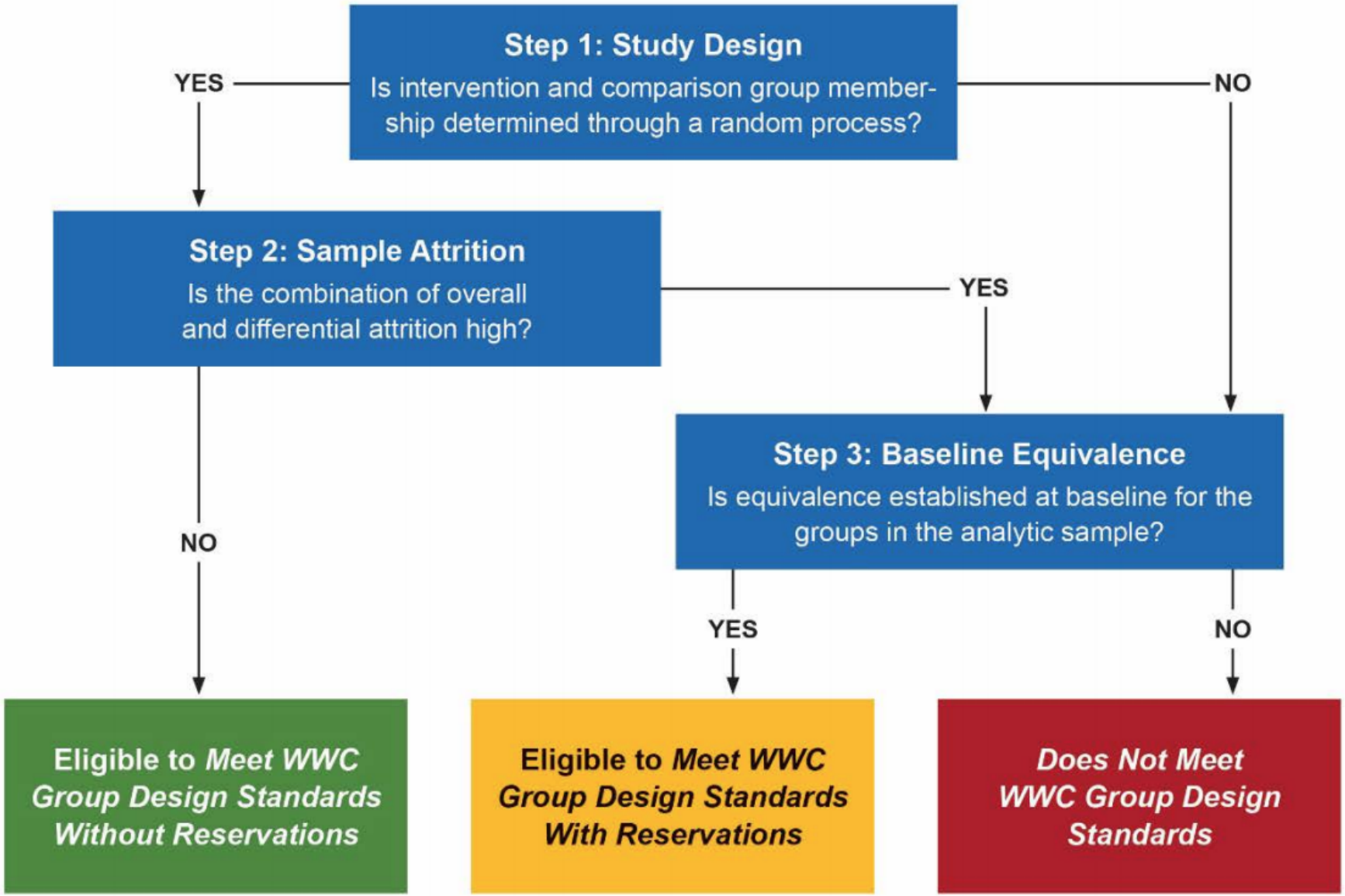
- Evidence standards and descriptions already adopted in federal legislation.
  - Secondary Education Act, No Child Left Behind, Every Student Succeeds Act
- Department of Education standards.
  - Strong causal evidence
  - Moderate causal evidence
  - Low causal evidence
  - High levels of specificity covering cluster-random assignment Figure, IVs Figure, and missingness/attrition Figure, and RDs Figure.

Evidence standards

- Evidence standards and descriptions already adopted in federal legislation.
  - Secondary Education Act, No Child Left Behind, Every Student Succeeds Act
- Department of Education standards.
  - Strong causal evidence
  - Moderate causal evidence
  - Low causal evidence
  - High levels of specificity covering cluster-random assignment Figure, IVs Figure, and missingness/attrition Figure, and RDs Figure.
- Department of Labor has adopted similar standards.

DoE evidence standards

Figure II.1. Study Ratings for Individual-Level RCTs and QEDs



Design

Overview of experimental design

- 2x2 factorial design with two treatments:
  - Evidence standard (low vs. high)
  - Whether evidence standards are explained to policymakers

Table 1: Treatment arms: 2x2 factorial design

	Lower Tier	**Higher Tier**
No information	Control	High and no info
Information	Low and info	High and info

Outcomes

- Stuff here

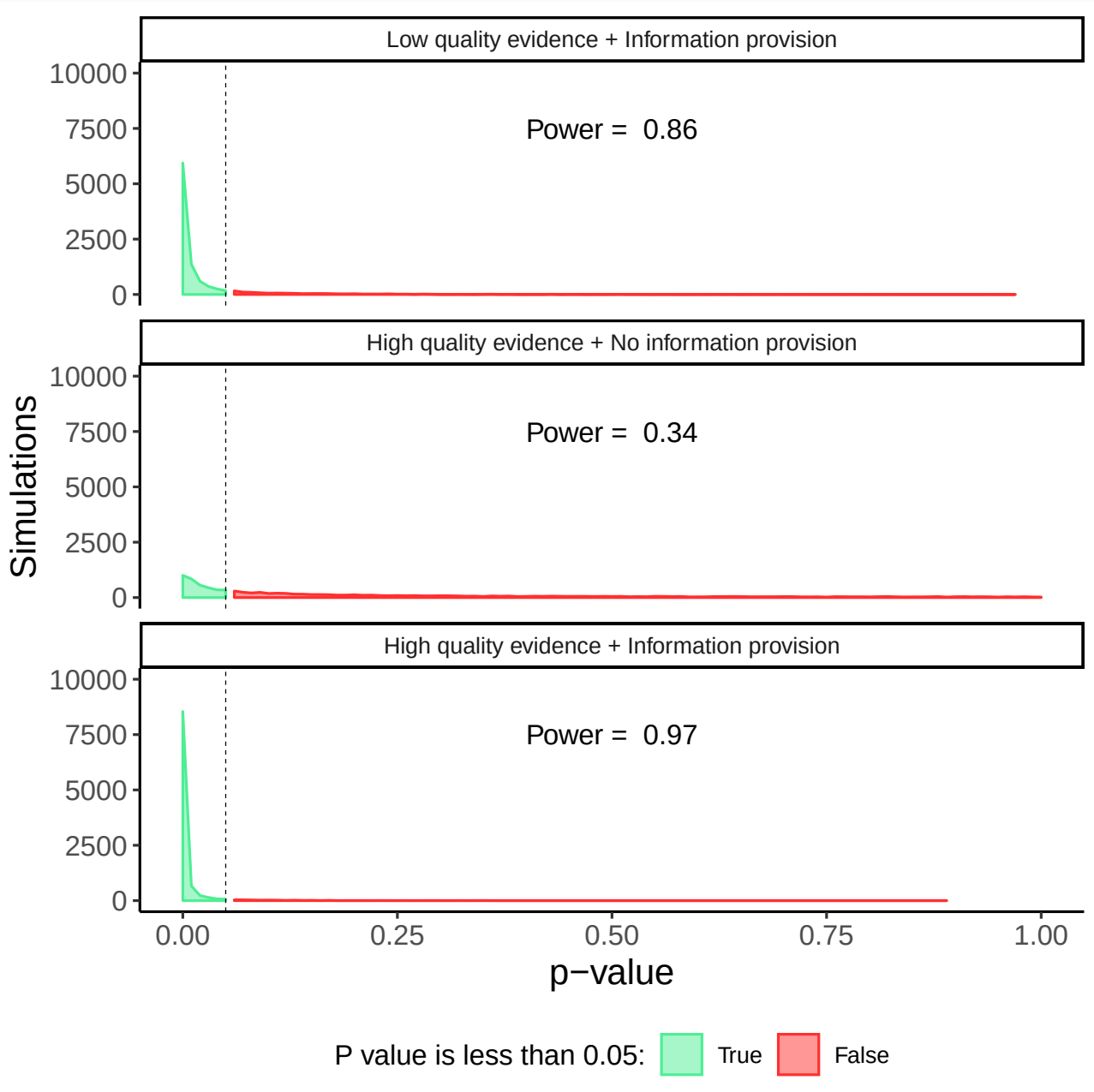
Treatment effect estimation

- Block random assignment.
- $ATE = \sum_{j=1}^{J} \frac{N_j}{N} ATE_j$
  - Where J is the number of blocks, blocks are indexed by j, and $\frac{N_j}{N}$ represents the share of subjects who belong to block j.
- P-values calculated using randomization inference.
- Control group = Low quality evidence + no information

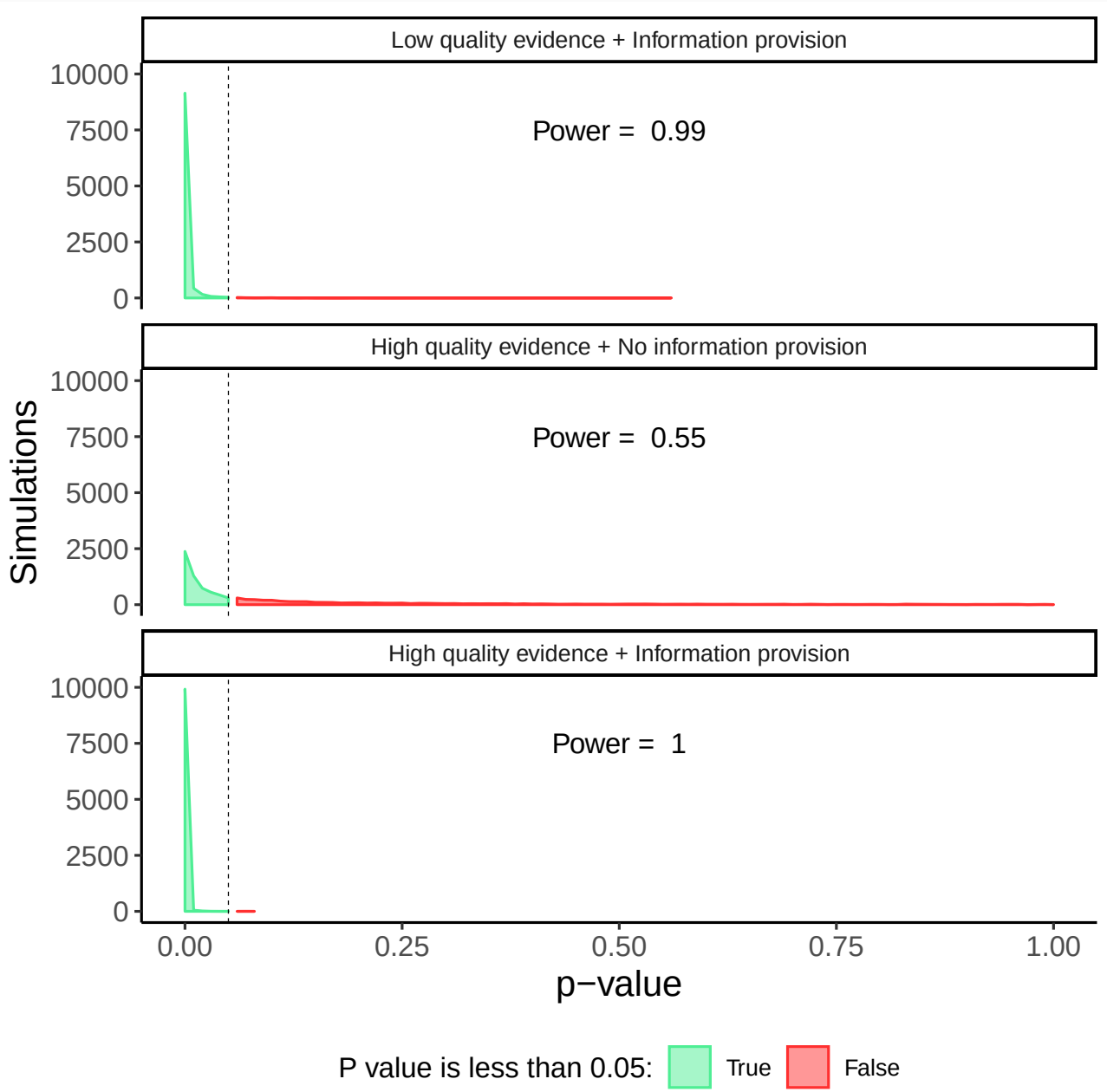
Heterogenous treatment effects (CATEs)

- Party
- ... ?

Power analysis: federal



Power analysis: state



Power analysis assumptions

- N = 535 (federal) and 1000 (state)
- Low quality evidence + information provision = -10%
- High quality evidence + no information provision = 5%
- High quality evidence + information provision = 12.5%
- Standard deviation = 0.08

Conclusion

Questions

- Use a neutral or partisan policy proposal?
  - Partisan policy proposal might allow us to test legislator's motivated reasoning, but power issues.
- Better outcome measurements?
- Suggestions for kind of organization to partner with? Is organizational partnering feasible?
- Federal, state, or local level?
- Other suggestions?

Supplemental material

Evidence tiers

ESSA's definition of "evidence-based" includes 4 levels of evidence. The top 3 levels require findings of a **statistically significant effect** on improving student outcomes or other relevant outcomes based on:

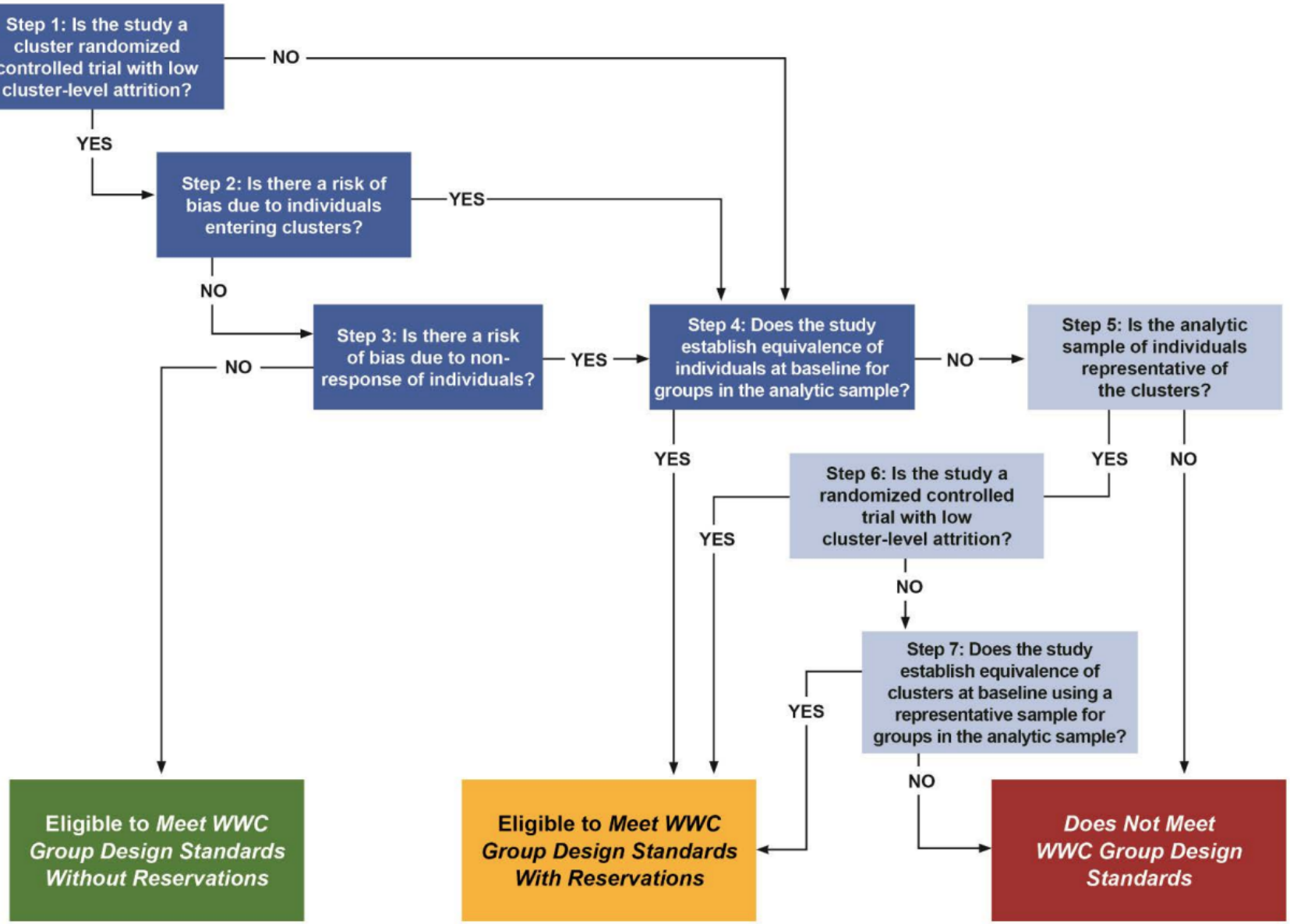
(1) Strong	• At least 1 well-designed and well-implemented **experimental** study (i.e., randomized)	**Required for school improvement plans funded by 7% set aside (Section 1003) & Eligible for a priority under 7 competitive grants**
(2) Moderate	• At least 1 well-designed and well-implemented **quasi-experimental** study (i.e., matched)	
(3) Promising	• At least 1 well-designed and well-implemented **correlational** study with statistical controls for selection bias	

The 4th level is designed for ideas that do not yet have an evidence base qualifying for the top 3 levels above. Given the requirement in the second bullet below to examine the effects of these ideas, this evidence-*building* level can be referred to as "under evaluation."

(4) "Under Evaluation"	• **Demonstrates rationale** based on high-quality research or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes • Includes **ongoing efforts to examine the effects** of such activity, strategy, or intervention	**Included for all other uses of "evidence-based"**

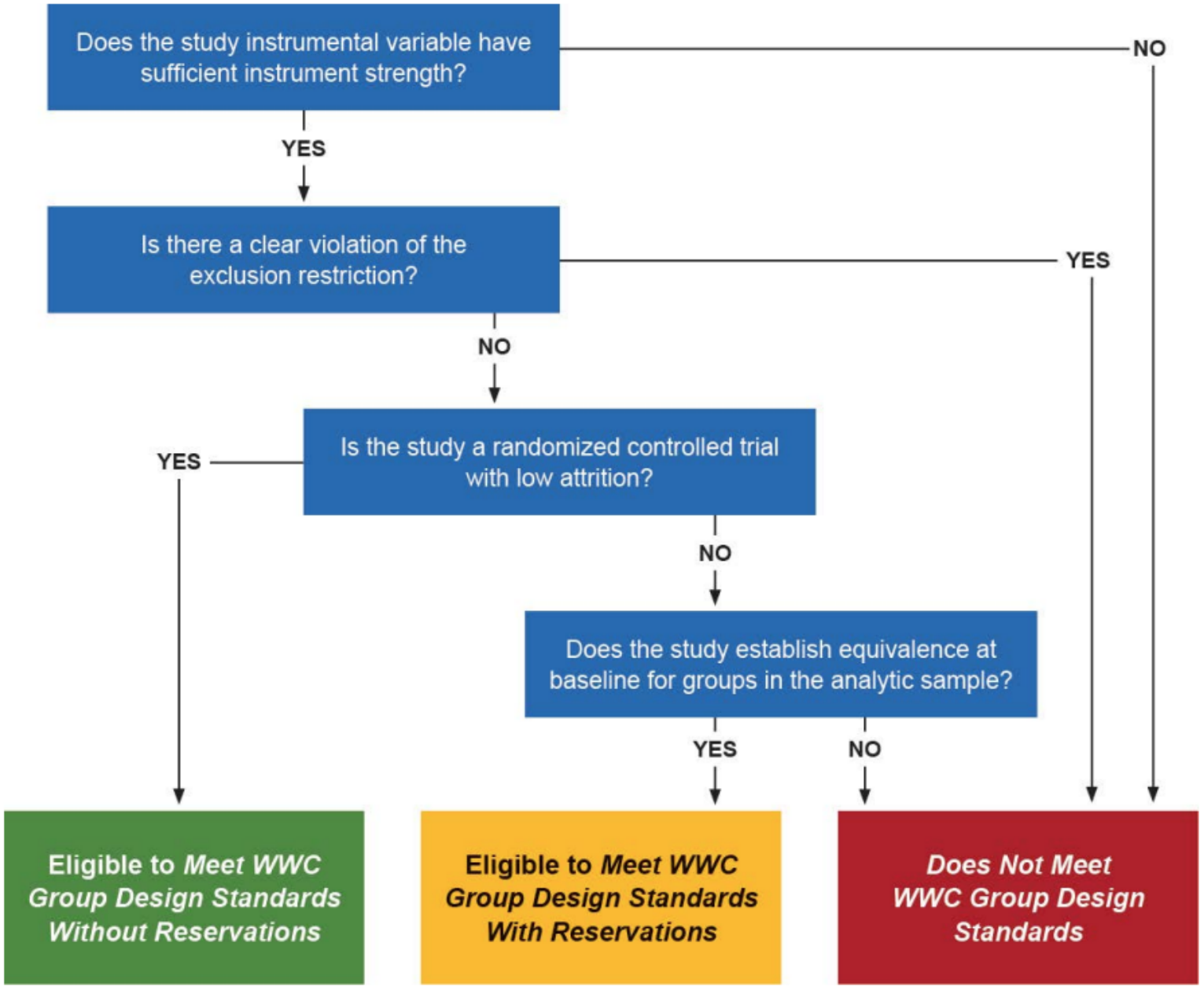
Evidence tiers: cluster random assignment

Figure II.4. Review Process for Cluster-Level Assignment Studies



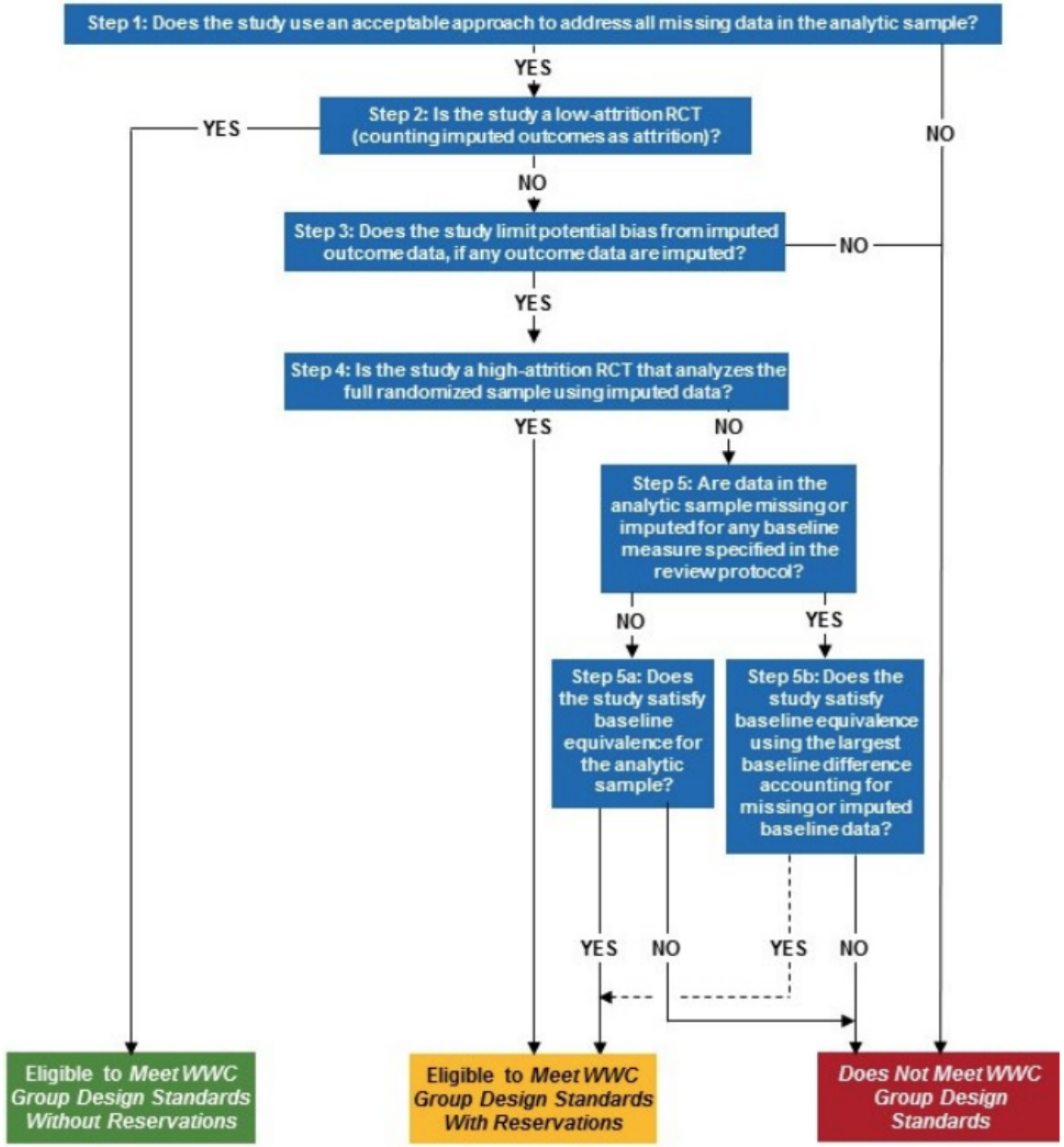
Evidence tiers: instrumental variables

Figure II.6. Review Process for Studies that Report a CACE Estimate



Evidence tiers: missingness and attrition

Figure II.5. Study Ratings for RCTs and QEDs with Missing Outcome or Baseline Data



Evidence tiers: regression discontinuity

Table III.1. RDD Study Ratings

Standard	To be rated *Meets WWC RDD Standards* ***Without*** *Reservations*, studies must:	To be rated *Meets WWC RDD Standards* ***With*** *Reservations*, studies must:
1: Integrity of the forcing variable	Completely satisfy	Partially satisfy
2: Sample attrition	Completely satisfy	Partially satisfy at least one of these two standards
3: Continuity	Completely satisfy	
4. Bandwidth/Functional form	Completely satisfy	Partially satisfy
5. Fuzzy RDD	Completely satisfy	Partially satisfy