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

Angèle Delevoye and Sōm Duchébaggè

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

- Literature on evidence use in policy-making, on relationship between science, researchers and policy-makers in a democracy
- Existing field/audit experiments reaching out to policy-makers

Figure

Evidence standards

- Evidence standards and descriptions already adopted in federal legislation.
    - Secondary Education Act 65, No Child Left Behind 01, Every Student Succeeds Act 2015

Evidence standards

- Evidence standards and descriptions already adopted in federal legislation.
  - Secondary Education Act 65, No Child Left Behind 01, Every Student Succeeds Act 2015

- Department of Education (DoE) standards under ESSA 2015:
  - 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 65, No Child Left Behind 01, Every Student Succeeds Act 2015

- Department of Education (DoE) standards under ESSA 2015:
  - 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.
- Other federal agencies have adopted similar standards Figure

DoE evidence standards under ASSA 2015

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"**

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

Treatment 1, Choice of policy - descriptive statistics of DoE's database

Treatment 2, Information on evidence standards

From DoL's CLEAR database:

"High Causal Evidence standards mean there is strong evidence that the effects estimated in this study are solely attributable to the intervention being examined. This does not necessarily mean that the study found positive impacts, only that the analysis meets high methodological standards and the causal impacts estimated, whether positive, negative, or null, are credible. Currently, only well-implemented randomized controlled trials can receive this rating"

"Low Causal Evidence standards mean there is little evidence that the effects estimated in the study are attributable to the intervention being examined, and other factors are likely to have contributed to the results. This does not imply that the study's results are not useful for some purposes, but they should be interpreted with caution. Causal studies that do not meet criteria for a high or moderate evidence rating receive this rating."

Outcomes

- Ideally: partner with a 3rd party organization and examine:
  1. Whether or not a meeting was established.
  2. Seniority of the individual with whom a successful meeting was granted (as in **?**).

Outcomes

- Ideally: partner with a 3rd party organization and examine:
  1. Whether or not a meeting was established.
  2. Seniority of the individual with whom a successful meeting was granted (as in **?**).
  - May also allow us to engage in participant observation (qualitative data)

Outcomes

- Ideally: partner with a 3rd party organization and examine:
  1. Whether or not a meeting was established.
  2. Seniority of the individual with whom a successful meeting was granted (as in **?**).
  - May also allow us to engage in participant observation (qualitative data)

- Alternatively: email response rates

Email example

Add screenshot of potential email

Treatment effect estimation

Primary effects (ATE)

- 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, ..., ?
- Note preregistration, multiple comparisons, and power.

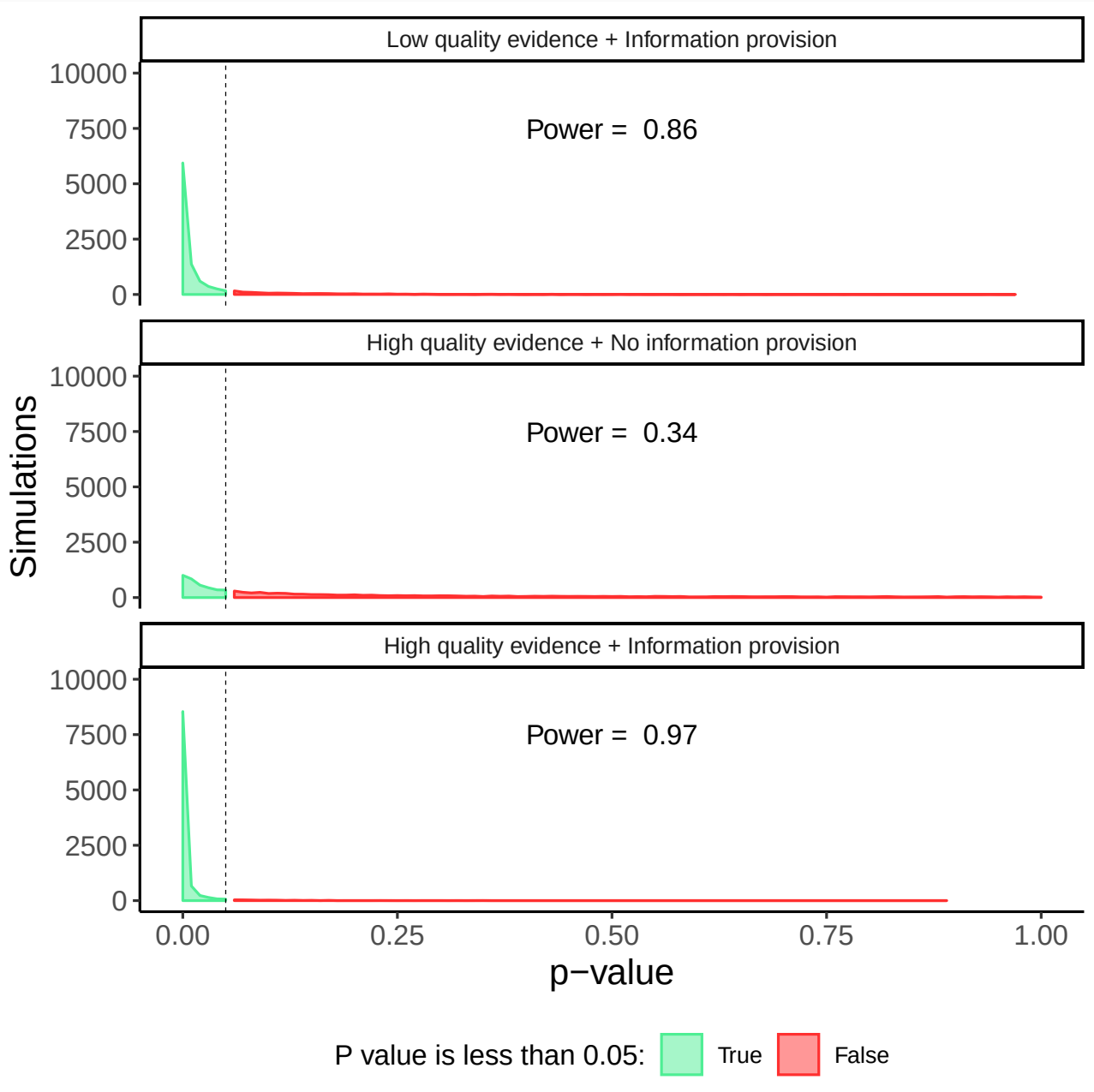
Power analysis assumptions

- N = 535 (federal) and 1000 (state)

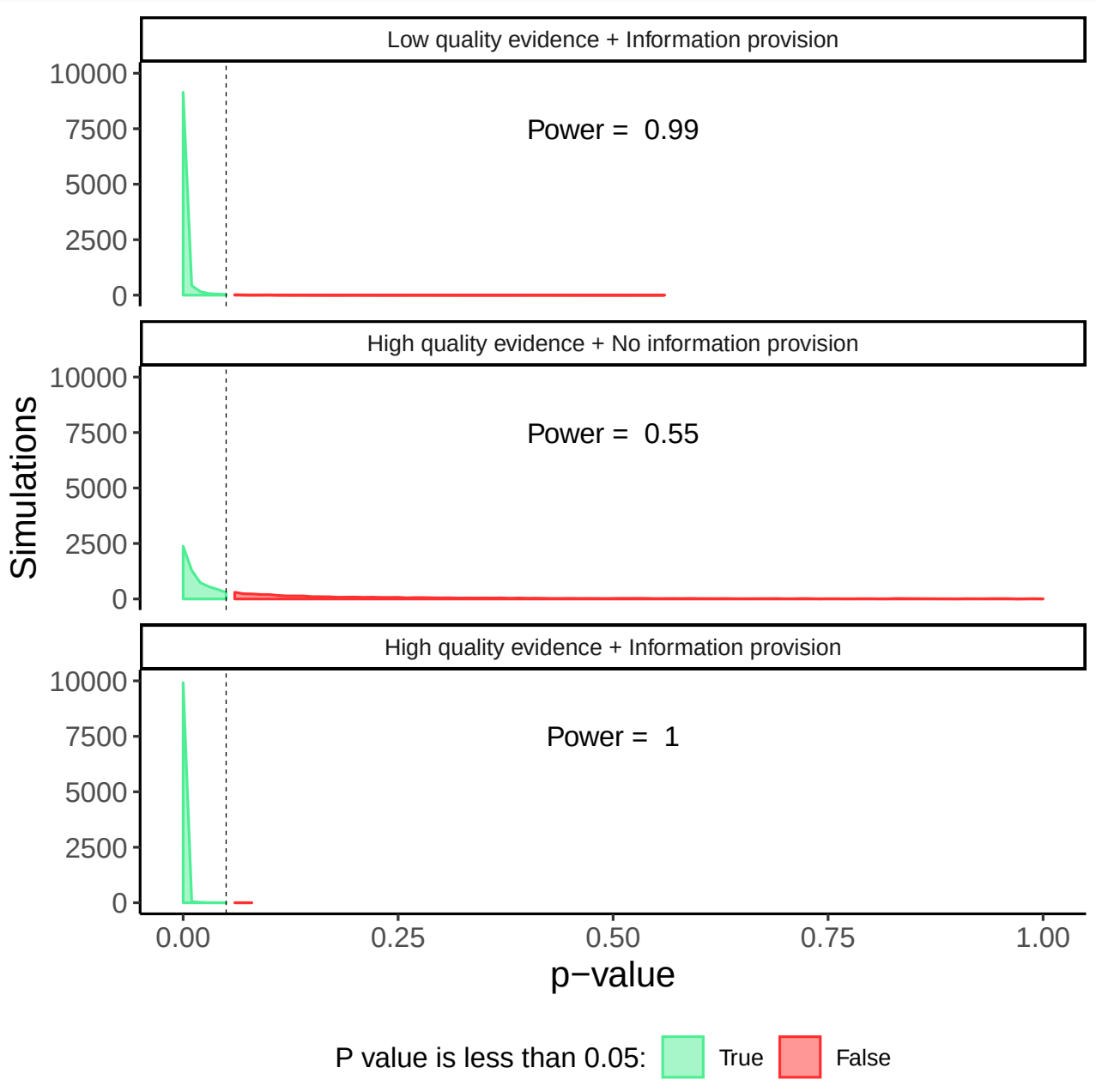
- Low quality evidence + information provision = -10%
- High quality evidence + no information provision = +5%
- High quality evidence + information provision = +12.5%

- Standard deviation = 0.08

Power analysis: federal



Power analysis: state



Conclusion

Timeline and questions

- Ideal timeline: pre-registration and initial contact with 3rd party organization by end of 2019, roll-out of the experiment in the first half of 2020 (political context)
- 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

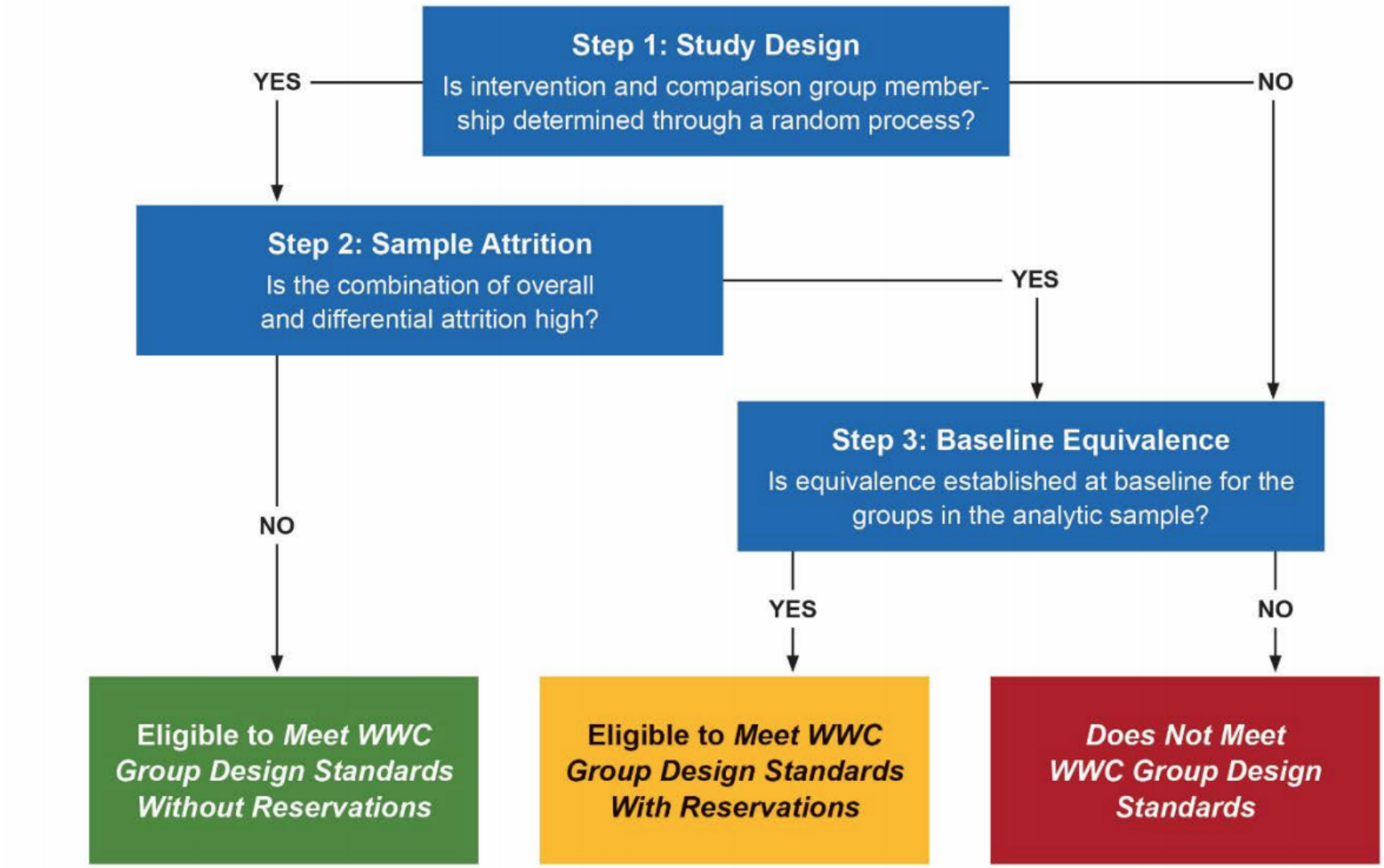
Existing field/audit experimens

Table 2: Audit experiments conducted with U.S. policy-makers

Reference	Federal/State	Arms	Treatment	Design	Outcor
?	State (New Hampshire)	1	Contacted by activists	Matched pairs (multimembers districts) Randomization within party and district stratas	Roll Ca
?	4,859 state legislators (44 states)	2x3	Black or white name and party (D/R/blank) of email sender	Block randomization by state, chamber, party, and whether legislator is up for reelection	Rate o to ema
?	US Congress 191 offices that had not yet sponsored bill	1	Reveal in email that prospective attendees had contributed to campaigns	Blocks of 3 offices: closest similarity on multiple covariates 1 treated, 2 control in each of the 64 blocks	Rate o to ema seniorit meetin
?	1,108 Canadian bureaucrats	2x2	Source of the policy information (academic, think tanks, research-based advocacy groups)	Sources in treatment groups were falsified Pre treatment survey for covariates	Credibi each o Based And ra
?	18 bills	1	Assigned to in-person briefings by a	Treatment assigned at legislator-bill	Cospon

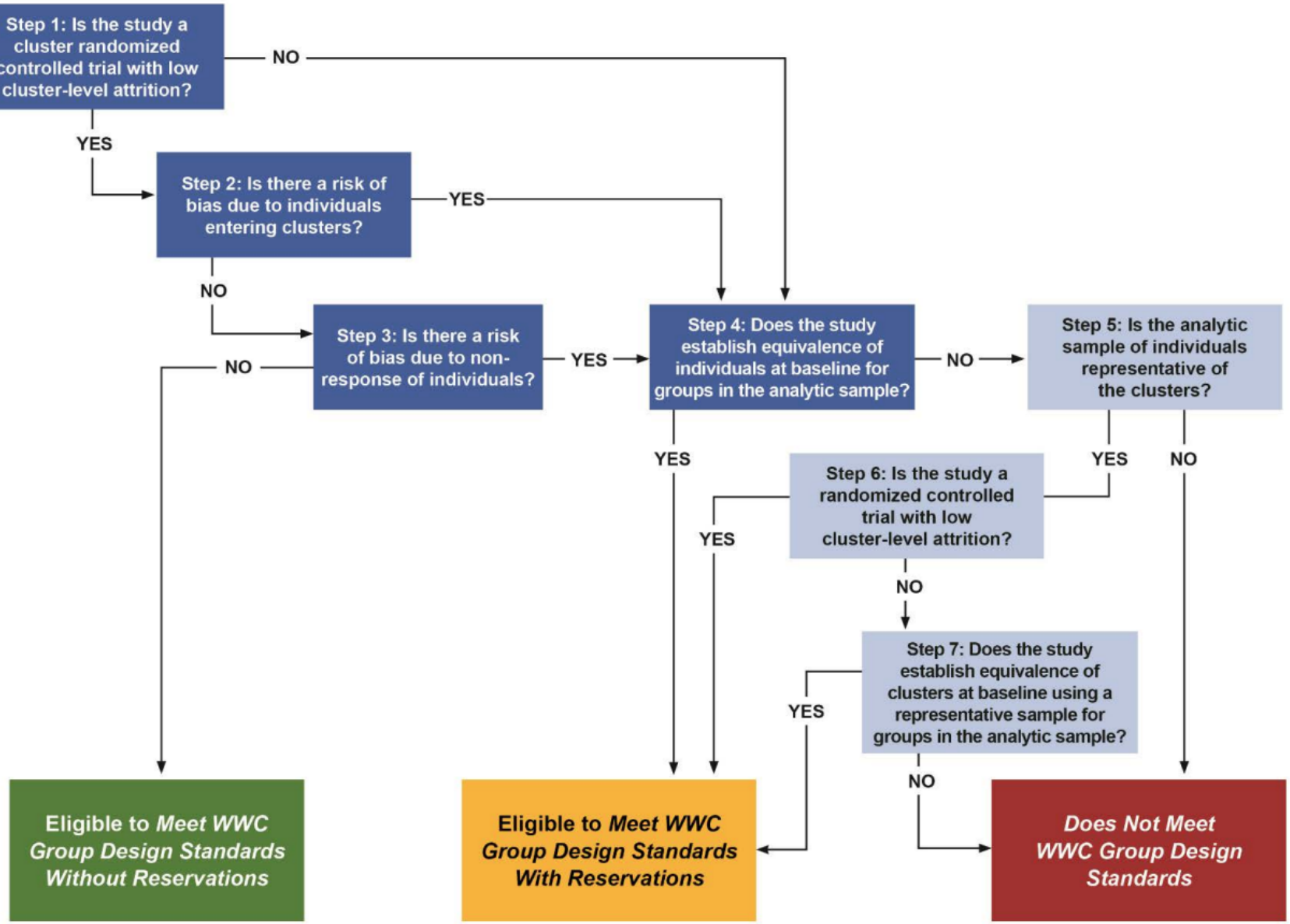
Evidence tiers

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



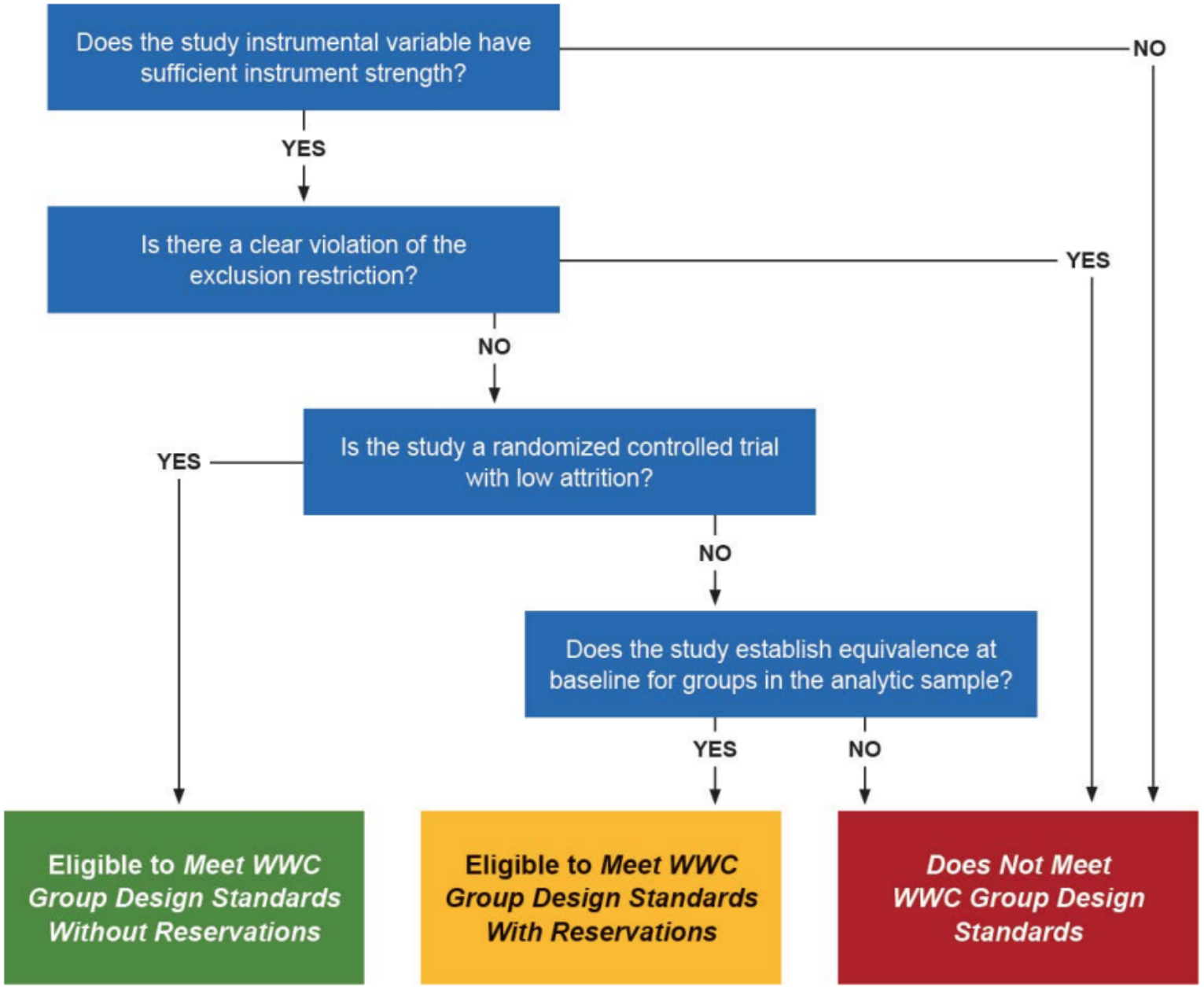
Evidence tiers: cluster random assignment

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



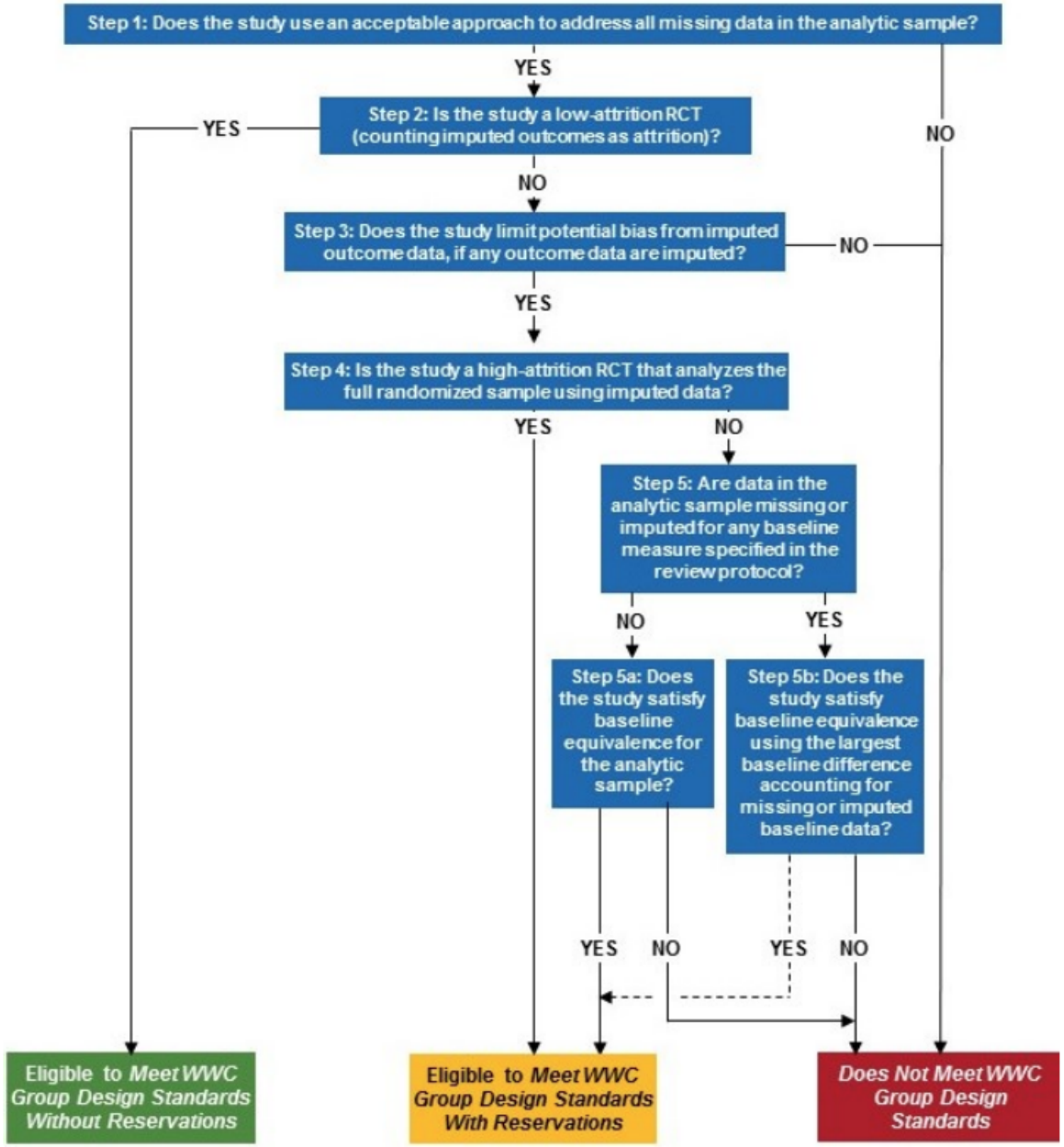
Evidence tiers: instrumental variables

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

Other federal evidence standards and databases

- Department of Labor (DoL)'s CLEAR's clearinghouse: evidence on on labor topics
- Corporation for National and Community Service (CNCS): evidence on what works in national service, social innovation, civic engagement, and volunteering
- U.S. Agency for International Development (USAID), YouthPower: evidence on what works in youth and peacebuilding, youth and health, youth and agriculture, food security, and nutrition
- US Departments of Agriculture and Defense's ClearingHouse for military family readiness: evidence on wide-ranging family and mental health issues.
- US Department of Health and Human services: multiple databases on programs whose purpose is to prevent and/or reduce delinquency or other problem behaviors in young people, teen pregnancy and substance prevention programs, etc.
- US Department of Justice: multiple databases on drugs and substance abuse, juveniles, crime and crime prevention, victims and victimization, law enforcement, technology and forensics, corrections and reentry, and courts