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

Angèle Delevoye and Trevor Incerti

29 May 2019

Introduction

Evidence-based policymaking: a bipartisan goal

H. R. 4174

One Hundred Fifteenth Congress
of the
United States of America

AT THE SECOND SESSION

Begun and held at the City of Washington on Wednesday, the third day of January, two thousand and eighteen

An Act

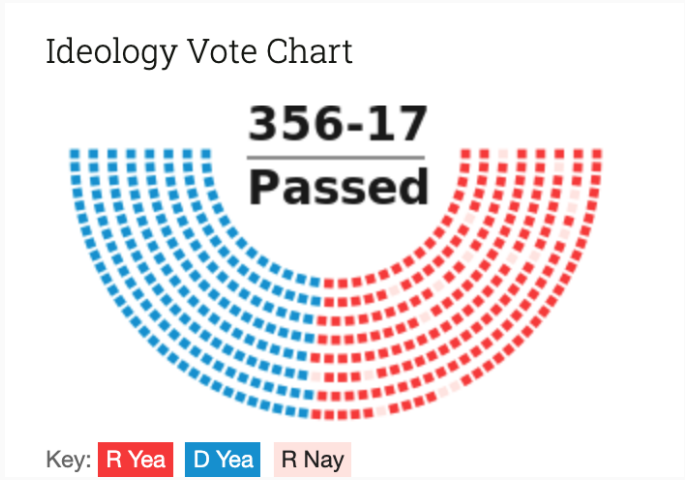
To amend titles 5 and 44, United States Code, to require Federal evaluation activities, improve Federal data management, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the "Foundations for Evidence-Based Policymaking Act of 2018".

(b) TABLE OF CONTENTS.—The table of contents for this Act is as follows:



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, and on the relationship between science, researchers and policy-makers.

- Existing field/audit experiments reaching out to policy-makers

Figure

Current use (?) of evidence standards in policymaking

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

Current use (?) of evidence standards in policymaking

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

- Department of Education (DoE) standards tiers under ESSA 2015:
  - Strong causal evidence
  - Moderate causal evidence
  - Low causal evidence
  - High levels of specificity covering cluster-random assignment Figure, IVs Figure, missingness/attrition Figure, and RDs Figure.

Current use (?) of evidence standards in policymaking

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

- Department of Education (DoE) standards tiers under ESSA 2015:
  - Strong causal evidence
  - Moderate causal evidence
  - Low causal evidence
  - High levels of specificity covering cluster-random assignment Figure, IVs Figure, missingness/attrition Figure, and RDs Figure.
- Other federal agencies have adopted similar standards Figure

Current use (?) of evidence standards in policymaking

From DOL 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."

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

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

Overview of DoE studies

158 interventions examining 49 outcomes

12 interventions with significant results and same outcome, but analyzed by studies with two different research designs.

	Intervention	Outcome
1	ACT/SAT Test Preparation and Coaching Programs	General academic achievement (high school)
2	Dual Enrollment Programs	Access and enrollment
3	Dual Enrollment Programs	Attainment
4	Knowledge is Power Program (KIPP)	English language arts achievement
5	Knowledge is Power Program (KIPP)	General Mathematics Achievement
6	Pre-K Mathematics	General Mathematics Achievement
7	READ 180®	Comprehension
8	READ 180®	Literacy achievement
9	Success for All®	Alphabetics
10	Success for All®	Comprehension
11	Teach for America (TFA)	English language arts achievement
12	Teach for America (TFA)	General Mathematics Achievement

Example: same intervention and different evidence tiers

Success for All: Longitudinal Effects of a Restructuring Program for Inner-City Elementary Schools

Nancy A. Madden, Robert E. Slavin, Nancy L. Karweit, Lawrence J. Dolan, and Barbara A. Wasik
The Johns Hopkins University

This article presents the effects of variations of a schoolwide restructuring program, Success for All, on student reading achievement and other outcomes in elementary schools serving large numbers of disadvantaged students. Success for All uses research-based preschool and kindergarten programs, beginning and intermediate reading programs in Grades 1–3, one-to-one tutoring for low-achieving students, family support programs, and other elements. A total of five Baltimore schools were studied over a period of 3 years (four schools) or 4 years (one school). Comparisons with matched students in matched schools indicated strong positive effects on most individually administered reading measures in most schools for students who have been in the program since first grade. Retentions in grade were also substantially reduced, and attendance increased over time.

Final Reading Outcomes of the National Randomized Field Trial of Success for All

Geoffrey D. Borman
University of Wisconsin–Madison
Robert E. Slavin
Johns Hopkins University and University of York
Alan C. K. Cheung
Hong Kong Institute of Education
Anne M. Chamberlain
Success for All Foundation
Nancy A. Madden
Bette Chambers
Johns Hopkins University and University of York

Using a cluster randomization design, schools were randomly assigned to implement Success for All, a comprehensive reading reform model, or control methods. This article reports final literacy outcomes for a 3-year longitudinal sample of children who participated in the treatment or control condition from kindergarten through second grade and a combined longitudinal and in-mover student sample, both of which were nested within 35 schools. Hierarchical linear model analyses of all three outcomes for both samples revealed statistically significant school-level effects of treatment assignment as large as one third of a standard deviation. The results correspond with the Success for All program theory, which emphasizes both comprehensive school-level reform and targeted student-level achievement effects through a multi-year sequencing of literacy instruction.

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 Kalla and Broockman (2016)).

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 Kalla and Broockman (2016)).
  - 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 Kalla and Broockman (2016)).
  - May also allow us to engage in participant observation (qualitative data)

- Alternatively: email response rates

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.

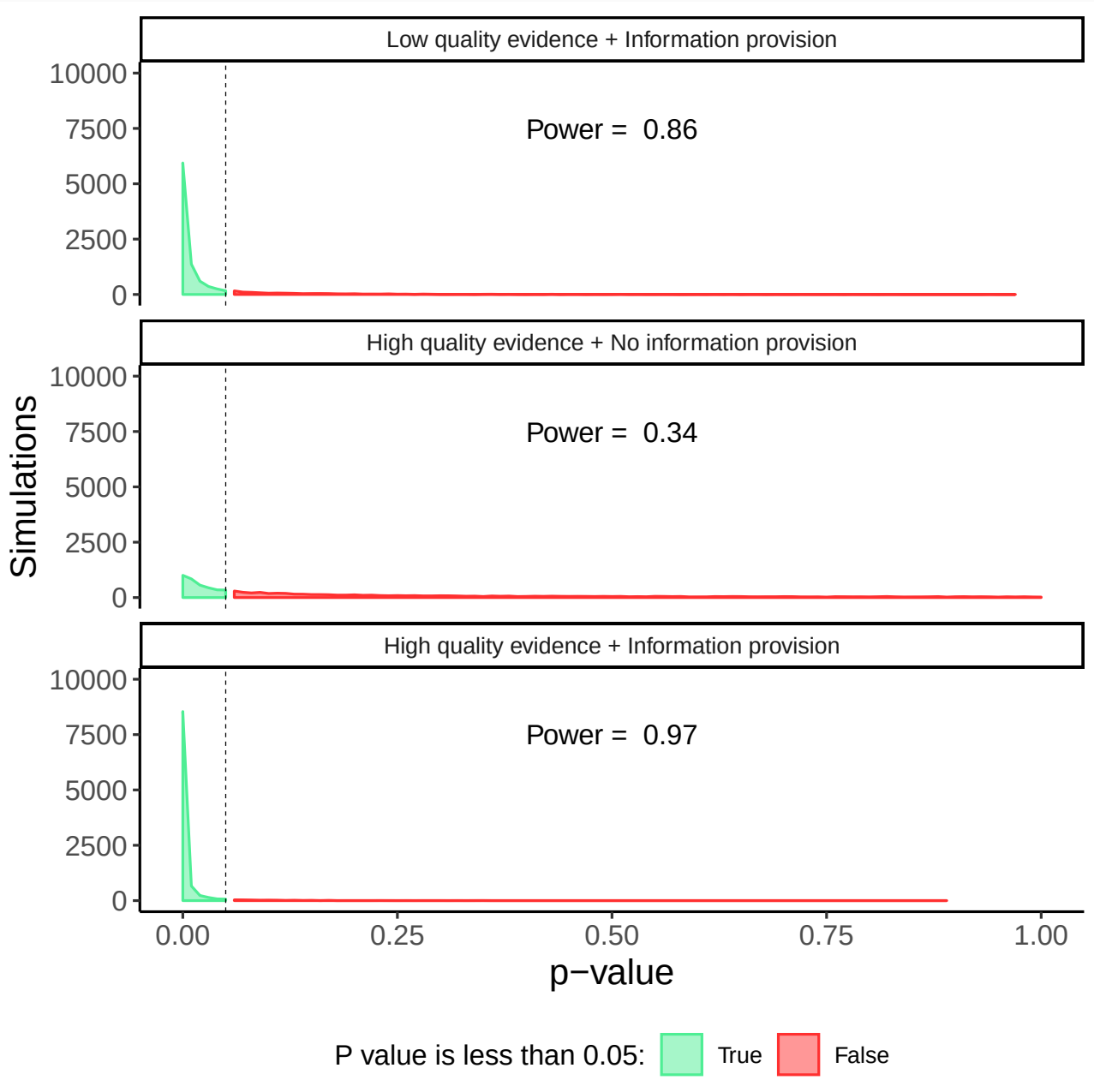
Power analysis assumptions

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

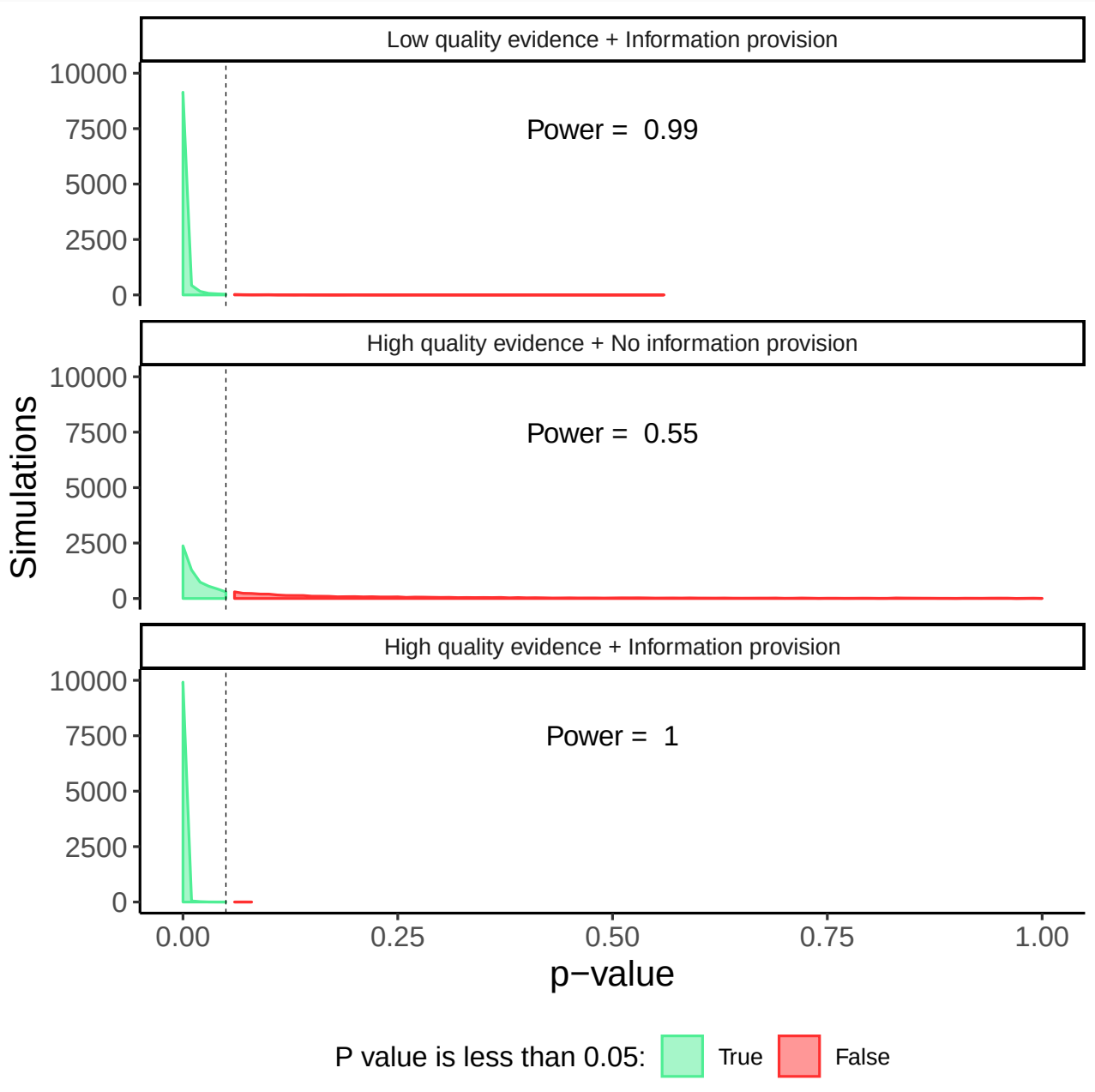
- Low quality evidence + information provision = -10%
- High quality evidence + no information provision = +5%
- High quality evidence + information provision = +12.5%

- 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?

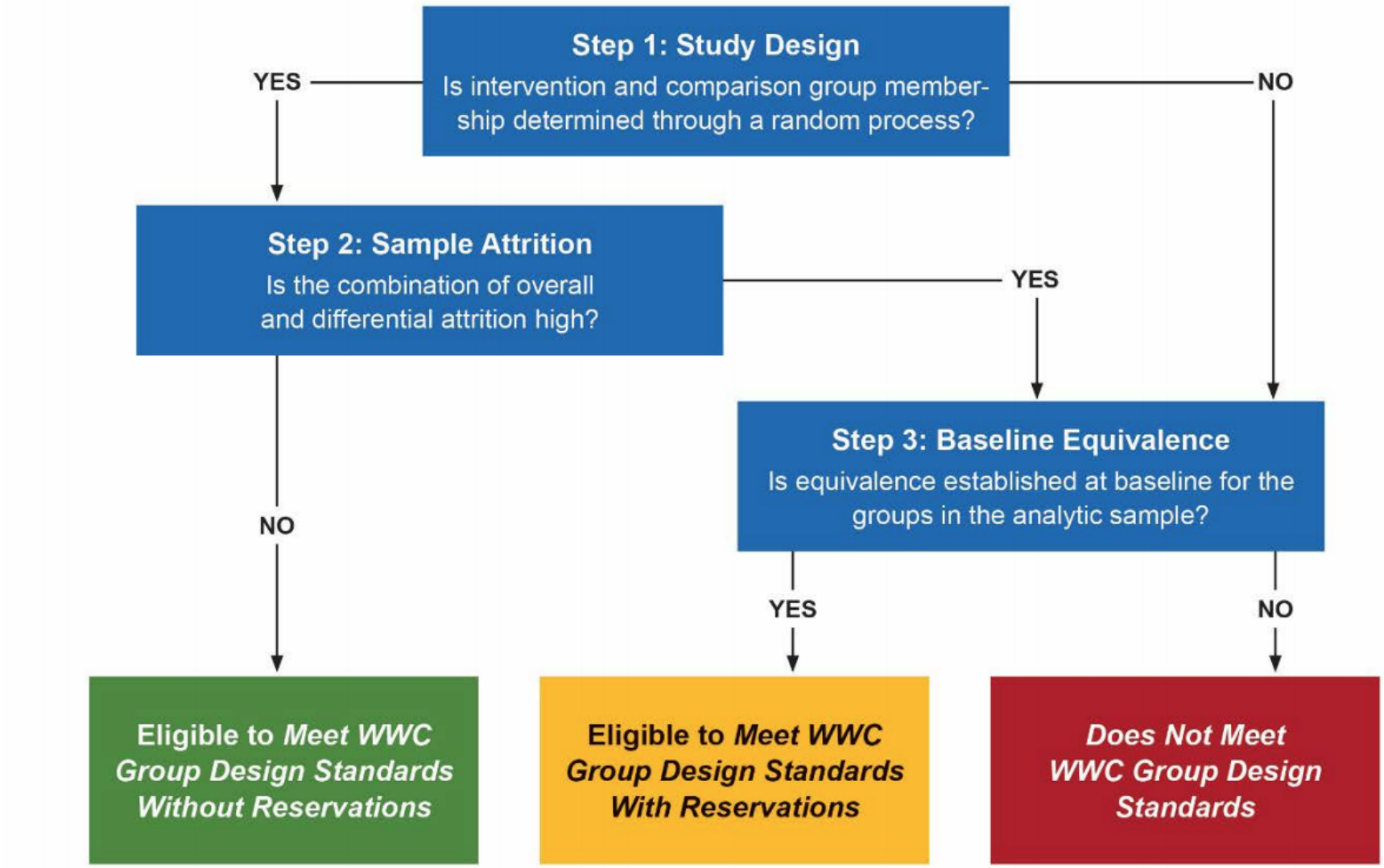
Supplemental material

Existing field/audit experiments

- Level: 5 state and local, 1 federal (Kalla & Broockman 2016). Between 70 state legislators in 1 state to 4,859 states legislators in 44 states (Butler & Broockman 2011)
- Treatments: contact by activists, ethnicity of email sender, contact by donators, in-person briefings, send district-specific survey results
- Outcomes: roll call votes, rate of response to emails, seniority of meeting, co-sponsorship of bills
- Design: block randomization on several covariates (state, chamber, party, reelection), matched pairs whenever possible
- Partnership with outside groups: 3 yes, 3 no

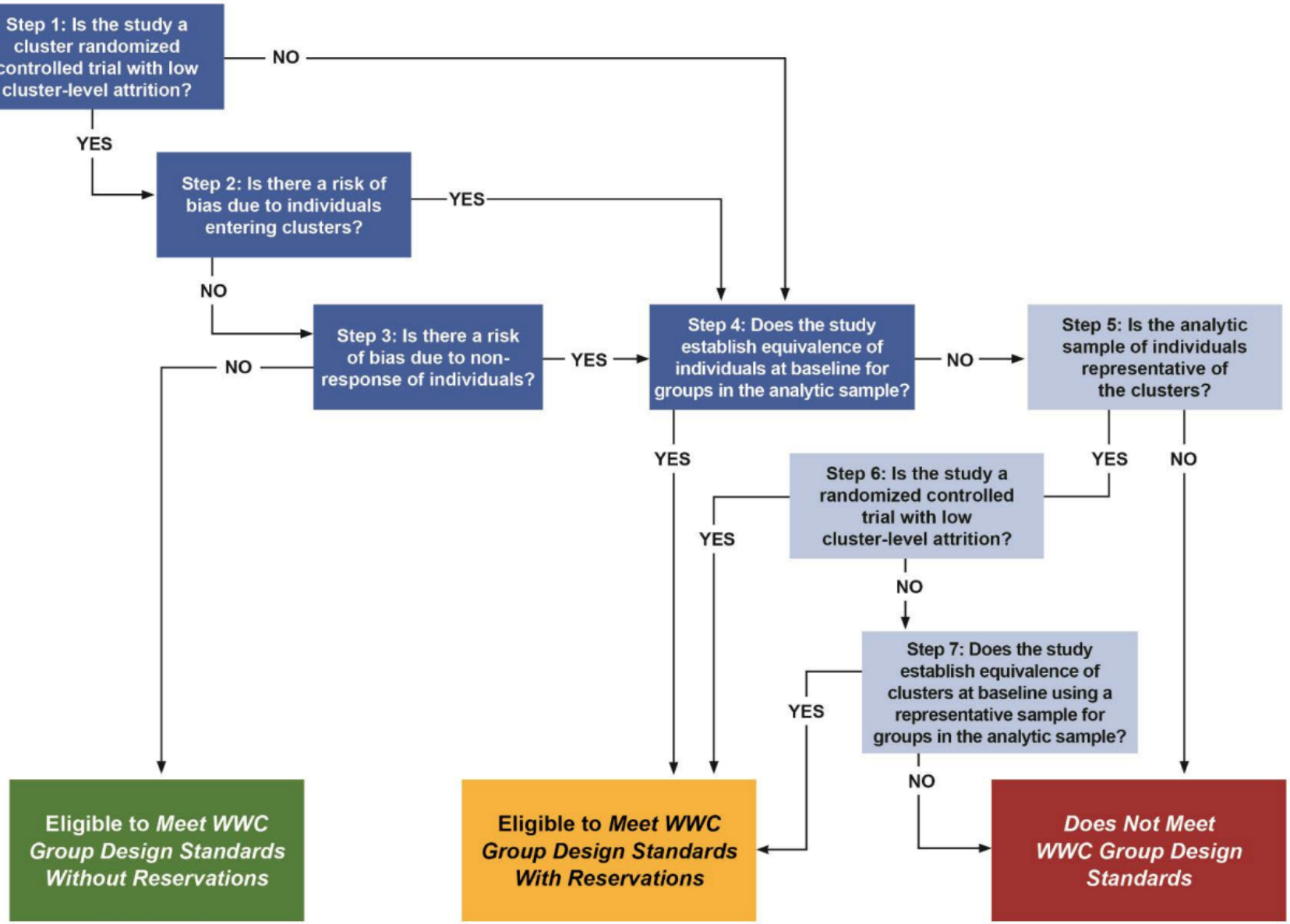
Evidence tiers

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



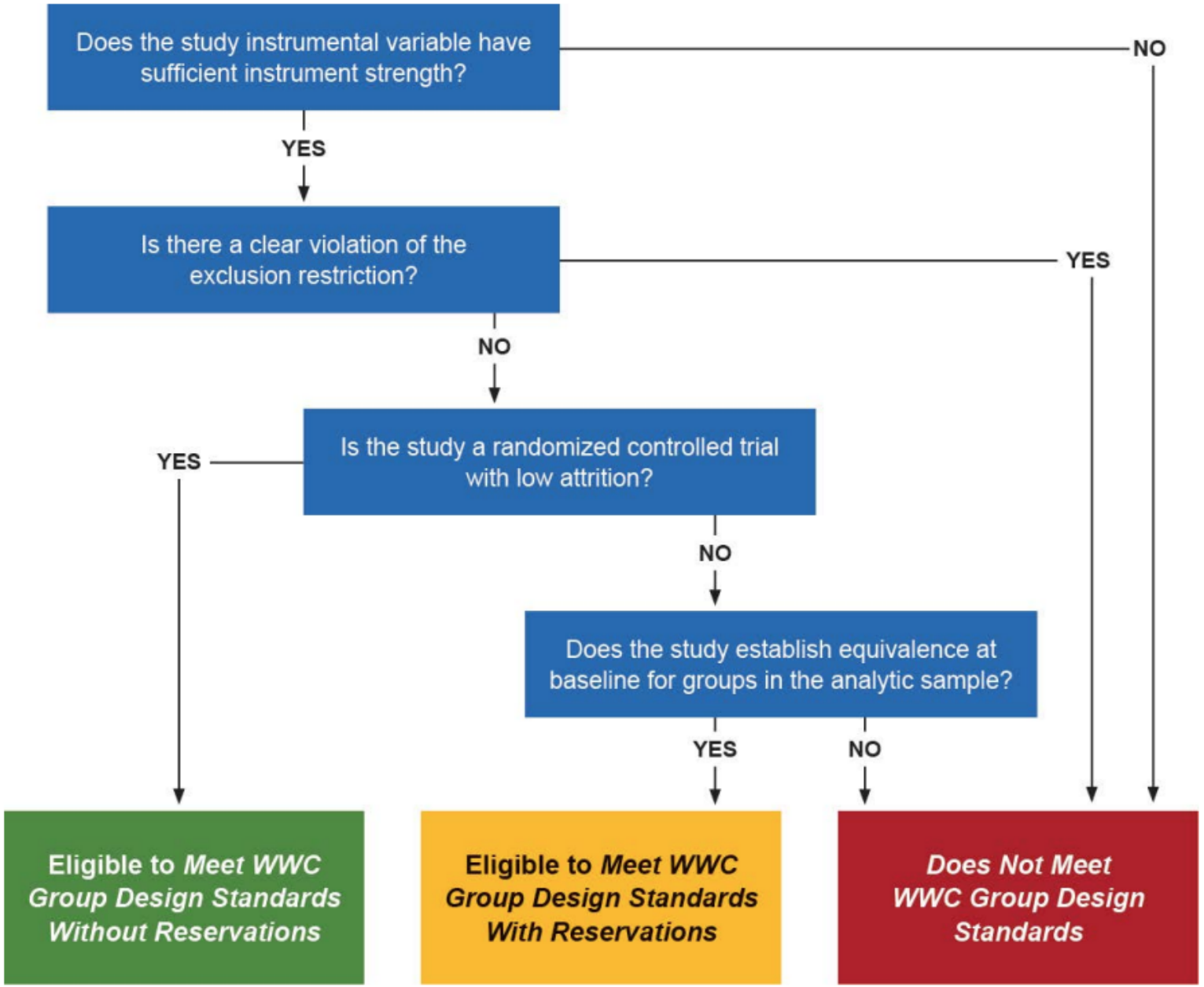
Evidence tiers: cluster random assignment

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



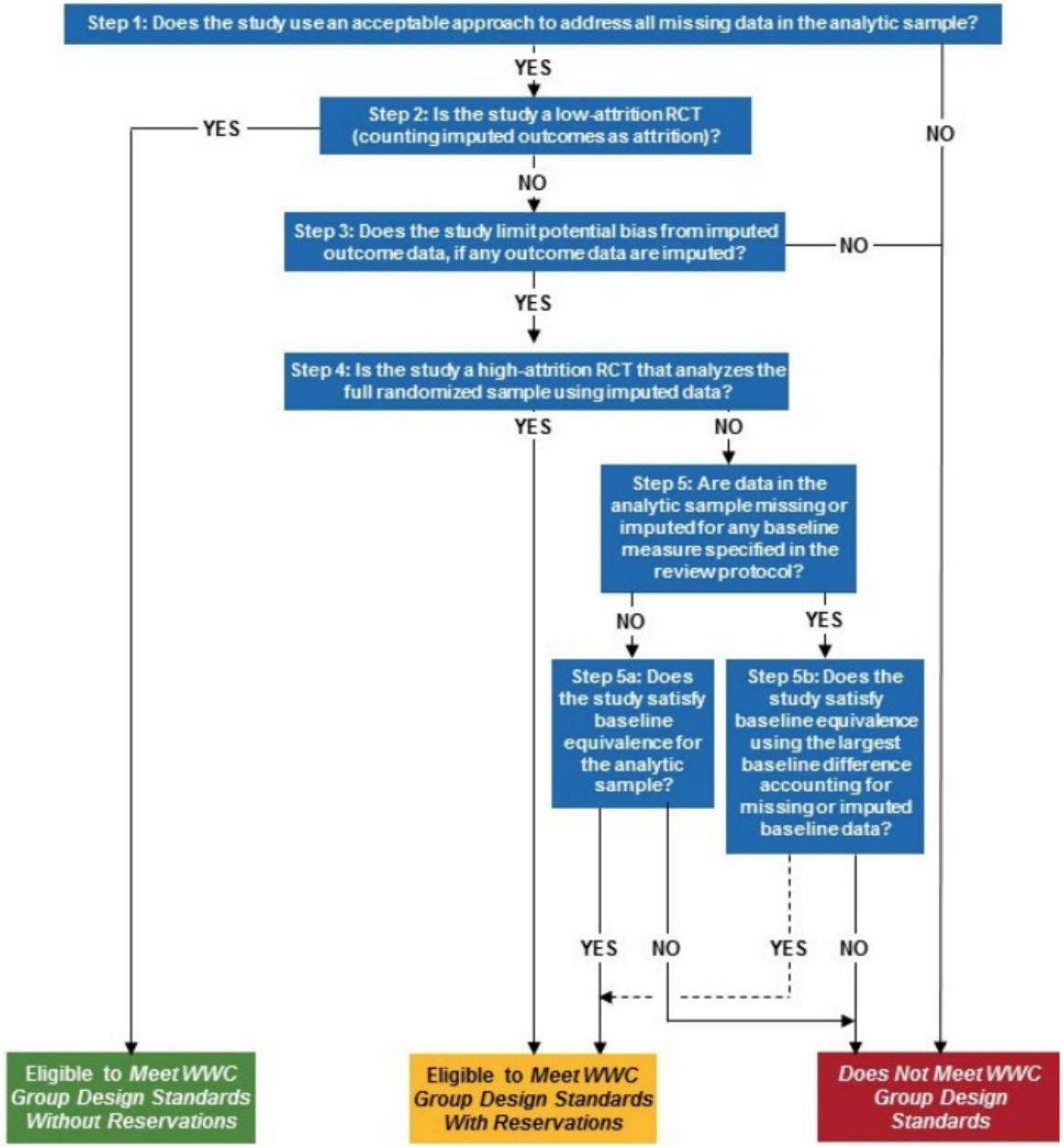
Evidence tiers: instrumental variables

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

References

Butler, D. M., & Broockman, D. E. (2011). Do politicians racially discriminate against constituents? a field experiment on state legislators. *American Journal of Political Science*, *55*(3), 463–477.

Kalla, J. L., & Broockman, D. E. (2016). Campaign contributions facilitate access to congressional officials: A randomized field experiment. *American Journal of Political Science*, *60*(3), 545–558.