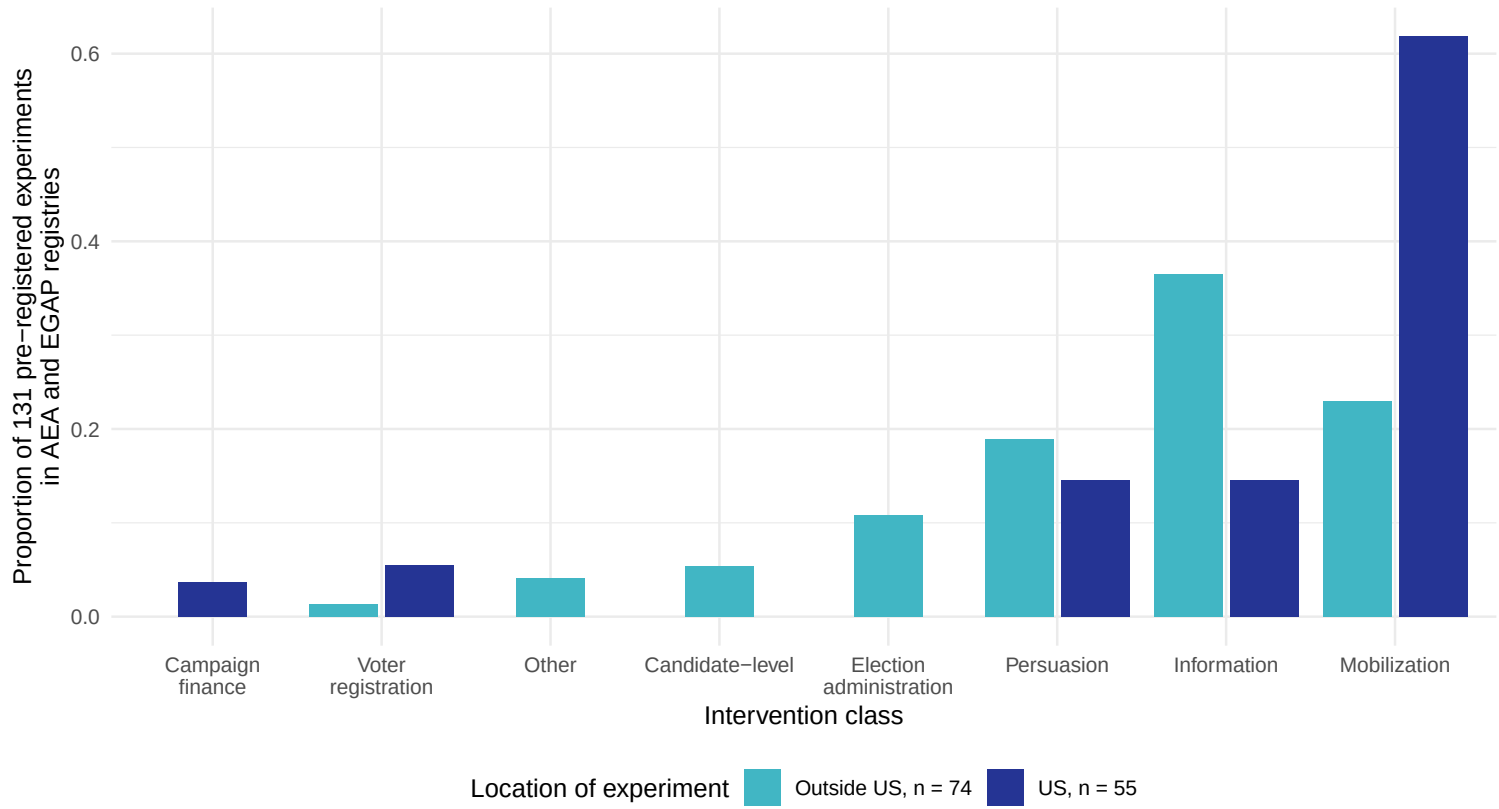
Ethics of Electoral Experiments: Design-Based Recommendations

Tara Slough—NYU

September 15, 2022

Electoral experiments as a tool

- Experiments on real elections represent a popular tool in studies of elections, political behavior, and political accountability.
- Developments in past ≈ 10 years:
  - Modal experiment is now outside US → wider variety of institutional contexts.
  - Some experiments conducted at massive scale.



The concern: downstream social harms

- Many of the treatments in survey of PAPs: low individual risk of harm.
- But by changing electoral behavior, we might change election outcomes.
- Changing the outcome of a contested election will harm some actors, e.g., the candidate made to lose and her supporters.
  - Downstream social impacts (Teele, 2013; Phillips, 2021; McDermott and Hatemi, 2020)
  - Often unpredictable (Baele, 2013; Carlson 2020)
  - Potential for disparate welfare impacts across groups (McDermott and Hatemi, 2020; Gubler and Selway, 2016; and Zimmerman 2016)
- APSA (2020) guidance (p. 15): interventions are of "minimal social risk if they are not done at a scale liable to alter electoral outcomes."

Surveying current practice

- Are we currently designing experiments with an eye to aggregate effects?
  - No: Aggregate electoral impact discussed in 2/129 PAPs
- Why do we need tools to do this better?
  - We need to consider mapping to the level of vote aggregation, the district → typically not unit of assignment or outcome measurement.
  - Ethical consideration is about changing any election outcome, not an average causal effect.
  - We need to think through these possibilities ex-ante.

Framework

Three key elements

1. The counterfactual to the experiment: What would have happened in the absence of a randomized experiment?
    - No intervention?
    - Partner conducts (likely) non-randomized intervention.
2. A model of voter response to the intervention:
    - Whose behavior might change in response → interference assumptions
    - A model of how voters respond.
3. A decision rule that compares potential aggregate electoral impact to predicted outcomes.

Experiments and their counterfactuals

- Critical question: who gets the intervention without absent the randomized experiment?

Stratum	Intervention	
Voter is...	$\pi(e)$	$\pi(\neg e)$
Always assigned to intervention	1	1
If-experiment assigned to intervention	1	0
If non-experiment assigned to intervention	0	1
Never assigned to intervention	0	0

- Typically if "always assigned" or "if non-experiment assigned" is non-empty → a partner is doing the intervention anyway.
  - But randomizing can have a distinct impact.
  - This is the impact we should be guarding against.
  - Distinct from current treatments of partners as a "get out of jail free" card.

Modeling voter behavior

- Interference assumption: which voters' behaviors can be affected by intervention?
  - Baseline: SUTVA + no within-cluster inference: intervention only impacts those directly assigned.
  - Extension: only SUTVA
  - Extension: no SUTVA → but we need to bound interference.

- How many votes change?
  - Manski (extreme value) bounds provide a maximally agnostic model of voter response to treatment.
  - Outcome defined as:

$$Y_i = \begin{cases} 1 & \text{vote for ex-ante marginal winner} \\ 0 & \text{else} \end{cases}$$

  - EV bounds will be conservative if assumption about interference is correct (or conservative).

Maximal aggregate electoral impact + decision rule

- Collectively we can characterize Maximal Aggregate Electoral Impact in district d from:
  1. Who is treated due to experiment? → counterfactual to experiment
  2. Whose behavior might be affected? → interference assumption
  3. How many votes are changed? → Manski bounds

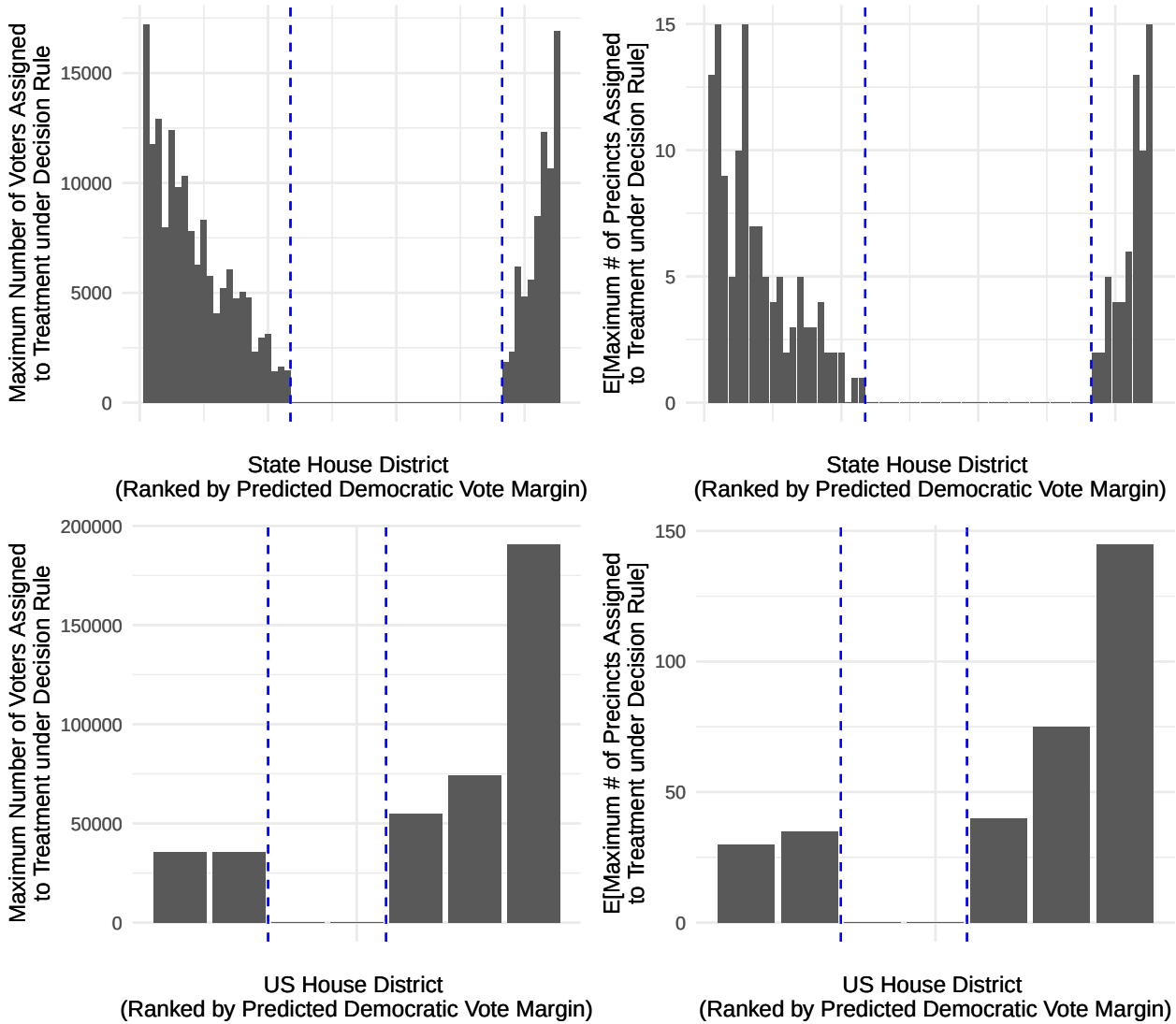
- We can also predict margin of victory.
  - Better prediction in some places than others.
  - But we can examine the predictive distribution.

- Decision rule in district d: Run an experiment iff:

$$2MAEI < \widehat{F}^{-1}(0.05)$$

What does this get us?

- Example from simulation on 2018 state house and US House elections in Colorado



Discussion

Five recommendations + tradeoffs for learning

- Five recommendations:
  1. Select treatments to improve the plausibility of **restricted interference**.
  2. Experiment in **FPTP** races.
  3. Implement interventions in **larger electoral districts**.
  4. Avoid implementing experimental interventions in close or **unpredictable** races.
  5. Reduce the number of subjects per district **assigned because of experiment**.
- Posit **tradeoffs** for:
  - The **types of treatments** we study experimentally
  - **Generalizability** of measured effects (#2-#4)
  - **Power** (#5, possibly #4)

Is the ethical objective too strict? Too permissive?

- Ethical objective is to avoid changing aggregate outcomes
- But there are other arguments that may lead to a stricter or more permissive approach to electoral experiments.

Justifictions for **stricter** rules	Justifictions for more **permissive** rules
1. Lack of consent	1. Intervention increases welfare
2. Self-determination	2. Election outcomes have many causes

- I view impetus to avoid changing aggregate outcomes as a useful default.

Conclusion

- Common criticism of electoral experiments: changing electoral outcomes → social harms
  - Extant ethical guidance generally suggests we shouldn't be changing elections around the world.
- ... but not reflected in the design of preregistered electoral experiments surveyed.
- I provide tools to design experiments that are unlikely to change electoral outcomes.
- Framework shows which design levers we can use to reduce aggregate electoral impact.
  - Study of "permissible" designs allows us to understand limits of what we can learn from electoral experimentation.

Thanks!