Target Trial Emulation for Health Policy Evaluation

Nicholas J. Seewald, PhD

Assistant Professor of Biostatistics
Department of Biostatistics, Epidemiology, and Informatics
University of Pennsylvania Perelman School of Medicine

13 December 2023
DBEI Epi Seminar
Joint with E.E. McGinty and E.A. Stuart

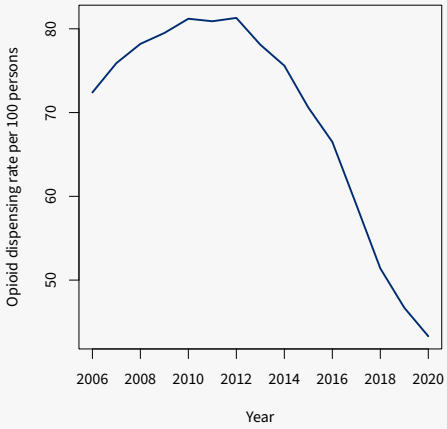
Perelman SCHOOL OF MEDICINE
University of Pennsylvania

Slides are Online

slides.nickseewald.com/dbeiepi2023.pdf

Motivating Example: Medical Cannabis Laws and Opioid Prescribing

- **4x** increase in opioid prescribing in U.S. from 1999-2012
  - Opioid prescribing for chronic non-cancer pain has played a meaningful role
- Getting better: prescribing down since 2012, but still ~3x higher than 1999



Dart, R. C. et al. (2015). *N. Engl. J. Med.*
https://www.cdc.gov/drugoverdose/rxrate-maps/index.html

Do Medical Cannabis Laws Change Opioid Prescribing?

- Cannabis is a potentially effective treatment for chronic non-cancer pain, but evidence is limited.
- Patients with chronic non-cancer pain are eligible to use cannabis under all existing state medical cannabis laws
- Some evidence of substitution among adults with chronic non-cancer pain

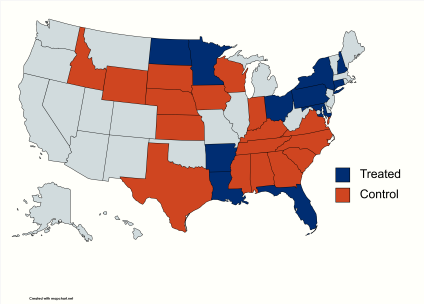
Question: What are the effects of state medical cannabis laws on receipt of opioid and non-opioid pain treatment among patients with chronic non-cancer pain, relative to what would have happened in the absence of such a law?

Bicket, M. C., Stone, E. M., and McGinty, E. E. (2023). *JAMA Netw. Open*.

Motivating Example: Medical Cannabis Laws and Opioid Prescribing

Our sample:

- 12 *treated* states that implemented a medical cannabis law between 2012 and 2019 and *do not also have recreational cannabis laws*
- 17 *comparison* states without medical or recreational cannabis laws

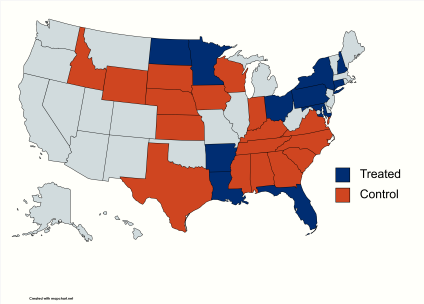


Motivating Example: Medical Cannabis Laws and Opioid Prescribing

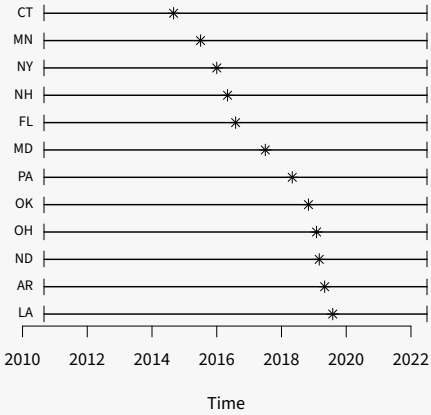
Our sample:

- 12 *treated* states that implemented a medical cannabis law between 2012 and 2019 and *do not also have recreational cannabis laws*
- 17 *comparison* states without medical or recreational cannabis laws

Goal: Estimate the effect of implementing a medical cannabis law on opioid prescribing outcomes, relative to what would have happened in the absence of treatment, among states that implemented such a law (an ATT).



"Staggered Adoption" of the Policy across States



States implemented medical cannabis laws at different times

Causal Inference for Policy Evaluation is Hard

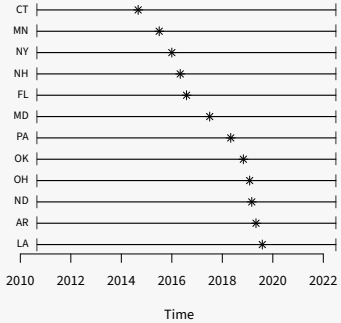
This is a causal question about the effects of a policy. But:

- Necessarily limited sample size
- Can't randomize
- Often high variability in "treatment" definitions
- Hard to isolate a particular policy's effects when other policies are in place.

An Early Attempt

Big Idea: Just toss everything into a "two-way fixed effects" regression model.

$$Y_{sit} = \underbrace{\alpha_s}_{\text{state fixed effects}} + \underbrace{\eta_t}_{\text{time fixed effects}} + \underbrace{\beta}_{\text{tx effect}} \mathbb{1}\{\text{state } s \text{ is treated at time } t\} + \epsilon_{sit}$$

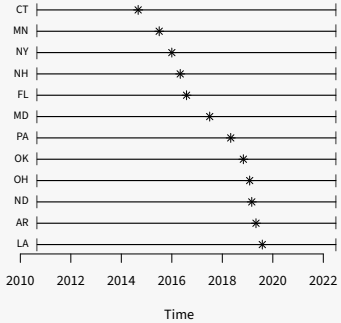


Goodman-Bacon, A. (2021). *Journal of Econometrics*.

An Early Attempt

Big Idea: Just toss everything into a "two-way fixed effects" regression model.

$$Y_{sit} = \underbrace{\alpha_s}_{\text{state fixed effects}} + \underbrace{\eta_t}_{\text{time fixed effects}} + \underbrace{\beta}_{\text{tx effect}} \mathbb{1}\{\text{state } s \text{ is treated at time } t\} + \epsilon_{sit}$$



This is (usually) a bad idea!

Biased under staggered policy adoption if there's a time-varying effect (basically always).

Goodman-Bacon, A. (2021). *Journal of Econometrics*.

Target Trial Emulation

A framework for thinking about non-experimental studies that enables stronger designs and facilitates causal inference.

- **Key Idea:** Think about the trial you would run if you could, then design a non-experimental study that gets as close to that as possible.
- Commonly used in epi, but broadly applicable.

Target Trial Emulation

A framework for thinking about non-experimental studies that enables stronger designs and facilitates causal inference.

- **Key Idea:** Think about the trial you would run if you could, then design a non-experimental study that gets as close to that as possible.
- Commonly used in epi, but broadly applicable.

Target trial emulation is a way to talk about non-experimental study design in a way familiar to trialists.

A Warning!

Policy trial emulation does not map well onto trial emulation in other contexts (like epi).

Health policy applications require different considerations than epidemiologic ones. Crucially:

1. Policies are cluster-level interventions
2. Policy evaluations require natural experiments

We have to make trade-offs. Keep this in mind throughout the talk!

Components of a Policy Trial Emulation

1. Scientific Question and Definitions of Exposure & Control
2. Units and Eligibility Criteria
3. Assignment Mechanism
4. Baseline / Time Zero
5. Outcomes and Follow-Up
6. Causal Estimand
7. Statistical Analysis

Components of a Policy Trial Emulation

1. Scientific Question and Definitions of Exposure & Control
2. Units and Eligibility Criteria
3. Assignment Mechanism
4. Baseline / Time Zero
5. Outcomes and Follow-Up
6. Causal Estimand
7. Statistical Analysis

We recommend explicit comparisons of the non-experimental study to a target trial on these 7 components.

Scientific Question and Definitions of Exposure & Control

Policy evaluations require natural experiments:

- Policies must be implemented before they can be studied
- Researchers don't decide what a policy does or who it affects

The definition of the exposure necessarily preceeds and shapes the scientific question.

Scientific Question and Definitions of Exposure & Control

Trials require clear definitions of what each randomized arm receives and must ensure consistent treatment delivery.

In a policy trial, we would

- implement the same policy in the same way in every treated unit
- do the same for controls (if control is a specific alternative policy) or “business as usual”

Scientific Question and Definitions of Exposure & Control

Trials require clear definitions of what each randomized arm receives and must ensure consistent treatment delivery.

In a policy trial, we would

- implement the same policy in the same way in every treated unit
- do the same for controls (if control is a specific alternative policy) or “business as usual”

In non-experimental policy evaluation, specifics of each policy can be quite heterogeneous.

Scientific Question and Definitions of Exposure & Control

Goal: Identify a class (or small number of classes) of qualitatively similar policies that will be the exposure(s).

- “Policy mapping”/“legal epidemiology”: systematic approach to understanding timing of policies and the granular rules within them
- Understand different versions and core components of the policy, then decide which are qualitatively similar.

Scientific Question and Definitions of Exposure & Control

In our medical cannabis law study:

- Exposure: “A state medical cannabis law permitting cannabis use among individuals with chronic non-cancer pain with cannabis available for patient purchase through dispensaries”
- Comparison: Absence of such a law over the entire 2010-2022 study period

---

McGinty, E. E. et al. (2023). *Ann Intern Med*.

Scientific Question and Definitions of Exposure & Control

"Confounding" policies may offer an alternative explanation for any observed effect.

- A strong policy trial emulation will precisely define exposure and comparison conditions to disentangle effect of interest.
- In medical cannabis law study, exposure was refined to a state medical cannabis law *and* absence of a recreational cannabis law throughout the entire study period.

Simultaneously-implemented policy bundles can only be studied in aggregate.

Units and Eligibility Criteria

Must consider

1. Units that could implement the policy (“policy-level”)
    - Fundamentally of interest
    - Would be randomized to (not) implement policy
    - In medical cannabis laws study: states with no medical cannabis law in 2010.
2. Units that would be affected by the policy if enacted (“individual-level”)

Units and Eligibility Criteria

Must consider

1. Units that could implement the policy (“policy-level”)
    - Fundamentally of interest
    - Would be randomized to (not) implement policy
    - In medical cannabis laws study: states with no medical cannabis law in 2010.
2. Units that would be affected by the policy if enacted (“individual-level”)
    - Contribute (possibly summary-level) data to analysis
    - In medical cannabis laws study: adults with chronic non-cancer pain diagnoses eligible to use medical cannabis under their state’s law.

Units and Eligibility Criteria

Policy evaluations should emulate *cluster-randomized* trials. Quality of emulation is partially determined by available data.

- "Group panel" data aggregated to policy-level is common
  - Might not be possible to restrict to target population $\rightarrow$ weaker trial emulation
  - Okay if aggregated from target population (e.g., everyone in a state) or in some contexts (e.g., state-month homicide counts)
- Individual-level data enables additional eligibility criteria
  - Ability to restrict to target population strengthens trial emulation

Policy-Level Eligibility Criteria

Choosing appropriate comparators is critical.

- Should consider contextual factors that may affect policy adoption and outcomes *differently over time*
- Selecting geographically distant controls alleviates spillover concerns
- Could use all units untreated at baseline OR require comparators remain untreated throughout the study
    - I know, but hear me out

Policy-Level Eligibility Criteria

Causal inference for policy evaluation requires careful trade-offs.

- **“Never-Treated” Controls**
  - Chosen using knowledge of policy status later in time – could lead to bias!
  - BUT, the control group remains constant over time
- **“Untreated at Baseline” Controls**
  - Avoids conditioning on post-treatment information
  - Allows the control group to change over time $\rightarrow$ is observed effect due to the policy or the changing comparator?

Choice needs to consider potential biases, follow-up time, need to retain sufficient units for estimation (only 50 states!), etc.

Individual-Level Eligibility Criteria

- Included indivduals should be from the target population and would be affected by the policy of interest if enacted.
- May choose to mimic high-quality retention efforts in an RCT by requiring “continuous presence” (e.g., continuous enrollment in health insurance claims)
  - Maybe not appropriate if exposure affects probability of continuous presence.
  - Not doing this allows patient case-mix to change over time, threatening internal validity (but improving external); weighting can help.
  - Done in medical cannabis laws study because infeasible that law would impact insurance enrollment.

Assignment Mechanism

Hypothetical Target Trial	Policy Trial Emulation Analogue
• Cluster-randomized • Possibly stratified • Possibly blinded (?) • Unconfounded on average	• Not randomized • Almost certainly unblinded • Could emulate cluster randomization • Affected by known and unknown state-level confounders

Baseline / Time Zero

Hypothetical Target Trial	Policy Trial Emulation Analogue
The time of randomization. • Recruitment & prep done prior, so policy can be implemented immediately.	When policy could start impacting outcomes. • e.g., when first cannabis dispensary opens in a state.

Baseline / Time Zero

Without randomization, baseline is complicated for comparison units. When could they have implemented the policy but did not?

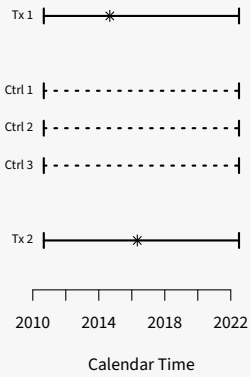
Poor definition of baseline for controls can lead to bias from conditioning on post-treatment information.

Especially complicated under staggered adoption. One solution is **serial trial emulation**:

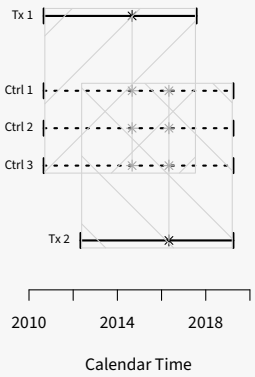
- Define baseline for each treated unit, then use those calendar times to define a series of baselines for controls.
- Creates multiple trial emulations, one per unique policy implementation date.

Serial Trial Emulation

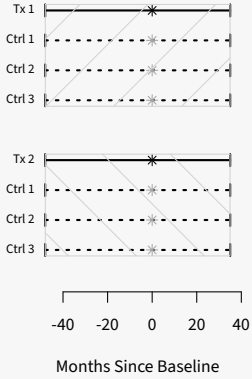
1. Identify Implementation Dates



2. Map Implementation Dates and Study Periods onto Controls



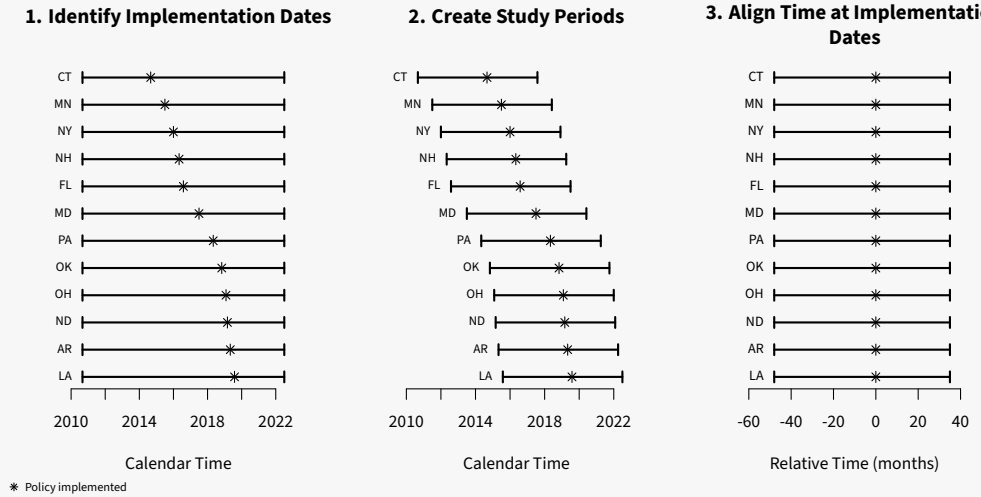
3. Create Unique Trials Aligned in Relative Time



✳ Policy implemented

N.J. Seewald, Target Trials for Policy Evaluation

Serial Trial Emulation in Our Medical Cannabis Study



Outcomes

Outcomes are interpreted at the policy level: they'll be proportions, means, etc. for each policy-level unit.

- Natural for group-panel data
- Individual-level data will be aggregated to the policy level

Can be prospectively designed in an RCT, but non-experimental policy evaluations are retrospective by nature.

Follow-Up

- RCTs typically have one (or few) pre-exposure measurements.
- In non-experimental context, validity of causal estimate relies on reasonably large number of pre-treatment measurement occasions.
  - Need to establish pre-policy outcome trends & anticipation effects
  - 4 years in medical cannabis law study
- Post-exposure follow-up should capture meaningful effects & changes therein (e.g., ramp-up)
  - 3 years in medical cannabis law study: balance need to look for ramp-up effects against need to avoid confounding laws

Causal Estimand

An *estimand* is a population-level quantity that statistically describes the treatment effect of interest.

Here, a causal quantity that describes the average difference between counterfactual outcomes in policy-level units under exposure and control.

- Answers questions about what would have happened under different states of the world (e.g., with and without the policy exposure of interest)

Causal Estimand

An *estimand* is a population-level quantity that statistically describes the treatment effect of interest.

Here, a causal quantity that describes the average difference between counterfactual outcomes in policy-level units under exposure and control.

- Answers questions about what would have happened under different states of the world (e.g., with and without the policy exposure of interest)

"The expected difference in the proportion of individuals receiving any opioid prescription in a given month, averaged over treated states and over three years, had the law been implemented versus had it not been implemented in those states."

Categories of Causal Estimand

1. **Average treatment effect (ATE)** compares expected counterfactual outcomes under treatment to those under control on average over the entire population: $\mathrm{E}\left[Y(1) - Y(0)\right]$.
2. **Average treatment effect among the treated (ATT)** compares observed outcomes in the treated group to what would have happened had it not been treated: $\mathrm{E}\left[Y(1) - Y(0) \mid A = 1\right]$.
3. **Average treatment effect among controls (ATC)** compares observed outcomes in the untreated group to what would have happened had it been treated: $\mathrm{E}\left[Y(1) - Y(0) \mid A = 1\right]$

Policy evaluations typically target the ATT: most feasible with fewest big conceptual jumps.

Statistical Analysis

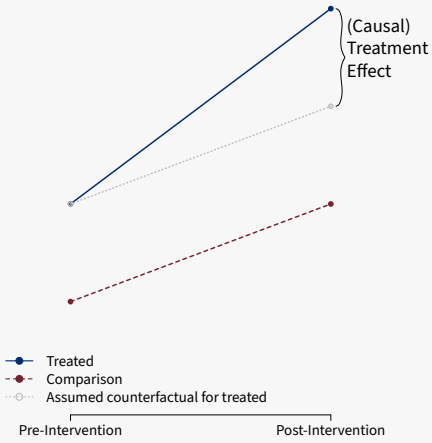
The cluster-randomized target trial can used “standard” analytic tools.

In non-experimental policy trial analogue:

- Methods typically use pre-baseline information from treated and control groups to extrapolate an estimate of treated group’s counterfactual outcomes under no treatment.
- Broad class of methods: difference-in-differences, synthetic controls, etc.
- Analytic approach should estimate the estimand under reasonable assumptions.

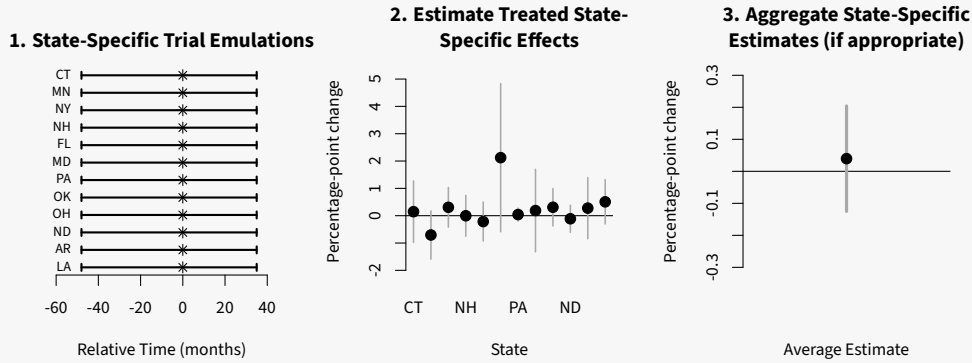
Difference-in-Differences

- Compare change in outcome over time between treated and comparison groups
- Under assumption that treated group would look like comparison group in absence of treatment, can estimate causal treatment effect
  - This is called the *(counterfactual) parallel trends assumption*



Stacked Difference-in-Differences

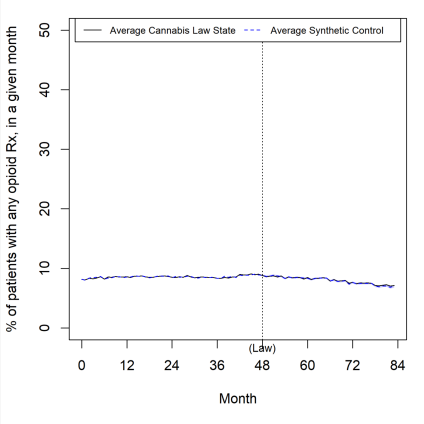
Uses diff-in-diff to estimate effects for each serial per-implementation-date trial emulation then aggregates them if appropriate.



Baker, A., Larcker, D. F., and Wang, C. C. Y. (2021).

Synthetic Controls

- Construct a weighted combination of control states that mimics the outcome trajectory of the treated state in the pre-treatment period.
- Use the “synthetic control” trajectory to estimate treated state’s counterfactual under no treatment.



Discussion

Explicit head-to-head comparison of target trial and a non-experimental policy evaluation helps identify threats to causal inference.

- Poorly-defined exposure inappropriately grouping different policies $\rightarrow$ estimate effect of some average policy that doesn't exist, ignores heterogeneity
- Failure to account for confounding policies $\rightarrow$ could estimate effect of wrong thing

Discussion

Strong agreement between trial emulation and target trial allows for use of causal language.

- Use “estimated effect” to acknowledge statistical and causal uncertainty
- Emphasize confidence intervals

Anecdotally, explicit comparisons and transparency about design and analysis have greatly improved understanding of our non-experimental studies and their results.

Acknowledgements

NIDA R01DA049789 (PI: McGinty)

Draft paper under review at *Annals of Internal Medicine* available on request.

seewaldn@pennmedicine.upenn.edu
nickseewald.com