Perelman SCHOOL OF MEDICINE
UNIVERSITY of PENNSYLVANIA

HP/CHIBE ANALYTICS MONTHLY LUNCH ZOOM MEETING

Target Trial Emulation for Evaluating Health Policy

Nicholas J. Seewald, Ph.D.
Assistant Professor of Biostatistics

21 January 2025

Follow Along!

doi.org/nmmw

nickseewald.com/chibeAnalytics2025

The Goal of Policy Evaluation

In general:

> *"What is the effect of [a policy] on [outcome(s) of interest] over [a defined period of time], relative to what would have happened in the absence of the policy?"*

Challenges of Policy Evaluation

Can be difficult to isolate policy of interest

Confounding by time

Heterogeneous policies

Small sample size

Designing for Policy Evaluation

High-quality study design helps alleviate concerns about

- Isolating the policy of interest
- Confounding by time
- Heterogeneous policies

Throughout, I'll advocate blending with qualitative methods to allow better understanding

- "Treatment" definition
- Implementation time
- Effects (or lack thereof)

State Medical Cannabis Laws & Opioid Rx

McGINTY ET AL. 2023, *ANNALS OF INTERNAL MEDICINE*

Cannabis is a potentially effective treatment for chronic noncancer pain.

Patients with chronic noncancer pain are eligible to use cannabis under all existing state medical cannabis laws.

Some evidence of substitution of cannabis for opioids among adults with chronic noncancer pain (Bicket et al. 2023, *JAMA Network Open*)

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

Target Trial Emulation (TTE)

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 analogue that gets as close as possible.

- Common in epidemiology, but broadly applicable
- *Not magic!* TTE *per se* does not guarantee quality.

Components of Policy Trial Emulation

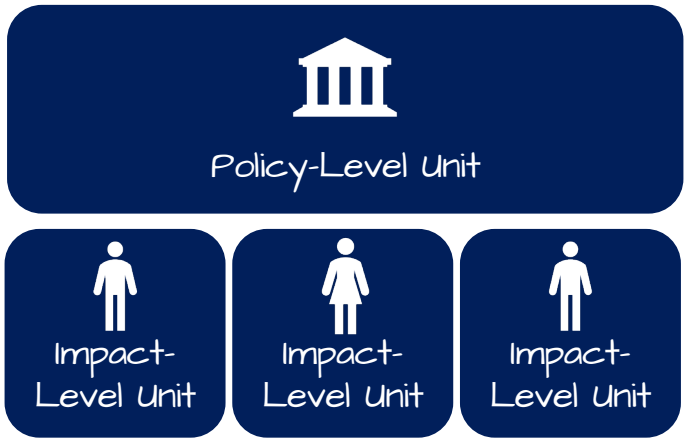
1. Units and eligibility criteria
2. Definitions of exposure and comparison conditions
3. Assignment mechanism
4. Baseline / time zero and follow-up
5. Outcomes
6. Causal estimand
7. Statistical analysis and assumptions

Units and Eligibility Criteria

Policy evaluations must consider

1. **"Policy-level"** units that could implement the policy or comparison condition
2. **"Impact-level"** units that the policy is designed to affect and on which outcomes are measured.

If policy- and impact-level units are different, policy evaluations would emulate *cluster-randomized* trials.



Units and Eligibility Criteria

Policy evaluations must consider

1. **"Policy-level"** units that could implement the policy or comparison condition
2. **"Impact-level"** units that the policy is designed to affect and on which outcomes are measured.

If policy- and impact-level units are different, policy evaluations would emulate *cluster-randomized* trials.

Policy-Level Units

In a **hypothetical policy trial**, policy-level units would be

- units that *could* implement the policy (states, organizations, etc.) monitored longitudinally

Eligibility criteria would be based only on pre-policy information:

- "has not implemented the policy before" or more complex (e.g., "has not previously implemented policies X, Y, Z")

Policy-Level Units

In a **policy trial emulation**, policy-level units would be

- units that *did* implement the policy or *did* implement the comparison condition
- at “time zero” / “study entry” (ideally), and
- monitored longitudinally

Eligibility criteria *should* be based only on pre-policy information:

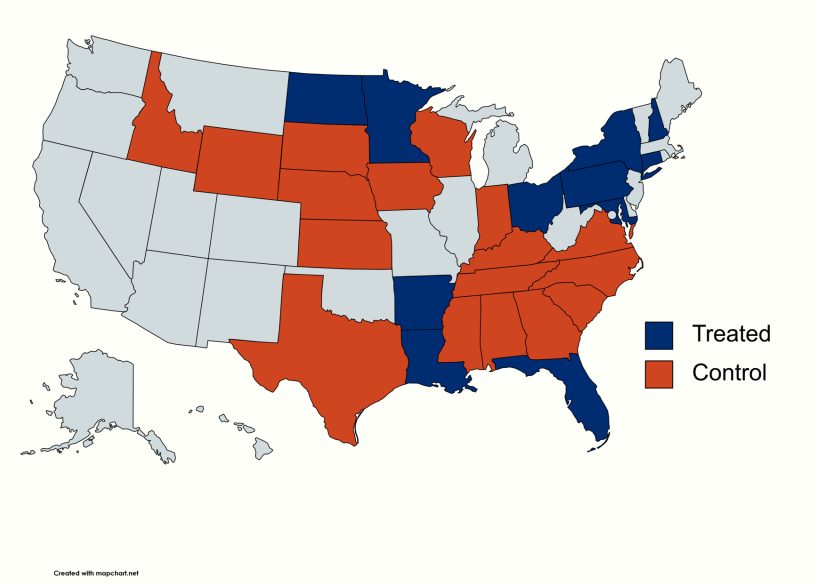
- “has not implemented the policy before” or more complex (e.g., “has not previously implemented policies X, Y, Z”)

State Medical Cannabis Laws & Opioid Rx

McGINTY ET AL. 2023, *ANNALS OF INTERNAL MEDICINE*

Policy-Level Units:

- 12 *treated* states that implemented a medical cannabis law between 2012 and 2019 and did not also have recreational cannabis laws.
- 17 *comparison* states without medical or recreational cannabis laws, 2010-2022



Impact-Level Units

In a **hypothetical policy trial**, impact-level units are those that the policy is designed to affect. Possibly

- the policy-level units themselves, *or*
- sub-units nested in policy-level units on which outcomes are measured, ideally from the population the policy is designed to affect.

Eligibility would be based only on pre-policy information:

- “Lives in state X” for policies that apply to everyone
- “Lives in state X and was diagnosed with Y before the policy”, etc.

Retention efforts if impact-level units followed longitudinally

Impact-Level Units

In a **policy trial emulation**, the same considerations apply.

Outcome data will ideally be available from impact-level units.

In state medical cannabis laws study,

> *U.S. adults with diagnosed with a qualifying chronic noncancer pain condition (low back pain, fibromyalgia, migraine, etc.) prior to their state implementing a medical cannabis law.*

Available Data Affects Emulation Quality

Quality of trial 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 (→ weaker study)
- Okay if aggregated from target population (e.g., everyone in a state) or in some contexts (e.g., state-month homicide counts)

Impact-level data enables additional eligibility criteria

- Can restrict to target population (→ stronger study)

Longitudinal Follow-Up of Impact-Level Units

In policy trial emulation, following impact-level units longitudinally vs. in repeated cross-sections changes the *sampling frame*.

"Continuous presence" requirement can mimic high-quality retention efforts in an RCT

- Maybe inappropriate if exposure affects probability of continuous presence
- Not requiring this allows patient case-mix to change over time
  - Threatens internal validity but improves external validity (weighting can help!)

Impacts generalizability

Definitions of Exposure & Comparison Conditions

In a **hypothetical policy trial**, we would

- have one policy that all implementing units are assigned to implement;
- similarly for controls if comparison condition is a specific alternative policy, or “business as usual”

But, in a **policy trial emulation**,

- specific details of each policy can be quite heterogeneous

Defining the Exposure

Use **qualitative methods** to identify a class (or small number of classes) of similar policies that will be the exposure(s).

- "Policy mapping" or "legal epidemiology" - systematic approach to understanding policies & rules
- Decide which core policy components are necessary for the study

Under high heterogeneity, could emulate a multi-arm trial.

Definition should be precise to help disentangle effects of interest & avoid confounding policies.

In the state medical cannabis laws study,

> *A state medical cannabis law permitting cannabis use among individuals with chronic noncancer pain with cannabis available for patient purchase through dispensaries.*

Defining the Comparison Condition

Best practices for trial emulation:

1. At time zero, the comparison group is every policy-level unit that has not been exposed at that time
2. If unexposed units become exposed later, censor their outcomes when they become exposed.

This ideal design isn't always practical for policy evaluations.

Choosing Comparators for Policy Evaluation

"Unexposed at Baseline" Comparators

- Avoids conditioning on post-treatment information
- Allows the comparison group to change (possibly meaningfully) over time.
  - Is an observed effect due to the policy or the changing comparison group?

"Never Exposed" Comparators

- Chosen using knowledge of future policy status – could lead to bias!
- Clearly not ideal in the target trial framework, but
- the comparison group remains unchanged over time.

Never-Exposed Comparators

Very commonly used in policy evaluations, but

- Studies that choose to use never-exposed comparators are subject to additional assumptions about the comparability of ever- and never-exposed units and are subject to bias.
- *This choice deviates from ideal target trial emulation.*

Options for redesigning the study:

- Change policy-level eligibility criteria to *de facto* exclude likely bad comparators (geography, urbanicity, etc.). Pay attention to remaining sample size!
- Limit the follow-up period to one in which good comparators exist.

Assignment Mechanism

Hypothetical Target Trial

- Cluster-randomized
- Possibly stratified
- Almost certainly unblinded
- Unconfounded

Policy Trial Emulation Analogue

- Not randomized
- (Usually) emulates cluster randomization
- Almost certainly unblinded
- Affected by known and unknown characteristics of policy-level units

Baseline / Time Zero

Hypothetical Target Trial

- Time of randomization
  - Recruitment/prep done prior, so implementation can happen right away.
- Easy to define for all units

Policy Trial Emulation Analogue

- Time at which the policy could start impacting outcomes
- Challenging to define for comparison units: when could they have implemented the policy, but didn't?

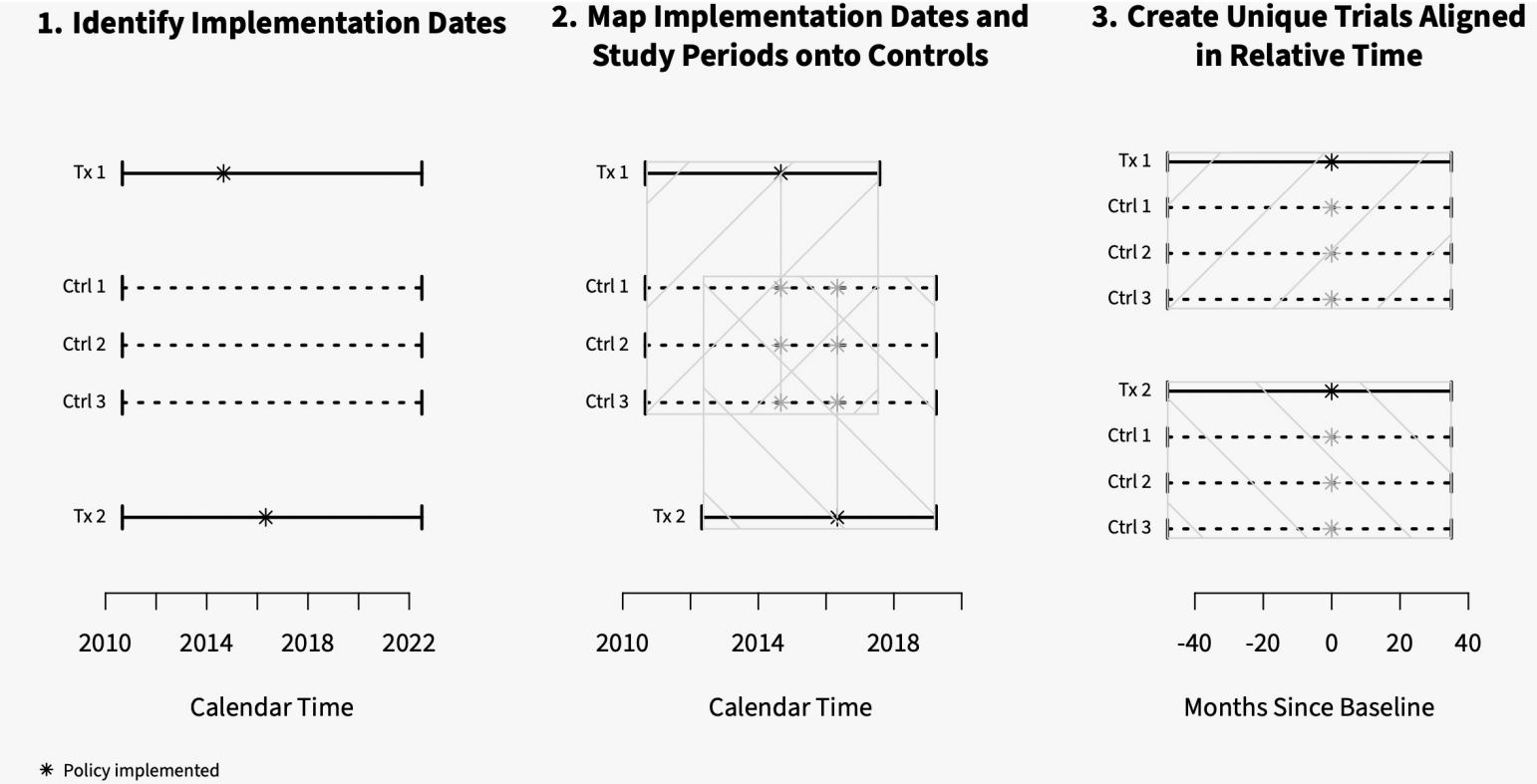
Baseline / Time zero

A bad definition can lead to bias (conditioning on post-treatment information)

Staggered adoption yields even more complexity. One solution is **serial trial emulation:**

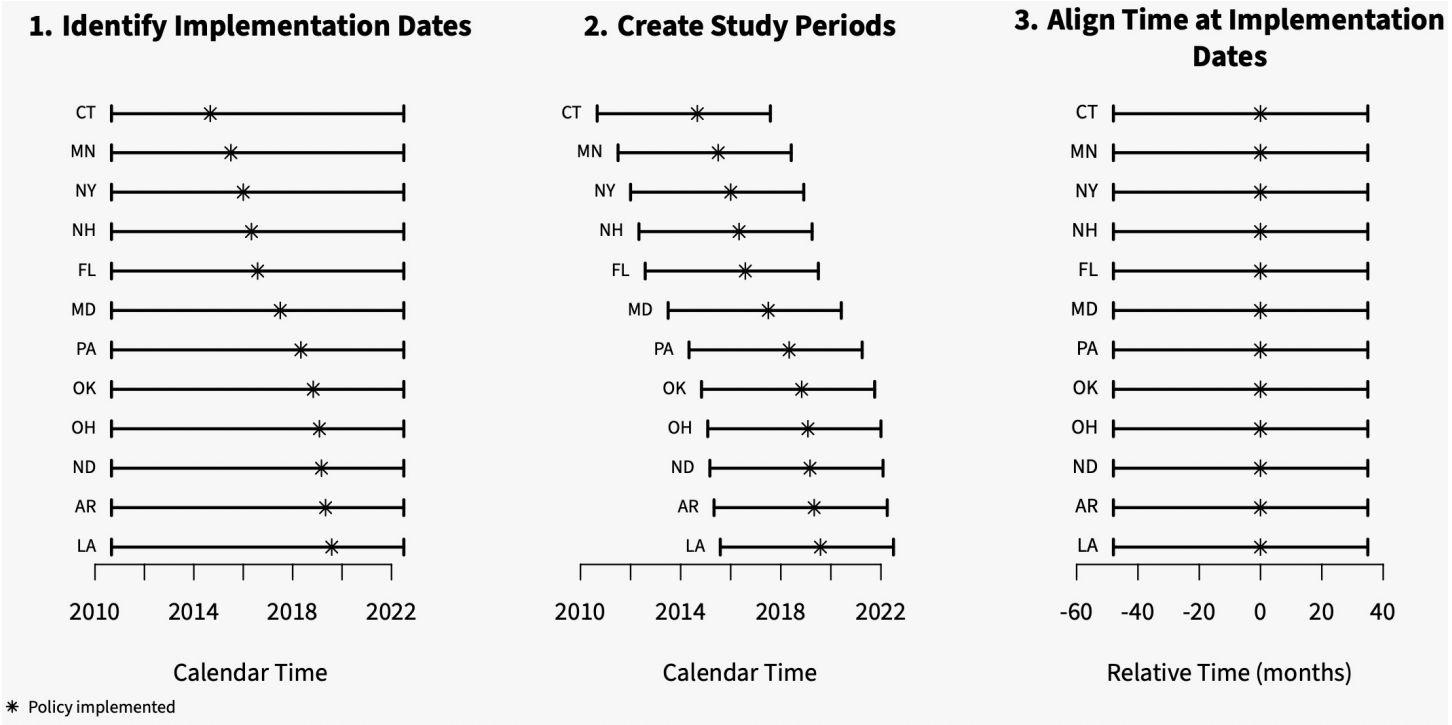
- Define baseline for each treated unit, then use those calendar times to define a series of baselines for comparators
- Creates multiple trial emulations, one per unique policy implementation date

Serial Trial Emulation



State Medical Cannabis Laws & Opioid Rx

McGINTY ET AL. 2023, *ANNALS OF INTERNAL MEDICINE*



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.

Validity of causal estimate in non-experimental study often relies on reasonably large number of pre-treatment measurement occasions.

Post-exposure follow-up should capture meaningful effects & changes therein.

In state medical cannabis laws study,

4 years pre-law and 3 years post-law

Causal Estimand

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

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

- Answers questions about what would have happened under different states of the world.

Categories of Causal Estimand

Average treatment effect (ATE) compares expected counterfactual outcomes under exposure to those under the comparison condition on average over the entire population

$$E[Y(1) - Y(0)]$$

Average treatment effect among the treated (ATT) compares observed outcomes in the exposed group to what would have happened had they been unexposed:

$$E[Y(1) - Y(0) \mid A = 1]$$

Average treatment effect among comparators (ATC) compares observed outcomes in the unexposed group to what would have happened had they been exposed:

$$E[Y(1) - Y(0) \mid A = 0]$$

Typically the target (by convention)

Why the emphasis on ATT?

Not always what we really want, or what's really of interest.

- Policymakers want to know what will happen in *their* state *if* they implement the policy
- ATT describes effects among states that already implemented

Targeted by convention and for feasibility

- Estimating counterfactuals under treatment for untreated states feels like a big conceptual jump
  - This is required to estimate ATE and ATC

Analysis in 60 Seconds

The **hypothetical policy trial** could use “standard” analytic tools for (cluster-)randomized trials

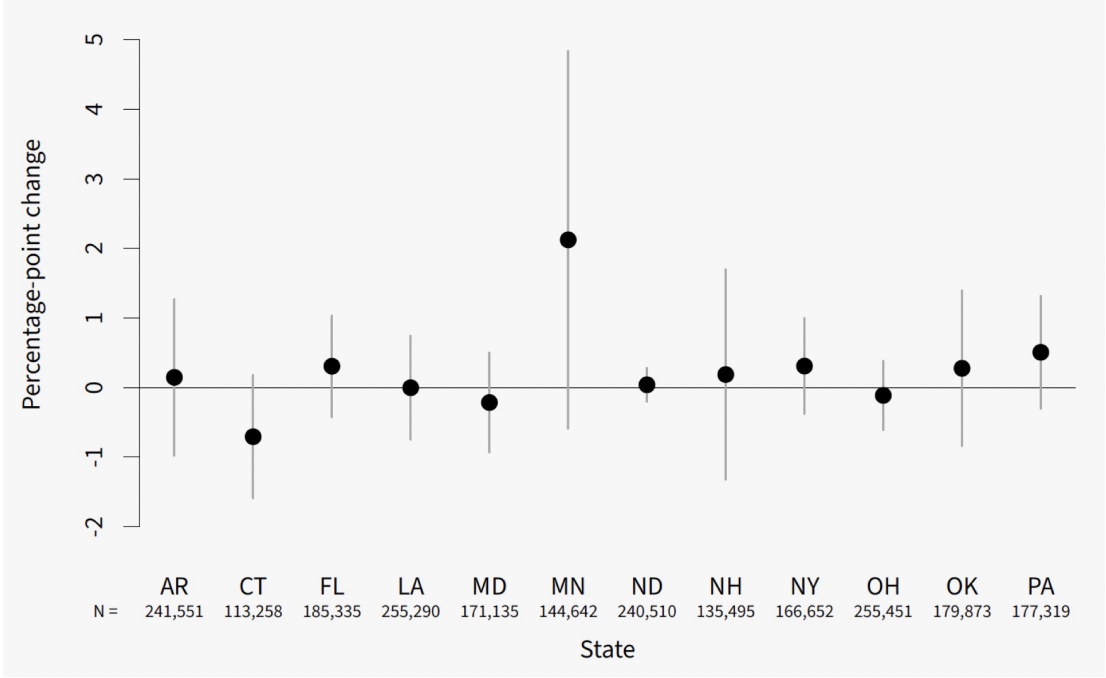
But, in a **policy trial emulation**,

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

- Well-reported policy trial emulations will discuss statistical/causal assumptions and their plausibility

State Medical Cannabis Laws & Opioid Rx

McGINTY ET AL. 2023, *ANNALS OF INTERNAL MEDICINE*



Change in proportion of chronic noncancer pain patients receiving any opioid prescription, per month, attributable to state medical cannabis law in first 3 years of implementation.

Augmented synthetic controls estimates of ATT averaged over post-law period.

Perelman School of Medicine University of Pennsylvania

Wrap-Up: Good Study Design is Critical

Policy trial emulation provides a framework for thinking about good policy evaluation study design

- Think about the trial you would run if you could, then try to get as close as possible.

Closer alignment between hypothetical target trial and non-experimental analogue improves communication

- Clearly articulate similarities & differences across all 7 components
- Helps readers understand design better
- Helps readers calibrate confidence in results

Wrap-Up: Good Study Design is not Magic

Using the policy trial emulation framework does not guarantee quality!

- An emulated trial is not a trial
- (Not even randomized trials guarantee quality)

There will always be trade-offs. The goal is to make reasonable decisions for your study that are guided by strong design principles.

Wrap-Up: Statistical tools alone are often insufficient

Methods rely on assumptions.

- Strong study design improves confidence that assumptions are satisfied!

Quantitative analysis without domain knowledge is dangerous.

Rigorous qualitative research is crucial to understanding the policy of interest and its effects (or lack thereof).

- Adds context that can't be gleaned from data
- Policies affect people. Understanding how improves your research.

Working together with experts across disciplines (including qualitative!) is challenging, but fun and rewarding.

doi.org/nmmw

nickseewald.com/chibeAnalytics2025