State-Level Opioid Policy Analyses: Moving Beyond the Classic Diff-in-Diff Model

Beth Ann Griffin, Megan Schuler, Elizabeth Stuart
Acknowledgments & Disclosures

• This work has been supported by the National Institute on Drug Abuse of the NIH through P50-DA046351
• No conflicts of interest
• Contributors to the OPTIC team:
  • Bradley Stein
  • Rosalie Liccardo Pacula
  • Rosanna Smart
  • Sara Heins
  • Courtney Kase
  • Seema Choksy Pessar
  • Mark Sorbero
  • Stephen Patrick
  • Elizabeth Stone
  • Joseph Pane
  • Elizabeth McNeer
  • David Powell
  • Mary Vaiana
  • Hilary Peterson
Roadmap for Day 2

Our focus today will be:

• Key methodological challenges for state-level policy evaluations
• Potential solutions
Genesis of this work

• As part of OPTIC, we conducted a literature review of opioid-related policy evaluation studies published in 2005-2018

• This literature review, as well as our applied work, informed a recent paper providing an overview of methodological challenges/issues encountered in opioid policy research
Today we will discuss the following key challenges as well as potential solutions for each

1) Obtaining high-quality opioid policy data
2) Obtaining high-quality opioid outcome data
3) Appropriately operationalizing and specifying opioid policies
4) Addressing confounding due to systematic differences between policy and non-policy states
5) Disentangling effects of concurrent policies
6) Identifying heterogeneous policy effects across time, states, and population subgroups
7) Overcoming limited statistical power to detect policy effects afforded by commonly-used methods
Challenge 1: Obtaining high-quality opioid policy data
Challenge 1: Obtaining high-quality opioid policy data

• One major challenge in opioid policy research is obtaining reliable, comprehensive opioid policy data

• Assembling accurate policy datasets across states is not necessarily trivial
  • May require detailed review of statutory and agency documents
  • “Enactment” vs. “implementation” date: may differ, since policies passed into law are not always implemented immediately
  • Policy happens within an existing legal framework, which is often relevant for understanding the policy of interest

• Growing field of legal epidemiology serves to bridge this gap between legal research and public health
Solution 1A: **Utilizing existing, publicly-available national databases**

- **OPTIC-Vetted Policy Data Sets** at RAND and USC
- **The Policy Surveillance Program** at Temple Law
- **National Conference of State Legislatures**: systematically characterized many opioid-related policies

State prescription drug monitoring programs (PDMP) policies:
- **Prescription Drug Abuse Policy System** (PDAPS)
- **National Alliance for Model State Drug Laws** (NAMSDL)
- Brandeis’ **PDMP Training & Technical Assistance Center**

- **CDC Public Health Law Program**: other state policies, including prescribing regulations

OPTIC-Vetted Policy Data Sets

One of OPTIC’s goals is to improve policymaking and the effectiveness of opioid policies by enhancing opioid policy science. As part of that effort, OPTIC is posting OPTIC-Vetted Policy Data Sets. The purpose of these data is to help the research and policy communities consistently define policy data variables, thereby improving policy data use, replicability and reproducibility of study findings.

The policy data areas currently available:

- State laws and regulations for prescribing Naloxone in conjunction with additional medications;
- State laws and regulations authorizing third party prescribing and lay administration of Naloxone;
- State laws providing protection from criminal sanctions to overdose victims or witnesses who seek emergency services;
- State laws and regulations governing the operation and use of prescription drug monitoring programs; and
- State laws and regulations sale and consumption of marijuana for therapeutic purposes.

How can OPTIC-Vetted Policy Data Sets contribute to better opioid policy science?

When researchers are assessing and comparing the effects of a given policy at the state level, it is essential to know the details of how individual policies are implemented. For example, prescription drug monitoring programs (PDMP) as they appear in state laws might look identical, but the functionality of the programs can differ by state. Our OPTIC-vetted data showcases those important distinctions for PDMPs and other opioid-related policies.

To create the OPTIC-Vetted Policy Data Sets, OPTIC investigators consulted with public health lawyers in synthesizing data sets and coordinating key elements of each policy that are important for influencing opioid outcomes based on theory and evidence. These variables will provide better guidance about which elements of laws are most important to adopt. For example, for PDMP to be effective, state laws must specify that doctors MUST access the PDMP before prescribing. However, even if the state law clearly states this requirement, if the PDMP is not updated daily but is updated weekly instead, it cannot be as effective.

Please cite the use of these data as "RAND-USC Schaeffer Opioid Policy Tools and Information Center (YEAR). OPTIC-Vetted (DATASET TITLE). Obtained from https://www.rand.org/health-care/centers/optic/resources/datasets.html on [DOWNLOAD DATE]."

Download the Datasets

Co-prescribing Naloxone Policy Data (.zip file)

Related Publication

Legal Requirements and Recommendations to Prescribe Naloxone
Roberta Lee-Huifigl, Samantha Chesney, Rosanna Smart
We conducted a systematic legal review of state laws that mandate or recommend that healthcare providers prescribe naloxone to patients with opioid prescriptions.
Health is impacted by a wide-ranging array of laws and policies. Search by public health topic or alphabetically to find legal maps and begin exploring the law.

Alcohol, Tobacco and Other Drugs

- Electronic Cigarette Laws
- Good Samaritan Overdose Prevention Laws
- Laws Authorizing Involuntary Commitment for Substance Use
- Local Medical Marijuana Laws in Washington State
- Local Recreational Marijuana Laws in Washington State
- Medical Marijuana Caregiver Rules
- Medical Marijuana Dispensaries
- Medical Marijuana Laws for Patients
- Medical Marijuana Product Safety Laws
- Morphine Equivalent Daily Dose (MEDD) Policies
- Naloxone Overdose Prevention Laws
- Recreational Marijuana Laws
- State Laws and Other Regulatory Policies Related to Pain Care
- Syringe Distribution Laws
- Syringe Possession Laws
- Syringe Service Program Laws
- Tobacco Pricing Strategies
Solution 1A: Utilizing existing, publicly-available national databases

• Important considerations when using these data, particularly across different databases

• Different entities may have different definitions/classifications of policies
  • e.g., the minimum regulatory threshold comprising an “active PDMP”

• Different databases may report different policy enactment and implementation dates

• Ideally, policy data are assembled and interpreted in conjunction with a public health lawyer

Solution 1B: *Using mixed-methods approaches*

- Qualitative interviews with key legislative and administrative leaders can be used to collect more detailed policy data

- McGinty and colleagues: conducted semi-structured interviews with key stakeholders regarding pill mill laws in Ohio and Tennessee
  

- Surveys of policymakers can also provide key insights
  
  
  
  
Challenge 2: Obtaining high-quality opioid outcome data
Challenge 2: Obtaining high-quality opioid outcome data

- Limited access to nationally-representative and reliable measures of important outcomes
  - opioid prescribing, prescription and illicit opioid misuse, opioid-related overdoses, OUD treatment utilization

- Opioid-related mortality is a commonly examined outcome
  - Profound variation in data quality and reporting practices
  - Guidance from CDC:
    - Data quality varies across states; quality of state data has changed and improved over time
    - Recommends that any sub-state analysis use overdose deaths, not opioid-related deaths

Solution 2A: Examining a set of outcomes to address limitations of individual measures

• The use of a set of opioid-related outcomes, rather than a single outcome, can effectively increase robustness of study findings
  • The opioid epidemic is characterized by numerous dimensions:
    • e.g., prescription opioid misuse, illicit opioid use, OUD, risky prescribing, OUD treatment utilization, among others.
  • Comprehensively assessing a policy’s impact will be improved by examining its effects across multiple dimensions

• Example: Validate primary outcome, self-reported treatment, using administrative treatment admission data
Solution 2B: Use corrected opioid-related fatal overdoses to address significant missingness

• Most prior estimates of opioid-related mortality rates/counts are understated because of high levels of missingness on the specific drugs that lead to the death.

• Ruhm (2018) provides corrected national estimates of opioid and heroin/synthetic opioid-involved counts and mortality rates, as well as how they have changed from 1999 to 2015.
  • Exploit characteristics associated with opioid-related overdoses to impute likelihood of an “unspecified” drug overdose truly being opioid-related.
  • After imputation, revised rates of opioid-related fatal overdoses were 20–35% higher for all years 1999–2015 compared to the original reported rates.


Solution 2C: Bolstering confidence in study findings by conducting sensitivity analyses

• One sensitivity analysis approach = consider an outcome(s) that is expected to be unaffected by the policy of interest
  • Null results can bolster support for significant effects for outcomes expected to be impacted by the policy

Examples of placebo tests include:
  • Impact of PDMPs on opioid prescribing: prescribing rates of statins, beta-blockers
  • Impact of naloxone programs on opioid-related deaths: motor vehicle-related deaths

• Alternatively, sensitivity analyses can be conducted in which the date of policy enactment is randomly varied
  • Null results can bolster support for significant results in correctly-specified analyses
Challenge 3: Appropriately operationalizing and specifying opioid policies
Challenge 3: Appropriately operationalizing and specifying opioid policies

• Currently, tremendous heterogeneity in terms of how studies classify and operationalize opioid policies
  • No PDMP vs. Any PDMP
  • No PDMP, mandatory access PDMP, vs. non-mandatory PDMP
  • Classify PDMPs by summing distinct components
  • Compare PDMPs deemed “robust” vs. non-robust

• Policy classifications should be theoretically / empirically motivated, rather than arbitrary
Empirical clustering methods such as latent class analysis (LCA) and latent transition analysis (LTA) can be used to characterize how policies differ across states and over time.

- LCA can be used to identify groupings of: (1) policies that are often adopted together (2) groupings of states with similar likelihoods of adopting multiple policy characteristics

- LTA can empirically identify how state opioid policies change over time
Solution 3A: **Empirically identifying similar opioid policies and policy trajectories**

- LTA to identify classes of states with similar PDMP characteristics
- Intervals 1, 2:
  1) **Weak** = states with no PDMP
  2) **Reactive** = fairly basic PDMPs with infrequent data updates
  3) **Proactive** = proactively provide unsolicited updates to users, more open access to law enforcement, and require more frequent data reporting

---

Solution 3A: Empirically identifying similar opioid policies and policy trajectories

• Interval 3:
  1) **Reactive** = fairly basic PDMPs with infrequent data updates
  2) **Proactive** = proactively provide unsolicited updates to users, more open access to law enforcement, and require more frequent data reporting
  3) **Cooperative** = similar to proactive, but also shared data with other states

• By using LTA, able to characterize how states transitioned to more robust PDMPs over time and to identify the combinations of policy provisions that characterized different types of state PDMP policy models
Challenge 4: Confounding due to systematic differences between policy and non-policy states
Challenge 4: Confounding due to systematic differences between policy and non-policy states

• Overwhelmingly, opioid policy studies are observational – subject to confounding concerns

• Ignoring pre-existing differences between policy and non-policy states may potentially bias true policy effect
  • e.g., states with higher rates of opioid-related mortality may be more likely to enact policies aimed at curbing mortality

• In our prior literature review, we found multi-state evaluation studies often did not extensively consider potential bias due to confounding
Solution 4: Using well-established statistical methods to address confounding

- Numerous statistical approaches can address selection bias concerns:
  - Standard regression adjustment
  - Propensity score methods
  - Synthetic control methods
  - Doubly robust DID estimators

- Unique considerations in the longitudinal / DID context
  - Daw and Hatfield emphasize variables associated with both policy and outcome trend are key confounders in DID studies, since identifying assumption relates to outcome trends
  - Conversely, variables strictly associated with both policy adoption and baseline outcome level do not bias DID studies
Standard regression adjustment

• DID, Detrended and AR models shown on day 1 are already doing this

  DID model: \( g(Y_{it}) = \gamma \cdot Z_{it} + \beta \cdot X_{it} + \rho_s + \sigma_t + \epsilon_{it} \)

  AR model: \( g(Y_{it}) = \gamma \cdot (Z_{it} - Z_{i,t-1}) + \beta \cdot X_{it} + \eta \cdot Y_{it-1} + \sigma_t + \epsilon_{it} \)

• In addition to controlling for policy effects, these models control for confounding via the use of fixed effects, \( X_{it} \), and for the AR model – the lagged value of the outcomes (\( Y_{it-1} \))

• Determining which variables should go into \( X_{it} \) is always a challenge
  • Ideally anything that is a true confounder should be included
  • Need to be careful of overfitting
Propensity score (PS) adjustment

• PS is defined to be the conditional probability of enacting a policy, given a fixed set of observed covariates (e.g., \( \text{Pr}(Z_{it}=1 \mid X_{it-1}) \))
  • Can be used to create balance between policy states that do and do not enact a policy via matching, weighting, or stratifying

• Multiple methods proposed to deal with time-varying confounding such as occurs in state-level policy evaluations (i.e., marginal structural models [MSM] + inverse probability of treatment weighting [IPTW])
  • While common in studies of individual-level treatments, rarely used for policy evaluation
  • An exception: McGinty et al. 2020 used MSM to estimate effects of Maryland’s Medicaid initiative to integrate physical health care into specialty mental health programs
  • Limited sample size inherent to state policy evaluations (i.e., max 50 states) makes estimation of high-quality ITPW for MSM challenging in years with few policy states
  • Methodological work needed regarding constraints of small sample sizes
Synthetic control methods (SCM)

• Originally developed for cases where there is only one state enacting a new policy by Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010, 2013)

• “Synthetic control” method reweights comparison states to make the pre-law trends as similar as possible to the state with the policy change

• Intuitively, the idea is that we might believe that their trends would have continued in similar ways if they were similar on pre-intervention factors
Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.
Synthetic control methods (SCM)

• **Benefits:**
  • Transparency: data-driven selection of comparison units
    • Can do this using just the pre-period data
    • Separation of “design” and “analysis” (Rubin, 2001)
  • Safeguard against extrapolation ensures similarity of intervention and comparison units

• **Challenges:**
  • Often difficult to obtain good “synthetic control” and other performance problems
Augmented synthetic control methods (SCM)

- Augmented synthetic controls, which combines synthetic controls + outcome model (Ben-Michael, Feller, and Rothstein, 2020), allows for improved performance and easier inference.

- Also demonstrates interesting connection between synthetic controls and propensity score weighting.

- Recently, the methods have been extended to handle staggered adoption and more than one treated unit (Ben-Michael, Feller & Rothstein, 2021)!

Doubly robust difference-in-difference estimators

• More on this later...
Simulation: Assessing the impact of confounding
Current objective: Examine performance when we have confounding

• Original simulation focused on “best case” scenario
  • “Treated” states that enacted the new policy were randomly selected so no systematic difference between treated and control states

• We have expanded our original simulations to consider
  • Relative performance of different methods in the presence of confounding (i.e., meaningful differences between states that do and do not enact the policy on key confounders)
Policy heterogeneity: Type of PDMP mandate

Type of mandate: Does the mandate apply to all prescribers and at least all initial opioid prescriptions?

- Yes (25)
- No (16)
- No mandate (10)

Data source = National Alliance for Model State Drug Laws and the PDMP Training and Technical Assistance Center at Brandeis University
Image credit = Pew Charitable Trust
Heterogeneity in state opioid overdose rates

Data source = CDC WONDER.
Image credit = NIDA
Build potential confounding into the simulations using the following four steps:

1. Select treated states
2. Simulate outcomes
3. Estimate policy effect
4. Compare models
Step 1: Selecting the treated states that enact the policy

• Data Structure: We have repeated state-level opioid-mortality data from all 50 states between 1999-2016 ($Y_{it} = \text{the outcome for state } i \text{ in year } t$)

• Need to generate treatment indicators ($Z_{it}$) for if state $i$ will enact policy in year $t$
  • Once $Z_{it}$ is turned on, it remains on for the rest of the time period

• Confounding generated by linear and nonlinear treatment selection models:

  \[ \text{Linear: } \text{Logit}(Pr(Z_{it} = 1)) = b_0 + b_1 C_{it} + b_2 X_{it} \]

  \[ \text{Nonlinear: } \text{Logit}(Pr(Z_{it} = 1)) = b_0 + b_1 C_{it} + b_2 X_{it} + b_3 C_{it}^2 + b_4 X_{it}^2 + b_5 (C_{it} \times X_{it}) \]

  where $X_{it}$ is state-level covariate and $C_{it}$ is some measure of prior outcome values
Step 2: Generating the outcomes in the presence of confounding

• For treated states, generate synthetic observation $Y_{it}^*$ with policy effect and confounding

$$Y_{it}^* = Y_{it} + a_1 C_{it} + a_2 X_{it} + \gamma^* Z_{it}$$

Original outcome  Confounding  Policy effect

• For non-treated states, $Y_{it}^*$ is the original outcome + confounding

• Also consider nonlinear confounding=$(a_1 C_{it} + a_2 X_{it} + a_3 C_{it}^2 + a_4 X_{it}^2 + a_5(C_{it} \times X_{it}))$

Examined both:
1. Confounding by levels, $C_{it}$ = prior 3 year-average opioid-mortality rate
2. Confounding by trends, $C_{it}$ = change in rate over the prior 3 years
Consider performance over a range of confounding models derived from real policy data

Effect Size Difference in Opioid-Mortality Rates between Treated and Control States

Parameters for Linear Confounding By Trends

<table>
<thead>
<tr>
<th></th>
<th>$b_0$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$a_1$</th>
<th>$a_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>-5</td>
<td>0.03</td>
<td>0.1</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Moderate</td>
<td>-5</td>
<td>0.05</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Large</td>
<td>-5</td>
<td>0.1</td>
<td>0.1</td>
<td>0.5</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Step 3: Estimating the policy effects

• We selected four models from our prior simulations
  
  • Linear and negative binomial 2-way fixed effects (FE) model (classic DID)
    \[ g(Y_{it}) = \gamma \cdot Z_{it} + \beta \cdot X_{it} + \rho_s + \sigma_t + \epsilon_{it} \]
  
  • Linear and negative binomial autoregressive (AR) model
    \[ g(Y_{it}) = \gamma \cdot (Z_{it} - Z_{i,t-1}) + \beta \cdot X_{it} + \eta \cdot Y_{it-1} + \sigma_t + \epsilon_{it} \]
  
• Also use augmented synthetic control methods (SCM) extended to handle staggered adoption and more than one treated unit (Ben-Michael, Feller & Rothstein, 2021)
Results when confounding is linear and by trends
Bias for all methods increases with greater confounding
Variance also increases as confounding gets larger.
Root Mean Square Error (RMSE) generally shows linear AR model as doing better

![Diagram showing RMSE for different models with varying levels of confounding]
Key recommendations for practice (so far)

• All methods are impacted by increasing levels of confounding
• Confounding by trends is always more challenging for these models
• AR linear model continues to be a competitive model due to increase in precision (lowest variance) across all scenarios
• Key limitations of current efforts:
  • Our data generating models might not capture true confounding that happens in real world, though we are working hard to derive realistic scenarios
  • Need to add in additional approaches (e.g., doubly robust DID; Goodman-Bacon decomposition approach; propensity score adjustment)
Challenge 5: Disentangling effects of concurrent policies
Challenge 5: Disentangling effects of concurrent policies

- We know states are not enacting policies in a vacuum
- Prior work identified correlation in types of policies adopted in various places
Challenge 5: Disentangling effects of concurrent policies

• In evaluation studies, typically there is one “primary” policy of interest
  • Goal: isolate the impact of this policy from that of any co-occurring policies

• Ignoring co-occurring policies in analyses results in model misspecification
  • Potentially biases effect estimate for the primary policy

• In our opioid policy literature review:
  • 66% did not address any co-occurring policies in the analytic design
  • 8% adjusted for a single co-occurring policy in the primary analyses
  • 20% accounted for 2+ co-occurring policies in the primary analyses
  • 5% accounted for co-occurring policies in a secondary or sensitivity analyses
Solution 5: Rigorously and transparently adjusting for co-occurring policies

• Think critically, be transparent about policies that co-occur with (and may confound) the primary policy of interest
  • May require a detailed review of the time-varying policy landscape

• Approaches:
  • Co-occurring policies may be controlled for as regression covariates
  • Restrict sample or study period
    • Restrict to subset of states that do not have any additional opioid policies during the study period
    • Define study period to end prior to adoption of potential confounding policy

• Also, co-occurring policies may yield synergistic policy effects
  • Additive or interactive effects rarely examined
Simulation: Impact of concurrently enacted policies
We expanded our simulation tool to include scenarios where two policies are concurrently enacted.

- Simulate concurrently enacted policies and their effects in real data
- Estimate policy effects using statistical models
- Compare model performance

5,000 trials for each of multiple conditions: N = 50 states

Study performance of 4 types of models (2-way fixed effects vs. autoregressive; linear vs. negative binomial)

Five performance measures: Bias, Variance, RMSE, Type I error, Type S error, Power
Simulation design: Concurrently enacted policies

1. Randomly sample a set of states that will enact both policies (n.trt = 5 and 30)
2. Randomly select 2 enactment dates assuming a fixed average length of time between them (yrs.apart = 0, 3, 6, 9 years)
3. Consider cases where policies occurred in fixed order (primary always first) versus random order
4. Assume policies have an additive effect on the outcome ($Y_{it}^* = Y_{it} + \gamma_1 Z_{1it} + \gamma_2 Z_{2it}$)
Model specifications considered

• Four key model types:

<table>
<thead>
<tr>
<th>GLM</th>
<th>Regression Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>2-way fixed effects</td>
</tr>
<tr>
<td>Linear</td>
<td>Autoregressive</td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>2-way fixed effects</td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>Autoregressive</td>
</tr>
</tbody>
</table>

• Correctly specified models: Control for indicators of both policies and estimate individual policy effects as well as joint effect of the 2 policies

• Misspecified models: Only control for one of the policies
Findings for linear autoregressive model when policies are ordered
Misspecified models: Relative bias is substantially worse in the case with ~0-1 years between enactment dates.
Correctly specified models: Relative bias much lower than in misspecified model and decreases as length of time between policy enactment dates increases.
Correctly specified models do better when policies do not occur in fixed order and as true treatment effect size increases for the policies.
Estimates of policy effects should consider confounding from co-occurring policies, particularly when enacted close together (0-1 years apart). If the co-occurring policy is a confounder, leaving it out will introduce unobserved confounding bias in the estimate of the primary policy. Bias ~ 75% if policies enacted in rapid succession (0-1 years apart). Less of a problem as length of time between enactment dates gets larger. Optimal model controls for both confounders: Bias < 5% if policies not ordered (regardless of length of time between the 2 policies). Bias = 10-25% if primary policy always comes first & within 0-1 years apart (harder for model to disentangle policy effects).
Propose the use of simple data checks prior to policy evaluations

• Compute length of time between enactment dates for each state
• Examine distribution to ensure there is enough time between enactment dates
  • With AR models – need > 3 years
  • With classic DID – need > 6 years

Final notes

- Key limitations of current work –
  - Still haven’t studied model performance when we have interaction effects between the concurrently implemented policies
  - Nor have we explored the potential impacts of having states included that enacted only 1 of the 2 policies

- Forthcoming R library (OPTIC.simRM): easy to implement our simulations with any type of repeated measures data
Challenge 6: Identifying heterogeneous policy effects across time & states
Challenge 6: Identifying heterogeneous policy effects across time & states

• In practice, many evaluation studies are conducted in the context of multiple states adopting the policy of interest at varying times ("staggered adoption")

• Often, no specific analytic considerations are made to address the variation in policy timing

• In this setting, a DID estimate can be obtained by using the 2-way FE model, which operationalizes the policy variable as a state-specific, time-varying indicator
Variation in policy timing: 3 groups

Image credit: Andrew Goodman-Bacon
How does the DID estimator generalize to this context?

• In a 2 group, multiple time period context, DID estimator makes 2 comparisons:
  1) difference between average outcome during the post-period and pre-period for the treated group
  2) difference between average outcome during the post-period and pre-period for the comparison group

• Often, no specific analytic considerations are made to address the variation in policy timing

• When policy timing differs across states, the following become less clear:
  • when is “pre-policy” and “post-policy”?
  • what is the control group?
Solution 6a: Goodman-Bacon decomposition

• Shows that, when policy timing varies, 2-way FE DID model produces an ATT estimate that can be decomposed into a weighted average of all possible 2-group, 2-time period DID estimators that can be constructed

• More specifically, DID estimate is a weighted average of:
  1) Compare early adopters (“treated”) to later adopters during periods when they are not yet treated (“control”)
  2) Compare later adopters (“treated”) to early adopters during period when the early adopters are continuously treated (“control”)
  3) Compare different timing groups (e.g., early adopters or later adopters) to the never-treated group, if there is one
Variation in policy timing: visualizing component 2x2 DID comparisons

Image credit: Andrew Goodman-Bacon
Variation in policy timing: visualizing component 2x2 DID comparisons

B. Late Group vs. Untreated Group

Units of $y$

$y^l_i$

$y^r_i$

PRE(l)

POST(l)

Time

$t^*_i$

Image credit: Andrew Goodman-Bacon
Variation in policy timing: visualizing component 2x2 DID comparisons
Variation in policy timing: visualizing component 2x2 DID comparisons
Solution 6a: Goodman-Bacon decomposition

• DID estimate can be decomposed into a weighted average of all possible 2-group, 2-time period (2x2) DID estimators

• Goodman-Bacon paper derives the nature of the weights

• Factors that influence weights:
  1) Larger size: bigger group
  2) Larger variance: treated closer to middle of the panel

• If some weights are large, a few 2x2 comparisons may dominate the overall DID estimate
Solution 6a: Goodman-Bacon decomposition

STATA package

• Developed package that graphs DID decomposition

• Allows users to assess:
  1) Which 2x2 DIDs matter the most? (determine relative weights)
  2) How different are the 2x2 DIDs? (effect heterogeneity)
Goodman-Bacon decomposition
Goodman-Bacon paper: Implication 1

• Shows that, when policy timing varies, 2-way FE DID model yields biased treatment effect estimate when treatment effects are heterogeneous across states
  • When treatment effects vary across states, OLS over-weights units with more variance in treatment status in order to achieve a more precise estimate of the treatment effect
  • States that are treated near the middle of the study period receive relatively more weight; states treated at the beginning or end of the study period receive relatively less weight
  • Sometimes weights may be negative

• Proposed solution: Alternative estimator that intentionally specifies weights
Goodman-Bacon paper: Implication 2

• Shows that DID estimates are biased when treatment effects change over time within state (e.g., policy effect grows or decays over time)

• Intuitively, this occurs because already treated units serve as controls in some of the 2x2 DIDs comprising the weighted average

• When treatment effects are not constant over time, using already treated units as controls necessarily biases estimates of the treatment effect (by introducing a term representing the change in the treatment effect on the already treated units)

• Proposed solution: 2-way FE estimators are not appropriate - alternative approaches (e.g., event study estimation) should be used
Solution 6b: Use methods designed for staggered adoption

• Another approach: Explicitly do a series of DID type designs, one at each of the policy adoption dates (“policy cohorts”)
• Avoids inadvertent conditioning on post-policy data
• Allows explicit averaging of the cohort specific effects
• Facilitates transparency regarding the comparison states/time points, and design elements to equate the policy and comparison states in the pre-periods
• Variations on this theme: Callaway and Sant’Anna (2021), Sun and Abraham (2020), Ben-Michael et al. (2021), Sant’Anna and Zhou (2020)
Additional notes on Callaway & Sant’Anna

• Approach yields a series of group-time Average Treatment Effect on the Treated (ATT) estimates
  • e.g., the estimated effect for the 2016 cohort, in 2017 (1 year out)
  • e.g., the estimated effect for the 2018 cohort, in 2020 (2 years out)

• These can then be averaged explicitly; often averaged across time and treated states (i.e., weight each cohort effect by the # of treated states)

• Confounding can be dealt with using propensity score weights
  • Estimated separately for each cohort, using only baseline data (before that cohort’s policy implementation date)

• Inference done using bootstrap
Additional considerations

• Can use either “not yet treated” or “never treated” states as the comparison pool for each cohort

• Need to have some never-treated states in the dataset
  • Otherwise, ATT not estimable for the final cohort

• Need to code a variable with the implementation year for each treated state (to identify treated cohorts)
Look at effect of “must access” PDMPs on prescription opioid related overdose death rates

5 cohorts (2012-2016), 3-5 treated states per cohort

Results shown adjust for unemployment rate; unadjusted results very similar
6a: Goodman-Bacon Decomposition

• Overall diff-in-diff effect estimate: 0.64 (95% CI: -0.21, 1.48)
• Decomposition:
  • Treated groups: 10%
  • Never-treated vs. treated groups: 89%
  • Within treated groups: 0%

• Stata code:
  • `xtset state_fips year`
  • `xtreg OpioidRate AnyPDMP pct_unemployed i.year, fe cluster(state_fips)`
  • `bacondecomp OpioidRate AnyPDMP pct_unemployed year, stub(Bacon_) cluster(state_fips)`
So what are the individual cohort effects (6b)?

Use Callaway and Sant’Anna (2020) approach to estimate effects separately for each timing cohort, and then aggregate

*did* package in R
Average effects by cohort and length of exposure
Average effects by time period and group
Conclusions

• Treatment and comparison states generally have similar outcomes during the pre-policy periods

• Overall effect: -0.98 (95% CI: -2.79, 0.83)
  • 2012 cohort shows largest effects
Needed Data Steps

# create version of policy variable that maps each state to its implementation year
states<-unique(data$state)
data$treat.PDMP=rep(0,nrow(data))
new.dat=c()
for(ss in 1:length(states)) {
  s.dat=data[data$state==states[ss],]
enact.yr=min(s.dat$year[s.dat$AnyPDMP==1])
s.dat$treat.PDMP=enact.yr
new.dat=rbind(new.dat,s.dat)}
data=new.dat

# set control states to 0
data$treat.PDMP[data$treat.PDMP=="Inf"]=0

## R code to run Callaway & Sant’Anna

library(did)
set.seed(1814)

# estimate group-time average treatment effects with covariates
mw.attgt <- att_gt(yname = "OpioidRate",
  gname = "treat.PDMP",
  idname = "state_fips",
  tname = "year",
  xformla = ~pct_unemployed,
  data = data,)

# summarize the results
summary(mw.attgt, na.rm=true)

# plot the results
ggdid(mw.attgt, ylim = c(-20,20))

# aggregate the group-time average treatment effects
agg gs <- aggte(mw.attgt, type="group")
summary(agg.gs)
ggdid(agg.gs)

# simple aggregation - weighted average of all group-time average tx effects with weights proportional to group size.
agg.simple <- aggte(mw.attgt, type="simple")
summary(agg.simple)

## calendar time effects - average effect of participating in the tx in a particular time period for all groups that participated in the treated at that time period
agg.ct <- aggte(mw.attgt, type="calendar")
summary(agg.ct)
ggdid(agg.ct)

## average effect by group
agg.gs <- aggte(mw.attgt, type="group")
summary(agg.gs)
ggdid(agg.gs)

# simple aggregation - weighted average of all group-time average tx effects with weights proportional to group size.
agg.simple <- aggte(mw.attgt, type="simple")
summary(agg.simple)
Challenge 7: Overcoming limited statistical power to detect policy effects afforded by commonly-used methods
Challenge 7: Overcoming limited statistical power to detect policy effects afforded by commonly-used methods

- As shown on day 1 – There are notable concerns that commonly-used statistical methods are underpowered
  - Power ranged from 5 to 35% for classic DID model, far lower than the 80% standard
  - Most methods can’t detect potential actionable state-level policy effects (e.g., 5-7%)

- Power analyses rarely conducted for DID & observational studies (Black et al., 2019)
  - Review of large set of economics articles - using DID and other statistical methods commonly used in the opioid policy literature - estimated median statistical power to be 18% (Ioannidis et al., 2017)
Challenge 7: Overcoming limited statistical power to detect policy effects afforded by commonly-used methods

• As discussed in Schell, Griffin, and Morral (2018), scientific fields built on studies with very low power that underestimate the statistical uncertainty in effect estimates will have a large proportion of findings that are inaccurate, inflated, and even often in the wrong direction.

• There is a pressing need for the field to develop more robust and powerful methods that can be used to help guide state-policy given the constrained sample sizes and effects sizes.
Solution 7: Utilizing Bayesian methods to improve robustness and better quantify uncertainty in estimates

• Recent application of Bayesian inference to gun policy research (Schell et al., 2020) highlights several important advantages of Bayesian methods in policy studies:
  • Bayesian methods can be used to estimate the probability that a specific policy is associated with an increase or a decrease in an outcome of interest.
    • e.g., gun policy work found evidence Child-Access Prevention laws associated with 97% probability that firearm deaths would decline within 6 years after enactment
    • This type of evidence helpful to policymakers facing yes/no decisions about which policies to enact
Solution 7: Utilizing Bayesian methods to improve robustness and better quantify uncertainty in estimates

• Bayesian inference with modestly informative priors mitigates limited power problems in frequentist models applied to state-level policy (Gelman and Tuerlinckx, 2000)

• Models also allow for a more detailed analysis of time-varying policy effects, as marginal effects can be calculated for each year after implementation
  • Parameters can be included that capture both the direct effect of the law on the outcome in that year as well as indirect effects arising from the presence of the law in prior years
Thanks!!

Beth Ann Griffin, Megan Schuler, Elizabeth Stuart