Evaluating the Effects of Regulatory Strengthening on Drinking Water Violations: A Synthetic Difference-in-Differences Approach

Chanheung Cho*
North Carolina State University

Younghyeon Jeon†
University of Florida

May 5, 2025

Abstract

We evaluate the impact of Wisconsin's 2018 regulatory initiative aimed at improving compliance with the Safe Drinking Water Act (SDWA). Using the synthetic difference-in-differences (SDID) method, we estimate the causal effect of the intervention on the number of SDWA violations among public water systems (PWSs). The analysis draws on 2014–2023 data from the EPA's Safe Drinking Water Information System and adjusts for state-level covariates including log real GDP, population growth, and site visit rates. We find that Wisconsin's intervention reduced the annual number of violations by approximately 965 relative to a synthetic control. Placebo tests confirm the robustness of the results. This study contributes to the environmental economics literature by providing new evidence on the effectiveness of strengthened state-level enforcement in the drinking water sector and highlights the usefulness of SDID methods for evaluating environmental policies.

JEL Codes: Q53; Q58; I18; C21
Keywords: Safe Drinking Water Act; Regulatory compliance; Drinking water violations; Synthetic difference-in-differences; Environmental policy evaluation

*Address: Partners Building II, Raleigh, NC 27695, US. E-mail: ccho5@ncsu.edu
†(Corresponding Author), Address: 104 Gainesvillee FL 32611, US. E-mail: yjeon1[at]ufl.edu

The authors thank participants at various conferences and seminars for useful comments and discussions.

1 Introduction

Access to safe and clean drinking water is a fundamental public health necessity, and the Safe Drinking Water Act (SDWA) has served as the cornerstone of drinking water protection in the United States since its passage in 1974 (U.S. Congress 2000). While the regulatory framework under SDWA has led to major improvements in drinking water quality, violations of drinking water standards remain common across public water systems (PWSs), particularly among smaller and under-resourced systems (Allaire et al. 2018). Violations may involve health-based breaches, such as exceeding maximum contaminant levels (MCLs) for dangerous contaminants, or non-health-based failures, such as monitoring, reporting, and public notification violations (EPA-SDWIS 2024). Persistent violations not only expose communities to public health risks but also erode trust in regulatory institutions and exacerbate environmental justice concerns, as lower-income and rural areas are often disproportionately affected (Konisky and Teodoro 2016).

Recent events have underscored the fragility of drinking water safety in the United States. Incidents such as the Flint water crisis and growing concerns about emerging contaminants have placed renewed emphasis on the effectiveness of regulatory oversight (Gray et al. 2017, Wang et al. 2022, Christensen et al. 2023). Against this backdrop, understanding the factors that influence regulatory compliance, and evaluating the effectiveness of interventions aimed at reducing violations, have become increasingly important research priorities in environmental economics and public policy.

This paper examines the causal impact of a major regulatory intervention implemented by the state of Wisconsin in 2018 to improve compliance with SDWA requirements. Facing persistent issues with monitoring, reporting, and treatment compliance across its PWSs, Wisconsin launched a comprehensive effort that combined strengthened enforcement, increased frequency of inspections and site visits, enhanced technical assistance, and greater coordination with federal authorities (Wisconsin Legislature 2018). Unlike most states, which maintained status quo compliance monitoring practices during the same period, Wisconsin's intervention provides a rare quasi-experimental opportunity to assess the effectiveness of intensified regulatory efforts on drinking water compliance outcomes (Greenstone and Gayer 2009).

Our primary research question is whether Wisconsin's 2018 regulatory initiative led to a reduction in the number of SDWA violations recorded among its public water systems, relative to a set of comparison states that did not implement similar interventions. To address this question, we apply the synthetic difference-in-differences (SDID) method proposed by Arkhangelsky et al. (2021), which combines the strengths of synthetic control (SC) and difference-in-differences (DID) approaches. The SDID estimator offers improved robustness against differential pre-trends and time-varying unobservables compared to conventional DID models, making it particularly well-suited for policy evaluation in environmental and public health contexts where interventions are staggered, localized, and potentially confounded by dynamic factors (Arkhangelsky et al. 2021).

We construct the synthetic control group from states that did not adopt major new drinking water regulations between 2014 and 2023, carefully excluding states that introduced significant interventions such as expanded lead testing mandates, Per- and Polyfluoroalkyl Substances (PFAS) regulations, or major enforcement reforms. Our dataset merges annual PWS-level violation data from the Environmental protection agency's (EPA) Safe Drinking Water Information System (EPA-SDWIS 2024) and EPA/State Drinking Water Dashboard (EPA 2025) with state-level covariates including log real Gross Domestic Product (GDP), population growth rates, and the proportion of systems receiving site visits, which serve as proxies for economic capacity, demographic pressure, and regulatory oversight intensity, respectively. By controlling for these confounders, we aim to isolate the causal effect of Wisconsin's policy changes from broader macroeconomic and regulatory trends.

The results show that Wisconsin experienced a substantial and sustained decline in the number of SDWA violations following the 2018 intervention. While Wisconsin initially exhibited higher violation counts compared to the control group, the gap narrowed after the policy change. The estimated average treatment effect on the treated (ATT) indicates that Wisconsin reduced its annual number of violations by approximately 965 relative to the synthetic control during the post-intervention period. These findings are robust to alternative donor pool specifications, placebo tests.

This study makes several contributions to the environmental economics and policy evaluation literatures. First, it provides new empirical evidence on the effectiveness of state-led regulatory strengthening in the drinking water sector, a domain that has received comparatively less attention than air pollution, energy policy, or climate regulation. Second, by applying the SDID method to the context of drinking water compliance, it advances the methodological toolkit available for evaluating environmental interventions under partial observability and non-random program adoption. Third, it informs ongoing policy debates about how decentralized governance structures—where both state and federal actors share responsibility for enforcement—can be leveraged to improve environmental and public health outcomes (Shimshack and Ward 2005, Grant and Grooms 2017). Finally, the findings have practical relevance for policymakers considering how to prioritize limited regulatory resources to achieve the greatest improvements in public water system performance.

The remainder of the paper proceeds as follows. Section 2 describes the background of this research. Section 3 outlines the empirical strategy and variable construction. Section 4 explains the data sources and descriptive statistics of the data. Section 5 presents the main results and robustness checks. Section 6 discusses policy implications and concludes.

2 Background

In February 2018, Wisconsin enacted a landmark piece of legislation aimed at mitigating drinking water contamination risks associated with lead service lines (River Network 2025). Known as 2017 Wisconsin Act 137, this statute was signed into law on February 21 and became effective the following day, February 22, 2018 (Wisconsin Legislature 2018). The Act established a legal and financial framework enabling public water utilities and local governments to offer financial assistance to property owners for the replacement of lead-containing customer-side water service lines (Wisconsin Legislature 2018). As such, the policy represents a major institutional effort to address a well-documented and persistent source of SDWA violations and potential health risks, particularly in aging urban infrastructure.

This legislative action is an important contextual element of Wisconsin's broader 2018 regulatory push to reduce violations among PWSs. It complements administrative reforms and enhanced enforcement by tackling one of the root causes of health-based violations: the continued presence of legacy lead infrastructure. Act 137 provides a statutory basis for local ordinances requiring property owners to replace lead lines and authorizes municipalities to finance such replacements through grants or loans facilitated by the water utility (Wisconsin Legislature 2018). Notably, it also allows for loan repayments to be collected as special charges on property tax bills, reducing the transaction costs of program implementation and encouraging uptake.

From a policy design perspective, Act 137 reflects a coordinated governance model: political subdivisions (cities, villages, towns) are granted authority to implement replacement mandates, while water utilities can use customer charges within the same jurisdiction to fund the assistance (Wisconsin Legislature 2018). The Public Service Commission of Wisconsin is tasked with reviewing and approving utility-led financial assistance programs to ensure consistency, equity, and consumer protection. The legislation stipulates safeguards, such as uniformity of grants or loans within customer classes, and limits on grant generosity (no more than 50% of customer-side replacement costs), preserving fiscal sustainability and fairness.

The legislative motivation behind Act 137 lies in the widespread acknowledgment of the public health risks posed by lead in drinking water. Numerous studies and high-profile events—most notably the Flint water crisis—highlight the dangers of even low levels of lead exposure, especially for children (Jakubowski 2011, Hanna-Attisha et al. 2016). In Wisconsin, many PWSs still operated with partial lead service lines (customer-side only), meaning that even if the utility-owned portion had been replaced, the customer's pipe could continue leaching lead into tap water. This structural fragmentation presented a regulatory blind spot and a recurring source of health-based SDWA violations.

By facilitating complete lead service line replacement, Act 137 aimed to resolve a major compliance barrier for utilities and property owners alike. Health-based violations, including exceedances of lead and copper action levels, are among the most serious under the SDWA, often resulting in significant enforcement actions and reputational damage (Allaire et al. 2018). Prior to Act 137, utilities were limited in their ability to offer financial help for private-side replacements, leading to delays and inconsistent remediation. The new law empowered utilities to coordinate full-line replacements, thereby removing one of the most persistent sources of noncompliance.

In economic terms, the policy can be understood as a mechanism to internalize the externalities of lead exposure by correcting a classic coordination failure: property owners lacked incentives or resources to replace their lines independently, while utilities faced regulatory and financial barriers to assisting. Act 137 addressed this by creating a legal and fiscal pathway for cooperation, increasing the likelihood of full lead service line replacement and thus reducing SDWA violations.

The passage of Act 137 should be seen as part of Wisconsin's comprehensive 2018 regulatory initiative to enhance PWS compliance with the SDWA. While the broader initiative included stepped-up enforcement, expanded technical assistance, and more frequent site visits, Act 137 provided the legislative muscle to eliminate a specific infrastructure-based cause of violations. Importantly, the law did not impose a top-down mandate from the state level, but rather empowered local governments to act, consistent with Wisconsin's decentralized governance structure.

In combination with the administrative reforms documented elsewhere in this paper, Act 137 bolstered the state's capacity to reduce violations in a sustainable and equitable manner. The expected outcomes included lower rates of lead-related MCL exceedances, improved public trust in drinking water systems, and enhanced ability of water utilities to fulfill their SDWA obligations. Because lead violations tend to be clustered in disadvantaged or older urban areas, the policy also had an equity-enhancing dimension.

3 Methodology

We estimate the causal effect of Wisconsin's regulatory intervention on PWS compliance outcomes using SDID estimator proposed by Arkhangelsky et al. (2021). This approach generalizes traditional DID and SC methods by flexibly reweighting both units and time period to relax the parallel trends assumption and improve robustness to latent confounders. Let Y_{it} denote the observed outcome (number of violations) for unit $i \in 1, ..., N$ at time $t \in 1, ..., T$. Let $W_{it} \in 0, 1$ denote the treatment indicator, where $W_{it} = 1$ if unit i exposed to treatment at time t and $W_{it} = 0$ otherwise [1].

[1] Technical interpretation of Synthetic Difference-in-Difference hereafter closely follows Arkhangelsky et al. (2021) and Clarke et al. 2024.

Following the Arkhangelsky et al. (2021), SDID estimates average treatment effect on the treated (ATT) τ by solving the weighted two-way fixed effects regression:

$$\left(\hat{\tau}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \arg \min_{\tau,\mu,\alpha,\beta} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau \right)^2 \hat{\omega}_i \hat{\lambda}_t, \tag{1}$$

where $\hat{\omega}_i$ are unit weights and $\hat{\lambda}_t$ are time weights optimized pre-treatment trajectories and period, respectively. The flexibility of procedure allows for shared temporal aggregate factors given the time-fixed effects β_t and unit fixed effect α_i (Clarke et al. 2024). Compared to standard DID, which assume equal weights across all units and times, SDID introduces a localization mechanism that downweights units and period that are poorly comparable to the treated observations (Bertrand et al. 2004, Goodman-Bacon 2021).

The unit weights $\hat{\omega}_i$ are chosen to minimize the discrepancy between the pre-treatment outcomes of the treated and control units:

$$\left(\hat{\omega}_0, \hat{\omega}\right) = \arg \min_{\omega_0 \in \mathbb{R}, \omega \in \Omega} \sum_{t=1}^{T_{pre}} \left(\omega_0 + \sum_{i=1}^{N_{control}} \omega_i Y_{it} - \frac{1}{N_{treated}} \sum_{i=N_{control+1}}^{N} Y_{it} \right)^2 + \zeta^2 T_{pre} \|\omega\|_2^2, \tag{2}$$

subject to $\omega \in \mathbb{R}_+^N$, with $\sum_{i=1}^{N_{control}} \omega_i = 1$. Here, ω_0 is an intercept term allowing for level shifts and ζ is a regularization parameter to avoid overfitting. Similarly, the time weights $\hat{\lambda}_t$ are computed by solving

$$\left(\hat{\lambda}_0, \hat{\lambda}\right) = \arg \min_{\lambda_0 \in \mathbb{R}, \lambda \in \Lambda} \sum_{i=1}^{N_{control}} \left(\lambda_0 + \sum_{t=1}^{T_{pre}} \lambda_t Y_{it} - \frac{1}{T_{post}} \sum_{t=T_0}^{T} Y_{it} \right)^2, \tag{3}$$

subject to $\lambda \in \mathbb{R}_+^T$, with $\sum_{t=1}^{T_{pre}} \lambda_t = 1$ [2]. The objective is to select pre-treatment periods that

[2] ζ is defined as $\zeta = (N_{treated} T_{post})^{1/4} \hat{\sigma}$, where $\hat{\sigma}^2 = (N_{control}(T_0 - 1))^{-1} \sum_{i=1}^{N_{control}} \sum_{t=1}^{T_0 - 1} (\Delta_{it} - \bar{\Delta})^2$ with $\Delta_{it} = Y_{it+1} - Y_{it}$ and $\bar{\Delta}$ the average first difference across control units (Arkhangelsky et al. 2021). In our study, we adopt the SDID estimation technique developed by Clarke et al. (2024), and it computes $\hat{\lambda}_0$ and $\hat{\lambda}$ by minimizing

$$\left(\hat{\lambda}_0, \hat{\lambda}\right) = \arg \min_{\lambda_0 \in \mathbb{R}, \lambda \in \Lambda} \sum_{i=1}^{N_{control}} \left(\lambda_0 + \sum_{t=1}^{T_{pre}} \lambda_t Y_{it} - \frac{1}{T_{post}} \sum_{t=T_0}^{T} Y_{it} \right)^2 + \zeta^2 N_{control} \|\lambda\|^2, \tag{4}$$

and assumes the very small regularization term with $\zeta = 1 \times 10^{-6} \hat{\sigma}$ to ensure the uniquness of time weight (Clarke et al. 2024).

best predict post-treatment outcomes for the control units, thereby ensuring the comparability of the pre-treatment period.

Under the SDID, the ATT estimator $\hat{\tau}$ can be interpreted as a weighted DID estimator:

$$\hat{\tau} = \hat{\delta}_{treated} - \sum_{i=1}^{N_{control}} \hat{\omega}\hat{\delta}_i \tag{5}$$

where the adjusted outcome differences are

$$\hat{\delta}_{treated} = \frac{1}{N_{treated}} \sum_{i=N_{control}+1}^{N} \left(\frac{1}{T_{post}} \sum_{t=T_0}^{T} Y_{it} - \sum_{t=1}^{T_{pre}} \hat{\lambda}_t Y_{it} \right), \tag{6}$$

and similarly for control units. This framework show that SDID combines ideas from SC (matching pre-trends) and DID(differencing out fixed effects), allowing for flexible deviation from strict parallel trends.

In this paper, we also includes the time-varying covariates X_{it} to improve the validity of SDID estimates by controlling for covariates prior to estimating treatment effects. When time-varying covariates X_{it} are included, the estimation needs to adopt two-step procedure in which covariates are first regressed out of the outcomes variable (Clarke et al. (2024)).

Therefore, we regress Y_{it} on X_{it} using the following model:

$$Y_{it} = X_{it}'\beta + \epsilon_{it}, \tag{7}$$

where β denotes the vector of coefficients associated with covariates (Clarke et al. 2024). We then compute residualized outcomes:

$$\hat{Y}_{it} = Y_{it} - X_{it}'\hat{\beta}. \tag{8}$$

The SDID estimation precedure is the applied to the residualized outcomes $\hat{Y}_{it}$ instead of the raw outcomes Y_{it}. This procedure effectively removes the variation in outcomes attributable to covariates, allowing SDID estimator to focus on the variation associated with treatment while maintaining the robustness properties of the method. Arkhangelsky et al. (2021) note, this residualization step is conceptually distinct from the covariate matching typically employed in sythetic control method, and it aligns more closely with standard regression adjustment procedure.

4 Data

Our analysis uses data from the Safe Drinking Water Information System (EPA-SDWIS 2024) and the EPA/State Drinking Water Dashboard (EPA 2025), covering the period from 2014 to 2023. These sources serve as the national databases of record for monitoring compliance with the SDWA and provide detailed information on PWS activities, including inspections, violations, and enforcement actions. Data are updated quarterly with a three-month lag, meaning that information for a given calendar year is finalized and incorporated into the database by April of the following year (EPA 2025).

The primary outcome variable is the annual number of SDWA violations at the state level. Violations are categorized according to EPA standards into health-based violations, acute health-based violations, monitoring and reporting violations, and public notification violations (EPA 2025). Health-based violations include breaches of MCLs, maximum residual disinfectant levels (MRDLs), or treatment technique (TT) requirements. Acute health-based violations are a subset of health-based violations that have the potential to cause immediate illness. Monitoring and reporting violations occur when systems fail to regularly monitor drinking water quality or fail to submit monitoring results as required. Public notification violations refer to failures to appropriately notify the public about risks to drinking water safety (EPA 2025).

The control group consists of U.S. states that did not implement major new drinking water regulations or interventions during the study period (Cho 2025). States that undertook significant policy changes, such as the adoption of lead testing programs or PFAS regulations between 2014 and 2023, were excluded to ensure comparability. The complete list of excluded states is provided in Appendix Table A1.

To adjust for confounding factors that could simultaneously influence both violation outcomes and policy adoption, the analysis includes several time-varying covariates. These are the log of real GDP, which captures state-level economic conditions; the population

	Pooled sample	Wisconsin	Control group
log Real GDP	12.005	10.900	12.050
Population growth (%)	0.696	0.331	0.710
Site Visit (%)	40.270	67.110	39.204
Number of Violations			
2014 ($T_0 - 4$)	3131.423	5305	3044.480
2018 (T_0)	3096.115	4216	3051.320
2022 ($T_0 + 4$)	4018.923	2753	4069.560

Table 1: Covariate and outcome means

growth rate, which reflects demographic pressures that could affect water system demand and operational strain; and the percentage of public water systems that received a site visit during the year, which serves as a proxy for regulatory oversight intensity. These covariates were selected based on their theoretical and empirical relevance to water system compliance dynamics [3].

Table 1 presents descriptive statistics for the pooled sample, Wisconsin as the treated unit, and the control group of states. Wisconsin exhibits lower economic output compared to the control group, with an average log real GDP of 10.90 relative to 12.05 in the control group. Population growth in Wisconsin is also notably slower, at 0.331% compared to 0.710% in the control group. In contrast, Wisconsin shows a much higher rate of site visits, with 67.11% of its public water systems receiving at least one site visit during the study period, compared to 39.20% in the control states.

Regarding violation outcomes, Wisconsin initially had a higher number of violations than the control group. In 2014, the number of violations in Wisconsin was 5,305, while the control group's average was 3,044.48. By 2018, the year of policy intervention, Wisconsin's violations declined to 4,216, whereas the control group averaged 3,051.32 violations. By 2022, Wisconsin

[3] Real GDP data are from the U.S. Bureau of Economic Analysis (U.S. Bureau of Economic Analysis 2024); population growth data are from the Federal Reserve Bank of St. Louis (Federal Reserve Bank of St. Louis 2024); and site visit data are obtained from the U.S. Environmental Protection Agency's State Drinking Water Dashboard EPA 2025.

had reduced its violations further to 2,753, while the control group's violations increased to 4,069.56. These descriptive patterns suggest that Wisconsin experienced a reduction in violations over the study period relative to other states, supporting the need for a formal causal analysis to quantify the effect of the intervention.

5 Results

Figure 1 presents the estimated unit-specific weights assigned to each control state under both the SC and SDID estimators. These weights reflect how much each untreated state contributes to constructing the counterfactual trend for Wisconsin's compliance outcomes in the absence of the policy intervention. Consistent with the properties of the SC method, the SC weights are highly sparse, with only a few control states receiving non-zero weight. In particular, Texas and Utah emerge as the dominant contributors, with Texas receiving a weight of over 0.5. This sparsity is a defining feature of SC, which seeks to match the pre-treatment outcome path of the treated unit as closely as possible, typically relying on a small number of donor units that best replicate the treated unit's pre-policy trajectory.

In contrast, the SDID estimator yields a markedly different weighting pattern. Under SDID, the weights are distributed more evenly across a broader set of control units, with no single state receiving a disproportionate share. All individual weights under SDID remain below 0.1, and most weights are positive but small. This more diffuse weighting structure is the result of two key modifications introduced by SDID relative to traditional SC. First, SDID incorporates regularization when solving for unit weights, discouraging sparsity and preventing the over-reliance on a few units (Arkhangelsky et al. 2021). Second, by integrating a two-way fixed effects structure (including unit fixed effects) into the outcome model, SDID partially absorbs permanent differences across states, reducing the burden on the weighting scheme to exactly replicate pre-treatment outcome levels (Arkhangelsky et al. 2021).

The differences in weighting behavior between SC and SDID are especially important in the context of this study, where we now observe an extended pre-treatment period from 2014 to 2017. While traditional SC relies on matching the pre-treatment levels of the outcome variable, it may still overfit short-term fluctuations, especially when a small number of donor

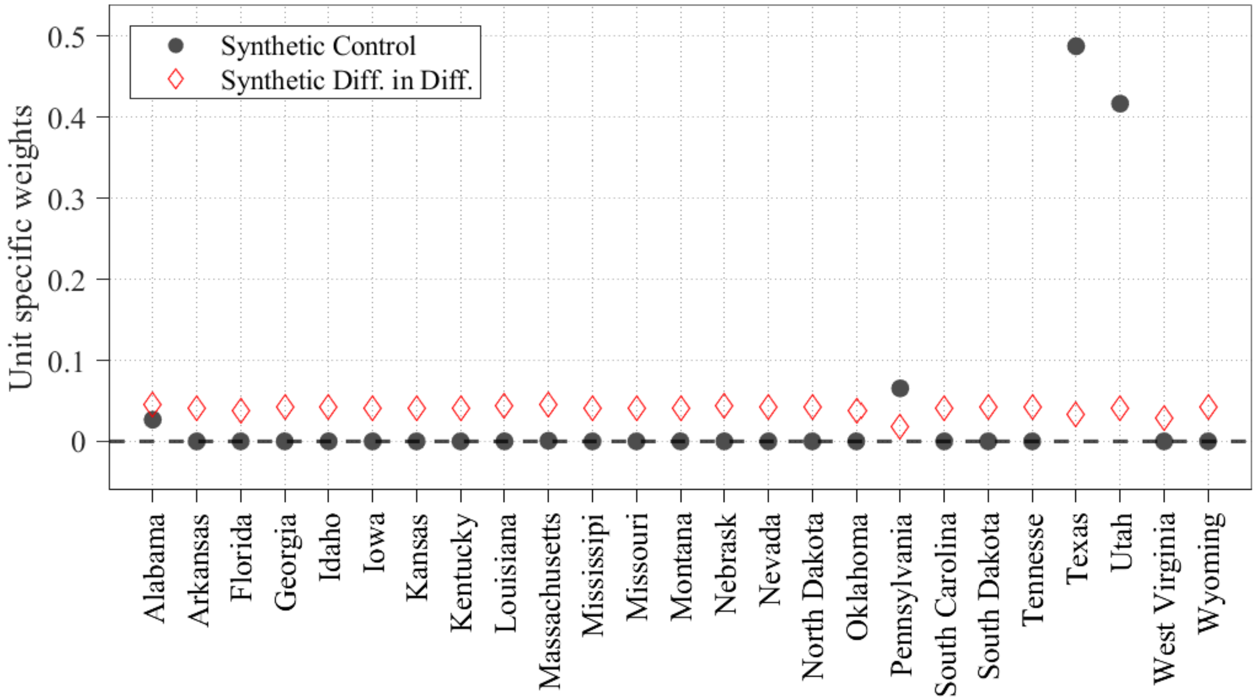

Figure 1: Estimated unit specific weights

units receive disproportionately large weights. In contrast, SDID mitigates this risk by placing greater emphasis on matching pre-treatment trends and not just levels. By incorporating both unit and time weights, SDID ensures that donor units with more stable and parallel trajectories to California's pre-policy trend are given priority, while regularization avoids overreliance on any single state.

Table 2 reports the estimated ATT for Wisconsin using three methods: SDID, SC, and traditional DID. Consistent with Arkhangelsky et al. (2021), the SDID estimate demonstrates both robustness and precision, showing a significant decrease of approximately 979 in the number of PWS violations, with a comparatively small standard error. In contrast, the SC estimate is positive but highly imprecise, reflecting the vulnerability of SC to overfitting or misalignment when the pre-treatment periods are noisy or the number of donor units is limited. The DID estimate is similar in sign and magnitude to SDID but exhibits a larger standard error, highlighting the limitations of DID under imperfect parallel trends (Abadie et al. 2010) [4]. These results illustrate one of the core advantages of SDID suggested by Arkhangelsky et al. (2021): by combining unit and time weighting with a two-way fixed

[4]The raw violation data used to assess parallel trends are presented in Appendix Figure A1.

	Synthetic Diff. in Diff	Synthetic Control	Diff. in Diff
ATT	-965.063	333.405	-958.246
Standard error	(501.927)	(1276.643)	(535.209)

Table 2: Estimates for average treatment effect on the treated (ATT) on Wisconsin. We employ the placebo-based standard error estimator. Placebo treatments in estimation is to control units and compute the distribution of placebo estimates $\hat{\tau}_p$ to approximate the sampling variability of the estimator. The variance estimate is given by $\hat{V}_{placebo}(\hat{\tau}) = \mathrm{Var}(\hat{\tau}_p)$, and a $(1-\alpha)$ level confidence interval is contructed as $\hat{\tau} \pm z_{\alpha/2}\sqrt{\hat{V}_{placebo}(\hat{\tau})}$, where $z_{\alpha/2}$ denotes the standard normal critical value (Arkhangelsky et al. 2021, Clarke et al. 2024).

effects structure, SDID reduces bias from latent confounders and achieves greater stability than either SC or DID alone. Building on these results, Figure 2 presents the estimated compliance trend for Wisconsin alongside the SDID-constructed counterfactual.

Figure 2 illustrates the trends in the number of violations for Wisconsin and the control group. The vertical dashed line marks the policy intervention year (T_0), corresponding to the implementation of regulatory changes. Before the intervention, Wisconsin consistently exhibits a higher number of violations compared to the control group, but the two trends do not run strictly parallel, indicating the presence of imperfect parallel trends. The visual departure from the strict parallel trends assumption highlights the need for a more flexible estimation approach such as the SDID method, which combines reweighting of units and time periods to improve causal inference (Arkhangelsky et al. 2021).

The observed number of violations declines relative to the counterfactual after 2018, suggesting that the SDID method captures a meaningful shift in compliance behavior following the policy intervention. Unlike traditional approaches that rely on rigid assumptions of level-matching or parallel trends, SDID is designed to flexibly match trends by estimating both unit and time weights. This design allows the estimator to construct a control trajectory that aligns with Wisconsin's pre-treatment dynamics without being overly influenced by short-term fluctuations or the levels of a few donor units, which can be a limitation of the SC method.

An important feature of the SDID estimator is its use of time weights, depicted by the bar chart at the bottom of Figure 2. These weights determine how much each pre-treatment year

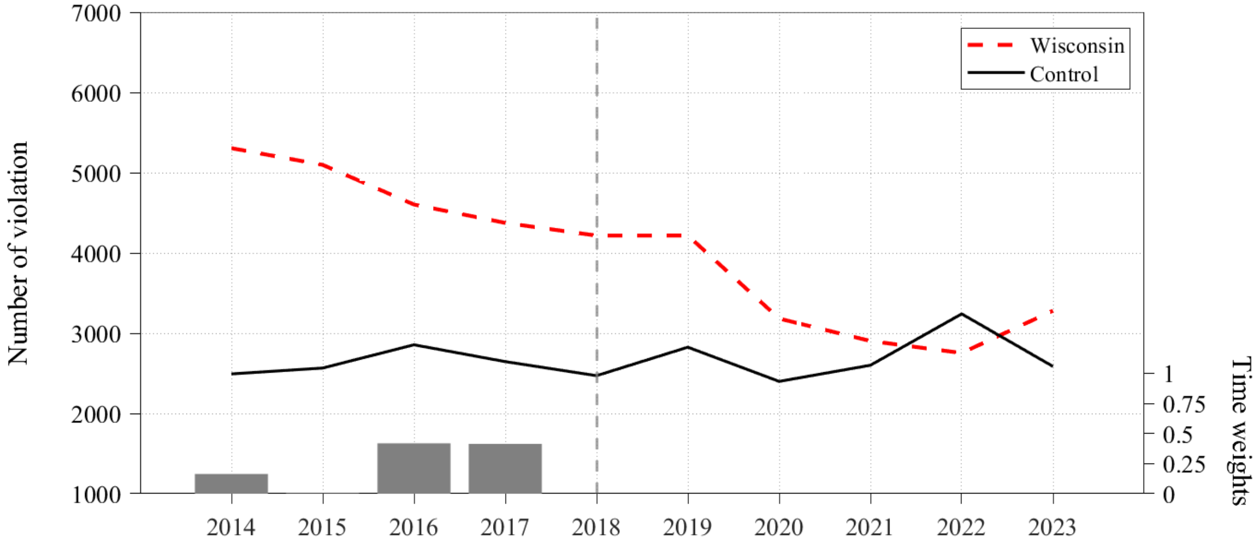

Figure 2: Comparison of estimators. The dashed red line represents the observed compliance trend in Wisconsin between 2014 and 2023. The solid black line traces the SDID estimate. The gray bars at the bottom indicate the time weights used in SDID, which reflect the relative importance assigned to each pre-treatment year in constructing the counterfactual trajectory.

contributes to aligning the synthetics with the treated unit. In this case, the estimator places the highest weight on 2016 and 2017, while assigning modest but non-negligible weights to earlier years such as 2014. This selective emphasis allows SDID to exploit the most relevant pre-treatment information while reducing the influence of years that may be less predictive due to noise or structural shifts.

The application of time weights is central to why the SDID counterfactual does not track Wisconsin as closely during the pre-treatment period as the SC method does. Rather than overfitting to match observed levels in all pre-treatment years, SDID prioritizes matching the trend trajectory, which is most informative for estimating the post-treatment counterfactual. This explains why the control line lies below Wisconsin even before the policy was implemented, yet still provides a valid and robust estimate of the treatment effect by ensuring trend similarity and down-weighting irrelevant noise.

To evaluate the credibility of the estimated treatment effect for California, we conduct a placebo-based falsification test using the Ratio of Mean Squared Prediction Error Ratio (RMSPE) (Abadie et al. 2010). Figure 3 presents the results of a placebo test using the absolute value of the log Ratio of Mean Squared Prediction Error ($|\log \text{RMSPE}|$) for Wisconsin

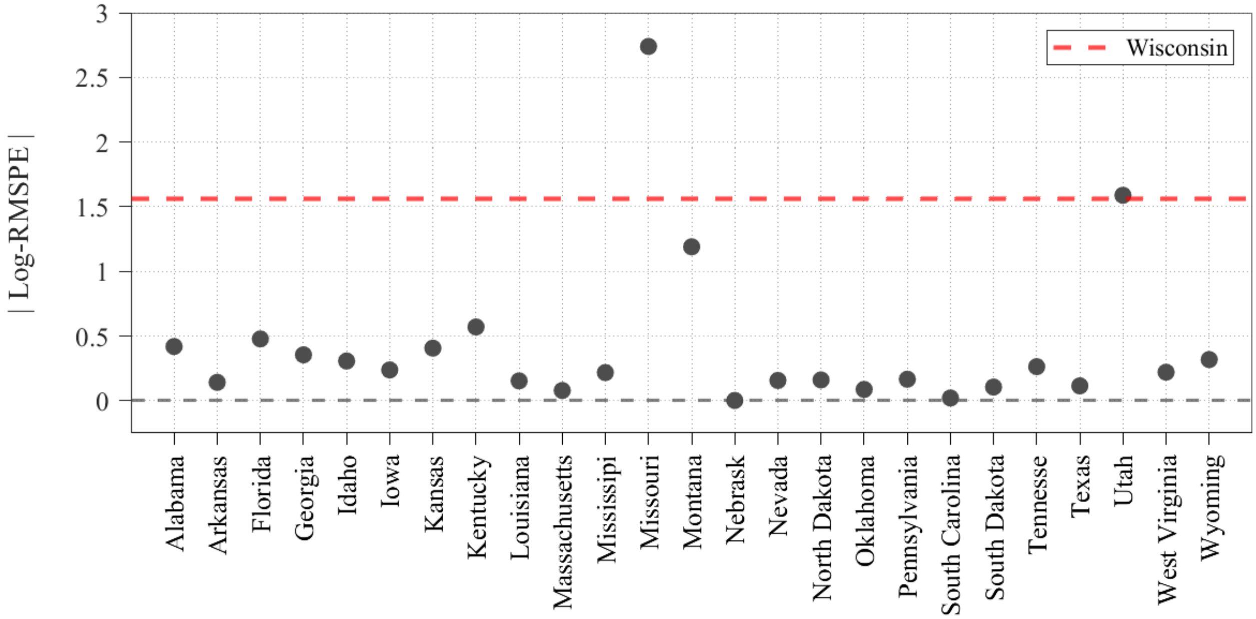

Figure 3: Log-RMSPE ratios across states. Each black dot represents a control state, and the red dashed line denotes California.

compared to control states [5]. Each black dot represents a control state's $|\log \text{RMSPE}|$, while the red dashed line indicates Wisconsin's value. The $|\log \text{RMSPE}|$ metric captures the relative change in prediction error between the pre- and post-treatment periods. Specifically, a $|\log \text{RMSPE}|$ value close to zero indicates that the model's predictive accuracy remained stable before and after the intervention, suggesting no substantial structural change. In contrast, a large $|\log \text{RMSPE}|$ implies a significant increase in prediction error after the intervention, which is consistent with a real underlying shift in outcomes due to treatment. Thus, higher $|\log \text{RMSPE}|$ values are interpreted as evidence of a treatment effect relative to the counterfactual trajectory.

The measure used in this analysis for each state j is as follows:

$$\left|\log \text{RMSPE}_j\right| = \left|\log\left(\frac{\frac{1}{T-(T_0-1)}\sum_{t=T_0}^{T}(Y_{jt}-\hat{Y}_{jt})^2}{\frac{1}{T_0-1}\sum_{t=1}^{T_0-1}(Y_{jt}-\hat{Y}_{jt})^2}\right)\right| \tag{9}$$

where Y_{jt} represents the observed outcome for state j in year t, specifically the number of violations. $\hat{Y}_{jt}$ is the counterfactual prediction for state j generated by the SDID estimator.

[5] Figure A2 in Appendix shows the raw differences between observed and synthetic control outcomes for Wisconsin and placebo states.

T_0 denotes the treatment year, and T is the last observed year. The numerator captures the mean squared prediction error in the post-treatment period, while the denominator captures the mean squared prediction error in the pre-treatment period.

The placebo test is designed to assess the validity of the SDID estimates by examining whether untreated states exhibit similar post-treatment deviations. If large treatment effects were common among untreated units, it would cast doubt on the causal interpretation of the Wisconsin estimates. However, the figure clearly shows that Wisconsin's $|\log \mathrm{RMSPE}|$ is substantially higher than that of any control state. The vast majority of control states cluster near zero, indicating minimal post-treatment deviations and reinforcing the stability of their pre- and post-treatment trajectories. Although a few control states, such as Missouri, Montana and Utah, show increases, none exhibit deviations comparable to that observed in Wisconsin.

The stark separation between Wisconsin and the control group provides strong evidence that the estimated treatment effect is unlikely to be driven by random variation or model overfitting. Instead, the placebo test results support the interpretation that the intervention induced a meaningful shift in compliance behavior specifically in Wisconsin. This strengthens the internal validity of the SDID estimates and bolsters the causal claim that the policy intervention effectively reduced violations among PWSs in Wisconsin.

To further examine the dynamic effects of Wisconsin's 2018 regulatory intervention, we conduct an event study analysis following the method proposed by Clarke et al. (2024). Figure 4 presents the estimated treatment effects for each year from 2014 to 2023, using the SDID framework. Each point estimate reflects the annual difference in SDWA violations between Wisconsin and its synthetic control, controlling for observed covariates. The blue diamonds indicate point estimates, and the shaded band represents the 95% confidence interval. The vertical dashed line at 2018 marks the intervention year, while the red dashed horizontal line at zero provides a reference for evaluating the null of no effect. The estimates prior to 2018 are close to zero, supporting the credibility of the identifying assumption that the treated and control units followed parallel trends in the absence of the policy change. After 2018, the estimates become increasingly negative, indicating a substantial and statistically significant reduction in violations relative to the synthetic control.

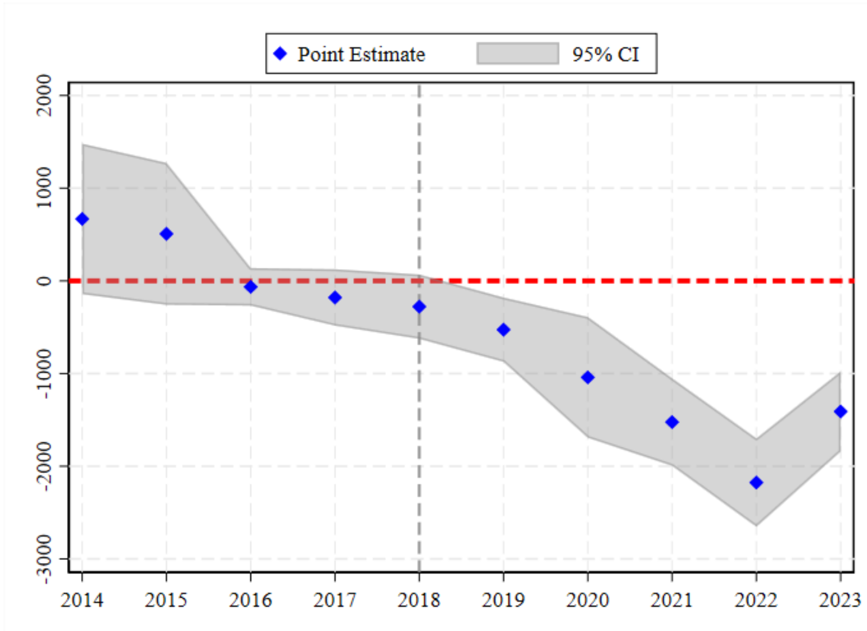

Figure 4: Event study. Vertical axis indicates that the number of violations.

This event study is constructed using a two-step procedure. First, the outcomes are residualized by regressing the outcome variable on covariates to obtain by Eq. (8).

Second, the year-specific treatment effects are estimated by comparing the residualized outcomes of the treated unit to a weighted average of the control units. The treatment effect at time t is given by:

$$\hat{\tau}_t = \frac{1}{N_{\text{treated}}} \sum_{i=1}^{N_{\text{treated}}} \left(\hat{Y}_{it} - \sum_{t=1}^{T_0 - 1} \hat{\lambda}_s \hat{Y}_{is} \right) - \sum_{i=1}^{N_{\text{control}}} \hat{\omega}_i \left(\hat{Y}_{it} - \sum_{t=1}^{T_0 - 1} \hat{\lambda}_s \hat{Y}_{is} \right), \tag{10}$$

where $\hat{\omega}_i$ and $\hat{\lambda}_s$ are unit and time weights estimated from the pre-treatment period, T_0 is the intervention year, and $N_{treated}$ is the number of treated units (Clarke et al. 2024). In this setting, Wisconsin is the only treated unit, and the time weights are optimized to reweight the pre-treatment path of the treated unit [6].

This dynamic pattern of treatment effects complements our main results and provides further evidence that the 2018 policy intervention had a persistent and growing impact on regulatory compliance in Wisconsin.

[6] Confidence intervals are computed using a cluster bootstrap.

6 Discussion

This study provides evidence that targeted regulatory interventions can significantly improve compliance outcomes among public utilities, even in the presence of imperfect pre-treatment trends. By applying the SDID estimator, we mitigate concerns about latent confounding and deviations from parallel trends that often complicate causal inference in observational settings. The findings offer important contributions to both the methodological literature on causal inference and the policy discourse surrounding environmental and public health regulations.

Methodologically, this study reinforces the utility of SDID as a credible alternative to traditional DID and SC approaches, particularly when pre-treatment comparability between treated and control units is imperfect. As emphasized by Arkhangelsky et al. (2021) and Clarke et al. (2024), SDID achieves robustness by reweighting both units and time periods while incorporating fixed effects, thus effectively controlling for both observable and latent sources of bias. The empirical application here demonstrates that these theoretical advantages translate into practical gains: the estimates obtained are more precise than those generated by DID or SC alone. This supports the broader view that causal inference frameworks should accommodate flexible counterfactual construction rather than rely on rigid assumptions like strict parallel trends or convex hull conditions.

The findings also contribute to the broader literature on regulatory compliance and policy evaluation. While previous studies have documented the challenges in enforcing environmental and public health regulations, particularly in decentralized settings, this study shows that carefully designed interventions can lead to sustained improvements in compliance. Importantly, the substantial reduction in violations observed in Wisconsin suggests that regulatory efforts targeting technical compliance issues can have immediate and meaningful impacts. This supports arguments in the policy literature advocating for proactive, rather than purely punitive, regulatory strategies (Coglianese and Kagan 2007).

Moreover, the results have implications for the design and evaluation of future regulatory interventions. First, the success of Wisconsin's program underscores the importance of targeted, state-level policies tailored to local conditions, rather than relying solely on broad federal mandates. Second, the methodology employed here illustrates the value of adopting more flexible econometric techniques in policy evaluation, particularly when dealing with complex real-world data that may violate traditional modeling assumptions. Policymakers and researchers alike should recognize that standard parallel-trends assumptions may often be implausible and that methodological innovations such as SDID can enhance the credibility of impact evaluations.

Nevertheless, several limitations warrant discussion. While SDID substantially relaxes the assumptions underlying DID and SC, it still relies on the availability of sufficient pre-treatment periods and appropriate donor units to construct reliable weights. If unobserved shocks contemporaneous with the intervention differentially affect the treated unit, even SDID estimates may be biased. Although placebo tests in this study provide reassurance against such threats, future work could benefit from explicitly modeling potential confounders or exploring extensions (Athey and Imbens 2017). Furthermore, the current analysis focuses on aggregate outcomes at the state level; disaggregated analyses by PWS size, ownership type, or violation category could yield further insights into the heterogeneity of treatment effects.

Future research could also explore the persistence of the observed compliance improvements over longer horizons, investigating whether regulatory impacts diminish, stabilize, or even amplify over time. Additionally, comparative analyses across multiple states implementing similar interventions would provide a broader understanding of contextual factors that condition regulatory effectiveness.

In conclusion, this study highlights both the promise of innovative causal inference methods like SDID and the potential of well-crafted regulatory interventions to improve public health outcomes. As policymakers grapple with increasingly complex environmental and public health challenges, evidence-based, methodologically rigorous evaluations will be crucial in guiding effective governance.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American Statistical Association*, 105, 493–505.

Allaire, M., H. Wu, and U. Lall (2018): "National trends in drinking water quality violations," *Proceedings of the National Academy of Sciences*, 115, 2078–2083.

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2021): "Synthetic difference-in-differences," *American Economic Review*, 111, 4088–4118.

Athey, S. and G. Imbens (2017): "The state of applied econometrics: Causality and policy evaluation," *Journal of Economic Perspectives*, 31, 3–32.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How much should we trust differences-in-differences estimates?" *The Quarterly journal of economics*, 119, 249–275.

Cho, C. (2025): "Human Right to Water Act and Drinking Water Compliance: A Synthetic Control Approach," .

Christensen, P., D. A. Keiser, and G. E. Lade (2023): "Economic effects of environmental crises: Evidence from Flint, Michigan," *American Economic Journal: Economic Policy*, 15, 196–232.

Clarke, D., D. Pailañir, S. Athey, and G. Imbens (2024): "On synthetic difference-in-differences and related estimation methods in Stata," *The Stata Journal*, 24, 557–598.

Coglianese, C. and R. A. Kagan (2007): *Regulation and regulatory processes*, Oxford University Press.

EPA (2025): "Enforcement and compliance history online: EPA/State Drinking Water Dashboard," Accessed: 2025-02-25.

EPA-SDWIS (2024): "Safe Drinking Water Information System (SDWIS) - Federal Reporting Services," Accessed: 2025-04-28.

Federal Reserve Bank of St. Louis (2024): "Annual Population Estimates by State," Accessed April 2025.

Goodman-Bacon, A. (2021): "Difference-in-differences with variation in treatment timing," *Journal of econometrics*, 225, 254–277.

Grant, L. E. and K. K. Grooms (2017): "Do nonprofits encourage environmental compliance?" *Journal of the Association of Environmental and Resource Economists*, 4, S261–S288.

Gray, S., A. Singer, L. Schmitt-Olabisi, J. Introne, and J. Henderson (2017):

"Identifying the causes, consequences, and solutions to the Flint Water Crisis through collaborative modeling," *Environmental Justice*, 10, 154–161.

Greenstone, M. and T. Gayer (2009): "Quasi-experimental and experimental approaches to environmental economics," *Journal of Environmental Economics and Management*, 57, 21–44.

Hanna-Attisha, M., J. LaChance, R. C. Sadler, and A. Champney Schnepp (2016): "Elevated blood lead levels in children associated with the Flint drinking water crisis: a spatial analysis of risk and public health response," *American journal of public health*, 106, 283–290.

Jakubowski, M. (2011): "Low-level environmental lead exposure and intellectual impairment in children—the current concepts of risk assessment," *International journal of occupational medicine and environmental health*, 24, 1–7.

Konisky, D. M. and M. P. Teodoro (2016): "When governments regulate governments," *American Journal of Political Science*, 60, 559–574.

River Network (2025): "State Policy Hub: Drinking Water," Accessed: 2025-02-25.

Shimshack, J. P. and M. B. Ward (2005): "Regulator reputation, enforcement, and environmental compliance," *Journal of Environmental Economics and Management*, 50, 519–540.

U.S. Bureau of Economic Analysis (2024): "Gross Domestic Product (GDP) by State," Accessed April 2025.

U.S. Congress (2000): *The Safe Drinking Water Act as Amended by the Safe Drinking Water Act Amendments of 1996: Public Law 104-182, August 6, 1996*, S. Prt. 106-59, Washington, DC: U.S. Government Printing Office, printed for the use of the Committee on Environment and Public Works, 106th Congress, 2nd Session.

Wang, R., X. Chen, and X. Li (2022): "Something in the pipe: the Flint water crisis and health at birth," *Journal of Population Economics*, 35, 1723–1749.

Wisconsin Legislature (2018): "2017 Wisconsin Act 137," Enacted on February 21, 2018. Effective on February 22, 2018. Relates to financial assistance for the replacement of lead-containing customer-side water service lines.

Appendix

State	Policy	Date	Description
Alaska	AK H 209	07.28.2016	committee studies rural water and sewer needs
Arizona	HB 2049	04.28.2017	expands grant eligibility for small water systems
	SB 1459	05.12.2016	assist low-income homeowners with well improvements
California	HR2W	09.25.2012	ensuring affordable, accessible, acceptable and safe water
Colorado	HB 1306	06.08.2017	funds lead testing in public schools
	HB 20-1119	06.29.2020	regulates PFAS storage, disposal, and firefighting foam
	SB 20-2018	06.29.2020	establishes PFAS fund for grants, takeback, and assistance
	HB 22-1358	06.07.2022	law mandates lead testing in schools, childcare
Connecticut	HB 5509	06.14.2018	protects vulnerable groups from sewer foreclosures
Delaware	HB 200	07.22.2021	funds clean water projects, prioritizing equity
Illinois	SB 550	01.17.2017	mandates lead testing, inventory, and notification
	SB 2146	08.23.2019	invests in clean water infrastructure and workforce training
	HB 0414	08.06.2021	creates low-income water and sewer assistance program
	HB 3739	01.01.2022	mandates full lead pipe replacement and assistance
Indiana	HB 1138	05.01.2023	preschools and childcare must test for lead
Maine	S.P. 64	06.21.2021	mandates PFAS monitoring, notification, and mitigation
	HP 113	07.15.2021	nation's first comprehensive PFAS product ban enacted
Maryland	SB 96	04.30.2019	prohibits tax sales for water bill liens
Michigan	HB 4342	10.24.2023	child care centers must label water safety
	SB 88	10.24.2023	child care centers must manage lead exposure
Minnesota	HF 1	10.21.2020	funds water infrastructure upgrades and protection
	HF 2310	05.24.2023	funds PFAS mitigation, bans, and regulations
New Hampshire	SB 309	07.10.2018	sets PFAS water standards, adds toxicologist
	HB 1264	07.23.2020	sets PFAS MCLs, funds programs, expands standards
New Jersey	SB 968/A2863	05.11.2021	law mandates lead level notifications quickly
	SB 994	09.13.2022	mandates utility affordability
New Mexico	SB 1	03.13.2023	facilitates regionalization of water utilities
New York	SB S8158	09.06.2016	schools must test for lead, provide aid
	VolA-5-5-1	08.26.2020	sets maximum contaminant levels for contaminants

State	Policy	Date	Description
North Carolina	HB 1087	07.01.2020	funds utilities, reviews, mergers, and projects
Ohio	HB 512	09.09.2016	strengthens Lead and copper testing requirement
	3745-81-84	05.01.2018	revised Lead and Copper Rule
	HB 166	11.01.2019	H2Ohio fund for water quality projects
Oregon	Water Vision	2019	improvements to our infrastructure and ecosystems
Rhode Island	SB 2298	06.24.2022	mandates PFAS testing, standards, and monitoring
	SB 0724	06.22.2023	revises PFAS contamination response
Vermont	Act 21	05.15.2019	regulation of poly-fluoroalkyl substances
	Act 139	07.06.2020	construction grants for public water improvement
Virginia	HJ538	02.24.2021	access to clean, potable, and affordable water
	HB 1257	01.01.2022	sets maximum contaminant levels
Washington	SB 6413	06.07.2018	bans PFAS firefighting foam, mandates disclosure
	SB 5135	07.28.2019	regulates priority toxic chemicals in products

Table A1: States Excluded from Analysis Due to Policy Interventions (2016–2023). HF: House File, HB: House Bill, SB: Senate Bill, PFAS: Perfluoroalkyl and Polyfluoroalkyl Substances. Data Source: River Network 2025, retrieved on April 29, 2025 (Cho 2025). We also excluded D.C. and Hawaii from the control group due to their structural dissimilarity to Wisconsin. D.C., as a city-state, lacks rural drinking water systems and exhibits administrative characteristics fundamentally distinct from continental states. Hawaii, being a geographically isolated island state, operates under water supply and enforcement systems that differ markedly from those on the mainland. Including these units would violate the synthetic control method's requirement for comparable untreated units and risk undermining the credibility of our causal estimates.

Treatment Group	**Control Group (25)**
Wisconsin	Alabama, Arkansas, Florida, Georgia, Idaho, Iowa, Kansas, Kentucky, Louisiana, Massachusetts, Mississippi, Missouri, Montana, Nebraska, Nevada, North Dakota, Oklahoma, Pennsylvania, South Carolina, South Dakota, Tennessee, Texas, Utah, West Virginia, Wyoming

Table A2: Treatment Group and Control Group States

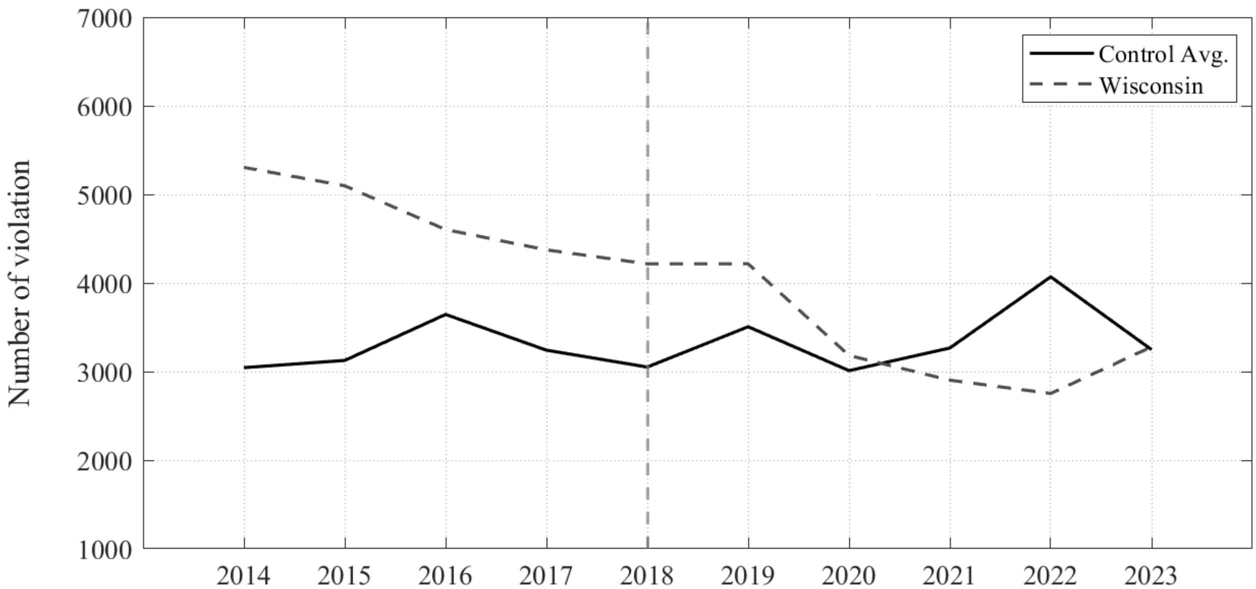

Figure A1: Trends in Number of Violations for Wisconsin and Control States (2011–2023)

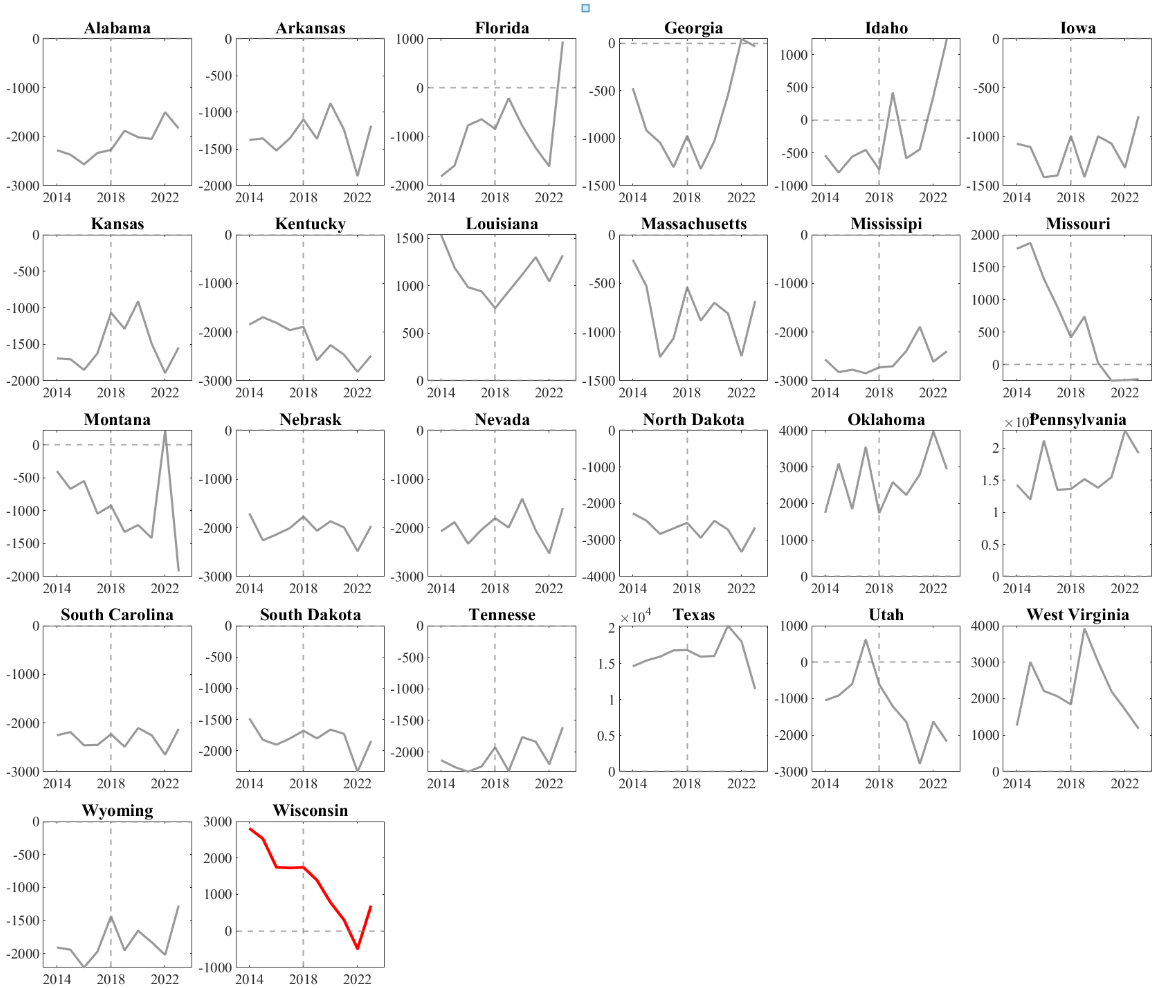

Figure A2: Estimated treatment-control gaps by state. This figure presents the difference between observed outcomes and synthetic control outcomes for Wisconsin and each placebo control state over the period 2014–2023. Each subplot corresponds to a different state, showing the evolution of the gap between the state's actual violations and the synthetic counterfactual, with the vertical dashed line indicating the intervention year (2018). The trajectory for Wisconsin is highlighted in red.