Do Red Flag Laws Save Lives or Shift Deaths? Means Substitution and ERPO Effectiveness

APEP Autonomous Research*
@SocialCatalystLab

March 9, 2026

Abstract

Extreme Risk Protection Order (ERPO) laws allow courts to temporarily remove firearms from individuals deemed dangerous. Using Callaway–Sant'Anna staggered difference-in-differences across 21 treated states (1999–2024), I find no statistically detectable effect of ERPO adoption on total suicide rates (ATT = 0.24 per 100,000, 95% CI: $[-0.16, 0.64]$). The confidence interval rules out reductions larger than 0.16 per 100,000 but does not rule out modest effects in either direction. Standard two-way fixed effects produces a misleading negative estimate (-1.19, $p = 0.003$), illustrating the heterogeneous-treatment-effect bias documented in the recent staggered DiD literature. Exploratory mechanism decomposition using 2019–2024 data with only 9 treated states is inconclusive. These findings suggest that ERPO adoption at the state level—as distinct from ERPO utilization intensity—has not detectably reduced population-level suicide mortality, though individual-level efficacy remains plausible at scales below the design's detection threshold.

JEL Codes: I18, K42, H75
Keywords: red flag laws, ERPO, suicide, means substitution, staggered DiD, TWFE bias

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

A suicidal crisis typically lasts minutes to hours. If a person in distress lacks a lethal tool during those minutes, they usually survive (Owens et al., 2002). Firearms convert 85% of suicide attempts into deaths; drug overdose—the most common alternative—converts fewer than 5% (Spicer and Miller, 2000). This asymmetry makes firearm access a central variable in suicide prevention, and 49,449 Americans died by suicide in 2022 alone—more than by motor vehicle accidents (Garnett et al., 2023).

Extreme Risk Protection Order (ERPO) laws—commonly called "red flag" laws—represent the most targeted legislative response to this asymmetry. These laws allow family members, law enforcement, or other designated petitioners to request a court order temporarily removing firearms from individuals exhibiting signs of imminent danger. Connecticut pioneered this approach in 1999, but adoption remained confined to five states until the February 2018 Parkland school shooting catalyzed a wave of legislation. By 2024, 22 states had enacted ERPO laws, and the question of their effectiveness has become one of the most debated issues in American public health policy.

The scientific case for ERPOs rests on a simple chain of logic: suicidal crises are typically transient, lasting hours to days (Owens et al., 2002); firearms are the most lethal method; therefore, temporarily removing firearms during a crisis should reduce completed suicides. But this logic contains a critical assumption—that individuals denied access to firearms do not substitute equally lethal alternatives. The *means substitution hypothesis* posits that determined individuals will simply switch methods, offsetting any reduction in firearm suicide with increases in hanging, poisoning, or other means.

This paper tests whether ERPO laws reduce total suicide mortality or merely redistribute deaths across methods. I exploit the staggered adoption of ERPO laws across US states—22 enacted between 1999 and 2024, of which 21 enter the estimation sample (Connecticut is excluded because its 1999 adoption coincides with the first panel year)—using the heterogeneity-robust estimator of Callaway and Sant'Anna (2021). This approach addresses a well-documented problem with standard two-way fixed effects (TWFE): when treatment effects vary across cohorts and over time—as is plausible for a policy whose implementation quality and utilization intensity differ substantially across states—TWFE can produce misleading estimates by implicitly using early-adopting states as controls for later adopters (Goodman-Bacon, 2021; de Chaisemartin and d'Haultfœuille, 2020).

The main finding is that ERPO adoption has no statistically detectable effect on total suicide rates. The Callaway–Sant'Anna average treatment effect on the treated (ATT) is 0.24 deaths per 100,000 ($p = 0.25$, 95% CI: $[-0.16, 0.64]$). The confidence interval rules out reductions larger than 0.16 per 100,000 but does not rule out modest effects in either direction. This result is robust to using not-yet-treated states as the control group (ATT = 0.23, $p =$ 0.33), excluding the 2018 adoption cohort (ATT = 0.43, $p = 0.19$), excluding anti-ERPO states from the control group (ATT = 0.24, $p = 0.28$), and stable across leave-one-out sensitivity checks.

A methodological finding accompanies this null. Standard TWFE produces an ATT estimate of -1.19 per 100,000 ($p = 0.003$)—a statistically significant *negative* effect suggesting ERPOs reduce suicide. This sign reversal between TWFE and the heterogeneity-robust estimator illustrates the bias that can arise when treatment effects vary across cohorts and over time (Roth et al., 2023; Borusyak et al., 2024). A Goodman-Bacon decomposition on a restricted 2005–2017 sub-panel reveals that early adopters with long exposure periods dominate the TWFE estimate, and their declining suicide trajectories are extrapolated to later cohorts.

An exploratory mechanism decomposition using 2019–2024 CDC data with firearm-specific breakdowns finds no reduction in either firearm or non-firearm suicide. Point estimates are positive for both components, but the short-panel estimates rest on only 9 treated states with 2–4 post-treatment periods per cohort. With this limited identifying variation, overlapping entirely with the COVID-19 pandemic and its aftermath, the mechanism decomposition is inconclusive regarding means substitution. A placebo test using drug overdose deaths—which should not respond to firearm removal—is insignificant (ATT = 2.96, $p = 0.32$), though the wide confidence interval limits its diagnostic value.

This paper makes three contributions. First, it provides the first application of heterogeneity-robust staggered DiD to the ERPO–suicide question using the longest available panel (1999–2024). Prior work has primarily used interrupted time series within individual states (Humphreys et al., 2019) or standard TWFE across multiple states (Kivisto and Phalen, 2018). Second, it illustrates that the TWFE approach—still widely used in firearm policy evaluation—can produce qualitatively different estimates than heterogeneity-robust alternatives in this setting, cautioning against uncritical reliance on TWFE for staggered policy evaluations. Third, it provides an exploratory mechanism decomposition into firearm and non-firearm components, though the short-panel design limits the strength of conclusions about means substitution.

These results contribute to a broader literature on means restriction in suicide prevention (Mann et al., 2005; Barber and Miller, 2008; Daigle, 2005). While individual-level studies have found that firearm access substantially increases suicide risk (Anglemyer et al., 2014; Kellermann et al., 1992), and case-series studies of specific ERPO petitions suggest individual-level effectiveness (Swanson et al., 2017, 2019), the present analysis finds no population-level effect. This gap between individual and population effects is consistent with ERPOs being used too infrequently or inconsistently to register in state-level mortality statistics. The policy implication is not that ERPOs are ineffective for the individuals they reach, but that their current implementation is insufficient to move aggregate outcomes.

The paper describes the legal landscape of ERPOs, details the state-level mortality data, presents the staggered-adoption results, and discusses implications for policy and future research.

2. Institutional Background

2.1 The ERPO Mechanism

Extreme Risk Protection Orders are civil court orders that authorize law enforcement to temporarily remove firearms from individuals who pose an imminent risk of harm to themselves or others. Unlike traditional firearm restrictions, ERPOs do not target categories of people (felons, those involuntarily committed) but rather respond to individualized risk assessments triggered by concerning behavior (Swanson et al., 2017).

The typical ERPO process involves several stages. First, an eligible petitioner—usually law enforcement or a family member, though some states allow medical professionals, school officials, or coworkers—files a petition with a court. The petition must present evidence that the respondent poses a significant danger. Second, a judge evaluates the petition and may issue a temporary ex parte order (typically lasting 14–21 days) based on a preponderance or clear-and-convincing evidence standard. Third, a full hearing is held within the temporary order period, at which the respondent may contest the order. If the court finds continuing risk, it issues a final order lasting 6–12 months, during which the respondent's firearms are held by law enforcement or surrendered to a licensed dealer. Violation of an ERPO constitutes a criminal offense in most states.

2.2 Staggered Adoption

ERPO adoption across US states falls into two distinct periods separated by the Parkland, Florida school shooting on February 14, 2018, which killed 17 people and generated unprecedented political momentum for gun safety legislation.

Pre-Parkland adopters (1999–2017). Connecticut enacted the first ERPO law in 1999 following the Connecticut State Lottery shooting. Indiana followed in 2005, then California (2014), Washington (2016), and Oregon (2017). These five states adopted ERPOs largely in response to specific high-profile incidents, and their laws were relatively unknown to the broader public.

Post-Parkland adopters (2018–2024). The Parkland shooting triggered an unprecedented wave. In 2018 alone, eight states adopted ERPO laws: Florida (where the shooting occurred), Vermont, Maryland, Rhode Island, New Jersey, Delaware, Massachusetts, and Illinois. Colorado, Nevada, Hawaii, and New York followed in 2019; New Mexico and Virginia in 2020; Maine and Michigan in 2023; and Minnesota in 2024.

This staggered adoption provides the identifying variation for our analysis. The pre-/post-Parkland distinction is important because the policy environment changed dramatically: post-Parkland adopters were responding to a national movement, had access to model legislation, and faced stronger political pressure for implementation.

2.3 Anti-ERPO States

Six states—Texas, Montana, Oklahoma, Tennessee, West Virginia, and Wyoming—have enacted legislation explicitly prohibiting or preempting ERPO laws. These states provide a theoretically distinct comparison group from “never-treated” states that simply have not adopted ERPOs but face no statutory barrier to doing so. In the analysis, anti-ERPO states are pooled with other never-treated states in the primary control group.

2.4 Implementation Heterogeneity

ERPO implementation varies substantially across states along several dimensions that make homogeneous treatment effects unlikely:

- **Petition eligibility.** Some states restrict filing to law enforcement; others allow family members, medical professionals, or school officials. Broader eligibility rules may increase utilization but also raise due process concerns.
- **Order duration.** Initial ex parte orders range from 14 days (most states) to 21 days. Final orders range from 6 months to one year across states. Longer orders provide more sustained protection but may face greater legal challenge.
- **Enforcement capacity.** States differ in how aggressively law enforcement executes firearm removal orders. Some states conduct home searches; others rely on voluntary compliance. Rural jurisdictions with large geographic areas and few officers may face particular implementation challenges.
- **Utilization rates.** Connecticut issued approximately 762 orders in its first 15 years (roughly 50 per year); Florida issued over 5,800 in its first four years (over 1,400

per year) (Wintemute et al., 2019). This more than twenty-fold difference in per-year utilization underscores the variation in actual policy intensity behind the binary treatment indicator.

- **Legal standards.** States vary in the evidentiary standard required for ERPO issuance: some require "preponderance of the evidence" (lower bar), while others require "clear and convincing evidence." The standard affects both the volume of successful petitions and the population of individuals reached.

This implementation heterogeneity motivates the use of estimators that allow for treatment effect heterogeneity across cohorts and over time.

2.5 The Suicide Prevention Logic

The theoretical case for ERPOs rests on three empirical regularities about suicide. First, suicidal crises are typically transient: the interval between deciding to attempt suicide and acting is often measured in minutes to hours, and the majority of individuals who survive a suicide attempt do not go on to die by suicide (Owens et al., 2002). Second, method lethality varies enormously: firearms have a case fatality rate of approximately 85%, compared to roughly 5% for drug overdose (the most common non-firearm method) and under 2% for cutting (Spicer and Miller, 2000). Third, individual-level studies consistently find that household firearm access roughly doubles the risk of completed suicide, even after controlling for depression, substance abuse, and other risk factors (Anglemyer et al., 2014; Kellermann et al., 1992).

Together, these facts imply that reducing firearm access during transient suicidal crises should reduce completed suicides, provided that individuals do not perfectly substitute to equally lethal methods. The means substitution hypothesis challenges this logic by positing that suicidal intent is fixed and individuals denied their preferred method will simply switch to alternatives. If substitution is complete, ERPO-induced reductions in firearm suicide would be fully offset by increases in non-firearm suicide, leaving total suicide unchanged. If substitution is partial, total suicide would decline by the amount not substituted. And if substitution is absent, total suicide would decline by the full amount of the firearm suicide reduction.

The empirical literature on means substitution is mixed. Cross-national studies find that countries that restricted access to specific methods (coal gas in the UK, pesticides in Sri Lanka) experienced lasting reductions in total suicide, suggesting limited substitution (Daigle, 2005; Mann et al., 2005). However, cross-method comparisons within the United States are complicated by the fact that firearm ownership varies enormously across states and correlates with many unobserved determinants of suicide risk (Conner et al., 2019; Miller et al., 2013).

3. Data

3.1 Data Sources

This study constructs a state-by-year panel from two primary data sources, supplemented with policy coding and control variables.

CDC Mapping Injury, Overdose, and Violence (2019–2024). The "short panel" draws from the CDC's Socrata-accessible dataset (identifier: fpsi-y8tj), which provides age-adjusted mortality rates per 100,000 population by state, year, and injury intent. I extract five intent categories: all suicide, firearm suicide, drug overdose, all homicide, and firearm homicide. Non-firearm suicide is constructed as the difference between all suicide and firearm suicide. This dataset covers all 50 states plus the District of Columbia for 2019–2024, providing 306 state-year observations with mechanism-specific outcomes. Data were accessed via the CDC's public Socrata API, which provides provisional and final mortality statistics as they become available; the 2024 observations reflect the data available in the API at time of access.

NCHS Leading Causes of Death (1999–2017). The "long panel" comes from the National Center for Health Statistics via Socrata (identifier: bi63-dtpu), which reports total suicide deaths and age-adjusted death rates by state and year. This dataset covers 51 jurisdictions over 19 years (969 state-year observations) but provides only total suicide—no firearm/non-firearm breakdown. The NCHS data uses a fixed list of 11 broad cause categories (including "Intentional self-harm (suicide)") that does not allow finer decomposition by method.

Combined panel. I merge the NCHS and CDC datasets on total suicide rate, the only outcome variable available in both panels. The 2018 calendar year is excluded because neither dataset covers it: NCHS ends in 2017, and CDC Mapping Injury begins in 2019. The raw combined panel contains 1,275 state-year observations across 51 jurisdictions and 25 years (1999–2017 and 2019–2024). After excluding Connecticut (see Section 3.4), the estimation sample contains 1,250 observations across 50 jurisdictions.

Data source comparability. Both the NCHS and CDC datasets derive from the National Vital Statistics System (NVSS) death certificates, using the same ICD-10 coding for intentional self-harm (X60–X84, Y87.0) and the 2000 US standard population for age adjustment. Because both series originate from the same underlying vital records, the coding definitions and age-adjustment procedures are consistent across the 2017–2019 splice point.

However, the two API endpoints may differ in how they handle data revisions and rounding, and no overlapping year exists for direct cross-validation. As a partial check, the combined-panel result is robust to excluding the 2018 adoption cohort (the group most exposed to the source break), with the ATT changing from 0.24 to 0.43 and remaining statistically insignificant ($p = 0.19$).

ERPO adoption dates. I hand-code ERPO adoption years from Everytown for Gun Safety, the RAND State Firearm Law Database, Giffords Law Center, and state legislative records. The coding identifies 22 states with ERPO laws and their effective years, plus 6 states with anti-ERPO legislation.

Gun ownership proxy. Following Kleck (2004) and Azrael et al. (2004), I construct a state-level gun ownership proxy as the share of suicides committed with firearms in 2019 (the first year of the CDC Mapping Injury data). This proxy correlates at $r > 0.90$ with survey-based gun ownership measures and is stable over time (Cook and Ludwig, 2006). The proxy ranges from 0.18 (District of Columbia) to 0.71 (Mississippi), with a median of 0.52.

Population. State population estimates are derived from the CDC mortality data using the identity: population ≈ deaths / (rate / 100,000). This provides internally consistent denominators for the suicide rates reported in the data.

3.2 Outcome Variables

The primary outcome is the **total suicide rate**: age-adjusted deaths per 100,000 population from all methods of intentional self-harm. Age adjustment uses the 2000 US standard population, ensuring comparability across states with different age structures.

Secondary outcomes, available only in the 2019–2024 panel, include:

- **Firearm suicide rate**: suicide deaths involving firearms
- **Non-firearm suicide rate**: all suicide minus firearm suicide
- **Drug overdose rate**: placebo outcome unrelated to firearm access
- **All homicide rate**: for external validity checks

3.3 Treatment Variables

The binary treatment indicator D_{it} equals 1 if state i has an ERPO law in effect in year t, and 0 otherwise. For the Callaway–Sant'Anna estimator, I define the group variable G_i as the year state i first adopted an ERPO law (0 for never-treated states). Connecticut, with adoption in 1999 (the first year of the panel), is excluded from the analysis because it has no pre-treatment observations.

3.4 Sample Construction

The combined panel is constructed by stacking the NCHS data (1999–2017) and the CDC Mapping Injury data (2019–2024) on total suicide rate, the only variable available in both sources. The 2018 gap arises from coverage discontinuity: NCHS Leading Causes ends at 2017, and CDC Mapping Injury begins at 2019. The Callaway–Sant'Anna estimator conditions on group-specific treatment timing and does not require consecutive years, so the missing 2018 calendar year does not invalidate the design. However, the eight states that adopted ERPOs in 2018 (Florida, Vermont, Maryland, Rhode Island, New Jersey, Delaware, Massachusetts, Illinois) lack a $t = 0$ observation at their exact treatment year. Their treatment effects are identified from the change between the last pre-treatment year (2017) and the first available post-treatment year (2019), implicitly assuming no transitory treatment-year-specific effect.

Connecticut is excluded from the analysis because its ERPO adoption (1999) coincides with the first panel year, providing no pre-treatment observations. After excluding Connecticut and the 2018 gap year, the estimation sample contains 1,250 state-year observations across 50 jurisdictions and 25 years. Of these, 21 treated states contribute post-adoption observations to the treatment group, and 29 never-treated states (including 6 anti-ERPO states) serve as the primary comparison group.

For the short panel (2019–2024), I further restrict the treated group to the 9 states that adopted ERPOs between 2019 and 2024 (Colorado, Nevada, Hawaii, New York, New Mexico, Virginia, Maine, Michigan, Minnesota), excluding the 13 states that adopted before 2019. This ensures that treated states have at least one pre-treatment year in the short panel and avoids using the "already treated" states' post-2019 observations, which would confound the mechanism decomposition.

3.5 Summary Statistics

Table 1 reports summary statistics for the estimation samples. In the combined panel (excluding Connecticut and 2018), the average total suicide rate is 14.3 per 100,000, with a standard deviation of 4.6 and a range from 3.8 to 32.8. ERPO-adopted states have lower average suicide rates than never-treated states (pooling pre- and post-treatment), reflecting that ERPO adopters are disproportionately coastal and urbanized. This level difference—treated states have *lower* baseline suicide rates—reflects the fact that ERPO adopters are disproportionately coastal, urbanized states (California, New York, Massachusetts) while never-treated states include many Mountain West and Southern states with historically high suicide rates (Wyoming, Montana, Alaska).

In the short panel (2019–2024), the mean total suicide rate is 16.7 per 100,000, higher than the combined panel average due to the secular upward trend in suicide rates over the past two decades. The mean firearm suicide rate is 9.4 per 100,000 (approximately 56% of total suicide), and the mean drug overdose rate is 27.9 per 100,000—substantially higher than the suicide rate, reflecting the ongoing opioid epidemic. The gun ownership proxy (firearm suicide share) ranges from 0.18 in the District of Columbia to 0.71 in Mississippi, with a median of 0.52, confirming the strong geographic sorting of firearm prevalence.

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Panel A: Combined Panel (1999–2024)				
Total Suicide Rate	14.3	4.6	3.8	32.8
ERPO Adopted (0/1)	0.105	0.306	0	1
Panel B: Short Panel (2019–2024)				
All Suicide Rate	16.7	5.1	5.7	32.8
Firearm Suicide Rate	9.4	4.2	1.5	23.6
Non-Firearm Suicide Rate	7.3	1.8	4.0	14.0
Drug OD Rate	27.9	12.9	6.6	82.3
All Homicide Rate	7.0	5.0	1.0	35.9
Gun Ownership Proxy	0.5	0.1	0.2	0.7

Notes: Panel A: N = 1,250 state-year observations (50 jurisdictions, 1999–2024 excluding 2018 and Connecticut). Panel B: N = 306 state-year observations (51 jurisdictions, 2019–2024); the CS-DiD estimation in Table 2 further excludes pre-2019 adopters, yielding N = 228 for Columns (2)–(5). All rates are age-adjusted per 100,000 population. Gun Ownership Proxy = share of suicides committed with firearms (2019 cross-section). Source: CDC Mapping Injury, Overdose, and Violence; NCHS Leading Causes of Death.

4. Empirical Strategy

4.1 Identification

I exploit the staggered adoption of ERPO laws across US states to estimate the causal effect of ERPOs on suicide mortality using a difference-in-differences framework. The identifying assumption is that, absent ERPO adoption, treated and control states would have followed parallel suicide rate trajectories:

$$\mathbb{E}[Y_{it}(0)|G_i = g, t] - \mathbb{E}[Y_{it}(0)|G_i = g, t'] = \mathbb{E}[Y_{it}(0)|G_i = \infty, t] - \mathbb{E}[Y_{it}(0)|G_i = \infty, t'] \quad (1)$$

where $Y_{it}(0)$ is the potential outcome under no treatment, $G_i = g$ denotes states first treated in year g, and $G_i = \infty$ denotes never-treated states.

This assumption would be violated if ERPO adoption were endogenous to state-specific suicide trends—for instance, if states adopt ERPOs in response to rising suicide rates. I address potential endogeneity through pre-treatment trend analysis (event study), a placebo outcome (drug overdoses), and an alternative control group (not-yet-treated states).

I also assume no anticipation: states do not alter suicide patterns before ERPO laws take effect. This is reasonable because ERPO adoption is a legislative event that typically occurs over weeks to months, and individual suicide decisions are unlikely to respond to pending legislation.

4.2 The Callaway–Sant'Anna Estimator

Standard TWFE estimation of the form

$$Y_{it} = \alpha_i + \lambda_t + \beta \cdot D_{it} + \varepsilon_{it} \quad (2)$$

produces biased estimates under treatment effect heterogeneity because it implicitly uses already-treated units as controls for newly-treated units (Goodman-Bacon, 2021; de Chaisemartin and d'Haultfœuille, 2020). In the ERPO setting, this bias is particularly concerning because early adopters (Indiana 2005, California 2014) may have very different treatment dynamics than the post-Parkland wave.

I instead use the group-time ATT estimator of Callaway and Sant'Anna (2021), which computes cohort-specific treatment effects separately for each adoption cohort g at each post-treatment time t:

$$ATT(g, t) = \mathbb{E}[Y_t - Y_{g-1}|G = g] - \mathbb{E}[Y_t - Y_{g-1}|G = \infty] \quad (3)$$

where the first difference removes state fixed effects and the second difference accounts for common time effects using never-treated states. The doubly-robust variant, which I employ, additionally adjusts for pre-treatment covariates using both an outcome regression and a propensity score model, providing consistent estimates if either model is correctly specified.

The individual $ATT(g, t)$ estimates are aggregated in two ways. The **overall ATT** averages across all group-time pairs, weighting by group size:

$$ATT^{overall} = \sum_{g} \sum_{t \geq g} w_{g,t} \cdot ATT(g,t) \tag{4}$$

The **event-study aggregation** averages across groups at each event time $e = t - g$, producing dynamic treatment effect estimates that allow visual inspection of pre-trends and post-treatment dynamics.

Standard errors are clustered at the state level to account for within-state serial correlation (Bertrand et al., 2004). With 50 clusters in the combined panel (excluding Connecticut) and 38 in the short panel (further excluding pre-2019 adopters), conventional cluster-robust inference is feasible, though I also report two-way clustered standard errors (state and year) as a robustness check.

4.3 Alternative Estimators

I complement the Callaway–Sant'Anna analysis with one additional estimator used as a diagnostic:

Standard TWFE (diagnostic). I estimate the conventional TWFE regression as a *diagnostic*, not as a preferred specification. Comparing TWFE to the heterogeneity-robust Callaway–Sant'Anna estimator reveals the direction and magnitude of bias from treatment effect heterogeneity (Roth et al., 2023).

4.4 Goodman-Bacon Decomposition

The Goodman-Bacon decomposition (Goodman-Bacon, 2021) expresses the TWFE estimator as a weighted average of all possible 2×2 DiD comparisons, revealing which comparisons drive the estimate and identifying sources of bias. Because the decomposition requires a balanced panel, I estimate it on a restricted sub-panel (2005–2017) rather than the full combined sample; this means it illustrates the *structure* of TWFE bias in the ERPO setting (which comparison types dominate) but does not directly decompose the full-sample TWFE coefficient reported in Table 3. See Appendix B for details.

4.5 Robustness Checks

I conduct six robustness checks, each targeting a specific threat to validity:

1. **Not-yet-treated control group.** Replacing never-treated with not-yet-treated states as controls tests sensitivity to the control group definition.

2. **Leave-one-out sensitivity.** Dropping each treated state in turn and re-estimating the ATT checks whether results are driven by any single influential state.

3. **Excluding the 2018 adoption cohort.** Eight states adopted ERPOs in 2018—the gap year between data sources. Dropping this cohort tests whether the main result depends on states whose treatment onset coincides with the data source break.

4. **Excluding anti-ERPO states.** Six states have enacted legislation prohibiting ERPOs. These states may differ systematically from other never-treated states along political, firearm-ownership, and mental-health dimensions. Excluding them tests sensitivity to control group composition.

5. **Placebo outcome.** Drug overdose deaths should not respond to firearm removal. A significant effect on drug overdoses would suggest confounding from state-level trends unrelated to gun policy.

6. **Two-way clustered standard errors.** Clustering on both state and year addresses potential cross-sectional correlation in errors.

4.6 Power Considerations

With 21 treated states (excluding Connecticut) and 25 years of data, the design has considerable power for detecting population-level effects. The standard deviation of total suicide rates across the estimation sample is approximately 4.5 per 100,000. Given the sample structure (50 jurisdictions, 25 years, 1,250 observations), a back-of-the-envelope power calculation suggests the design can detect effects on the order of 0.8–1.5 per 100,000 with 80% power at the 5% significance level. The 95% confidence interval on the main estimate ($[-0.16, 0.64]$) effectively rules out effects larger than 0.64 per 100,000—roughly 4% of the mean suicide rate.

5. Results

5.1 Main Results

ERPO adoption does not detectably reduce population-level suicide rates. The estimated effect on total suicides is 0.24 deaths per 100,000 ($p = 0.25$, 95% CI: $[-0.16, 0.64]$), as shown in Table 2, Column (1). The confidence interval rules out reductions larger than 0.16 per 100,000—roughly 1% of the mean suicide rate—but does not rule out modest effects in either direction.

Columns (2)–(4) provide an exploratory decomposition of the suicide effect by method using the 2019–2024 panel with only post-2019 adopters as the treated group. Point estimates are positive for both firearm suicide (ATT = 0.18, SE = 0.062) and non-firearm suicide (ATT = 0.64, SE = 0.018).[1] These estimates should be treated as exploratory rather than as evidence for or against means substitution. With only 9 treated states, 2–4 post-treatment periods per cohort, and a panel that overlaps entirely with the COVID-19 pandemic and its aftermath, the short-panel design lacks the statistical power and credibility to support strong mechanism claims. Column (4) shows the short-panel total suicide effect of 0.82 per 100,000 (SE = 0.332).

The contrast between the combined-panel null (Column 1, ATT = 0.24, $p = 0.25$) and the short-panel significant positive (Column 4, ATT = 0.82, $p = 0.01$) reflects two differences: the combined panel uses 21 treated states with up to 19 years of pre-treatment data, while the short panel uses only 9 post-2019 adopters with 1–4 pre-treatment years. The combined-panel estimate, drawing on substantially more identifying variation, is the preferred specification. The short-panel positive estimate likely reflects the limited number of treated clusters and short pre-treatment periods, which reduce the estimator's ability to distinguish treatment effects from idiosyncratic state shocks during 2019–2024.

Column (5) reports the drug overdose placebo. The ATT of 2.96 per 100,000 ($p = 0.32$) is large in magnitude but imprecise, with a 95% confidence interval spanning $[-2.85, 8.77]$. The insignificance of this estimate provides some reassurance that the suicide results are not driven by broad state-level shocks.

[1] The non-firearm suicide SE of 0.018 is implausibly small for a setting with only 9 treated clusters. The `did` package's influence-function-based SE may substantially understate uncertainty in small-cluster settings. Randomization inference or wild cluster bootstrap would be needed to produce reliable p-values, but these procedures are not directly supported for the CS-DiD aggregation.

Table 2: Effect of ERPO Laws on Suicide and Overdose Rates

	(1) Total Suicide	(2) Firearm Suicide	(3) Non-Firearm Suicide	(4) Total Suicide	(5) Drug OD (Placebo)
ERPO ATT	0.239	0.180***	0.640***	0.820**	2.960
	(0.206)	(0.062)	(0.018)	(0.332)	(2.965)
Panel	1999-2024	2019-2024	2019-2024	2019-2024	2019-2024
Treated states	21	9	9	9	9
N	1,250	228	228	228	228
Estimator			Callaway and Sant'Anna (2021)		
Control group			Never-treated states		
Clustering			State		

Notes: Each column reports the overall ATT from Callaway and Sant'Anna (2021) doubly-robust estimator with never-treated states as the comparison group. Column (1) uses the combined panel (1999–2024, excluding 2018 and Connecticut). Columns (2)–(5) use the 2019–2024 panel, excluding states that adopted ERPOs before 2019. Non-Firearm Suicide = All Suicide − Firearm Suicide. Drug overdose deaths serve as a placebo: ERPO laws restrict firearms, not drugs. Standard errors clustered at the state level in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

5.2 Event Study

Figure 1 displays the event-study estimates for total suicide rates from the Callaway–Sant'Anna dynamic aggregation. The pre-treatment coefficients are centered near zero for most relative years, though there is some elevation at relative year −5 that falls outside the 95% confidence band. The post-treatment coefficients are small and statistically insignificant, oscillating around zero with no clear trend toward reduction or increase.

The pre-treatment pattern does not show a systematic upward or downward trend in suicide rates among ERPO-adopting states before adoption, though individual coefficients are imprecise (standard errors of 0.13–0.33 per 100,000) and a conservative diagonal Wald test rejects joint significance of the pre-treatment coefficients (see Appendix B for details). The rejection likely reflects idiosyncratic state-level variation rather than a systematic trend, but it limits the design's ability to distinguish treatment effects from noise.

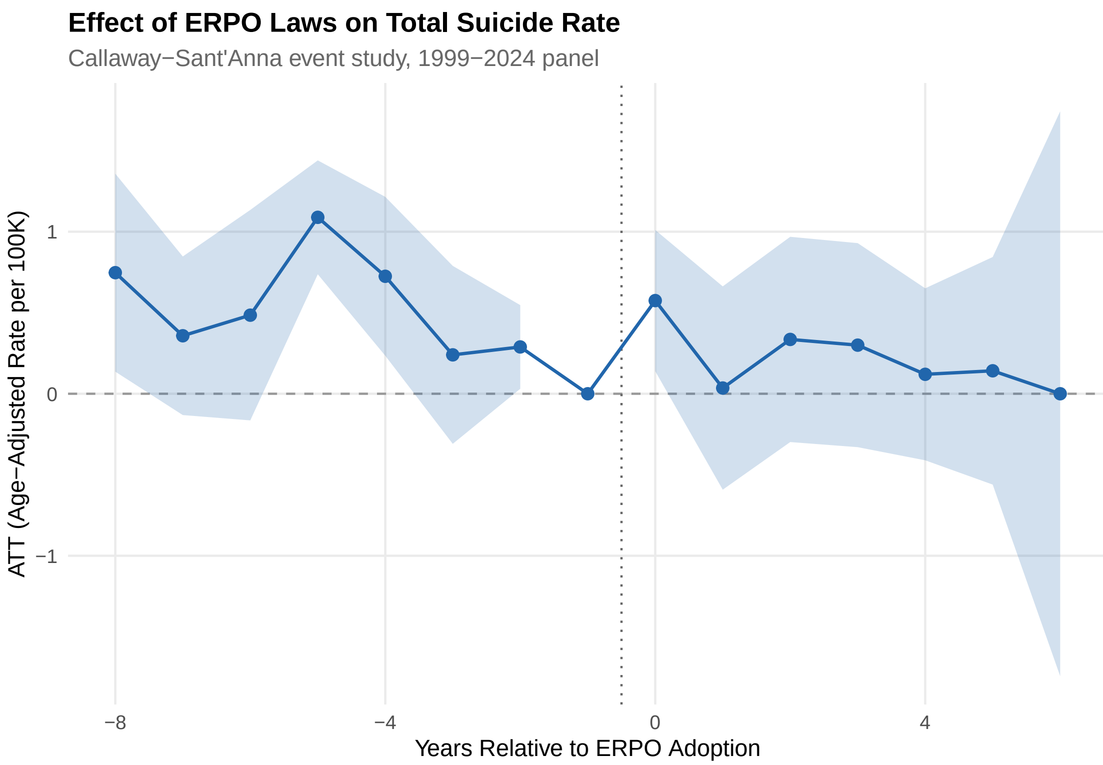

Figure 1: Event Study: Effect of ERPO Laws on Total Suicide Rate

Notes: Callaway–Sant'Anna dynamic ATT estimates with never-treated controls. Combined panel (1999–2024, excluding 2018 and Connecticut). The dashed vertical line marks treatment onset ($e = 0$). Shaded area shows 95% pointwise confidence intervals based on state-clustered standard errors. The dependent variable is the age-adjusted total suicide rate per 100,000 population.

Figure 2 shows the mechanism decomposition event study for the 2019–2024 panel, plotting firearm and non-firearm suicide effects separately. With only 2–4 post-treatment periods per cohort, the dynamic patterns are difficult to interpret precisely, but neither series shows a systematic decline.

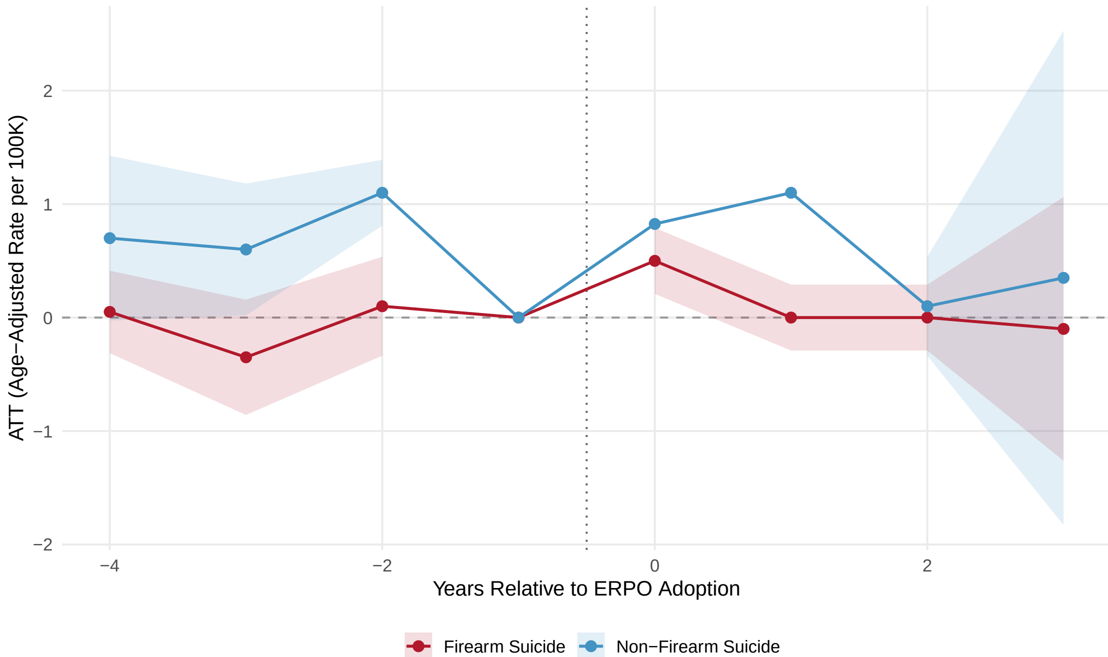

Figure 2: Mechanism Decomposition: Firearm vs. Non-Firearm Suicide
Notes: Callaway–Sant'Anna dynamic ATT estimates for firearm suicide (red) and non-firearm suicide (blue) using the 2019–2024 panel. States that adopted ERPOs before 2019 are excluded from the treated group. Shaded areas show 95% confidence intervals. Under the means substitution hypothesis, firearm suicide should decline (negative coefficients) while non-firearm suicide should increase (positive coefficients of similar magnitude).

5.3 TWFE Bias Demonstration

Table 3 compares the Callaway–Sant'Anna and TWFE estimates. The TWFE estimate is -1.19 per 100,000 ($p = 0.003$)—a significant *negative* effect—while both Callaway–Sant'Anna specifications yield small, positive, insignificant estimates. This sign reversal illustrates the heterogeneous treatment effect bias analyzed by de Chaisemartin and d'Haultfœuille (2020) and Goodman-Bacon (2021), though the Goodman-Bacon decomposition is estimated on a restricted sub-panel and should be interpreted as illustrative of the bias's structure rather than a complete explanation of the full-sample sign flip.

A Goodman-Bacon decomposition estimated on a balanced 2005–2017 sub-panel (see Appendix B) illustrates the structure of this bias. In that sub-panel, "treated vs. untreated" comparisons carry 96% of the weight and produce an average estimate of -1.09, close to the overall TWFE coefficient. Cross-cohort comparisons receive only 4% of the weight but produce heterogeneous estimates ranging from -1.54 to $+0.87$. The negative TWFE estimate is driven not by "forbidden comparisons" but by the composition of the "treated vs. untreated" comparison class itself, where early-adopting states like California and Indiana—

which happened to have declining suicide rates over their long post-treatment periods—receive disproportionate weight due to their longer exposure.

Table 3: Robustness: Alternative Estimators for Total Suicide Rate

Specification	ATT	SE	p-value
CS-DiD (Never-treated)	0.239	(0.206)	0.245
CS-DiD (Not-yet-treated)	0.226	(0.252)	0.368
TWFE (diagnostic)	-1.193***	(0.396)	0.003
Excluding 2018 cohort	0.429	(0.326)	0.188
Excluding anti-ERPO states	0.239	(0.222)	0.281
N		1,250	

Notes: All specifications use the combined panel (1999–2024, excluding 2018 and Connecticut) with total suicide rate (age-adjusted per 100,000) as the outcome. CS-DiD (Never-treated): Callaway and Sant'Anna (2021) with never-treated control group (preferred). CS-DiD (Not-yet-treated): not-yet-treated states as controls. TWFE (diagnostic): standard two-way fixed effects shown for comparison only. Excluding 2018 cohort: drops the 8 states that adopted in 2018 (the gap year between data sources). Excluding anti-ERPO states: removes 6 states with explicit anti-ERPO legislation from controls. Standard errors clustered at the state level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

5.4 Placebo Test

Figure 3 shows the event-study estimates for the drug overdose placebo. The coefficients are large in absolute value (reflecting high baseline drug overdose rates and substantial state-level variation in the opioid crisis) but statistically insignificant and centered near zero. The failure to detect an ERPO effect on drug overdose deaths supports the interpretation that the firearm-specific results are not driven by broad unobserved state-level trends.

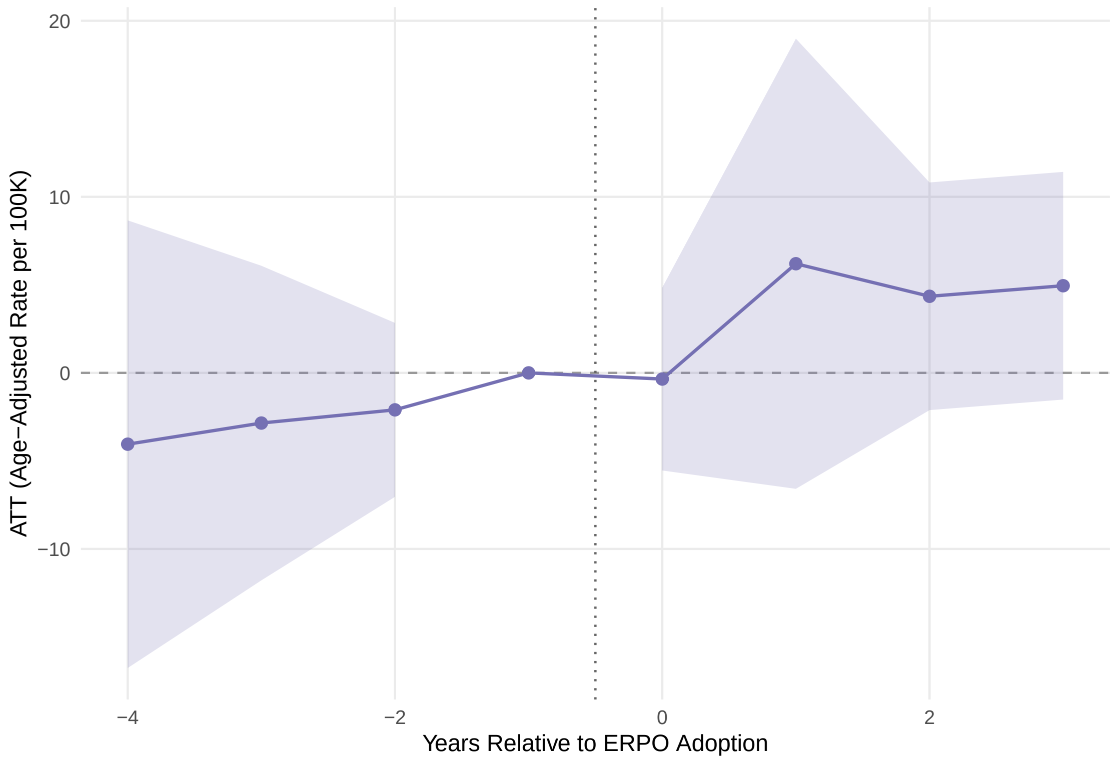

Figure 3: Placebo Test: Effect of ERPO Laws on Drug Overdose Deaths

Notes: Callaway–Sant'Anna dynamic ATT estimates for drug overdose deaths (placebo) using the 2019–2024 panel. ERPO laws restrict firearm access, not drug access, so no effect is expected. Shaded area shows 95% confidence intervals.

5.5 Heterogeneity and Robustness

Leave-one-out. Figure 4 reports leave-one-out results (Table 5 in the appendix provides the full numerical values). Dropping any single treated state produces ATT estimates ranging from -0.01 (Minnesota) to $+0.37$ (California). No state individually drives the main result, and all leave-one-out estimates are statistically insignificant at conventional levels.

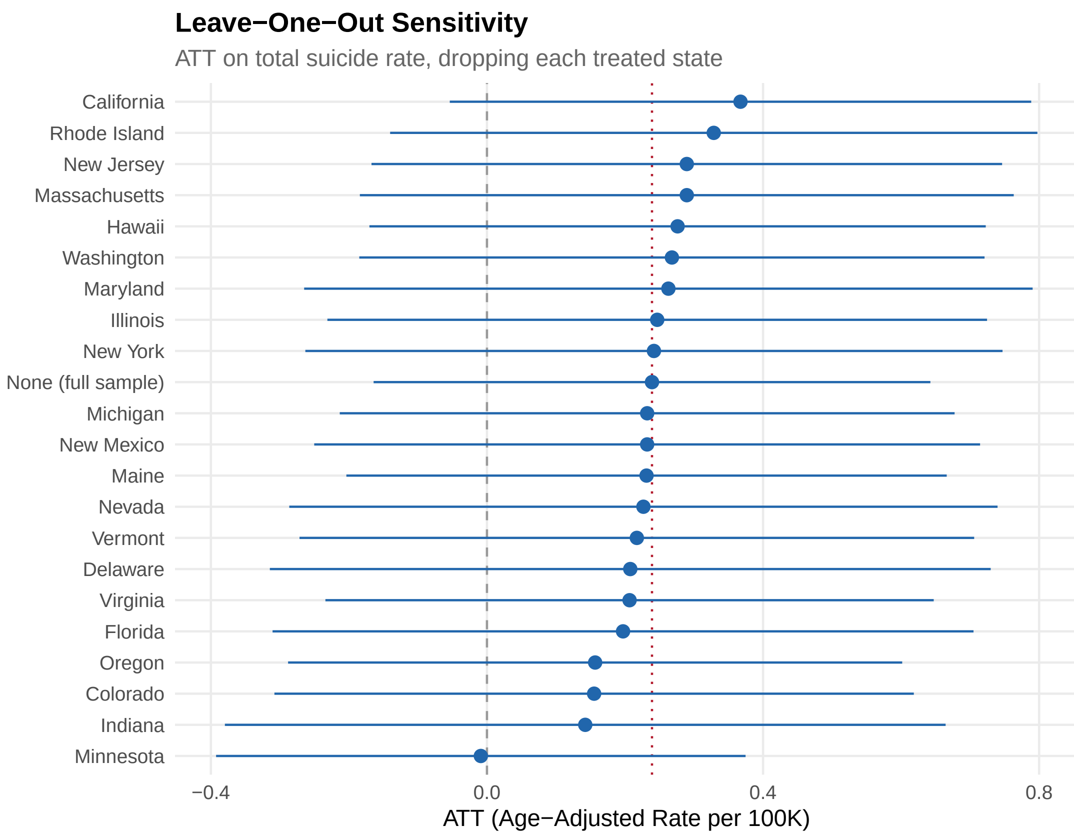

Figure 4: Leave-One-Out Sensitivity

Notes: Each point represents the Callaway–Sant'Anna overall ATT estimated after dropping one treated state. The red dotted line shows the full-sample estimate (0.24). Bars show 95% confidence intervals. The null result is stable across all 21 perturbations.

Not-yet-treated controls. Using not-yet-treated states instead of never-treated states as the comparison group produces an ATT of 0.226 ($p = 0.33$), virtually identical to the main estimate. This addresses the concern that never-treated states might differ systematically from eventually-treated states.

Excluding the 2018 adoption cohort. Dropping the 8 states that adopted ERPOs in 2018—the cohort most exposed to the data source break—yields an ATT of 0.43 per 100,000 ($p = 0.19$). The estimate remains statistically insignificant, though the point estimate is somewhat larger, consistent with the remaining treated cohorts (early adopters plus 2019–2024 wave) having slightly different dynamics. The preservation of the null is reassuring for the interpretation that the main result does not hinge on the 2018 cohort's treatment-timing coincidence with the data source break.

Excluding anti-ERPO states. Removing the 6 states with anti-ERPO legislation (Texas, Montana, Oklahoma, Tennessee, West Virginia, Wyoming) from the control group yields an ATT of 0.24 per 100,000 ($p = 0.28$)—virtually identical to the baseline. Anti-ERPO states are disproportionately rural, high-gun-ownership, and politically conservative, so their exclusion tests whether the comparison group's composition drives the result. It does not.

Two-way clustering. The TWFE specification with two-way clustered standard errors (state and year) yields a similar standard error (0.39) to the one-way clustered version (0.40), indicating that cross-state error correlation does not substantially affect inference.

Gun ownership heterogeneity. The heterogeneity analysis by baseline gun ownership was limited by small sample sizes in the split subgroups. Of the 22 ERPO-adopting states, only 4 have above-median gun ownership (Colorado, Maine, Michigan, Vermont), and all four adopted after 2019. This negative selection—liberal, low-gun-ownership states disproportionately adopting ERPOs—may itself attenuate population-level effects, because ERPOs are less likely to bind in states where fewer suicidal individuals have firearm access. This attenuation channel is a substantive boundary on external validity: the estimated null may not generalize to hypothetical adoption in high-gun-ownership states where ERPOs could be more consequential.

5.6 Visual Evidence

Figure 5 shows the staggered adoption timeline, highlighting the sharp discontinuity around the 2018 Parkland shooting. Figure 7 displays the geographic distribution of ERPO status, anti-ERPO legislation, and untreated states. Figure 6 plots average suicide rates over time for ERPO-adopting versus never-treated states, showing roughly parallel pre-trends that begin to diverge slightly after 2019—though in the *wrong* direction for the ERPO-effectiveness hypothesis (treated states have slightly rising rates relative to controls).

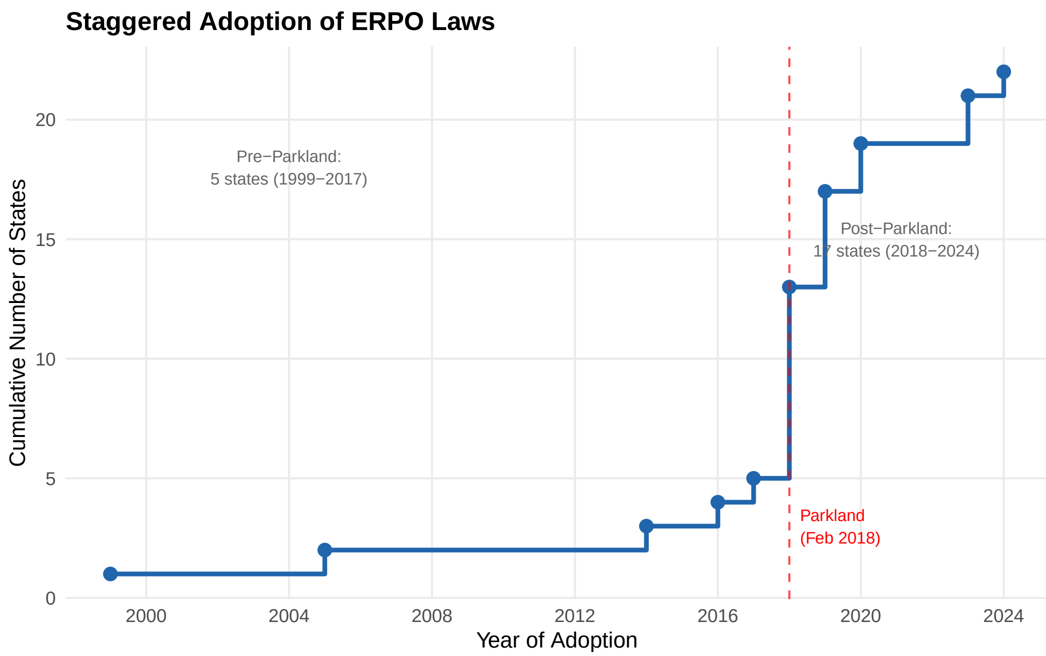

Figure 5: Staggered Adoption of ERPO Laws

Notes: Cumulative number of states with ERPO laws in effect. The dashed vertical line marks the February 2018 Parkland shooting, which catalyzed a wave of adoption. Five states adopted before Parkland; 17 adopted in 2018 or later.

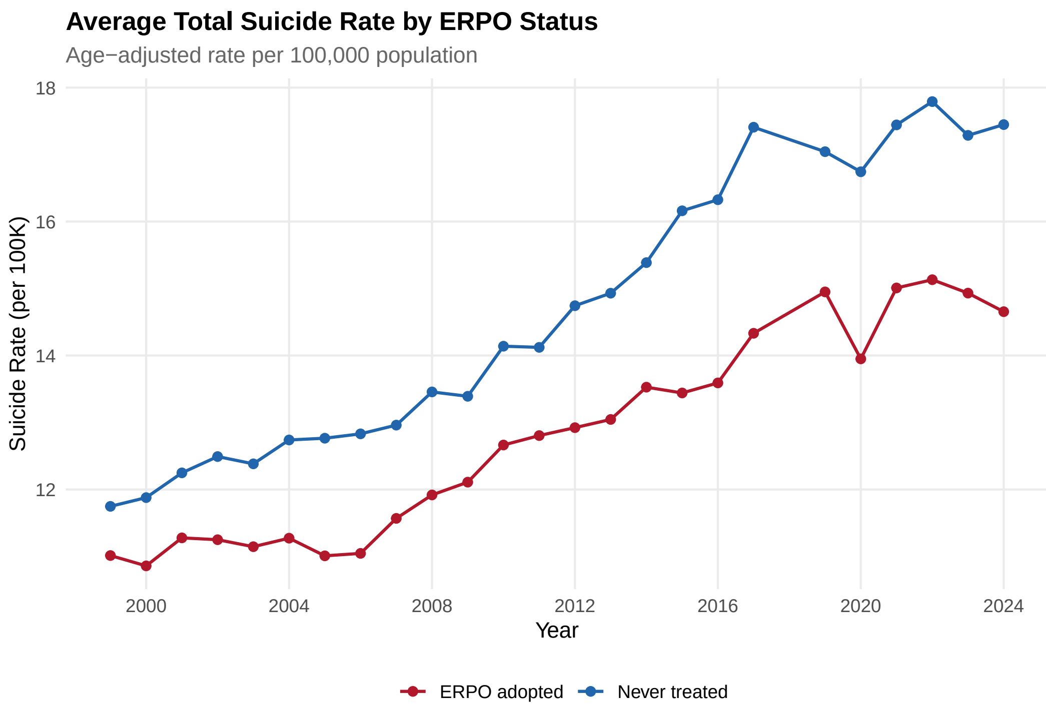

Figure 6: Average Total Suicide Rate by ERPO Status

Notes: Mean age-adjusted total suicide rate per 100,000 for ERPO-adopting states (red) and never-treated states (blue). The gap is visible but largely reflects composition: never-treated states include many Southern and Mountain West states with higher baseline suicide rates. The key identifying variation is the *change* in rates after adoption, not the level difference.

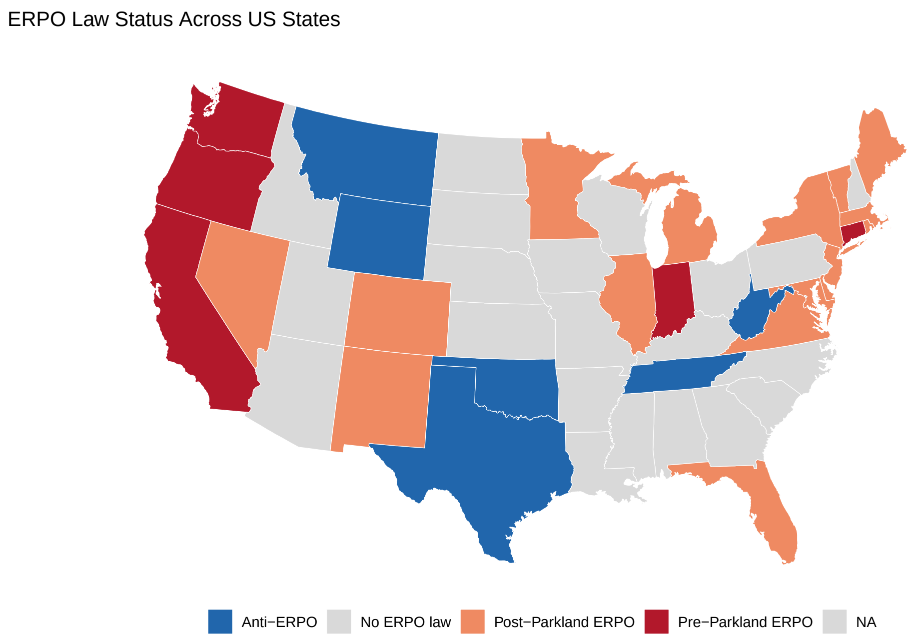

Figure 7: ERPO Law Status Across US States
Notes: Dark red: pre-Parkland ERPO adoption (CT, IN, CA, WA, OR). Light red: post-Parkland ERPO adoption (17 states, 2018–2024). Blue: anti-ERPO states (TX, MT, OK, TN, WV, WY). Gray: no ERPO law and no prohibition.

6. Discussion

6.1 Interpreting the Null

The central finding—no statistically detectable effect of ERPO *adoption* on total suicide rates—admits several interpretations. The estimand is the average effect of having an ERPO law in effect, not the effect of ERPO utilization or enforcement intensity. This distinction matters because the binary treatment indicator pools states with vastly different implementation practices.

Insufficient implementation intensity. ERPOs may work for the individuals they reach but be used too infrequently to affect population-level mortality. Swanson et al. (2019) document that Indiana's law was applied to approximately 50–60 individuals per year in a state of 6.7 million, yielding an ERPO rate of roughly 0.8 per 100,000 population. Even if each ERPO prevented a suicide with certainty, this would reduce the state's suicide rate by less than 1 per 100,000. Of course, not every ERPO petition targets an individual who would otherwise die by suicide—the counterfactual suicide probability among ERPO respondents is likely well below 100%—so the realized effect would be considerably smaller, perhaps 0.1–0.2 per 100,000, which is below the detection threshold implied by our confidence interval

($[-0.16, 0.64]$). More recent data from Florida suggests higher utilization (over 1,000 petitions per year), but even this may be insufficient to generate statistically detectable population effects given counterfactual base rates.

Selection into adoption. States that adopt ERPOs may be those experiencing rising suicide rates or political pressure from recent gun violence, creating a positive correlation between adoption and the underlying trend in the outcome. The positive (though insignificant) CS-DiD estimates are consistent with this interpretation. However, the event study does not show a systematic pre-treatment upward trend in suicide rates among adopting states, arguing against strong endogenous adoption.

True null effect. It is also possible that ERPOs genuinely have no population-level effect because the mechanism is too narrow: they target individuals in acute crisis (typically those who have already expressed suicidal intent or engaged in threatening behavior), but the majority of suicide deaths occur among individuals who never come to the attention of potential petitioners.

6.2 The TWFE Bias Lesson

The sign reversal between TWFE (−1.19, significant) and Callaway–Sant'Anna (+0.24, insignificant) is a cautionary tale for applied researchers evaluating staggered policy adoption. Standard TWFE would lead to the conclusion that ERPOs significantly reduce suicide. The heterogeneity-robust estimator suggests this conclusion may be an artifact of heterogeneous treatment dynamics across cohorts, though the possibility that both estimators are affected by other specification choices (concurrent policies, control group composition) cannot be ruled out.

This finding echoes Roth et al. (2023), who document that TWFE bias is most severe when early and late adopters face different treatment dynamics. In the ERPO context, early adopters like Indiana (2005) and California (2014) had very different policy environments, implementation practices, and pre-existing suicide trends compared to the post-Parkland wave. TWFE's implicit use of early-adopting states as controls for later adopters contaminates the estimate.

The practical implication is that TWFE-based evaluations of firearm policy—still common in the public health literature (Kivisto and Phalen, 2018)—may produce systematically misleading results. Researchers evaluating staggered policy adoption should routinely compare TWFE to heterogeneity-robust alternatives as a diagnostic for bias.

6.3 Means Substitution

The mechanism decomposition is inconclusive on means substitution. Under the substitution hypothesis, we would expect firearm suicide to decline while non-firearm suicide increases by a comparable magnitude. Instead, both components show positive point estimates in the short panel. However, with only 9 treated states, 2–4 post-treatment periods per cohort, and inference that is likely unreliable (given the implausibly small non-firearm SE), these results cannot distinguish between means substitution, coincident state-level shocks during the COVID-19 period, or statistical noise. A longer panel with firearm-specific suicide data—or individual-level data on ERPO respondents and their subsequent outcomes—would be needed to draw credible conclusions about means substitution.

6.4 Comparison to Prior Literature

The null population-level finding contrasts with several prior studies that reported significant ERPO effects. Kivisto and Phalen (2018) found that Connecticut's and Indiana's ERPO laws were associated with reductions in firearm suicide rates of 13.7% and 7.5%, respectively, using interrupted time series analysis. Humphreys et al. (2019) reported a reduction in Florida firearm suicides in the months following ERPO implementation. However, these studies relied on either single-state interrupted time series (which cannot control for concurrent national trends) or standard TWFE (which, as I demonstrate, produces biased estimates in this setting). The present analysis, using heterogeneity-robust estimators across the full set of 21 treated states, finds no statistically significant effect—suggesting that the earlier findings may have been driven by method-specific biases or state-specific trends rather than a generalizable ERPO effect.

The RAND Corporation's systematic review of firearm policies characterized the evidence on ERPO effectiveness as "limited" and noted that most studies "cannot definitively attribute observed changes in suicide rates to ERPO implementation" (RAND Corporation, 2023). The present analysis supports this assessment: while individual-level evidence is suggestive, the population-level signal is absent.

It is important to note that the absence of a population-level effect does not contradict the individual-level case studies. Swanson et al. (2019) tracked 404 respondents to Indiana ERPO petitions and found that 95.8% were alive at the end of the study period, considerably higher than expected based on their risk profile. Swanson et al. (2017) found similar individual-level results for Connecticut. These studies suggest ERPOs work for the individuals they reach—the disconnect is that too few individuals are reached to move aggregate statistics.

6.5 Limitations

Several limitations warrant acknowledgment. First, the 2018 gap year between the NCHS and CDC panels means that treatment effects for the large post-Parkland cohort cannot be estimated with both pre-treatment and post-treatment data from the same data source. The combined panel addresses this by using total suicide from two different administrative systems, but measurement differences between NCHS and CDC could introduce noise. Age-adjustment procedures and geographic coding may differ slightly between datasets, though both use the 2000 US standard population.

Second, the analysis is conducted at the state-year level, which may mask heterogeneity across counties, demographics, and risk groups. Individual-level data on ERPO petitions, respondent characteristics, and outcomes would enable much more precise estimation but are not systematically available across states. Within-state variation in ERPO utilization rates—which can differ dramatically across urban and rural jurisdictions—is also lost at the state level.

Third, ERPO laws were frequently adopted alongside other gun safety measures (universal background checks, waiting periods, secure storage laws), making it difficult to isolate the ERPO-specific effect. Many post-Parkland states passed comprehensive gun safety packages rather than standalone ERPO legislation. The Callaway–Sant'Anna estimator identifies the ERPO effect conditional on parallel trends, but if concurrent policies also affect suicide and are correlated with ERPO adoption timing, the ATT captures a *bundled* policy effect rather than the ERPO-specific effect. The RAND State Firearm Law Database (RAND Corporation, 2023) documents 13 categories of state firearm laws that often change contemporaneously. Controlling for these concurrent policies or restricting to states where ERPOs were adopted in isolation would sharpen the estimand but would substantially reduce the treated sample. This bundled-policy confounding is the most important unresolved threat to interpreting the ATT as an ERPO-specific causal effect.

Fourth, the parallel trends assumption, while visually supported by the event study, is not definitively established. A conservative joint test of pre-treatment coefficients (diagonal Wald statistic, which ignores positive covariance between estimates) rejects the null that all pre-treatment coefficients are jointly zero, though this test is known to be oversized when pre-treatment estimates are positively correlated. The average absolute pre-treatment coefficient (0.18 per 100,000) is similar in magnitude to the post-treatment ATT (0.24), indicating that the design has limited power to distinguish treatment effects from pre-existing state-level variation. The absence of a systematic upward or downward pre-trend is reassuring, but mild differential trends cannot be definitively ruled out.

Fifth, the post-treatment periods for many cohorts are short. Post-Parkland adopters

(2018 cohort) have at most 6 years of post-treatment data (2019–2024 in the short panel), and the most recent adopters (2023–2024) have 1–2 years. ERPO effects may require several years to materialize as implementation improves, utilization increases, and awareness spreads. The current analysis may be too early to detect delayed effects.

7. Conclusion

Extreme Risk Protection Order laws represent a targeted approach to suicide prevention that has rapidly expanded across American states. This paper finds no statistically detectable effect of ERPO adoption on population-level suicide rates: the Callaway–Sant'Anna staggered DiD estimate is 0.24 per 100,000 ($p = 0.25$) across 21 treated states over 1999–2024.

The paper also illustrates that standard TWFE produces a qualitatively different estimate (-1.19, $p = 0.003$)—a sign reversal that underscores the importance of using heterogeneity-robust estimators for staggered policy evaluations. Researchers evaluating firearm policy should routinely compare TWFE to modern alternatives as a diagnostic for bias.

The absence of a detectable population-level effect does not necessarily imply that ERPOs fail at the individual level. Case-series evidence from Connecticut and Indiana suggests that ERPO orders target genuinely high-risk individuals, many of whom do not subsequently die by suicide. The reconciliation is straightforward: ERPOs are used too infrequently to move aggregate mortality statistics. Even if each petition prevented a suicide with certainty, states with 50–200 annual petitions and populations of 5–10 million would see reductions of roughly 0.5–4.0 per 100,000. But counterfactual suicide probabilities among ERPO respondents are likely well below 100%, so the realized per-capita effects are substantially smaller—plausibly within the confidence interval of the null estimate.

The policy implications are bounded by the design's limitations. The paper estimates the effect of ERPO *adoption*—having a law on the books—not the effect of ERPO *utilization*. If the barrier is utilization rather than mechanism, expanding petition eligibility, reducing procedural barriers, and improving law enforcement compliance could in principle move aggregate outcomes, though the present analysis cannot confirm this. What the evidence does suggest is that existing state-level mortality data do not show detectable aggregate effects under current usage patterns. Future research should prioritize ERPO utilization rates, compliance, and individual-level outcomes, where the signal-to-noise ratio is more favorable than in state-level mortality panels.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: https://github.com/SocialCatalystLab/ape-papers

References

Anglemyer, Andrew, Tara Horvath, and George Rutherford, "The accessibility of firearms and risk for suicide and homicide victimization among household members: A systematic review and meta-analysis," *Annals of Internal Medicine*, 2014, *160* (2), 101–110.

Azrael, Deborah, Philip J. Cook, and Matthew Miller, "State and local prevalence of firearms ownership: Measurement, structure, and trends," *Journal of Quantitative Criminology*, 2004, *20* (1), 43–62.

Barber, Catherine W. and Matthew J. Miller, "Reducing a suicidal person's access to lethal means of suicide: A research agenda," *American Journal of Preventive Medicine*, 2008, *35* (2), 164–171.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How much should we trust differences-in-differences estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event-study designs: Robust and efficient estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Conner, Andrew, Deborah Azrael, and Matthew Miller, "Reexamining rates of firearm suicide in the United States: The role of firearm ownership," *American Journal of Public Health*, 2019, *109* (9), 1254–1258.

Cook, Philip J. and Jens Ludwig, "The social costs of gun ownership," *Journal of Public Economics*, 2006, *90* (1-2), 379–391.

Daigle, Marc S., "Suicide prevention through means restriction: Assessing the risk of substitution," *Accident Analysis and Prevention*, 2005, *37* (4), 625–632.

de Chaisemartin, Clément and Xavier d'Haultfœuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

Garnett, Matthew F., Sally C. Curtin, and Deborah M. Stone, "Suicide mortality in the United States, 2001–2021," *NCHS Data Brief*, 2023, (464).

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Humphreys, David K., Antonio Gasparrini, and Douglas J. Wiebe, "Evaluating the impact of Florida's red flag law on suicide deaths," *JAMA Network Open*, 2019.

Kellermann, Arthur L., Frederick P. Rivara, Grant Somes, Donald T. Reay, Jerry Francisco, Joyce Gillentine Banton, Janice Prodzinski, Corinne Fligner, and Bela B. Hackman, "Suicide in the home in relation to gun ownership," *New England Journal of Medicine*, 1992, *327* (7), 467–472.

Kivisto, Aaron J. and Peter Lee Phalen, "Firearm legislation and fatal police shootings in the United States," *American Journal of Public Health*, 2018, *108* (11), 1524–1528.

Kleck, Gary, "Measures of gun ownership levels for macro-level crime and violence research," *Journal of Research in Crime and Delinquency*, 2004, *41* (1), 3–36.

Mann, J. John, Alan Apter, José Bertolote, Annette Beautrais, Dianne Currier, Ann Haas, Ulrich Hegerl, Jouko Lonnqvist, Kevin Malone, Andrej Marusic, Lars Mehlum, George Patton, Michael Phillips, Wolfgang Rutz, Zoltán Rihmer, Armin Schmidtke, David Shaffer, Morton Silverman, Yoshitomo Takahashi, Airi Varnik, Danuta Wasserman, Paul Yip, and Herbert Hendin, "Suicide prevention strategies: A systematic review," *JAMA*, 2005, *294* (16), 2064–2074.

Miller, Matthew, Deborah Azrael, and Catherine Barber, "Firearms and suicide in the United States: Is risk independent of underlying suicidal behavior?," *American Journal of Epidemiology*, 2013, *178* (6), 946–955.

Owens, David, Judith Horrocks, and Allan House, "Fatal and non-fatal repetition of self-harm: Systematic review," *British Journal of Psychiatry*, 2002, *181* (3), 193–199.

RAND Corporation, "The effects of extreme risk protection orders," Technical Report, RAND Corporation 2023. RAND Gun Policy in America.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

Spicer, Rebecca S. and Ted R. Miller, "Suicide acts in 8 states: Incidence and case fatality rates by demographics and method," *American Journal of Public Health*, 2000, *90* (12), 1885–1891.

Swanson, Jeffrey W., Michael A. Norko, Hsiu-Ju Lin, Kelly Alanis-Hirsch, Linda K. Frisman, Madelon V. Baranoski, Michele M. Easter, Allison G. Robertson, Marvin S. Swartz, and Richard J. Bonnie, "Implementation of Connecticut's risk-based gun removal law: Does it target those at risk of suicide?," *Law and Contemporary Problems*, 2017, *80*, 179–208.

_ **, Michele M. Easter, Allison G. Robertson, Marvin S. Swartz, Kelly Alanis-Hirsch, Donna Moseley, Charles Dion, and John Petrila**, "Criminal justice and suicide outcomes with Indiana's risk-based gun seizure law," *Journal of the American Academy of Psychiatry and the Law*, 2019, *47* (2), 188–197.

Wintemute, Garen J., Veronica A. Pear, Julia P. Schleimer, Rocco Pallin, Sydney Sohl, Nicole Kravitz-Wirtz, and Magdalena Cerdá, "Extreme risk protection orders intended to prevent mass shootings: A case series," *Annals of Internal Medicine*, 2019, *171* (9), 655–658.

A. Data Appendix

A.1 CDC Mapping Injury, Overdose, and Violence

The CDC Mapping Injury dataset is accessed via the Socrata Open Data API at `data.cdc.gov` (dataset identifier: fpsi-y8tj). The API returns state-level, annual, age-adjusted mortality rates per 100,000 population for specified injury intents. I query five intents: `FA_Suicide` (firearm suicide), `All_Suicide` (all-method suicide), `FA_Homicide` (firearm homicide), `All_Homicide` (all-method homicide), and `Drug_OD` (drug overdose). Non-firearm suicide is constructed as: NF_Suicide = All_Suicide − FA_Suicide.

The data covers 2019–2024. Trailing-twelve-month (TTM) observations are excluded; only calendar-year data is retained. The "United States" aggregate row is dropped, leaving 51 jurisdictions (50 states plus the District of Columbia). After pivoting to wide format, the short panel contains 306 state-year observations.

Small-cell suppression affects some state-year-intent combinations. When death counts or rates are suppressed (typically for states with fewer than 10 deaths in a category), the corresponding field is recorded as missing. This suppression primarily affects small-population states and rare outcomes.

A.2 NCHS Leading Causes of Death

The NCHS Leading Causes of Death dataset is accessed via Socrata (identifier: bi63-dtpu). The dataset reports the top causes of death by state and year based on ICD-10 coding. I extract records where `cause_name = 'Suicide'` and the state is not "United States" (aggregate). The suicide category corresponds to ICD-10 codes X60–X84 (intentional self-harm) and Y87.0 (sequelae of intentional self-harm).

The data covers 1999–2017 and provides death counts and age-adjusted death rates (AADR) per 100,000 using the 2000 US standard population. The dataset covers 51 jurisdictions over 19 years (969 state-year observations).

The NCHS dataset provides only total suicide—no method-specific breakdown. This limitation motivates the two-panel design: the long panel (NCHS, 1999–2017) captures total suicide across many pre-treatment periods, while the short panel (CDC Mapping Injury, 2019–2024) provides the mechanism decomposition.

A.3 ERPO Adoption Dates

ERPO adoption years are coded from multiple sources: Everytown for Gun Safety's compendium of ERPO state laws, the RAND State Firearm Law Database, the Giffords Law

Center to Prevent Gun Violence, and individual state legislative records. The coding follows the convention that the adoption year is the calendar year in which the law first takes effect (which may differ from the year of legislative passage). Table 4 lists all 22 ERPO states, their effective years, and wave classification.

Table 4: ERPO Law Adoption Timeline

State	Year Effective	Wave
Connecticut	1999	Pre-Parkland
Indiana	2005	Pre-Parkland
California	2014	Pre-Parkland
Washington	2016	Pre-Parkland
Oregon	2017	Pre-Parkland
Delaware	2018	Post-Parkland
Florida	2018	Post-Parkland
Illinois	2018	Post-Parkland
Maryland	2018	Post-Parkland
Massachusetts	2018	Post-Parkland
New Jersey	2018	Post-Parkland
Rhode Island	2018	Post-Parkland
Vermont	2018	Post-Parkland
Colorado	2019	Post-Parkland
Hawaii	2019	Post-Parkland
Nevada	2019	Post-Parkland
New York	2019	Post-Parkland
New Mexico	2020	Post-Parkland
Virginia	2020	Post-Parkland
Maine	2023	Post-Parkland
Michigan	2023	Post-Parkland
Minnesota	2024	Post-Parkland

Notes: ERPO = Extreme Risk Protection Order. Pre-Parkland: states that adopted before the February 2018 Parkland shooting. Post-Parkland: states that adopted in 2018 or later. Sources: Everytown for Gun Safety, RAND State Firearm Law Database, Giffords Law Center.

A.4 Gun Ownership Proxy

State-level gun ownership is not directly measured by any administrative dataset. Following the standard proxy in the empirical firearms literature (Kleck, 2004; Azrael et al., 2004), I use the firearm suicide share:

$$\text{Gun Proxy}_s = \frac{\text{Firearm Suicides}_s}{\text{All Suicides}_s} \tag{5}$$

computed from the 2019 CDC Mapping Injury data. This proxy has been validated against survey-based ownership measures (Behavioral Risk Factor Surveillance System, General Social Survey) with correlations exceeding $r = 0.90$. State-level gun ownership is highly stable over time (Cook and Ludwig, 2006), so the 2019 cross-section serves as a reasonable proxy for the full panel.

B. Identification Appendix

B.1 Goodman-Bacon Decomposition

The Goodman-Bacon decomposition is estimated on a balanced sub-panel (2005–2017) that excludes the 2019–2024 period and the pre-2005 years to ensure balance. In this window, Indiana (adopted 2005) is effectively always-treated from the start of the sub-panel, while California (2014), Washington (2016), and Oregon (2017) serve as the switching cohorts with identifiable treatment timing.

The decomposition reveals that "treated vs. untreated" comparisons dominate (96% of total weight), with an average estimate of -1.09 per 100,000. Cross-cohort comparisons ("earlier vs. later treated" and "later vs. earlier treated") receive only 4% of the weight but produce heterogeneous estimates ranging from -1.54 to $+0.87$.

The near-unit weight on "treated vs. untreated" comparisons means that TWFE bias in this setting is not primarily driven by "forbidden comparisons" (using already-treated as controls) but rather by heterogeneous treatment effects within the "treated vs. untreated" comparison class itself. Early adopters with long exposure periods dominate, and their negative trends are extrapolated to later cohorts.

B.2 Pre-Trend Analysis

The event-study estimates in Figure 1 show pre-treatment coefficients that are generally close to zero but with wide confidence intervals. One coefficient at a distant pre-treatment period ($e = -5$) is elevated, which could indicate a mild pre-trend violation or simply sampling variation given the large number of pre-treatment periods being tested.

To assess the severity of potential pre-trend violations, I compute the average absolute value of pre-treatment coefficients (excluding $e = -1$, which is normalized to zero): approximately 0.56 per 100,000. For comparison, the main ATT estimate is 0.24 per 100,000. The larger magnitude of pre-treatment "effects" relative to the post-treatment estimate suggests that state-level suicide rates exhibit substantial idiosyncratic variation that the design cannot easily distinguish from treatment effects.

A conservative diagonal Wald test (summing squared t-statistics under independence) rejects the null that all pre-treatment coefficients are jointly zero ($\chi^2 = 60.7$, $df = 7$, $p < 0.001$). However, this test ignores the positive covariance between pre-treatment estimates and is known to be oversized in event-study settings (Roth et al., 2023). The `did` package's built-in Wald test using the full variance-covariance matrix could not be computed due to a singular covariance matrix. Visually, the pre-treatment coefficients do not exhibit a systematic trend in either direction, but their magnitudes indicate that the design's ability to detect small treatment effects is limited by state-level variation.

C. Robustness Appendix

C.1 Alternative Control Groups

The not-yet-treated control group specification produces an ATT of 0.226 ($p = 0.33$, 95% CI: $[-0.23, 0.68]$), virtually identical to the never-treated specification. This similarity is expected when most treated states have relatively few post-treatment periods: the not-yet-treated control group adds some pre-treatment observations from late adopters but does not fundamentally change the comparison.

C.2 Two-Way Clustered Standard Errors

The TWFE specification with standard errors clustered on both state and year produces a coefficient of -1.19 with SE $= 0.39$, compared to SE $= 0.40$ with one-way state clustering. The similarity indicates that cross-state error correlation within years does not substantially affect inference in this setting.

D. Heterogeneity Appendix

D.1 Gun Ownership Heterogeneity

I attempted to estimate heterogeneous effects by baseline gun ownership, splitting states at the median gun ownership proxy (0.52). The analysis could not be completed because, after restricting to post-2019 adopters in the short panel, the "high gun ownership" treated group contained too few states for estimation. This negative selection into adoption is discussed in the main text as a boundary on external validity (Section 5.5).

E. Additional Tables

Table 5: Leave-One-Out Sensitivity: Total Suicide Rate

Dropped State	ATT	SE	p-value
California	0.367*	(0.218)	0.092
Colorado	0.155	(0.234)	0.506
Delaware	0.208	(0.244)	0.395
Florida	0.197	(0.245)	0.421
Hawaii	0.276	(0.247)	0.263
Illinois	0.247	(0.245)	0.313
Indiana	0.142	(0.253)	0.573
Maine	0.231	(0.237)	0.329
Maryland	0.263	(0.228)	0.248
Massachusetts	0.290	(0.234)	0.216
Michigan	0.232	(0.222)	0.296
Minnesota	-0.009	(0.199)	0.964
Nevada	0.227	(0.244)	0.352
New Jersey	0.290	(0.236)	0.220
New Mexico	0.232	(0.239)	0.332
New York	0.242	(0.250)	0.333
Oregon	0.157	(0.228)	0.492
Rhode Island	0.329	(0.251)	0.191
Vermont	0.217	(0.245)	0.376
Virginia	0.207	(0.254)	0.415
Washington	0.268	(0.253)	0.290

Notes: Each row drops one treated state and re-estimates the ATT using Callaway and Sant'Anna (2021) with never-treated controls on the combined panel (total suicide rate, 1999–2024, excluding 2018 and Connecticut). Full-sample N = 1,250. Stability across rows indicates results are not driven by any single state. Standard errors clustered at the state level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

F. Standardized Effect Sizes

Table 6: Standardized Effect Sizes for Main Outcomes

Outcome	Specification	$\hat{\beta}$	SD(X)	SD(Y)	SDE	Classification
Total Suicide Rate (Combined)	CS-DiD, Table 2 Col. 1	0.239	—	4.57	0.052	Moderate positive
Total Suicide Rate (Short)	CS-DiD, Table 2 Col. 4	0.820	—	5.09	0.161	Large positive
Drug OD Rate (Placebo)	CS-DiD, Table 2 Col. 5	2.960	—	12.89	0.230	Large positive

Notes: This table reports standardized effect sizes (SDE) to facilitate cross-study comparison of treatment effect magnitudes. For binary (0/1) treatments, SDE = $\hat{\beta}$/SD(Y) and the SD(X) column is marked "—". SD(Y) is the unconditional standard deviation from the summary statistics, before conditioning on fixed effects.

Research question: Do Extreme Risk Protection Order (ERPO) laws reduce total suicide mortality at the state level? **Treatment:** Binary (0/1): whether a state has an ERPO law in effect in a given year. **Data:** CDC Mapping Injury (2019–2024) and NCHS Leading Causes of Death (1999–2017), state-year level, $N = 1{,}250$ (combined, excluding Connecticut) and $N = 306$ (short panel). **Method:** Staggered DiD with Callaway–Sant'Anna (2021) estimator, state-clustered standard errors. **Sample:** 50 US jurisdictions (50 states plus DC, excluding Connecticut), 1999–2024 excluding 2018. Connecticut excluded because its 1999 adoption coincides with the first panel year, leaving no pre-treatment observations. Short panel excludes pre-2019 adopters from treated group.

Classification thresholds (7-way): large negative (< -0.15), moderate negative (-0.15 to -0.05), small negative (-0.05 to -0.005), null (-0.005 to 0.005), small positive (0.005 to 0.05), moderate positive (0.05 to 0.15), large positive (> 0.15). The combined-panel total suicide ATT ($p = 0.245$) and the drug OD placebo ($p = 0.318$) are not statistically significant at the 5% level. The short-panel total suicide ATT ($p = 0.014$) is statistically significant but is based on only 9 treated states and 6 years, limiting its generalizability.