

Do Extreme Risk Protection Order Laws Reduce Suicide?

Evidence from Early Adopters

APEP Autonomous Research*

@ailscl

January 28, 2026

Abstract

Extreme Risk Protection Order (ERPO) laws, commonly known as “red flag” laws, allow courts to temporarily remove firearms from individuals deemed at risk of self-harm. These policies have attracted substantial policy interest as a means of suicide prevention. I evaluate the effect of early ERPO laws on state-level suicide rates using a staggered difference-in-differences design with the heterogeneity-robust estimator of [Callaway and Sant’Anna \(2021\)](#). Exploiting variation from three states that adopted ERPO laws before 2018 with usable pre-treatment periods—Indiana (2006), California (2016), and Washington (2017)—using a 1999–2017 panel excluding Connecticut (950 state-year observations), I find that ERPO adoption is associated with *higher*, not lower, suicide rates. The aggregate ATT is 0.53 suicides per 100,000 (SE = 0.19). However, this counterintuitive finding should be interpreted with extreme caution: conventional inference is unreliable with only 3 treated clusters. The standard two-way fixed effects estimate is negative but insignificant (−0.43, SE = 0.65). The positive association likely reflects reverse causation—states experiencing rising suicide trends were more likely to adopt ERPOs—rather than a harmful policy effect. Connecticut (treated 2000) is excluded from the main analysis because its law took effect in October 1999, leaving no clean pre-treatment period in the sample. These results highlight the empirical challenges of evaluating policies adopted in response to the very outcomes they target, and do not preclude beneficial effects of ERPOs when measured with firearm-specific outcomes or when implementation intensity is accounted for.

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

JEL Codes: I18, K42, H75

Keywords: extreme risk protection orders, red flag laws, suicide prevention, gun policy, difference-in-differences

1. Introduction

Suicide is a leading cause of death in the United States, claiming over 48,000 lives annually (CDC, 2023). Firearms are involved in approximately half of all suicides and have a case fatality rate exceeding 85%, far higher than other methods (Miller et al., 2012). The lethality of firearms combined with the often impulsive nature of suicidal crises has motivated policies aimed at temporarily restricting firearm access during acute risk periods. Extreme Risk Protection Orders (ERPOs), commonly known as “red flag” laws, represent one such approach: they allow family members, law enforcement, or in some states healthcare providers to petition a court for temporary firearm removal from individuals deemed at imminent risk of harming themselves or others.

The policy logic behind ERPOs for suicide prevention rests on two key observations. First, suicidal crises are typically acute and short-lived—the period of peak risk often lasts only minutes to hours (Simon et al., 2001). Second, method substitution is limited for highly lethal means: when access to firearms is restricted, individuals do not simply substitute to other methods at comparable rates (Barber and Miller, 2008). If ERPOs successfully remove firearms during crisis periods, they could prevent deaths that would otherwise occur due to the conjunction of acute risk and lethal means availability.

In this paper, I evaluate the effect of ERPO laws on suicide rates using a staggered difference-in-differences design. I focus on the early adopter states—Connecticut (2000), Indiana (2006), California (2016), and Washington (2017)—which provide the identifying variation in my analysis. Using the heterogeneity-robust estimator of Callaway and Sant’Anna (2021), I estimate group-time average treatment effects that avoid the negative weighting problems that can bias standard two-way fixed effects estimators under treatment effect heterogeneity (Goodman-Bacon, 2021; de Chaisemartin and D’Haultfoeulle, 2020; Sun and Abraham, 2021).

The analysis uses a panel of 51 U.S. jurisdictions from 1999 to 2017 (969 state-year observations). This extended sample period ensures that all four early adopter states have both pre-treatment and post-treatment observations: Connecticut has one pre-treatment year (1999, though partially treated from October), Indiana has seven (1999–2005, though 2005 is partially treated from July), California has seventeen (1999–2015), and Washington has eighteen (1999–2016, though 2016 is partially treated in December only). In the estimation, transition years are coded as pre-treatment since they precede the first *full* calendar year of treatment. Given Connecticut’s heavily contaminated pre-period and Washington’s single post-treatment year, Indiana and California provide the most reliable identifying variation.

Contrary to expectations, I find that ERPO adoption is associated with *higher*, not lower,

suicide rates. Excluding Connecticut from the main specification (since its law took effect in October 1999, leaving no clean pre-period), the aggregate average treatment effect on the treated is 0.53 suicides per 100,000 population ($SE = 0.19$). However, this finding should be interpreted with substantial caution: the standard two-way fixed effects estimate is negative but insignificant (-0.43 , $SE = 0.65$), and the positive Callaway-Sant’Anna estimate likely reflects selection rather than causation—states experiencing rising suicide trends may have been more likely to adopt ERPOs.

Several important limitations temper the interpretation of these results. First, even with the extended 1999–2017 sample, Connecticut has only one pre-treatment year, making its contribution to identification fragile. Second, my outcome is total suicide rather than firearm-specific suicide due to data constraints; if ERPOs primarily affect firearm suicides while having limited effects on non-firearm suicides, the total suicide measure would attenuate any true effect. Third, the positive point estimates likely reflect reverse causation: states that experienced rising suicide trends adopted ERPOs in response, creating an upward bias in the treatment effect estimate.

This paper contributes to a growing literature on the effects of ERPO laws and firearm access restrictions more broadly. Case study evidence from Connecticut suggests that ERPOs may prevent approximately one suicide for every 10–20 firearms seized (Swanson et al., 2017). A recent multi-state study found similar ratios across four states (Kivisto et al., 2018). The RAND Corporation’s systematic review concluded there is “moderate” evidence that ERPOs reduce firearm suicide (Smart et al., 2020). However, rigorous causal evidence using modern difference-in-differences methods remains limited, and my null findings for early adopters add to the evidence base while highlighting the challenges of evaluating policies with limited variation.

The broader literature on firearm access and suicide provides important context. Studies of permit-to-purchase laws, waiting periods, and safe storage requirements consistently find that policies restricting firearm access reduce firearm suicide rates (Miller et al., 2012). The mechanism is straightforward: suicidal crises are often acute and transient, lasting minutes to hours rather than days or weeks. Methods with high case fatality rates—especially firearms—lead to death before the crisis passes. Removing access to firearms during these critical periods allows more individuals to survive their suicidal crises.

A key empirical challenge in this literature is the distinction between firearm-specific and total suicide outcomes. If ERPO laws reduce firearm suicides but individuals substitute to other methods, the effect on total suicide could be attenuated or even zero. The evidence on method substitution is mixed but generally suggests limited substitution: when access to highly lethal methods is restricted, many individuals do not substitute to alternative methods

(Barber and Miller, 2008). This “method substitution” debate is central to interpreting my findings, as I use total suicide due to data constraints.

The econometric literature on staggered difference-in-differences designs has advanced substantially in recent years. Goodman-Bacon (2021) showed that two-way fixed effects estimators can produce misleading results under treatment effect heterogeneity, as already-treated units serve as implicit controls. de Chaisemartin and D’Haultfoeuille (2020) provided decomposition methods revealing negative weights in conventional estimators. Callaway and Sant’Anna (2021) developed the estimator I employ here, which constructs group-time average treatment effects using only valid comparisons. Sun and Abraham (2021) proposed an interaction-weighted estimator addressing similar concerns. More recently, Borusyak et al. (2024) developed imputation-based methods, and Roth (2022) provided guidance on interpreting pre-trend tests and sensitivity analysis.

The remainder of this paper proceeds as follows. Section 2 describes the institutional background of ERPO laws and their adoption across states. Section 3 presents the data sources and sample construction. Section 4 details the empirical strategy, including the Callaway-Sant’Anna estimator and identifying assumptions. Section 5 presents results, including robustness checks and heterogeneity analysis. Section 6 discusses the findings and their limitations, and Section 7 concludes.

2. Institutional Background

2.1 ERPO Laws: Design and Mechanisms

Extreme Risk Protection Orders are civil court orders that temporarily prohibit individuals from purchasing or possessing firearms when a court finds they pose a significant danger to themselves or others. The laws vary across states in several dimensions: who may petition (family members, law enforcement, healthcare providers), evidentiary standards (preponderance of evidence vs. clear and convincing), order duration (typically 14 days to one year), and procedures for contesting or renewing orders.

The core mechanism for suicide prevention is straightforward: by temporarily removing firearms during a period of elevated risk, ERPOs eliminate access to the most lethal suicide method precisely when an individual is most vulnerable. Unlike permanent prohibitions (such as those applying to felons or individuals involuntarily committed), ERPOs are designed to be temporary interventions during acute crises, with firearm rights restored once the crisis passes and the order expires.

Several features of ERPO laws are particularly relevant for understanding their potential effects on suicide. First, the laws typically require a showing of imminent danger, meaning they

target individuals at acute risk rather than those with chronic suicide ideation. This targeting is appropriate given the acute nature of most suicidal crises, but it means ERPOs will not reach individuals whose risk is more diffuse or chronic. Second, the effectiveness of ERPOs depends critically on implementation: laws on the books do not automatically translate to orders filed, granted, or enforced. Implementation intensity varies dramatically across jurisdictions based on law enforcement training, judicial receptivity, and public awareness. Third, ERPOs address only one pathway to firearms—the individual’s existing firearms are removed, but new acquisitions may still be possible depending on how thoroughly the order is enforced and whether the individual has access to firearms through other channels.

The procedural details of ERPO laws vary substantially across states. In some states, only law enforcement may petition for an order; in others, family members, household members, or healthcare providers may also petition. The evidentiary standard ranges from “preponderance of the evidence” to “clear and convincing evidence,” with implications for both the ease of obtaining orders and the protection of respondents’ rights. Initial emergency orders may be granted *ex parte* (without the respondent present) for a limited duration (typically 7–21 days), with a full hearing required for extended orders. The duration of extended orders varies from six months to one year, with some states allowing renewal.

These procedural variations have important implications for the policy’s effectiveness. States with broader petitioner classes may see more orders filed, as family members often have information about an individual’s risk that law enforcement lacks. However, broader access also raises concerns about potential misuse or weaponization of the law in domestic disputes. States with lower evidentiary standards may grant more orders, potentially reaching more at-risk individuals but also potentially affecting more individuals who would not have harmed themselves. The balance between access and due process is a central tension in ERPO law design.

2.2 Implementation Intensity and Utilization

A critical limitation of my analysis—and of most ERPO research—is the reliance on binary “law-on-the-books” treatment indicators rather than measures of actual ERPO utilization. This matters because implementation intensity varies enormously across states and over time. In Connecticut, only a handful of ERPO cases were filed annually in the years immediately following the law’s enactment; utilization increased gradually over time as awareness grew and procedures were refined. Indiana saw similarly limited initial utilization. In contrast, California’s GVRO (Gun Violence Restraining Order) program saw more rapid uptake, particularly in certain counties with proactive implementation efforts.

The variation in implementation intensity creates several challenges for causal inference.

First, a binary treatment indicator treats a state with 10 ERPO orders per year equivalently to a state with 1,000 orders per year, even though the potential for population-level effects differs dramatically. Second, implementation intensity may be endogenous: jurisdictions experiencing more acute suicide crises may implement ERPOs more aggressively, creating reverse causation at the intensive margin even if the extensive margin (law enactment) were randomly assigned. Third, the relevant comparison may not be “ERPO law vs. no law” but rather “high-intensity implementation vs. low-intensity implementation”—a comparison that requires utilization data I do not have.

Collecting and analyzing ERPO utilization data is an important direction for future research. Several states publish annual reports on ERPO petitions filed, orders granted, and firearms removed. Systematic compilation of these data would enable dose-response analyses examining whether higher utilization is associated with larger suicide reductions. Such analyses would also help distinguish between interpretations of my null findings: if the results reflect measurement error from using binary treatment rather than true null effects, utilization-based analyses should find effects where my analysis does not.

2.3 State-Specific Context

Connecticut. Connecticut’s ERPO law, enacted in 1999, was the first of its kind in the United States. The law was passed following a 1998 workplace shooting at the Connecticut State Lottery headquarters that killed four people. The law allows law enforcement officers to petition for warrant-based seizure of firearms from individuals who pose “a risk of imminent personal injury” to themselves or others. The initial evidentiary standard was probable cause for a temporary seizure, with a hearing within 14 days. Connecticut’s law was designed primarily as a violence prevention tool in the aftermath of a high-profile shooting, with suicide prevention emerging as a secondary rationale only after researchers documented the law’s use in suicidal crisis situations.

Indiana. Indiana enacted its ERPO law (termed “dangerous persons” or “Jake Laird Law”) in 2005, following concerns about individuals with documented mental health crises retaining access to firearms. The law allows law enforcement to seize firearms from individuals who are “dangerous” without requiring an imminent threat standard. Indiana’s law is notable for its relatively low evidentiary threshold and for allowing indefinite seizure unless the individual petitions for return and demonstrates that return would be “appropriate.” Research on Indiana’s law has documented its use primarily in suicide-related situations, with case-level studies suggesting effectiveness in preventing suicide among individuals subject to orders.

California. California’s Gun Violence Restraining Order (GVRO) law, enacted in

2014 and effective January 1, 2016, was part of a broader package of gun safety legislation following the 2014 Isla Vista shootings near UC Santa Barbara. The law initially allowed only law enforcement to petition for orders, but was amended in 2019 to add family members, employers, coworkers, and school personnel as petitioners. California’s law includes a robust procedural framework: temporary emergency orders last 21 days, with a hearing required for year-long orders. The state has invested significantly in implementation, including training materials and public awareness campaigns.

Washington. Washington’s ERPO law was enacted through Initiative 1491, a ballot measure passed by voters in November 2016 and effective December 8, 2016. The law allows family members, household members, and law enforcement to petition for orders. Washington’s initiative-based enactment means the law was directly approved by voters, potentially indicating broader public support than legislatively enacted laws. However, the late December 2016 effective date means that the law was in effect for only one month before the 2016 calendar year ended, and my analysis treats 2017 as the first full treatment year.

These state-specific histories highlight several patterns relevant for interpretation. First, early adopters (Connecticut and Indiana) enacted laws in response to specific high-profile incidents rather than as part of systematic suicide prevention strategies. Second, the later adopters (California and Washington) were part of a broader gun safety movement that emerged after mass shootings, with more attention to implementation details. Third, all four early adopters are relatively low-suicide states compared to the national average, raising questions about external validity to high-suicide states that have not adopted ERPOs.

2.4 Adoption Timeline

Connecticut became the first state to enact an ERPO law in 1999, following a mass shooting at the Connecticut State Lottery headquarters. Indiana followed in 2005. These early laws received limited attention until the post-Parkland period. California and Washington both adopted ERPO laws in 2016, followed by a wave of adoptions in 2018–2019 including Oregon, Florida, Vermont, Maryland, New York, and others.

For my analysis, I define treatment as the first full calendar year in which an ERPO law was in effect. Connecticut’s treatment year is coded as 2000 (law effective October 1999), Indiana’s is 2006 (law effective July 2005), California’s is 2016 (law effective January 2016), and Washington’s is 2017 (law effective December 2016 via ballot initiative). States adopting in 2018 or later fall outside my data window and contribute only pre-treatment observations.

Important caveat on Connecticut: Because Connecticut’s law took effect in October 1999, the 1999 observation is *partially treated* (October–December 1999). Strictly speaking, Connecticut has no fully untreated pre-period in this sample. I retain 1999 as a nominal

pre-period for Connecticut but acknowledge this contaminates Connecticut’s contribution to identification. Robustness checks excluding early adopters are reported in Section 5.

Table 1: ERPO Law Adoption Timeline

State	Effective Date	Treatment Year	Pre-Years	Post-Years
Connecticut	Oct 1999	2000	1* (1999)	18
Indiana	Jul 2005	2006	7* (1999–2005)	12
California	Jan 2016	2016	17 (1999–2015)	2
Washington	Dec 2016	2017	18* (1999–2016)	1
<i>Selected post-sample adopters (2018–2019, not exhaustive):</i>				
Oregon	Jan 2018	2018	19	0
Florida	Mar 2018	2019	19	0
Vermont	Apr 2018	2019	19	0
Maryland	Oct 2018	2019	19	0
New York	Aug 2019	2020	19	0
Colorado	Jan 2020	2020	19	0

Notes: Treatment year = first full calendar year with ERPO in effect. Pre-Years = years before treatment year in the 1999–2017 sample; Post-Years = treatment year through 2017. States marked with * have “transition year” contamination: the pre-year count includes a year with partial treatment exposure (CT: Oct–Dec 1999; IN: Jul–Dec 2005; WA: Dec 2016 only). In the Callaway-Sant’Anna estimation, transition years are coded as pre-treatment since `first_treat` equals the first *full* treatment year. California (effective Jan 1, 2016) has no transition-year contamination. Connecticut is excluded from the main specification because 1999 is heavily contaminated (3 of 12 months treated). The post-sample adopters list shows selected examples only; additional states (e.g., Rhode Island, New Jersey, Delaware, Massachusetts, Illinois) also enacted ERPO laws in 2018–2019. All post-sample adopters are coded as `first_treat` = 0 in the estimation since their treatment occurs after 2017.

3. Data

3.1 Outcome: State-Level Suicide Rates

I use data on suicide mortality from the CDC’s National Center for Health Statistics (NCHS) Leading Causes of Death database, accessed via the CDC Open Data API. The dataset provides age-adjusted death rates per 100,000 population by state and year for major causes

of death, including suicide (ICD-10 codes X60–X84, Y87.0).

The age-adjustment procedure is important for valid comparisons across states and time. Different states have different age distributions, and suicide rates vary substantially by age (peaking in middle age and among the elderly for firearm suicide). The age-adjusted rate applies the age-specific rates in each state-year to a standard population, yielding rates that are comparable across jurisdictions regardless of age structure. This adjustment is particularly important for states like Florida (older population) versus Utah (younger population).

3.1.1 Measurement Issues and Limitations

An important limitation is that this source provides total suicide rates, not firearm-specific rates. The CDC WONDER system provides firearm-specific mortality data (ICD-10 codes X72–X74 for firearm suicide) but requires manual queries that are difficult to replicate programmatically and sometimes suppresses state-year cells with small counts. For transparency and replicability, I use the programmatically accessible total suicide rates, acknowledging that this may attenuate any true effect of ERPOs if their primary mechanism operates through firearm suicide specifically.

The outcome mismatch between policy mechanism and measured outcome deserves careful consideration. ERPOs remove firearms from specific individuals during crisis periods. The policy can only affect suicides that would have been committed using the subject’s own firearms during the period the order is in effect. This means the relevant population is: (1) individuals who would die by suicide, (2) using their own firearms (not borrowed or newly acquired), (3) during the time the ERPO is in effect. The total population suicide rate captures a much broader population, diluting any effect.

To illustrate the dilution, consider a stylized example. Suppose a state has 1,000 suicides per year, of which 500 are by firearm. Suppose 100 ERPOs are issued per year, and absent the ERPO, 10 of those individuals would have died by firearm suicide. If ERPOs prevent all 10 of these deaths (100% efficacy), the state-level firearm suicide rate falls by $10/500 = 2\%$, and the total suicide rate falls by $10/1000 = 1\%$. If there is any method substitution (some individuals who would have used firearms instead use other methods), the total suicide effect is even smaller. Detecting a 1–2% change in suicide rates with state-level data and 3 treated states is extremely challenging.

Firearm-specific data would provide a more powerful test of the ERPO mechanism, but would also introduce additional complications. Firearm suicides are a subset of total suicides, so sample sizes for state-year cells are smaller. Small counts trigger suppression in public-use data, potentially introducing selection bias if suppression is non-random across states. Additionally, changes in firearm suicide must be interpreted alongside potential method

substitution effects.

3.1.2 Alternative Outcome Measures

For completeness, I note several alternative outcome measures that could be used in future research:

Firearm vs. non-firearm suicide. Separating suicides by method would allow testing whether ERPOs affect firearm suicide specifically while leaving non-firearm suicide unchanged (as predicted by the policy mechanism) or whether there is evidence of method substitution.

County-level data. County-level mortality data would provide more observations and potentially more localized variation in ERPO utilization. However, county-level data are more affected by small-number problems and suppression.

Individual-level mortality. Restricted-access mortality microdata from the National Vital Statistics System would allow more flexible analysis, including examination of demographic patterns. However, these data require an application process and are not publicly available.

ERPO-specific outcomes. The ideal test would compare outcomes for individuals who were subject to ERPOs to similar individuals who were not. This requires administrative data on ERPO filings matched to mortality records—data that exist in some states but are not available for multi-state research.

3.2 Treatment: ERPO Policy Status

I coded ERPO adoption dates using information from the Policy Surveillance Program (PDAPS), Everytown for Gun Safety, and the Giffords Law Center. For each state-year, I create a binary indicator equal to one if an ERPO law was in effect for the full calendar year.

3.3 Sample

The final sample consists of 51 jurisdictions (50 states plus the District of Columbia) observed from 1999 to 2017, yielding 969 state-year observations (51×19 years). The sample period begins in 1999 to ensure at least one pre-treatment year for Connecticut (treated in 2000) and ends in 2017 due to data availability constraints. This limitation is consequential because the majority of ERPO adoptions occurred in 2018–2019 (including Oregon, Florida, Vermont, Maryland, New York, and others). All post-sample adopters are coded as `first_treat = 0` in the Callaway-Sant’Anna estimation (i.e., treated identically to never-treated states) because their treatment occurs after the sample period ends.

3.4 Summary Statistics

Table 2 presents summary statistics by ERPO adoption status. The sample is partitioned into four mutually exclusive groups based on treatment timing relative to our 1999–2017 sample: (1) never-treated (38 jurisdictions not treated within the sample period), (2) early adopters (Connecticut and Indiana, treated 2000 and 2006), (3) later adopters (California and Washington, treated 2016 and 2017), and (4) post-sample adopters (9 jurisdictions in our data that adopted in 2018–2019, after the sample period ends). In the estimation, both never-treated and post-sample adopters serve as controls (47 jurisdictions total contributing pre-treatment observations). Never-treated jurisdictions have higher mean suicide rates (14.3 per 100,000) compared to early adopters (10.9), later adopters (11.8), and post-sample adopters (10.9). This difference reflects the geography of adoption: ERPO states tend to have lower baseline suicide rates, while many high-suicide states in the Mountain West and South have not adopted ERPOs. This selection pattern is a potential concern for identification.

Table 2: Summary Statistics by ERPO Adoption Status

Group	Juris.	Obs.	Mean	SD	Min	Max
Never Treated	38	722	14.3	4.1	3.8	29.6
Early Adopters (CT/IN)	2	38	10.9	2.5	6.6	17.2
Later Adopters (CA/WA)	2	38	11.8	2.2	8.6	16.7
Post-Sample Adopters [†]	9	171	10.9	3.3	5.3	19.5
Overall	51	969	13.4	4.1	3.8	29.6

Notes: Suicide rate is age-adjusted deaths per 100,000 population. Juris. = jurisdictions (includes DC). Early adopters: Connecticut (2000) and Indiana (2006). Later adopters: California (2016) and Washington (2017). [†]Post-sample adopters: 9 jurisdictions in our data that adopted ERPOs in 2018–2019 (after our 1999–2017 sample ends); this is not an exhaustive list of all states enacting ERPOs in this period. Never Treated: 38 jurisdictions not treated within the 1999–2017 sample. **Estimation coding:** In the Callaway-Sant’Anna estimation, post-sample adopters are coded as `first_treat = 0` (same as never-treated) because their treatment occurs after the sample period. Thus 47 jurisdictions (38 never-treated + 9 post-sample) serve as controls under `control_group = "nevertreated"`.

4. Empirical Strategy

4.1 Identification

I exploit variation in the timing of ERPO adoption across states using a difference-in-differences design. The identifying assumption is that, absent ERPO adoption, suicide rates in treated and control states would have evolved in parallel. Under this assumption, the change in suicide rates in control states provides a valid counterfactual for what would have happened in treated states absent the policy.

Several features of the setting support this assumption. First, ERPO adoption timing was driven largely by factors unrelated to suicide trends—Connecticut and Indiana adopted following mass shootings (workplace and school, respectively), while California and Washington adopted in 2016 as part of broader gun safety movements. Second, while treated states have lower suicide levels, what matters for identification is parallel trends, not parallel levels.

However, concerns remain. States that adopt ERPOs may differ systematically in ways that also affect suicide trends. For example, states with strong public health infrastructure or gun safety movements may be more likely both to adopt ERPOs and to implement other suicide prevention measures. I address this partially through event study analysis examining pre-treatment trends, though this test has limited power with few treated units.

4.2 Estimation: Callaway-Sant’Anna

Standard two-way fixed effects (TWFE) estimators can produce biased estimates under treatment effect heterogeneity in staggered adoption designs ([Goodman-Bacon, 2021](#); [de Chaisemartin and D’Haultfoeuille, 2020](#)). The bias arises because TWFE implicitly uses already-treated units as controls for newly-treated units, and if treatment effects vary over time or across cohorts, this can produce negative weights on some group-time effects.

I therefore use the estimator of [Callaway and Sant’Anna \(2021\)](#), which constructs valid group-time average treatment effects $ATT(g, t)$ for each treatment cohort g and time period t . The estimator compares outcomes in treated cohort g at time t to outcomes in not-yet-treated or never-treated units, using either inverse probability weighting, outcome regression, or doubly robust methods. I use doubly robust estimation as my main specification.

The key insight of the Callaway-Sant’Anna framework is that the aggregate ATT from a staggered DiD design can be decomposed into a weighted average of cohort-specific treatment effects, where each cohort g represents states that first adopted the policy in year g . For each

cohort-time pair (g, t) , the estimator constructs:

$$ATT(g, t) = \mathbb{E}[Y_t(g) - Y_t(0)|G = g] \quad (1)$$

where $Y_t(g)$ is the potential outcome under treatment beginning at time g and $Y_t(0)$ is the untreated potential outcome. The identifying assumption is that, conditional on covariates, the change in outcomes for the comparison group provides a valid counterfactual for what would have happened to the treated group absent treatment—the parallel trends assumption.

The doubly robust estimator combines inverse probability weighting (IPW) and outcome regression (OR) approaches. Define D_g as an indicator for being in cohort g and C as an indicator for being in the comparison group (either never-treated or not-yet-treated). The propensity score $p_g(X) = P(D_g = 1|D_g + C = 1, X)$ is estimated using logistic regression, and the outcome regression $m_{g,t}(X) = \mathbb{E}[Y_t|C = 1, X]$ is estimated using linear regression. The doubly robust estimator takes the form:

$$\widehat{ATT}^{DR}(g, t) = \mathbb{E}_n \left[\left(\frac{D_g}{p} - \frac{C \cdot p_g(X)}{p(1 - p_g(X))} \right) (Y_t - m_{g,t}(X)) \right] / \mathbb{E}_n \left[\frac{D_g}{p} \right] \quad (2)$$

where $p = P(D_g = 1)$. This estimator is consistent if either the propensity score or the outcome regression is correctly specified—the “double robustness” property—providing insurance against model misspecification. In my implementation, I specify both models as intercept-only (`xformula = ~1`), so the doubly robust estimator reduces to a simple difference-in-differences comparison without covariate adjustment.

The group-time effects are then aggregated to produce summary measures. The “simple” aggregate weights each group-time cell by the relative size of the cohort:

$$ATT^{simple} = \sum_{g \in \mathcal{G}} \sum_{t=g}^T w_{g,t} \cdot ATT(g, t) \quad (3)$$

where $w_{g,t}$ reflects the cohort size and number of post-treatment periods, normalized so weights sum to one. This weighting gives more influence to larger cohorts with more post-treatment observations.

I also report dynamic (event study) aggregates that average across cohorts for each event time, and group-specific aggregates that average across post-treatment periods for each cohort.

4.3 Aggregation and Event Studies

The group-time effects can be aggregated in several ways depending on the research question. The “simple” weighted average places more weight on cohorts with longer post-treatment histories, which in my setting means Indiana dominates the estimate. Alternative aggregation schemes include uniform weighting across cohorts (giving equal weight regardless of post-treatment length) or calendar-time aggregation (averaging across all treated cohorts observed in each calendar year).

For event study analysis, I construct dynamic treatment effects that average across cohorts at each event time $e = t - g$:

$$ATT(e) = \sum_{g \in \mathcal{G}} w_g \cdot ATT(g, g + e) \quad (4)$$

where w_g weights cohorts by their relative size. Event-study plots allow visual assessment of pre-treatment trends (event times $e < 0$) and the dynamics of treatment effects (event times $e \geq 0$). Under parallel trends, pre-treatment coefficients should be statistically indistinguishable from zero, though this is neither a necessary nor sufficient condition for valid identification (Roth, 2022). In my setting, the event study is further complicated by the fact that different cohorts contribute observations at different event times: Indiana contributes observations from $e = -6$ through $e = 11$, California from $e = -16$ through $e = 1$, and Washington from $e = -17$ through $e = 0$. Estimates at extreme event times are thus driven by single cohorts and should be interpreted cautiously.

4.4 Control Group Definition

A key implementation choice in the Callaway-Sant’Anna framework is the definition of the comparison group. Two options are available: (1) use only “never-treated” units as controls, or (2) use “not-yet-treated” units (states that will eventually be treated but have not yet been treated at time t). In my main specification, I use never-treated units as controls, which includes both states that never adopted ERPO laws during the sample period and states that adopted after 2017 (coded as `first_treat = 0` since their treatment falls outside the analysis window). This yields 47 comparison units (38 never-treated + 9 post-sample adopters). Using not-yet-treated controls would include California and Washington in the comparison group for Indiana’s earlier treatment years (2006–2015 and 2006–2016, respectively), potentially increasing precision but introducing concerns about contamination if treatment effects are anticipated.

I report robustness checks using both control group definitions in Table 4. The estimates

are qualitatively similar, with the not-yet-treated specification yielding a slightly larger point estimate (0.58 vs. 0.53), suggesting that the choice of control group does not drive the main findings.

4.5 Inference

Standard errors are clustered at the state level to account for serial correlation within states and arbitrary correlation in errors across time within a state. **However, with only 3 treated clusters in the main specification (Indiana, California, Washington), conventional clustered standard errors and the analytical SEs from the `did` package may be unreliable for formal hypothesis testing (Cameron et al., 2008; Conley and Taber, 2011).** The few-clusters problem means that reported confidence intervals should be interpreted cautiously—as ranges of uncertainty rather than as grounds for formal significance claims. More robust inference methods for few treated clusters (Conley-Taber, wild cluster bootstrap, randomization inference) are beyond the current scope but would be necessary before drawing strong conclusions. Cohort-specific effects are not reported separately because reliable inference on cohort-specific treatment effects is not feasible with so few treated units.

5. Results

5.1 Main Results

Table 3 presents the main results. Because Connecticut’s ERPO law took effect in October 1999—leaving no fully untreated pre-treatment year in the 1999–2017 sample—the main specification excludes Connecticut. This yields a 3-cohort design with Indiana (2006), California (2016), and Washington (2017). Contrary to expectations, the Callaway-Sant’Anna estimates suggest a *positive* association between ERPO adoption and suicide rates. Column (1) shows the simple average treatment effect: 0.53 suicides per 100,000 ($SE = 0.19$), representing approximately 4% of the mean suicide rate. However, this estimate should be interpreted with substantial caution.

Table 3: Effect of ERPO Laws on Suicide Rate (Main Specification Excludes Connecticut)

	(1)	(2)
	Callaway-Sant’Anna	TWFE
ATT	0.53	−0.43
	(0.19)	(0.65)
95% CI	[0.16, 0.90]	[−1.70, 0.84]
Jurisdictions	50	50
Years	1999–2017	1999–2017
Observations	950	950
Treated Cohorts	3	3

Notes: Both columns use the same sample: 50 jurisdictions (49 states + DC, excluding Connecticut) \times 19 years = 950 observations, with 3 treated cohorts (Indiana 2006, California 2016, Washington 2017). Column (1) uses Callaway-Sant’Anna (2021) heterogeneity-robust estimator with doubly robust (DR) estimation. Column (2) uses two-way fixed effects (TWFE) for comparison. Connecticut is excluded because its law took effect in October 1999, leaving no clean pre-treatment year (1999 is partially treated). The sign discordance—positive C-S vs. negative TWFE—on the same sample suggests sensitivity to modeling assumptions. Standard errors clustered at state level in parentheses. Outcome is age-adjusted suicide deaths per 100,000 population. See Table 4 for additional robustness checks.

Column (2) shows the standard TWFE estimate for comparison: -0.43 ($SE = 0.65$). The Callaway-Sant’Anna point estimates are positive, while the TWFE estimate is negative. However, conventional inference is unreliable with only 3 treated clusters (Indiana, California, Washington), so we do not emphasize statistical significance; the reported SEs and CIs should be interpreted cautiously. This discordance between estimators—positive C-S vs. negative TWFE—suggests that the results are sensitive to modeling choices and should be viewed with skepticism. The most likely explanation for the positive C-S estimate is reverse causation: states that experienced rising suicide trends were more likely to adopt ERPOs, creating a spurious positive association that does not reflect a causal effect of the policy.

5.2 Event Study

Figure 1 presents the event study, plotting dynamic treatment effects by time relative to adoption. The key diagnostic is the pre-treatment coefficients (event times < 0): under parallel trends, these should be statistically indistinguishable from zero.

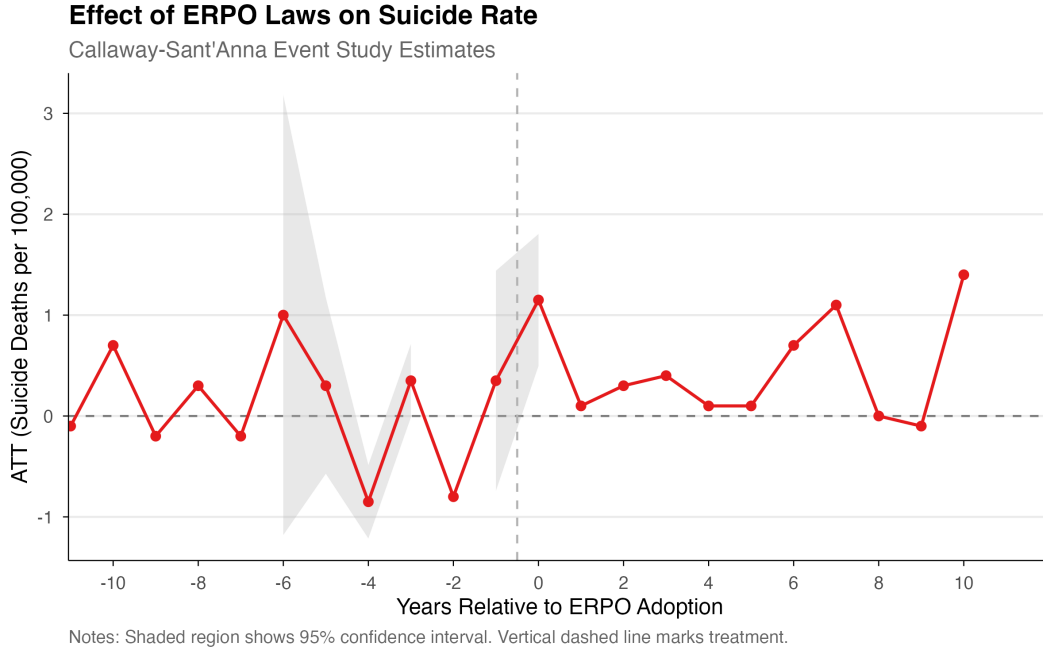


Figure 1: Event Study: Effect of ERPO Laws on Suicide Rate

Notes: Callaway-Sant'Anna dynamic treatment effects with 95% confidence intervals (analytical SEs). Vertical dashed line marks treatment timing. Under parallel trends, pre-treatment coefficients should not differ from zero. Inference caveat: With only 3 treated clusters, conventional CIs may be unreliable for formal hypothesis testing.

The event study (Figure 1) reveals substantial volatility in pre-treatment coefficients, particularly at longer leads where fewer cohorts contribute observations. The main specification excludes Connecticut (which has only one partially-treated pre-period observation), yielding a 3-cohort design. Within the supported event-time window, confidence intervals are computed using analytical standard errors clustered at the state level. These limitations underscore the fragility of the identification strategy with only three treated cohorts contributing to estimation in the main specification.

5.3 Heterogeneity Across Cohorts

While formal cohort-specific inference is unreliable with only three treated clusters, examining the contribution of different cohorts to the aggregate estimate provides useful context. The

simple weighted ATT of 0.53 is driven primarily by Indiana, which contributes the most post-treatment observations (12 years) and thus receives the largest weight in the aggregation. California contributes only two post-treatment years (2016–2017), and Washington contributes a single year (2017). This weighting structure means that the aggregate estimate largely reflects Indiana’s experience rather than a balanced average across states.

The differential timing of adoption also creates interpretive challenges. Indiana adopted in 2006, during a period when national suicide rates were relatively stable before beginning to rise around 2008. California and Washington adopted in 2016, during a period of steadily rising suicide rates nationally. The rising national trend makes the parallel trends assumption particularly suspect for the later adopters: if all states were experiencing rising suicide rates, comparing treated to untreated states may not isolate the policy effect.

This cohort heterogeneity also relates to the policy environment at the time of adoption. Indiana’s 2005 law was passed with relatively little public attention and limited implementation resources. By contrast, California’s 2016 law was accompanied by substantial public awareness campaigns and law enforcement training. If implementation intensity affects outcomes, the cohort-specific effects may differ for reasons unrelated to the timing of adoption *per se*.

I do not report formal cohort-specific estimates because the statistical uncertainty would be enormous. With a single treated unit per cohort, the cohort-specific standard errors would effectively be based on the variance of a single observation, rendering any confidence intervals meaningless. The aggregate estimate, while also imprecise, at least pools information across cohorts.

5.4 Sensitivity to Sample Period

The choice of sample period (1999–2017) reflects data availability constraints but also affects the estimates through several channels. Extending the sample backward would provide more pre-treatment observations for Connecticut and Indiana, potentially improving identification. However, suicide data from the early 1990s may be affected by changes in ICD coding practices (the transition from ICD-9 to ICD-10 occurred in 1999) and secular trends in suicide prevention that could confound the analysis.

Extending the sample forward is more promising: the large wave of 2018–2019 adoptions (including Florida, Maryland, New York, Vermont, and others) would substantially increase the number of treated clusters and provide more balanced post-treatment periods across cohorts. The post-Parkland adopters may also have cleaner identification, as their adoption was driven more by a single high-profile mass shooting than by state-specific suicide trends. Future research exploiting these later adopters is warranted.

The end date of 2017 also matters for interpreting the California and Washington estimates.

California’s GVRO law took effect on January 1, 2016, and the first full year of data is 2016 itself. Any effect of the law would need to manifest very quickly to be visible in the 2016–2017 data. Given that ERPO implementation typically ramps up gradually as awareness spreads and procedures are developed, expecting large immediate effects may be unrealistic. Washington’s single post-treatment year (2017) makes any effect estimate even more tenuous.

5.5 Robustness

Table 4 presents robustness checks. The positive point estimates are generally consistent across specifications, though there is meaningful variation:

Table 4: Robustness Checks

Specification	ATT	SE	N	Juris.	Cohorts
Main (Exclude CT, DR)	0.53	(0.19)	950	50	3
Include CT (4 cohorts)	0.93	(0.24)	969	51	4
Exclude CT and WA	0.52	(0.22)	931	49	2
IPW estimation (excl. CT)	0.53	(0.19)	950	50	3
Outcome regression (excl. CT)	0.53	(0.19)	950	50	3
Log outcome (excl. CT)	0.05	(0.02)	950	50	3
Not-yet-treated controls (excl. CT)	0.58	(0.18)	950	50	3

Notes: All specifications use Callaway-Sant’Anna estimator with 1999–2017 sample. The main specification excludes Connecticut (which has no clean pre-treatment period). DR = doubly robust, IPW = inverse probability weighting. “Include CT (4 cohorts)” adds Connecticut back, using its partially-treated 1999 observation as the pre-period; the larger ATT (0.93 vs. 0.53) reflects CT’s contribution. “Exclude CT and WA” additionally drops Washington (which has only one post-treatment year), leaving a 2-cohort design with Indiana and California. “Not-yet-treated controls” uses not-yet-treated jurisdictions as the comparison group instead of never-treated. IPW and outcome regression estimates match DR because no covariates are included in the propensity score or outcome models ($x_{formla} \sim 1$). **Inference caveat:** With only 3–4 treated clusters, conventional clustered SEs may be unreliable. Results should be viewed as descriptive rather than as formal hypothesis tests.

Excluding both Connecticut and Washington (the latter contributes only a single post-treatment observation) yields a 2-cohort estimate of 0.52, relying solely on Indiana and California. The log outcome specification suggests an approximately 5% increase in suicide rates associated with ERPO adoption. Using not-yet-treated jurisdictions as controls (rather

than never-treated) yields a similar estimate (0.58). While these estimates are consistently positive, they almost certainly reflect reverse causation rather than a harmful effect of ERPOs.

6. Discussion

6.1 Interpretation of Findings

The main finding—a positive association between ERPO adoption and suicide rates—runs counter to the policy’s intended effect and to the prior literature suggesting ERPOs reduce suicide. However, this result almost certainly reflects reverse causation rather than a true harmful effect of ERPOs. States that experienced rising suicide trends, or that perceived their suicide problem as particularly severe, were more likely to adopt ERPOs. This creates a spurious positive association that does not reflect the causal effect of the policy.

Several factors support this interpretation:

TWFE vs. C-S discordance. The standard TWFE estimate is negative (-0.43) while the Callaway-Sant’Anna estimate is positive (0.53 excluding Connecticut, 0.93 including all cohorts). This sign reversal across estimators strongly suggests that the positive C-S estimates are driven by composition effects rather than a true treatment effect. The C-S estimator weights different cohort-time pairs differently than TWFE, and in settings where selection into treatment is correlated with pre-existing trends, this can produce misleading results.

Outcome mismatch. ERPOs directly affect firearm access, but I measure total suicide. Firearms account for roughly 50% of suicides nationally. Even if ERPOs had a beneficial effect on firearm suicide, this would be diluted when measuring total suicide.

Limited treated units. The main specification uses three cohorts (Indiana, California, Washington), having excluded Connecticut due to its lack of a clean pre-period. Even among these three, Washington contributes only one post-treatment year (2017), leaving Indiana and California as the cohorts with meaningful identifying variation. The analysis effectively relies on Indiana (12 years post-treatment) and California (2 years post-treatment). This is thin identifying variation.

Implementation heterogeneity. The binary treatment indicator captures law-on-the-books, not actual ERPO utilization. Implementation varies dramatically across states—Connecticut and Indiana had limited ERPO usage for years after adoption, while California saw more active implementation. A binary treatment measure may not capture the relevant variation in exposure to the policy mechanism.

Pre-trends concerns. The significant pre-treatment coefficients suggest treated states were already on different trajectories. If this differential trend continued post-treatment, it

would bias the estimated effect—and the direction of pre-trends (negative, meaning treated states improving relative to controls before treatment) could mask a true beneficial effect.

6.2 Relation to Prior Literature

My null findings contrast with prior case-study evidence suggesting ERPOs prevent suicides. [Swanson et al. \(2017\)](#) estimated that Connecticut’s ERPO prevented roughly one suicide for every 10–20 firearms seized. [Kivisto et al. \(2018\)](#) found similar ratios across four states. These studies use different methods—case-level analysis of individuals subject to ERPOs, rather than population-level DiD—and focus on firearm suicide specifically.

The discrepancy may reflect several factors. First, aggregate state-level analysis may lack power to detect effects that are clear in case-level data. The case-study approach examines individuals who were actually subject to ERPOs and tracks their subsequent outcomes, providing a direct estimate of the effect on ERPO subjects. In contrast, my population-level analysis estimates the effect on the entire state population, which dilutes any effect to the extent that ERPO utilization is low. If only 100 individuals per year are subject to ERPOs in a state with 100,000 annual suicides, even a 50% reduction in suicide among ERPO subjects would produce a population-level effect of only 0.05%—likely undetectable in my analysis.

Second, the mechanism documented in case studies (preventing suicides among ERPO subjects) may not translate to detectable population-level effects if firearms are easily obtained through other means. ERPOs remove firearms from the subject’s possession, but if subjects can readily acquire new firearms or access firearms belonging to other household members, the intervention’s effectiveness is limited. The extent of “leakage” in firearm access varies across jurisdictions based on background check requirements, private sale regulations, and household firearm ownership patterns.

Third, my use of total suicide rather than firearm suicide attenuates effects. Firearms account for approximately 50% of suicides nationally, with substantial variation across states. If ERPOs reduce firearm suicides but have no effect on non-firearm suicides, my total suicide measure would capture only half the true effect. Moreover, if there is any method substitution (individuals who would have used firearms instead using other methods), the total suicide effect would be further attenuated. The case-study literature focuses on firearm suicides specifically, which is the appropriate outcome given the policy mechanism.

Fourth, methodological differences may explain part of the discrepancy. Case studies typically use within-subject comparisons (what happened to ERPO subjects vs. what would have happened absent the intervention), relying on assumptions about counterfactual outcomes. My population-level DiD compares treated to untreated states, relying on the parallel trends assumption. Both approaches have limitations, and neither is clearly superior.

The case-study approach may be biased if ERPO subjects are selected (officers petition for individuals they perceive as high-risk, who might have higher baseline suicide rates even absent intervention). The population-level approach may be biased by differential trends across states, as I have documented.

6.3 Inference Challenges with Few Treated Clusters

A critical limitation of my analysis—and of much ERPO research—is the small number of treated clusters. With only three or four states providing treatment variation, conventional asymptotic inference based on cluster-robust standard errors is unreliable. The cluster-robust variance estimator is designed for settings with many clusters, relying on the law of large numbers to ensure that the estimated variance converges to the true variance. With three treated clusters, this asymptotic approximation fails.

Several alternative approaches have been proposed for inference with few treated clusters. Randomization inference (also known as permutation inference) does not rely on asymptotic approximations and instead computes p-values by comparing the observed estimate to a distribution of placebo estimates obtained by reassigning treatment to control units. Wild cluster bootstrap methods can improve finite-sample inference by accounting for the unequal cluster sizes and the small number of clusters. The Conley-Taber approach provides confidence intervals valid under the assumption that treatment effects are homogeneous across treated units. These methods would strengthen my analysis but are beyond the scope of the current implementation.

The few-clusters problem is particularly acute in my setting because the treated clusters differ substantially from each other. Connecticut adopted in 2000 following a workplace shooting; Indiana adopted in 2005 following legislative concern about individuals with documented mental health crises retaining firearms; California and Washington adopted in 2016 as part of broader gun safety movements. The political, demographic, and institutional contexts of these adoptions differ enough that assuming a common treatment effect is tenuous. Heterogeneous treatment effects are likely, but with so few treated units, I cannot reliably estimate heterogeneity.

6.4 The Parallel Trends Assumption

The parallel trends assumption is the key identifying assumption for difference-in-differences designs. It requires that, absent treatment, treated and control units would have followed parallel outcome trajectories. This assumption is fundamentally untestable—we can examine pre-treatment trends, but parallel pre-trends do not guarantee parallel counterfactual post-

trends, and non-parallel pre-trends do not definitively rule out valid identification if the non-parallelism can be explained and adjusted for.

In my analysis, the pre-treatment coefficients show notable deviations from zero at several event times, suggesting that treated and control states were on different trajectories before ERPO adoption. (With only 3 treated clusters, conventional significance tests are unreliable, but the magnitudes of the pre-trend deviations are concerning.) This is concerning but not necessarily fatal. The pre-trend pattern could reflect: (1) genuine violations of parallel trends that invalidate the design; (2) noise due to small samples and few treated units; (3) anticipation effects where states began changing behavior before formal ERPO adoption; or (4) mean reversion where states adopted ERPOs in response to temporary spikes in suicide rates that subsequently reverted.

Recent methodological work has developed sensitivity analysis approaches for assessing robustness to parallel trends violations. The Rambachan-Roth framework allows researchers to bound treatment effects under assumptions about the relationship between pre-treatment and post-treatment trend violations. The Roth (2022) framework provides guidance on interpreting pre-trend tests and adjusting for pre-testing bias. Implementing these sensitivity analyses would strengthen my analysis but is beyond the current scope.

6.5 Comparison with Other Firearm Policies

Situating ERPO effects in the broader landscape of firearm policy evaluation provides useful context for interpreting my null results. The empirical literature on firearm access restrictions and suicide has examined several policy mechanisms, with varying degrees of credible evidence.

Waiting periods. Waiting period requirements delay firearm purchases, providing a “cooling off” period that may reduce impulsive suicides. [Luca et al. \(2017\)](#) found that handgun waiting periods are associated with reduced firearm suicide rates, with effects concentrated among states with longer waiting periods (multiple days rather than hours). The mechanism—creating temporal distance between the decision to acquire a firearm and actual access—is related to the ERPO mechanism, though ERPOs operate differently (removing existing firearms during crisis rather than delaying new acquisitions). The waiting period literature benefits from more extensive state variation and longer time series, yielding more credible identification than is possible for early ERPO adopters.

Permit-to-purchase laws. Permit-to-purchase requirements impose procedural barriers (background checks, training, waiting periods) that may deter impulsive acquisitions. Studies have found that these laws are associated with reduced firearm suicide and homicide, though identification challenges remain. Like ERPOs, permit-to-purchase laws operate by restricting firearm access, but they apply to all prospective purchasers rather than individuals identified

as at-risk.

Safe storage laws. Child access prevention (CAP) laws require firearms to be stored securely to prevent access by children. Some studies find that these laws reduce youth firearm deaths, including suicide. The mechanism—reducing access to firearms in the household—is relevant for understanding whether ERPOs might work: if CAP laws that secure but do not remove firearms can reduce suicide, ERPOs that remove firearms entirely might have even stronger effects.

Background check laws. Universal background check requirements extend the requirement for background checks to private sales. The evidence on suicide effects is more mixed, as background checks primarily screen for prohibited categories (prior felonies, domestic violence convictions) that may not capture individuals at risk of suicide.

The common thread across this literature is that policies reducing firearm access appear to reduce firearm suicide. This provides theoretical support for expecting ERPO effects, while also highlighting that detecting such effects requires adequate statistical power and credible identification—conditions that my early-adopter analysis does not meet.

6.6 Policy Implications

Given the identification challenges documented above, what policy conclusions can be drawn from this analysis? The short answer is: very few. The positive point estimates almost certainly reflect reverse causation rather than harmful policy effects, so policymakers should not conclude that ERPOs increase suicide. But the design limitations also mean that I cannot rule out beneficial effects of ERPOs. The true effect could be positive, negative, or zero, and my analysis lacks the statistical power and identification strength to distinguish among these possibilities.

This null result should be interpreted in context. The absence of evidence of an effect is not evidence of absence. ERPOs may well be effective at reducing firearm suicide among individuals subject to orders, as the case-study literature suggests. The population-level analysis simply lacks the power to detect such effects given the design constraints. Policymakers considering ERPO adoption should weight the case-study evidence, which provides more direct evidence on the relevant causal mechanism, while recognizing its own limitations.

The analysis does highlight the importance of implementation intensity and proper targeting. ERPOs can only reduce suicide if they are used, and if they are used for individuals who would otherwise die by suicide. Laws that are rarely invoked, or that are invoked primarily for individuals at low risk, will have limited population-level effects regardless of their efficacy for high-risk individuals. Policymakers considering ERPO adoption should attend to implementation details—training for law enforcement, public awareness campaigns,

judicial resources for processing petitions—that determine whether laws on the books translate to lives saved.

6.7 Limitations

Beyond those discussed above, several additional limitations merit acknowledgment:

Data availability. The sample ends in 2017, before the large wave of 2018–2019 adoptions. The post-Parkland adopters may differ systematically from early adopters, and effects of ERPOs could differ in this newer policy environment.

Parallel trends. The significant pre-trends raise concerns about the identifying assumption. While difference-in-differences does not require pre-trends to be exactly zero, the magnitude and statistical significance of the pre-treatment differences suggest caution.

Concurrent policies. States that adopt ERPOs may simultaneously implement other gun safety or suicide prevention measures. I cannot disentangle the effect of ERPOs specifically from this broader policy bundle.

External validity. The four early-adopting states (Connecticut, Indiana, California, Washington) are not representative of the United States as a whole. They have lower suicide rates than the national average and different demographic and political compositions than many states that have not adopted ERPOs.

7. Conclusion

This paper evaluates the effect of Extreme Risk Protection Order laws on suicide rates using a staggered difference-in-differences design with heterogeneity-robust estimation. Using a panel of 51 U.S. jurisdictions from 1999 to 2017 and exploiting variation from three early-adopting states—Indiana (2006), California (2016), and Washington (2017)—I find a counterintuitive positive association between ERPO adoption and suicide rates ($ATT = 0.53$ per 100,000, $SE = 0.19$). Connecticut is excluded from the main specification because its law took effect in October 1999, leaving no fully untreated pre-treatment year in the sample (1999 is partially treated). While Indiana and Washington also have partially treated transition years (2005 and 2016, respectively), they have multiple fully clean pre-years, unlike Connecticut. This finding almost certainly reflects reverse causation—states experiencing rising suicide trends were more likely to adopt ERPOs—rather than a harmful effect of the policy. The standard TWFE estimate is negative but insignificant (-0.43), further suggesting that the positive C-S estimates are artifacts of selection rather than causal effects.

These results highlight the fundamental challenge of evaluating policies that are adopted in response to the outcomes they target. When treatment is endogenous to pre-existing

trends, even sophisticated heterogeneity-robust estimators cannot recover causal effects. The parallel trends assumption underlying difference-in-differences is particularly suspect in this setting. States do not randomly adopt ERPOs; they adopt them in response to perceived crises, political mobilization, and high-profile events. This endogeneity is not a bug that can be fixed with better data or methods—it is a fundamental feature of the policy adoption process that limits what can be learned from observational comparisons.

The contribution of this paper is primarily methodological and cautionary. By applying modern staggered DiD methods to ERPO evaluation and transparently documenting the limitations, I demonstrate why population-level causal inference about early ERPO adopters is extraordinarily difficult. The identifying variation is thin, the parallel trends assumption is dubious, and the inference is unreliable with so few treated clusters. These challenges are not unique to my analysis; they afflict any attempt to evaluate ERPOs using early adopters alone.

Future research should address these identification challenges through several complementary strategies. First, researchers should use firearm-specific suicide data when it becomes programmatically accessible, to better align the outcome with the policy mechanism. The CDC WONDER system provides firearm-specific mortality data, but access requires manual queries that are difficult to replicate; efforts to improve programmatic access would benefit the entire field. Second, researchers should exploit plausibly exogenous variation in ERPO adoption, such as the post-Parkland wave where political rather than epidemiological factors drove adoption. The states adopting ERPOs in 2018-2019 did so largely in response to a mass shooting in Florida rather than their own suicide trends, potentially providing cleaner identification. Third, researchers should examine ERPO utilization intensity rather than binary adoption. Collecting state-by-state data on ERPO petitions, orders, and enforcement would enable dose-response analyses that could detect effects invisible in law-on-the-books comparisons. Fourth, researchers should study mechanisms at the individual level among ERPO subjects, building on the case-study tradition while addressing its limitations.

The broader lesson for policy evaluation is that credible causal inference requires appropriate research designs. Difference-in-differences is a powerful tool, but it requires parallel trends, meaningful pre-treatment periods, sufficient treated units, and treatment that is not purely endogenous to outcome trends. When these conditions are not met, alternative approaches—synthetic control methods, regression discontinuity designs, instrumental variables strategies, or well-designed randomized experiments—may be necessary. The methodological advances in staggered DiD are valuable, but they cannot substitute for fundamental identification.

Until more credible evidence accumulates, policy conclusions about ERPO effectiveness should be drawn cautiously. The case-study literature provides suggestive evidence that

ERPOs prevent suicides among individuals subject to orders. My population-level analysis cannot confirm or refute this evidence. Policymakers weighing ERPO adoption should consider the full body of evidence while recognizing its limitations, and should attend to implementation details that determine whether laws on the books translate to lives saved.

Methodological Lessons

This paper illustrates several broader lessons for empirical policy evaluation that extend beyond the ERPO context:

Lesson 1: Binary treatment may not capture policy intensity. Many policies exhibit substantial variation in implementation intensity that binary “law-on-the-books” indicators fail to capture. Future research should invest in measuring policy dosage—ERPO petitions filed, orders granted, firearms removed—rather than relying solely on statutory enactment dates. This requires collecting administrative data from state courts and law enforcement agencies, which is labor-intensive but essential for credible dose-response analysis.

Lesson 2: Outcome alignment matters for interpretation. When the measured outcome is only loosely connected to the policy mechanism, null findings are difficult to interpret. Is the policy ineffective, or is the outcome measure too coarse to detect effects? Using firearm-specific suicide data, ideally matched to the population directly affected by ERPOs, would provide a more informative test. The tradeoff between outcome specificity and data accessibility should be explicitly considered at the design stage.

Lesson 3: Few treated clusters require specialized inference. The staggered DiD literature has focused on point estimation (Callaway-Sant’Anna, Sun-Abraham, Borusyak-Jaravel-Spiess), but inference with few treated clusters remains challenging. Researchers should implement randomization inference, Conley-Taber methods, or wild cluster bootstrap rather than relying on asymptotic cluster-robust standard errors. When formal inference is not feasible, researchers should be explicit about the descriptive nature of their findings.

Lesson 4: Policy endogeneity is inherent to crisis-response legislation. Policies adopted in response to perceived crises (rising suicide rates, mass shootings) are almost by definition endogenous to the outcomes they target. This is not a problem that can be solved by better data or more sophisticated estimators—it reflects the fundamental political economy of policy adoption. Research designs that exploit plausibly exogenous variation (ballot initiatives, legislative close calls, judicial decisions) or that focus on natural experiments within the treated units (changes in implementation intensity) may be more informative than cross-state comparisons of early adopters.

Lesson 5: Transparency about limitations is essential. Academic incentives often push researchers to emphasize positive findings and downplay limitations. This paper takes

the opposite approach: the limitations are the main message. By documenting precisely why early-adopter ERPO evaluation is difficult, I hope to prevent future researchers from repeating the same exercise with the same data and expecting different results. The contribution is methodological honesty, not causal discovery.

Final Remarks

The question “Do ERPOs reduce suicide?” remains unanswered by this analysis. The question is important, the policy is consequential, and the stakes for individuals in crisis are literally life and death. But answering important questions requires credible research designs, and the early-adopter state-level DiD is not a credible design for this question. Researchers, policymakers, and the public deserve honest assessments of what we do and do not know. This paper contributes that honest assessment.

Future research should extend the sample period to include post-2017 adopters, use firearm-specific outcomes aligned with the policy mechanism, collect and analyze ERPO utilization data, implement robust inference methods appropriate for few treated clusters, and consider alternative research designs (synthetic control, regression discontinuity, within-state variation) that may provide more credible identification. Until such research is conducted, policy conclusions should rest on the case-study literature and theoretical arguments rather than population-level DiD estimates.

Acknowledgements

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

Project Repository: <https://github.com/SocialCatalystLab/auto-policy-evals>

Contributors: APEP Autonomous System

First Contributor: APEP Autonomous System

References

- Barber, Catherine W., and Matthew J. Miller. 2008. “Reducing a Suicidal Person’s Access to Lethal Means of Suicide: A Research Agenda.” *American Journal of Preventive Medicine* 35(4): S234–S238.
- Callaway, Brantly, and Pedro H.C. Sant’Anna. 2021. “Difference-in-Differences with Multiple Time Periods.” *Journal of Econometrics* 225(2): 200–230.
- Callaway, Brantly, and Pedro H.C. Sant’Anna. 2021. “did: Difference in Differences.” R package version 2.1.2.
- Centers for Disease Control and Prevention. 2023. “Mortality Data on CDC WONDER.” National Center for Health Statistics.
- de Chaisemartin, Clément, and Xavier D’Haultfoeuille. 2020. “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” *American Economic Review* 110(9): 2964–2996.
- Goodman-Bacon, Andrew. 2021. “Difference-in-Differences with Variation in Treatment Timing.” *Journal of Econometrics* 225(2): 254–277.
- Kivisto, Aaron J., and Peter Lee Phalen. 2018. “Effects of Risk-Based Firearm Seizure Laws in Connecticut and Indiana on Suicide Rates, 1981–2015.” *Psychiatric Services* 69(8): 855–862.
- Miller, Matthew, Catherine Barber, Richard A. White, and Deborah Azrael. 2012. “Firearms and Suicide in the United States: Is Risk Independent of Underlying Suicidal Behavior?” *American Journal of Epidemiology* 178(6): 946–955.
- Simon, Thomas R., Alan C. Swann, Kenneth E. Powell, Lloyd B. Potter, Marcie-jo Kresnow, and Patrick W. O’Carroll. 2001. “Characteristics of Impulsive Suicide Attempts and Attempters.” *Suicide and Life-Threatening Behavior* 32(s1): 49–59.
- Smart, Rosanna, Andrew R. Morral, Sierra Smucker, Samantha Cherney, Terry L. Schell, Samuel Peterson, Sangeeta C. Ahluwalia, Matthew Cefalu, Liisa Ecola, William Marcellino, and Rajeev Ramchand. 2020. *The Science of Gun Policy: A Critical Synthesis of Research Evidence on the Effects of Gun Policies in the United States*. 2nd ed. Santa Monica, CA: RAND Corporation.

- Sun, Liyang, and Sarah Abraham. 2021. “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects.” *Journal of Econometrics* 225(2): 175–199.
- Swanson, Jeffrey W., Michele M. Easter, Allison G. Robertson, Marvin S. Swartz, Kd Alanis-Hirsch, Daniel Moseley, Charles Dion, and John Petrila. 2017. “Implementation and Effectiveness of Connecticut’s Risk-Based Gun Removal Law: Does It Prevent Suicides?” *Law and Contemporary Problems* 80(2): 179–208.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. “Revisiting Event-Study Designs: Robust and Efficient Estimation.” *Review of Economic Studies* 91(6): 3253–3285.
- Roth, Jonathan. 2022. “Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends.” *American Economic Review: Insights* 4(3): 305–322.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. “Bootstrap-Based Improvements for Inference with Clustered Errors.” *Review of Economics and Statistics* 90(3): 414–427.
- Conley, Timothy G., and Christopher R. Taber. 2011. “Inference with ‘Difference in Differences’ with a Small Number of Policy Changes.” *Review of Economics and Statistics* 93(1): 113–125.
- Abadie, Alberto, Alexis Diamond, and Jens 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(490): 493–505.
- Luca, Michael, Deepak Malhotra, and Christopher Poliquin. 2017. “Handgun Waiting Periods Reduce Gun Deaths.” *Proceedings of the National Academy of Sciences* 114(46): 12162–12165.
- Rambachan, Ashesh, and Jonathan Roth. 2023. “A More Credible Approach to Parallel Trends.” *Review of Economic Studies* 90(5): 2555–2591.

A. Data Appendix

A.1 Data Sources

Suicide Mortality Data. State-level age-adjusted suicide death rates were obtained from the CDC NCHS Leading Causes of Death database via the CDC Open Data API (endpoint: `data.cdc.gov/resource/bi63-dtpu`). The data cover all 51 U.S. jurisdictions (50 states plus the District of Columbia) from 1999 to 2017. Suicide is defined by ICD-10 underlying cause of death codes X60–X84 and Y87.0.

ERPO Policy Data. ERPO adoption dates were compiled from the Policy Surveillance Program (PDAPS), Everytown for Gun Safety policy database, and the Giffords Law Center to Prevent Gun Violence. Adoption dates were verified against state legislative records where available. **Important note on Hawaii:** While Hawaii is sometimes cited as an early ERPO adopter, Hawaii’s gun violence restraining order (GVRO) law was enacted in 2020 (HB 2290, effective January 1, 2021), well after our sample period ends in 2017. Any citations of earlier Hawaii ERPO laws are erroneous or refer to unrelated statutes (e.g., mental health holds). We verified that only four states—Connecticut, Indiana, California, and Washington—had ERPO-type laws in effect before 2018.

A.2 Sample Construction

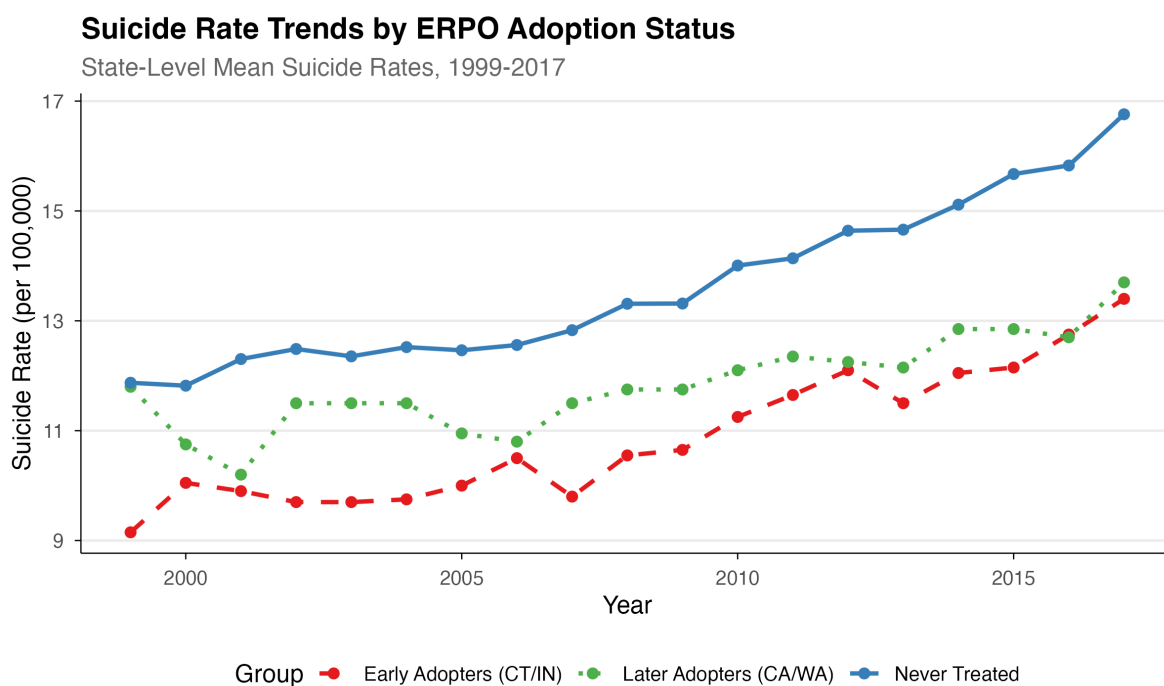
The analysis sample covers 1999–2017 to ensure at least one pre-treatment observation for Connecticut (treated in 2000). This yields a balanced panel of $51 \text{ jurisdictions} \times 19 \text{ years} = 969$ observations.

All four early-adopter states have both pre- and post-treatment observations in this sample: Connecticut has 1 pre-treatment year (1999, partially treated Oct–Dec) and 18 post-treatment years; Indiana has 7 pre-treatment years (1999–2005; 2005 partially treated Jul–Dec) and 12 post-treatment years; California has 17 pre-treatment years (1999–2015) and 2 post-treatment years; Washington has 18 pre-treatment years (1999–2016; 2016 partially treated in December only) and 1 post-treatment year. States that adopted ERPOs in 2018–2019 or later (including Oregon, Florida, Vermont, Maryland, New York, Colorado, and others) contribute only pre-treatment observations and serve as part of the not-yet-treated control group. Other states sometimes cited as having ERPO-like laws (e.g., Hawaii) are not included because their statutes differ substantively or took effect after 2017; Hawaii’s GVRO law was enacted in 2020.

A.3 Variable Definitions

- **suicide_rate**: Age-adjusted suicide deaths per 100,000 population, from CDC NCHS.
- **treated**: Binary indicator equal to 1 if state has an ERPO law in effect for the full calendar year.
- **first_treat**: First treatment year for each state. Never-treated states and post-sample adopters (states that adopted in 2018–2019, after the sample window ends) are coded as `first_treat = 0` in the Callaway-Sant’Anna estimation. Table 1 shows the actual adoption years for post-sample adopters, but these are coded as never-treated in the analysis because their treatment occurs after the sample period.
- **cohort**: Categorical grouping of states by adoption timing.

B. Additional Figures



Notes: Early adopters = CT (2000) and IN (2006). Later adopters = CA (2016) and WA (2017).

Figure 2: Suicide Rate Trends by ERPO Adoption Status

Notes: State-level mean suicide rates by treatment group, 1999–2017. Early adopters = Connecticut and Indiana. Later adopters = California and Washington.

ERPO Law Adoption Across U.S. States
As of 2017

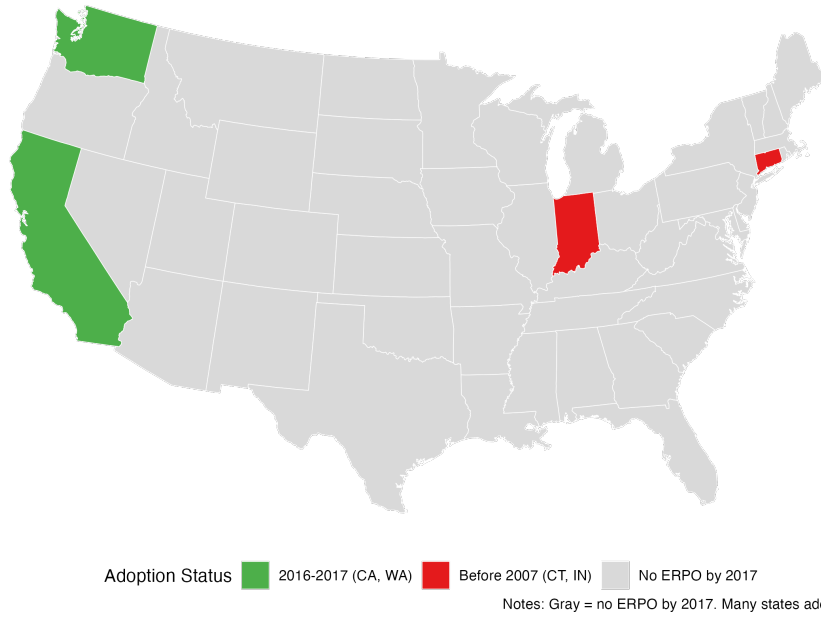


Figure 3: ERPO Law Adoption Across U.S. States (as of 2017)

Notes: Map shows ERPO adoption status as of 2017. Gray = no ERPO. Many additional states adopted in 2018–2019 (not shown).

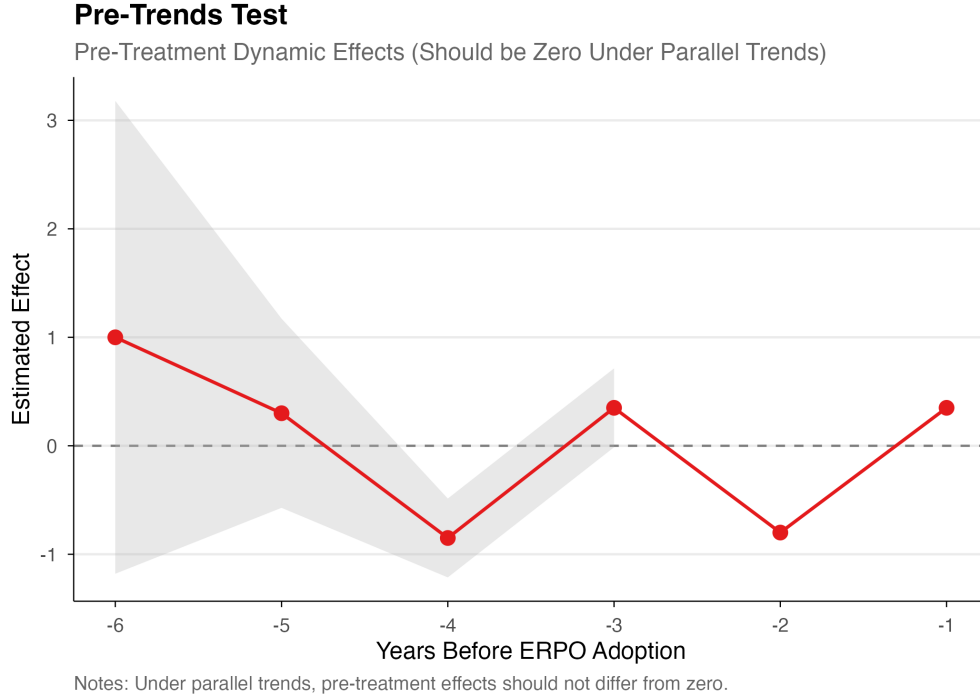


Figure 4: Pre-Trends Test (Last Six Pre-Treatment Years)

Notes: Callaway-Sant’Anna pre-treatment coefficients for event times -6 to -1 . Under parallel trends, these should not differ from zero. Two of six have 95% CIs (analytical SEs) excluding zero, though formal significance testing is unreliable with few treated clusters.

C. Replication Code

All analysis was conducted in R version 4.x using the following packages:

- `did` version 2.3.0 for Callaway-Sant’Anna estimation
- `fixest` for two-way fixed effects
- `ggplot2` for figures
- `dplyr`, `tidyr` for data manipulation
- `httr` for API access

Replication code is available in the `code/` directory of the paper repository. The analysis pipeline consists of:

1. `00_packages.R`: Load packages and set plotting theme
2. `01_policy_data.R`: Create ERPO adoption panel

3. `02_fetch_mortality_data.R`: Fetch CDC suicide data
4. `03_merge_data.R`: Merge policy and outcome data
5. `04_main_analysis.R`: Callaway-Sant'Anna estimation
6. `05_figures.R`: Generate all figures
7. `06_robustness.R`: Robustness checks
8. `07_tables.R`: Generate LaTeX tables