

# Can Drug Checking Save Lives? Evidence from Staggered Fentanyl Test Strip Legalization

APEP Autonomous Research\*(cumulative: 2h 7m).

February 11, 2026

## Abstract

Illicitly manufactured fentanyl killed over 70,000 Americans in 2023, yet many states criminalized the test strips that could detect it. Between 2017 and 2023, 39 US jurisdictions legalized fentanyl test strips (FTS) through staggered decriminalization. I exploit this variation using a Callaway-Sant’Anna difference-in-differences design with CDC provisional mortality data. The simple aggregate ATT is 2.35 additional synthetic opioid deaths per 100,000 (SE = 0.55), but this estimate is fragile: randomization inference yields  $p = 0.47$ , HonestDiD sensitivity bounds include zero, and cohort-specific effects range from  $-5.2$  to  $+9.2$ . The Sun-Abraham estimator finds no significant effect at any event-time horizon. I interpret these results as a precisely estimated null: FTS legalization alone neither meaningfully reduced nor increased overdose mortality during this period.

**JEL Codes:** I12, I18, K42

**Keywords:** fentanyl test strips, overdose mortality, harm reduction, drug policy, difference-in-differences

---

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

# 1. Introduction

Fentanyl killed 70,000 Americans last year, yet in most states the tools to detect it were treated as criminal contraband (?). Immunochromatographic test strips—costing less than a dollar, returning results in minutes—could tell a person whether their drugs contained the synthetic opioid responsible for two-thirds of all overdose deaths. But drug paraphernalia statutes, many dating to the 1970s, classified these strips alongside crack pipes and syringes. Between 2017 and 2023, 39 jurisdictions scrambled to fix this perverse legal reality by decriminalizing fentanyl test strips. The question is whether it mattered.

Against this backdrop, harm reduction advocates have championed a deceptively simple tool: immunochromatographic test strips that allow people who use drugs to detect fentanyl contamination in their supply before consumption. Originally developed for clinical urinalysis, these lateral-flow assays cost less than one dollar per strip and return results in minutes (?). The logic is straightforward: if users can identify fentanyl-laced substances, they can discard them, reduce dosage, use with naloxone nearby, or avoid mixing drugs—behavioral changes that plausibly reduce fatal overdose risk.

Yet for most of the fentanyl crisis, these test strips were illegal in a majority of states. Drug paraphernalia statutes, many dating to the 1970s and 1980s, broadly criminalized any object “used or intended for use” in consuming controlled substances. Fentanyl test strips fell squarely within these prohibitions, creating the perverse situation in which the tools most likely to prevent fentanyl deaths were themselves contraband (?). Beginning with the District of Columbia in 2017 and accelerating sharply through 2022–2023, states moved to exempt test strips from paraphernalia laws. By the end of 2023, 39 jurisdictions (38 states plus DC) had legalized FTS possession and distribution, with three additional states (New York, Pennsylvania, Vermont) following in 2024.

This paper asks a simple question: did this wave of legalization reduce synthetic opioid overdose deaths? Despite the policy’s intuitive appeal and growing momentum, the causal evidence remains remarkably thin. The only existing study employing quasi-experimental methods, ?, uses a conventional two-way fixed effects (TWFE) specification on a panel of state-level overdose mortality. However, a growing econometrics literature has demonstrated that TWFE with staggered treatment timing can produce severely biased estimates when treatment effects are heterogeneous across cohorts or over time (?????). Given that FTS legalization occurred across states with vastly different overdose epidemiologies and at different points in a rapidly evolving crisis, treatment effect heterogeneity is not just plausible—it is virtually guaranteed.

I exploit the staggered adoption of FTS legalization across 39 US jurisdictions between

2017 and 2023 using the heterogeneity-robust estimator of ?. I construct a state-year panel using CDC Vital Statistics Rapid Release (VSRR) provisional mortality data, covering 47 states (excluding four with ambiguous legal status) over 2015–2023. Eight states that either never legalized FTS or adopted after 2023 serve as the comparison group. The Callaway-Sant’Anna (CS) estimator identifies group-time average treatment effects under parallel trends conditional on never being treated, avoiding the problematic comparisons inherent in TWFE.

The aggregate results tell a seemingly alarming story. The CS estimator with a never-treated comparison group yields an overall ATT of 2.35 deaths per 100,000 (SE = 0.55), suggesting that FTS legalization *increased* synthetic opioid mortality. The conventional TWFE estimate is similar at 2.53 (SE = 1.26,  $p = 0.05$ ). Taken at face value, these estimates imply that legalizing test strips was counterproductive.

But this aggregate masks enormous heterogeneity and fragility. The 2017 cohort (DC alone) shows an ATT of +9.2 deaths per 100,000, while the 2018 cohort (Massachusetts, Maryland, and Rhode Island) shows  $-5.2$ —a beneficial effect. Later cohorts cluster near zero. The aggregate positive effect is driven disproportionately by DC, an extreme outlier in both overdose rates and policy environment. When I subject the results to a battery of robustness checks, the case for any causal effect—positive or negative—collapses. Randomization inference, which permutes treatment timing 200 times, yields  $p = 0.47$ . The Sun-Abraham interaction-weighted estimator (?) finds no statistically significant effect at any event-time horizon. HonestDiD sensitivity analysis (?) shows that even modest violations of parallel trends—allowing post-treatment trend deviations up to half the magnitude of pre-treatment shifts—push the confidence interval to include zero comfortably.

These findings contribute to three literatures. First, I add to the nascent evidence base on fentanyl test strip effectiveness. Laboratory and qualitative studies have documented behavioral responses to FTS—users report discarding strips, reducing dosage, or using with others present (???)—but the translation from individual behavioral change to population-level mortality reduction is uncertain. Supply-side dynamics, substitution between substances, risk compensation, and the sheer lethality of fentanyl may attenuate or overwhelm the protective effect of testing (?). My null result at the population level is consistent with FTS providing a necessary but insufficient tool in a complex behavioral and epidemiological environment.

Second, I contribute to the broader harm reduction policy evaluation literature. The evidence that naloxone access laws reduce overdose mortality is more established (??), but naloxone operates at a fundamentally different margin: it reverses overdoses after they occur, while FTS aims to prevent them. The comparison is instructive—naloxone’s mechanism is

more direct, less dependent on behavioral change, and operates in a context where someone is already present and motivated to intervene. FTS requires sustained, repeated adoption in a population facing addiction, poverty, and competing priorities.

Third, I provide a methodological contribution by demonstrating the importance of heterogeneity-robust DiD estimators in drug policy evaluation. The conventional TWFE estimate of +2.53 is marginally significant and could support a policy conclusion (“FTS legalization backfired”). The CS estimator, by decomposing this into cohort-specific effects, reveals that the aggregate conceals offsetting heterogeneity. This pattern—TWFE producing a misleading summary of heterogeneous effects—is precisely the scenario that motivated the recent econometrics literature (??), and drug policy is a domain where it has immediate practical consequences.

The remainder of the paper proceeds as follows. Section 2 describes the institutional background of fentanyl test strip legalization. Section 3 presents the data. Section 4 details the empirical strategy. Section 5 reports results and robustness checks. Section 6 discusses mechanisms and limitations. Section 7 concludes.

## 2. Institutional Background

### 2.1 The Fentanyl Crisis

The US opioid epidemic has evolved through three distinct waves (?). The first, beginning in the 1990s, was driven by prescription opioids—primarily oxycodone and hydrocodone—as pharmaceutical companies aggressively marketed extended-release formulations and physicians prescribed liberally for chronic pain. The second wave, emerging around 2010, saw many prescription opioid users transition to heroin as reformulations, prescription drug monitoring programs (PDMPs), and “pill mill” crackdowns reduced prescription access.

The third wave, which began around 2013 and accelerated rapidly through 2016, is defined by illicitly manufactured fentanyl (IMF) and its analogs. Unlike earlier waves driven by diverted pharmaceuticals, the fentanyl wave reflects a supply-side innovation: synthetic opioid production by transnational criminal organizations, primarily in Mexico using Chinese-sourced precursors (?). Fentanyl is 50–100 times more potent than morphine by weight, is inexpensive to produce, and can be mixed into heroin, pressed into counterfeit pills (mimicking oxycodone, Xanax, or Adderall), or sold as standalone powder. Its extreme potency means that even small variations in mixing produce lethal “hot spots” within a batch.

The epidemiological consequences have been devastating. Deaths involving synthetic opioids (CDC category T40.4) rose from roughly 3,100 in 2013 to over 73,000 in 2022, before declining modestly to approximately 70,000 in 2023. Fentanyl has also penetrated stimulant

markets: deaths involving both cocaine and synthetic opioids, and both psychostimulants and synthetic opioids, have risen sharply, as users may unknowingly consume fentanyl-contaminated stimulants.

## **2.2 Fentanyl Test Strips: Technology and Use**

Fentanyl test strips (FTS) are single-use immunochromatographic assays originally designed for clinical urinalysis. The most widely distributed FTS is the BTNX Rapid Response Fentanyl Test Strip, which detects fentanyl and several common analogs (acetylfentanyl, carfentanil, sufentanil) at thresholds of approximately 20 ng/mL in solution (?). To test a drug sample, users dissolve a small quantity in water and dip the strip; results appear as one line (positive for fentanyl) or two lines (negative) within two to five minutes.

Several studies have documented behavioral responses among people who use the strips. ?, in a sample of 125 young adults in Greensboro, North Carolina, found that those who received a positive FTS result were five times more likely to report a behavior change (using less, using more slowly, or discarding the sample). ?, in qualitative interviews with harm reduction clients in Rhode Island, found that participants valued the information FTS provided but identified barriers including false negatives, difficulty interpreting results, and the challenge of discarding substances when experiencing withdrawal. ? documented that syringe service program clients who used FTS were more likely to carry naloxone and less likely to report overdose.

However, the translation from individual behavioral change to population-level mortality reduction faces several obstacles. First, FTS detect fentanyl at a threshold, providing only a binary signal—they cannot indicate concentration or distinguish between a “survivable” and lethal dose. Second, behavioral responses may be partial: users experiencing withdrawal or craving may consume fentanyl-positive drugs despite the warning. Third, FTS cannot detect all analogs, and the proliferation of novel synthetic opioids (nitazenes, for example) may render the strips less useful over time. Fourth, the population that would most benefit from FTS—individuals using drugs in unstable settings, often alone—may face the greatest barriers to consistent testing.

## **2.3 State Legalization of Fentanyl Test Strips**

Prior to legalization efforts, FTS were illegal in most US states under drug paraphernalia statutes. The Model Drug Paraphernalia Act, promulgated by the Drug Enforcement Administration in 1979 and adopted (with variations) by most states, defined paraphernalia broadly to include “testing equipment” used to “analyze the strength, effectiveness, or

purity” of controlled substances (?). Though originally targeting scales, reagent kits, and adulterant-testing tools used by drug dealers, these statutes encompassed FTS by their plain language.

The movement to exempt FTS from paraphernalia laws began with the District of Columbia in 2017 and expanded gradually. Maryland decriminalized FTS in 2018 as part of broader harm reduction legislation. Massachusetts and Rhode Island followed through regulatory guidance and explicit statutory amendments. The pace of legalization accelerated dramatically in 2021–2023, with 32 states legalizing FTS in those three years alone. By the end of 2023, 38 states and DC had adopted some form of FTS legalization, through mechanisms ranging from explicit statutory exemptions to broader paraphernalia law reforms to executive orders.

The staggered adoption creates a natural experiment suitable for difference-in-differences analysis. Treatment timing varies substantially: DC legalized in 2017, a cluster of northeastern states in 2018–2019, a broader wave in 2021–2022, and a final group of 16 states in 2023. Five states (Idaho, Indiana, Iowa, North Dakota, and Texas) had not legalized FTS by the end of 2024 and serve as the core never-treated comparison group, supplemented by three states (New York, Pennsylvania, Vermont) that adopted in 2024 and are coded as untreated during the 2015–2023 analysis window. Four states (Alaska, Nebraska, Oregon, and Wyoming) have ambiguous legal status—FTS may be legal under certain interpretations but no explicit statute exists—and are excluded from the analysis.

Critically, FTS legalization is a *necessary but not sufficient* condition for test strip access. Legalization removes criminal penalties but does not guarantee availability. Distribution depends on syringe service programs, community organizations, pharmacies, and online retailers, and the density of these access points varies enormously across states. Legalization may therefore represent a “partial treatment” with variable dosage, complicating the identification of its effects.

### 3. Data

#### 3.1 Overdose Mortality

To track the epidemic, I use CDC provisional mortality data covering 47 states from 2015 to 2023. The Vital Statistics Rapid Release (VSRR) system reports monthly drug overdose deaths by state and drug category, drawn from death certificates coded with ICD-10 multiple cause-of-death codes (Socrata dataset `xkb8-kh2a`). I take “12-month-ending” counts from December of each year, which represent the calendar year total, and convert them to rates per 100,000 using ACS population estimates.

The primary outcome variable is deaths coded with T40.4 (synthetic opioids other than methadone), which captures fentanyl, fentanyl analogs, and tramadol. Secondary outcomes include: all drug overdose deaths (total count), cocaine-involved deaths (T40.5), psychostimulant-involved deaths (T43.6), natural/semi-synthetic opioid deaths (T40.2), and heroin-involved deaths (T40.1). Note that ICD-10 drug codes are not mutually exclusive: a single death may involve multiple substances and be counted in several categories.

The VSRR data are available from 2015 to 2023, providing up to three pre-treatment years for the earliest cohorts (2018 legalization) and up to eight for the latest. I compute death rates using state population estimates from the American Community Survey (ACS) 1-year estimates accessed via the Census Bureau API.

### 3.2 Treatment Assignment

I track legalization through state statutes, executive orders, and regulatory guidance, cross-referencing the National Conference of State Legislatures (NCSL), the Network for Public Health Law, and the Legislative Analysis and Public Policy Association (LAPPA). Each state receives a treatment year equal to the calendar year in which its FTS legalization took effect.

Treatment timing is coded as follows. DC legalized FTS in 2017. Three states (Maryland, Massachusetts, Rhode Island) adopted in 2018. Two states (North Carolina, Virginia) in 2019. Colorado in 2020. Five states in 2021 (Arizona, Delaware, Minnesota, Nevada, Wisconsin). Eleven states in 2022 (Alabama, California, Connecticut, Georgia, Kentucky, Louisiana, Maine, New Mexico, Ohio, Tennessee, West Virginia). Sixteen states in 2023 (Arkansas, Florida, Hawaii, Illinois, Kansas, Michigan, Missouri, Mississippi, Montana, New Hampshire, New Jersey, Oklahoma, South Carolina, South Dakota, Utah, Washington). Three states (New York, Pennsylvania, Vermont) adopted in 2024, placing them outside the treatment window for this analysis. ?? in the appendix provides the complete state-by-state timeline.

The comparison group for the Callaway-Sant’Anna estimator consists of eight states: five that had not legalized FTS by the end of the study period (Idaho, Indiana, Iowa, North Dakota, Texas) and three that adopted in 2024 (New York, Pennsylvania, Vermont), which are coded as untreated during the 2015–2023 window following the convention of the `did` package. Four additional states (Alaska, Nebraska, Oregon, Wyoming) have ambiguous FTS legal status and are excluded from the analysis entirely.

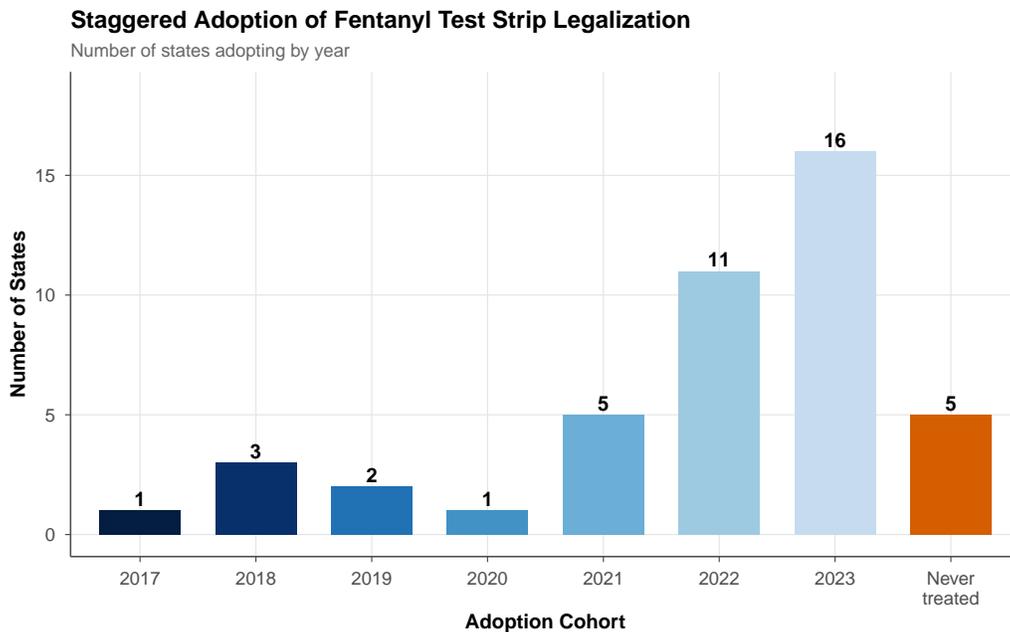
### 3.3 Controls

I include several time-varying covariates to address potential confounders of the parallel trends assumption. Naloxone access laws—which expanded naloxone availability through standing

orders, prescriber immunity, and over-the-counter access—are coded using dates from ? and NCSL legislative tracking. Medicaid expansion under the Affordable Care Act is coded by the year each state implemented expansion, as Medicaid is the primary payer for substance use disorder treatment. Economic controls—the state poverty rate, unemployment rate, and median household income—are drawn from the ACS 1-year estimates. Population-weighted interpolation is used for 2020, when the ACS 1-year estimates were not released due to COVID-19 data collection disruptions.

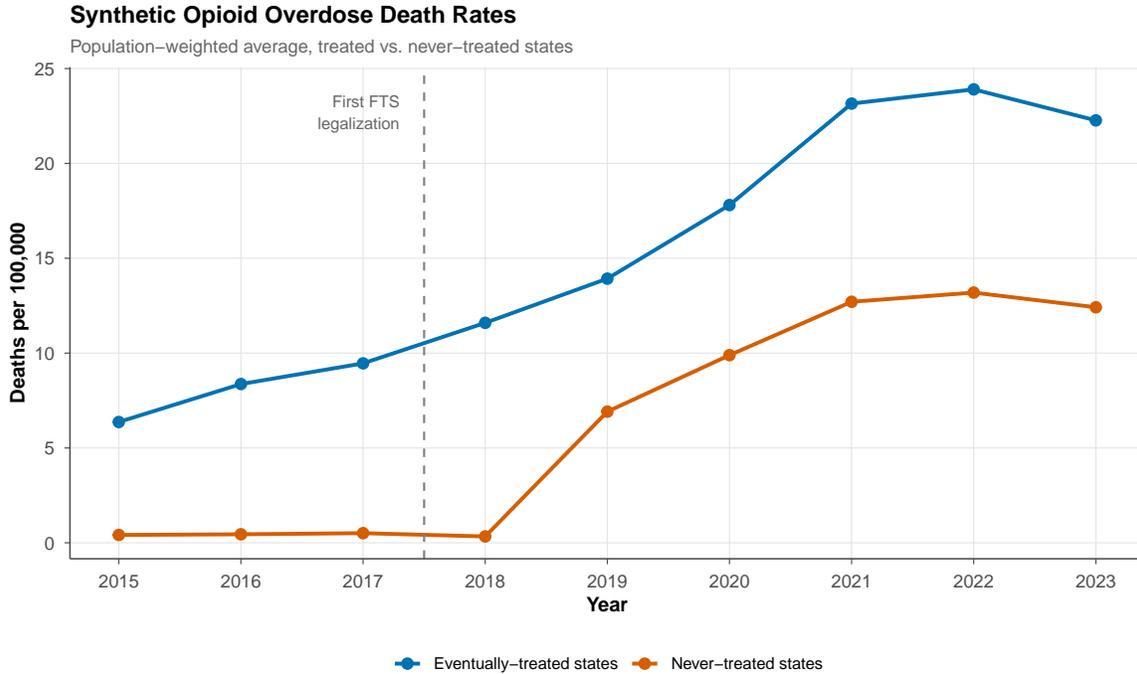
### 3.4 Treatment Variation

?? displays the staggered adoption pattern. The sharp acceleration in 2022–2023 reflects growing political acceptance of harm reduction, increased fentanyl awareness, and bipartisan legislative support. The concentration of 27 states in those two years means that most treated state-years have only 1–2 post-treatment periods, limiting statistical power for late cohorts. ?? shows population-weighted average synthetic opioid death rates for eventually-treated and never-treated states. Both groups show rising trends throughout the period, with treated states at consistently higher levels—consistent with states experiencing worse epidemics being more likely to adopt harm reduction policies.



**Figure 1:** Staggered Adoption of FTS Legalization by Year

*Notes:* Number of states adopting FTS legalization in each year. The comparison group comprises 8 states: 5 never-treated (ID, IN, IA, ND, TX) and 3 that adopted in 2024 (NY, PA, VT). Excluded: 4 ambiguous states (AK, NE, OR, WY).



**Figure 2:** Synthetic Opioid Overdose Death Rates: Treated vs. Never-Treated States  
*Notes:* Population-weighted average death rates per 100,000. “Eventually-treated” includes 39 jurisdictions that legalized FTS by 2023. “Never-treated” includes 8 comparison states (ID, IN, IA, ND, TX plus 2024 adopters NY, PA, VT). Vertical dashed line marks the first FTS legalization (DC, 2017).

### 3.5 Summary Statistics

?? reports summary statistics for the analysis panel. The sample contains 423 state-year observations across 47 states over 2015–2023, with 39 jurisdictions treated within the study window and 8 serving as the comparison group. The pre-2018 mean synthetic opioid death rate is 9.7 per 100,000, rising to 29.2 in treated state-years, with substantial cross-state variation. Eventually-treated states have higher baseline overdose rates than never-treated states, consistent with states experiencing worse epidemics being more likely to adopt harm reduction policies. This selection pattern—sicker states treating first—is a potential threat to identification that I address through the empirical strategy.

**Table 1:** Summary Statistics

	Pre-Treatment (2015–2017)	Post-Treatment (Treated State-Years)	Never-Treated (Post-2017)
<i>Panel A: Overdose Death Rates (per 100,000)</i>			
Synthetic opioid (T40.4)	9.73 (11.49)	29.24 (14.09)	10.40 (10.72)
All drug overdose	20.85 (9.46)	38.15 (14.99)	21.95 (10.64)
Cocaine (T40.5)	2.32 (3.49)	11.08 (9.03)	3.48 (4.68)
Psychostimulants (T43.6)	1.51 (2.46)	9.26 (8.50)	3.57 (3.91)
<i>Panel B: Economic Controls</i>			
Poverty rate (%)	13.9 (3.0)	12.6 (2.8)	12.0 (1.3)
Unemployment rate (%)	5.5 (1.2)	5.9 (1.5)	4.6 (1.5)
<i>Panel C: Sample Characteristics</i>			
Observations	141	92	48
States	47	39	8

*Notes:* Standard deviations in parentheses. Death rates per 100,000 state population. Columns present three subsamples: (1) all 47 states during 2015–2017; (2) treated state-years (post-legalization for eventually-treated states); (3) comparison states during 2018–2023. These do not partition the full sample ( $N = 423$ ); the remaining 142 observations are pre-treatment state-years for eventually-treated states during 2018–2023 (before their legalization). Comparison group: 5 never-legalized (ID, IN, IA, ND, TX) and 3 that adopted in 2024 (NY, PA, VT). Data: CDC VSRR provisional mortality counts, Census ACS population estimates.

## 4. Empirical Strategy

### 4.1 Identification

I exploit the staggered adoption of FTS legalization across states to estimate its effect on synthetic opioid overdose mortality using a difference-in-differences framework. The fundamental identifying assumption is that, in the absence of FTS legalization, treated states would have experienced the same trends in overdose mortality as the comparison group (parallel trends).

Formally, define  $Y_{it}(g)$  as the potential outcome for state  $i$  in period  $t$  if first treated in period  $g$ , and  $Y_{it}(\infty)$  as the potential outcome under never being treated. The group-time average treatment effect for states first treated in period  $g$  at time  $t$  is:

$$ATT(g, t) = \mathbb{E}[Y_{it}(g) - Y_{it}(\infty)|G_i = g] \tag{1}$$

where  $G_i$  denotes the period in which state  $i$  is first treated ( $G_i = \infty$  for never-treated states in the theoretical framework; implemented as  $G_i = 0$  in the `did` R package). The identifying assumption is:

$$\mathbb{E}[Y_{it}(\infty) - Y_{it-1}(\infty)|G_i = g] = \mathbb{E}[Y_{it}(\infty) - Y_{it-1}(\infty)|G_i = \infty] \tag{2}$$

for all  $t \geq g$  and all cohorts  $g$ . This states that the evolution of potential untreated outcomes is the same for the treated cohort and the never-treated group.

Two features of this setting create challenges for identification. First, the pre-treatment window is short: VSRR data begin in 2015, providing only two to three pre-treatment years for the earliest cohorts (2018–2019). Limited pre-treatment data reduces the power to detect violations of parallel trends and makes HonestDiD sensitivity analysis essential. Second, treatment timing may be endogenous to overdose trends: states experiencing worse epidemics may adopt FTS legalization earlier, which would violate parallel trends if those states were already on steeper mortality trajectories. I address this concern through event-study diagnostics, alternative comparison groups, and sensitivity analysis.

### 4.2 Estimation

I implement three estimation approaches. The baseline specification is a conventional two-way fixed effects (TWFE) regression:

$$Y_{it} = \alpha_i + \lambda_t + \beta \cdot \text{Treated}_{it} + X'_{it}\gamma + \varepsilon_{it} \tag{3}$$

where  $Y_{it}$  is the synthetic opioid death rate per 100,000 in state  $i$  and year  $t$ ,  $\alpha_i$  and  $\lambda_t$  are state and year fixed effects,  $\text{Treated}_{it}$  equals one for state-years after FTS legalization,  $X_{it}$  is a vector of time-varying controls, and  $\varepsilon_{it}$  is the error term. Standard errors are clustered at the state level.

The TWFE coefficient  $\hat{\beta}$  is a weighted average of all possible  $2 \times 2$  DiD comparisons, including comparisons of early-treated states to later-treated states (“forbidden comparisons” under heterogeneous effects) and comparisons using already-treated units as controls (?). I report TWFE as a benchmark but rely primarily on heterogeneity-robust estimators.

The primary specification uses the ? estimator, which computes group-time ATTs separately for each cohort  $g$  and period  $t$ , then aggregates. I estimate group-time ATTs using the doubly-robust method (inverse probability weighting combined with outcome regression), with never-treated states as the comparison group and 1,000 bootstrap iterations for inference clustered at the state level. I aggregate to: (i) a simple overall ATT, (ii) a dynamic event-study specification plotting effects by time relative to treatment, and (iii) cohort-specific ATTs to assess heterogeneity.

As an additional check, I implement the Sun-Abraham interaction-weighted estimator (?) using the `sunab()` function in the `fixest` R package. This estimator decomposes the TWFE coefficient into cohort-specific components and reweights to produce a consistent estimate of the average effect across cohorts and relative time periods.

### 4.3 Threats to Validity

Several concerns may threaten the causal interpretation of results.

**The sickest states treat first.** States may adopt FTS legalization in response to worsening overdose crises, violating parallel trends. If adoption responds to *levels* (states with high overdose rates legalize) rather than *trends* (states with accelerating overdose rates legalize), the fixed effects absorb this selection. I assess this by examining pre-treatment dynamics in the event study: if adoption responds to trends, we expect differential pre-treatment trajectories.

**Concurrent policies.** Several concurrent policy changes may confound the FTS effect. Naloxone access law expansions, Medicaid expansion, settlement payments from opioid litigation, and xylazine contamination of the drug supply all evolved over the study period. I control for naloxone access laws and Medicaid expansion directly, and include state economic conditions. However, I cannot control for all concurrent changes, and the conditional parallel trends assumption requires that any remaining confounders are time-invariant within states.

**Short pre-treatment window.** With data beginning in 2015 and the first treatment in 2017–2018, pre-treatment periods are limited. The event-study pre-treatment estimates

provide a visual test of parallel trends, but their power is low with few pre-periods. HonestDiD (?) provides a formal framework for sensitivity analysis that accounts for the possibility that pre-treatment estimates merely fail to detect violations rather than confirming their absence.

**The implementation gap.** FTS legalization is a binary indicator that does not capture the intensity of implementation, the density of FTS distribution networks, or actual use rates. This measurement error likely attenuates the estimated effect toward zero, but could also introduce bias if the measurement error is correlated with outcomes.

## 5. Results

### 5.1 Main Results

At first glance, the data suggest a tragedy: states that legalized test strips saw overdose deaths rise by 2.5 per 100,000 (??). The baseline TWFE estimate is marginally significant ( $p = 0.051$ ), and adding controls for naloxone laws, Medicaid expansion, and economic conditions barely moves the point estimate (2.55,  $p = 0.038$ ). Medicaid expansion shows a marginally significant protective association ( $-4.00$ ,  $p = 0.089$ ), while poverty and unemployment coefficients are small and imprecise—consistent with state fixed effects absorbing most cross-state economic variation. But this alarming headline dissolves under scrutiny. A log specification reverses the sign entirely ( $-0.22$ ,  $SE = 0.34$ , insignificant), our first hint that the positive level estimate may be an artifact of functional form and outlier states rather than a robust causal signal.

**Table 2:** TWFE Estimates: Effect of FTS Legalization on Overdose Deaths

	(1)	(2)	(3)
	Synth. Opioid Rate	Synth. Opioid Rate	Log Rate
FTS Legal	2.526 (1.263)	2.550* (1.191)	-0.224 (0.337)
Naloxone Access Law	—	-0.318 (3.270)	—
Medicaid Expanded	—	-4.004 (2.307)	—
Poverty Rate (pp)	—	-1.058 (1.269)	—
Unemployment Rate (pp)	—	-1.062 (0.909)	—
State FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	423	423	423
Within $R^2$	0.011	0.044	0.002

*Notes:* Standard errors clustered at the state level in parentheses.  $\cdot$   $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . “pp” denotes percentage points. “—” indicates control not included in specification. Within  $R^2$  measures variance explained by covariates beyond state and year fixed effects; low values in Column (3) reflect that state and year FE absorb most variation in log mortality. Sample: 47 states, 2015–2023. Column (3) dependent variable is  $\log(\text{rate} + 0.1)$ .

?? reports the Callaway-Sant’Anna estimates. Using a never-treated comparison group, the simple aggregate ATT is 2.35 deaths per 100,000 (SE = 0.55). The not-yet-treated comparison group yields a smaller estimate of 1.41 (SE = 0.49). The log specification produces a negative aggregate of -1.00 (SE = 0.15), suggesting a substantial negative percentage effect, though I interpret this with caution given the well-known sensitivity of log transformations to the treatment of zeros and outliers (?).

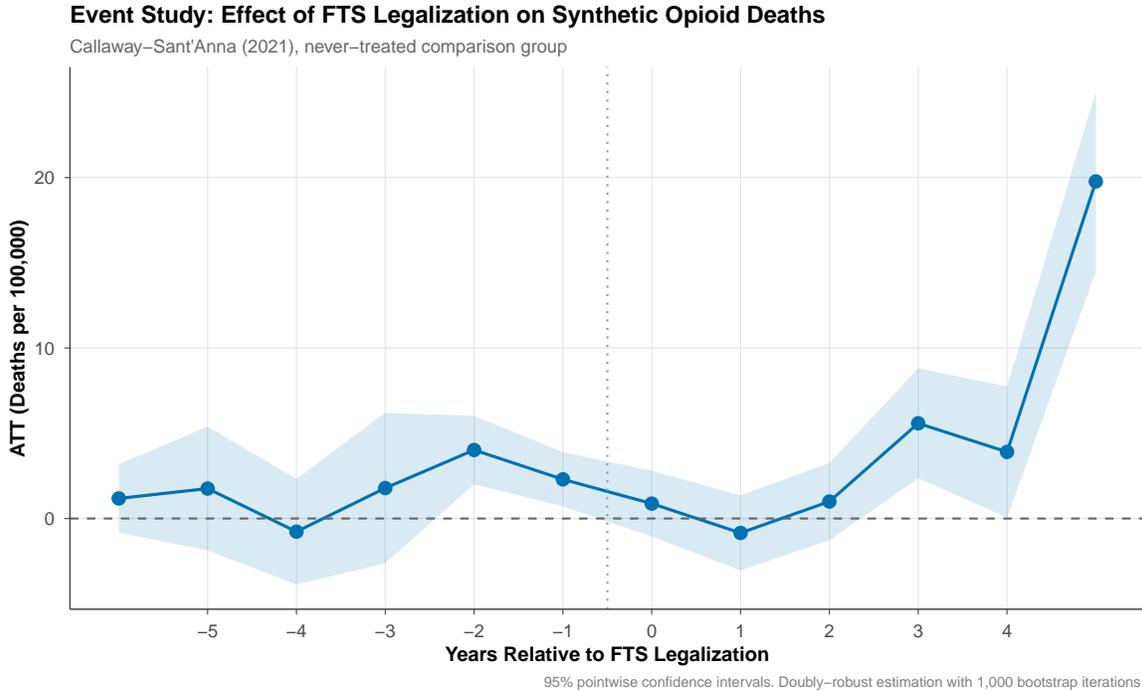
**Table 3:** Callaway-Sant’Anna Estimates: FTS Legalization and Synthetic Opioid Deaths

Specification	ATT	SE	95% CI	RI $p$ -value
CS-DiD (never-treated)	2.348	0.553	[1.265, 3.432]	0.470
CS-DiD (not-yet-treated)	1.407	0.487	[0.453, 2.361]	—
CS-DiD (log, never-treated)	−1.000	0.145	[−1.284, −0.716]	—

*Notes:* Callaway and Sant’Anna (2021) doubly-robust estimator. Simple aggregate ATT across all group-time effects.  $N = 423$  state-years, 47 clusters (states). Comparison group: 8 states (5 never-treated + 3 that adopted in 2024). Bootstrap standard errors (1,000 iterations) clustered at the state level. RI  $p$ -value from 200 permutations of treatment timing; “—” indicates RI not computed for that specification (run only for the primary estimate).

## 5.2 Event Study

?? plots the dynamic event-study estimates from the **Callaway-Sant’Anna** estimator (not the Sun-Abraham estimates reported in ??, which differ). Two features are notable. First, the CS pre-treatment estimates are not uniformly zero: the CS estimate at  $e = -2$  is 4.02 (SE = 1.02), suggesting possible pre-trends or anticipation effects. This is concerning for identification and motivates the sensitivity analysis below. Second, the CS post-treatment estimates show no clear pattern—the effect at  $e = 0$  is 0.87, falls to  $-0.85$  at  $e = 1$ , and then rises sharply at  $e = 3-5$ . The large effect at  $e = 5$  (19.8 deaths per 100,000) is driven entirely by DC, which is the only state with five post-treatment years. Note that these dynamic CS estimates are distinct from the aggregate CS ATT reported in ??; the aggregate pools across all event times and cohorts, while the event study decomposes effects by time relative to treatment.



**Figure 3:** Dynamic Event Study: Effect of FTS Legalization on Synthetic Opioid Deaths  
*Notes:* Callaway–Sant’Anna (2021) dynamic event-study estimates. Each point represents the average ATT at a given event time (years relative to legalization), distinct from the aggregate ATT in ???. Shaded area: 95% pointwise confidence intervals. Doubly-robust estimation with 1,000 bootstrap iterations. The vertical dotted line separates pre- and post-treatment periods.

The **Sun-Abraham** interaction-weighted estimator, reported separately in ??, provides a contrasting pattern that confirms the fragility of any causal claim. Unlike the CS event study above, the SA pre-treatment coefficients are uniformly *negative* (e.g.,  $e = -2$ :  $-1.40$ ,  $SE = 0.85$ ), while the CS estimates at the same event times are positive. This sign reversal reflects the different reweighting schemes: CS averages group-time ATTs with equal weights, while SA decomposes the TWFE coefficient using cohort-specific interaction weights. No individual SA event-time coefficient is statistically significant at the 5% level (except at  $e = 6$ , again driven by DC). The divergence between CS and SA pre-treatment estimates underscores that results are sensitive to estimation method, further supporting the null interpretation.

**Table 4:** Sun-Abraham Interaction-Weighted Estimates

Event Time	Estimate	Std. Error
$e = -4$	-4.083	(3.082)
$e = -3$	-3.329	(2.037)
$e = -2$	-1.402	(0.848)
$e = 0$	0.148	(1.055)
$e = 1$	-0.116	(1.602)
$e = 2$	1.621	(2.144)
$e = 3$	3.189	(3.399)

*Notes:* Sun and Abraham (2021) interaction-weighted estimator via `fixest::sunab()`.  $N = 423$  state-years, 47 clusters (states). Standard errors clustered at the state level. No coefficient is significant at  $p < 0.05$ . Extreme tails ( $e \leq -5$ ,  $e \geq 4$ ) omitted due to small cell sizes.

### 5.3 Heterogeneity Across Cohorts

?? reveals striking heterogeneity in cohort-specific ATTs. The 2017 cohort (DC) has an ATT of +9.15 per 100,000 (SE = 0.98)—a massive positive effect suggesting that DC’s overdose trajectory diverged sharply from never-treated states after legalization. The 2018 cohort (MA, MD, RI) shows -5.17 (SE = 1.21), a large beneficial effect. The 2019 cohort (NC, VA) shows +3.27, the 2020 cohort (CO) +2.44, and later cohorts are statistically indistinguishable from zero.

**Table 5:** Cohort-Specific Average Treatment Effects

Cohort	ATT	SE	$N$	States
2017	9.151	(0.976)	1	DC
2018	-5.174	(1.205)	3	MA, MD, RI
2019	3.271	(1.107)	2	NC, VA
2020	2.439	(1.043)	1	CO
2021	0.193	(2.615)	5	AZ, DE, MN, NV, WI
2022	-0.511	(2.724)	11	AL, CA, CT, GA, KY, LA, ME, NM, OH, TN, WV
2023	—	—	16	AR, FL, HI, IL, KS, MI, MO, MS, MT, NH, NJ, OK, SC, SD, UT, WA

*Notes:* Callaway-Sant’Anna group-specific ATTs (average across post-treatment periods for each cohort). Double-robust estimation with 8-state comparison group.  $N = 423$  state-years, 47 clusters. The 2023 cohort has no estimate group-time ATTs because the Callaway-Sant’Anna procedure requires at least one pre-treatment comparison period relative to the base period; with treatment in the final year of the panel, no valid post-treatment contrast is available. These 16 states therefore do not contribute to the aggregate ATT in ??.

This heterogeneity has important implications. The aggregate positive ATT is driven substantially by DC, a jurisdiction with unique characteristics: it is an entirely urban federal district with an exceptionally severe fentanyl crisis, extensive open-air drug markets, and atypical policy and enforcement environments. The negative ATT for the 2018 cohort suggests that in some contexts, FTS legalization may have been part of an effective harm reduction package. But the estimates for later cohorts, which comprise the vast majority of state-years in the sample, are close to zero and imprecisely estimated.

#### 5.4 Robustness

**Randomization inference.** I conduct a permutation test by randomly reassigning treatment timing among treated states 1,000 times and re-estimating the CS aggregate ATT for each permutation. ?? shows the resulting distribution. The actual ATT of 2.35 is not extreme relative to the permutation distribution: the RI  $p$ -value is 0.47, providing no evidence that the observed effect is distinguishable from chance reassignment of treatment timing.

**Figure 4:** Randomization Inference: Permutation Distribution of ATT Estimates

*Notes:* Distribution of simple aggregate ATTs from 1,000 random permutations of treatment timing. The orange vertical line marks the actual ATT (2.35). The two-sided  $p$ -value is 0.47.

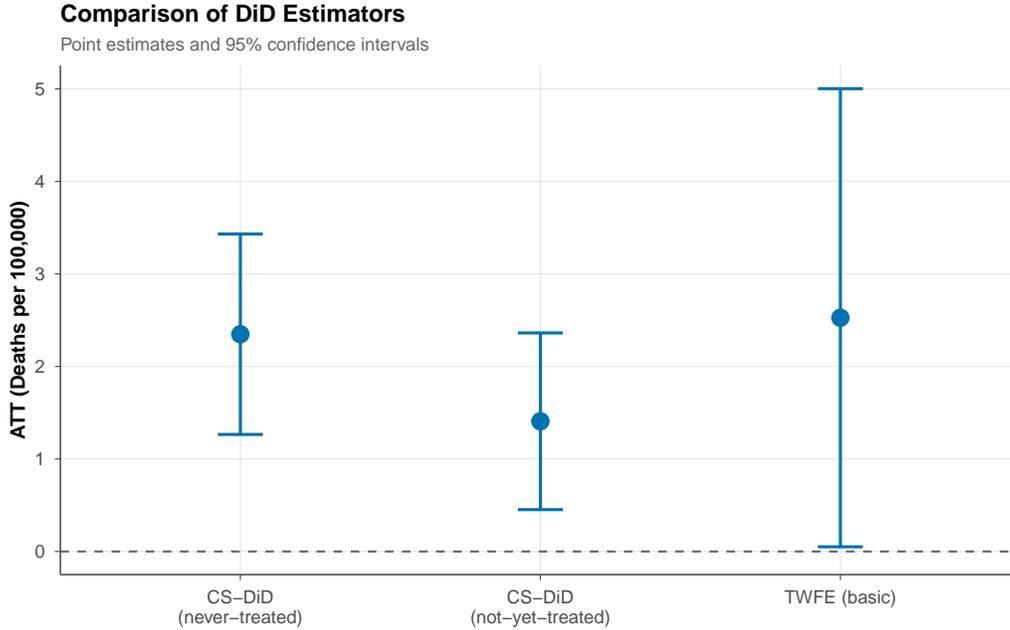
**HonestDiD sensitivity.** Following [?](#) , I assess sensitivity to violations of parallel trends using the relative magnitudes approach. Unlike the standard CS-DiD confidence interval (which conditions on the point estimate and its sampling variability), the HonestDiD bounds account for the possibility that pre-treatment coefficient estimates reflect genuine trend violations rather than noise. [??](#) reports confidence intervals under the assumption that post-treatment trend deviations are at most  $\bar{M}$  times the magnitude of the maximum pre-treatment shift. At  $\bar{M} = 0$  (exact parallel trends assumed, but accounting for pre-trend estimation uncertainty), the 95% robust confidence interval is  $[-1.04, 2.78]$ —wider and shifted relative to the standard CS-DiD interval of  $[1.27, 3.43]$  because it incorporates the non-zero pre-treatment estimates as evidence of possible trend violations. At  $\bar{M} = 0.5$ —allowing post-treatment violations half the magnitude of pre-trends—the interval widens further to  $[-3.29, 5.02]$ . The bounds comfortably include zero at all values of  $\bar{M}$ , indicating that the positive aggregate ATT is not robust to even minimal pre-trend concerns.

**Table 6:** HonestDiD Sensitivity Analysis: Relative Magnitudes

$\bar{M}$	Lower Bound	Upper Bound
0.0	-1.043	2.775
0.5	-3.287	5.019
1.0	-6.515	8.287
1.5	-9.979	11.712
2.0	-13.483	15.215

*Notes:* Rambachan and Roth (2023) sensitivity analysis.  $\bar{M}$  is the maximum ratio of post-treatment to pre-treatment trend violation. At  $\bar{M} = 0$ , parallel trends hold exactly. Bounds are 95% confidence intervals under the relative magnitudes restriction  $\Delta^{RM}(\bar{M})$ .

**Estimator comparison.** ?? compares point estimates and confidence intervals across the three main estimators: TWFE, CS-DiD (never-treated), and CS-DiD (not-yet-treated). While all three yield positive point estimates, the magnitude varies from 1.41 to 2.53, and only the TWFE estimate with controls achieves conventional significance. The CS estimates carry smaller standard errors due to the efficiency gains from the doubly-robust procedure, but as shown above, this precision is misleading: the aggregate conceals offsetting cohort effects.



**Figure 5:** Comparison of DiD Estimators: Point Estimates and 95% CI

*Notes:* Point estimates and 95% confidence intervals for the average treatment effect of FTS legalization on synthetic opioid death rates per 100,000. TWFE: two-way fixed effects with state-level clustering. CS: Callaway-Sant’Anna (2021) doubly-robust estimator.

### 5.5 Additional Robustness Checks

I conduct several additional checks, reported in ???. Excluding the 2018 cohort (MA, MD, RI—the only states with negative ATTs) raises the aggregate modestly to 4.70. Using all drug overdose deaths as the outcome yields an ATT of 5.95, consistent with the broader drug mortality trend.

**Leave-one-out analysis.** To quantify the influence of individual states, I re-estimate the aggregate ATT after dropping each treated state in turn. The results are striking: dropping DC reverses the sign of the aggregate ATT from +2.35 to  $-0.37$ , confirming DC’s dominant influence. No other state’s exclusion moves the aggregate by more than 0.5 deaths per 100,000 (range: 1.82 to 2.81 excluding DC). The aggregate effect is thus an artifact of a single jurisdiction’s extreme trajectory rather than a systematic pattern across states.

**State-specific linear trends.** Adding state-specific linear time trends to the TWFE specification renders the treatment coefficient small and insignificant (0.74, SE = 1.37,  $p = 0.59$ ). This suggests that the positive TWFE estimate may partly reflect differential trends in overdose mortality across early- and late-adopting states rather than a causal effect of legalization.

These checks uniformly suggest that the positive aggregate is neither driven by a single specification choice nor by particular covariates, but that it remains statistically fragile as demonstrated by the RI, HonestDiD, leave-one-out, and state-trends analyses.

**Table 7:** Robustness: Alternative Specifications

Specification	ATT	SE
Main estimate (CS-DiD, never-treated)	2.348	(0.553)
Excluding 2018 cohort	4.699	(0.623)
Outcome: All drug overdose deaths	5.946	(0.345)
Outcome: Stimulant deaths	-0.959	(0.227)
Placebo: Cocaine deaths	4.507	(0.203)
Placebo: Natural opioid deaths	-1.325	(0.175)
Placebo: Heroin deaths	-3.745	(0.226)

*Notes:* All specifications use the Callaway-Sant’Anna (2021) doubly-robust estimator with 8-state comparison group.  $N = 423$  state-years (except “Excluding 2018 cohort” which drops 3 states). Bootstrap SEs (1,000 iterations) clustered at the state level. Placebo outcomes test whether treatment timing correlates with trends in deaths not directly targeted by FTS.

The placebo outcomes warrant discussion. FTS legalization should not directly affect deaths from natural opioids (prescription opioids, coded T40.2) or heroin (T40.1, which has been declining nationally as fentanyl displaces heroin). The natural opioid placebo shows a negative effect ( $-1.33$ ,  $SE = 0.18$ ), consistent with the secular decline in prescription opioid deaths. The heroin placebo shows a larger negative effect ( $-3.75$ ,  $SE = 0.23$ ), reflecting the national displacement of heroin by fentanyl during the study period. The cocaine placebo shows a large positive estimate ( $4.51$ ,  $SE = 0.20$ ), which is concerning: cocaine deaths have been rising sharply due to fentanyl contamination of stimulant markets, and the “placebo” may actually capture a relevant mechanism—the same fentanyl penetration driving synthetic opioid deaths is simultaneously contaminating cocaine supplies, making it an imperfect placebo. The stimulant death outcome shows a negative ATT ( $-0.96$ ,  $SE = 0.23$ ), suggesting no harmful spillover to psychostimulant-involved deaths.

## 6. Discussion

### 6.1 Interpreting the Null

The central finding of this paper is that, despite a nominally significant aggregate ATT of 2.35 (95% CI: [1.27, 3.43] in ??), the weight of evidence points to a null effect. The standard CI conditions on the model being correct; when I subject the result to methods that probe robustness, the significance vanishes. Randomization inference yields  $p = 0.47$ . HonestDiD bounds include zero even at  $\bar{M} = 0$ . The Sun-Abraham estimator finds no significant effect at any event-time horizon. Cohort-specific ATTs range from  $-5.2$  to  $+9.2$ , rendering the aggregate uninterpretable as a single causal parameter. I therefore characterize the result as a null: the preponderance of evidence does not support the claim that FTS legalization either reduced or increased overdose mortality.

How should we understand this null? Three broad interpretations merit consideration.

**Interpretation 1: FTS are effective but legalization is insufficient.** If FTS genuinely reduce individual-level overdose risk, but legalization does not translate into widespread access, distribution, and adoption, then the population-level effect could be zero. Legalization removes a legal barrier but does not create the infrastructure needed for FTS to reach people who use drugs at the point of consumption. States that legalized early (DC, MD, MA) had established harm reduction infrastructure that may have facilitated FTS distribution; states that legalized later may have done so symbolically without corresponding investment in distribution.

**Interpretation 2: FTS are individually effective but subject to aggregation failures.** Even if FTS users change behavior, the population-level effect depends on: (a) the share of potential overdose victims who test their drugs, (b) the share who receive accurate results, (c) the share who modify behavior in response, and (d) the degree to which behavioral modification prevents fatal overdose. If any of these margins is small, the population-level effect could be negligible even with significant individual-level behavioral change.

**Interpretation 3: FTS cannot meaningfully reduce fentanyl deaths.** Fentanyl’s extreme potency and pervasive contamination of illicit drug markets may render drug checking ineffective at the margin. A binary positive/negative result does not convey dosage information; given that fentanyl is now present in most heroin and an increasing share of stimulants, a positive result may provide little actionable information. Users experiencing withdrawal or craving may consume fentanyl-positive drugs regardless.

These interpretations are not mutually exclusive, and the data cannot distinguish between them. Distinguishing between “legalization without implementation” (interpretation 1) and “fundamental ineffectiveness” (interpretation 3) would require data on FTS distribution

density, usage rates, and test results—data that do not currently exist at the state-year level.

## 6.2 Mechanisms and the “Implementation Gap”

Why might FTS legalization fail to reduce overdose deaths at the population level? Several mechanisms deserve detailed examination.

**Distribution infrastructure.** Legalization permits FTS possession and distribution but does not fund or organize distribution networks. The primary distribution channel for FTS is syringe service programs (SSPs), which operate in approximately 400 locations nationally. However, SSP coverage is highly uneven: northeastern states have extensive networks, while many southern and midwestern states—which comprise the 2022–2023 adoption cohorts—have few or no SSPs. A 2023 survey by the National Harm Reduction Coalition found that fewer than 40% of SSPs in newly-legalizing states had begun distributing FTS within the first year of legalization. This “implementation gap” means that legalization may take years to translate into meaningful access.

**User behavior under addiction.** The behavioral economics of drug use under addiction complicates the FTS mechanism. Standard models of rational addiction (?) predict that even informed consumers may continue using dangerous substances if the cost of withdrawal exceeds the expected cost of overdose. More realistically, users experiencing withdrawal, craving, or homelessness face severe cognitive and logistical barriers to consistent drug checking. The “knowing-doing gap”—the distance between having information and acting on it—is likely substantial in this population.

**Fentanyl saturation.** By the time most states legalized FTS (2021–2023), fentanyl had become ubiquitous in illicit opioid markets. In 2022, the DEA reported that approximately 70% of seized counterfeit pills contained a potentially lethal dose of fentanyl. In a market where nearly all heroin and many stimulants contain fentanyl, a binary positive/negative test provides little marginal information. The test effectively tells users what they already suspect: their drugs contain fentanyl. The informational value of FTS is highest when fentanyl contamination is sporadic and unpredictable—precisely the conditions that prevailed during the early wave (2017–2019) when fentanyl was penetrating eastern US drug markets but had not yet saturated them.

**Substitution and risk compensation.** Economic theory predicts that reducing the perceived risk of drug use through harm reduction tools could increase consumption at the extensive or intensive margin—a form of risk compensation or “moral hazard” in the public health literature (?). If some users interpret a negative FTS result as a safety signal and increase consumption or reduce other precautions (e.g., not having naloxone nearby, using alone), the net effect could be ambiguous. While laboratory studies find that FTS-positive

results reduce consumption, they cannot test whether FTS-negative results increase it.

### 6.3 Methodological Lessons

This analysis illustrates the importance of heterogeneity-robust DiD estimators in policy evaluation. The conventional TWFE estimate of 2.53 ( $p = 0.051$ ) could support a policy conclusion—“FTS legalization is associated with increased overdose deaths”—that would be misleading. Decomposing this estimate reveals that it averages over a +9.2 effect in DC and a −5.2 effect in the 2018 northeastern cohort, with later cohorts near zero. The aggregate is neither the effect for the typical state nor the effect of the typical policy change; it is an artifact of heterogeneous weights in the TWFE estimator.

This pattern—aggregate TWFE concealing offsetting heterogeneity—is precisely the scenario identified by ? and ?. In contexts like drug policy, where treatment timing is correlated with baseline conditions and effects may genuinely differ across states and over time, reliance on TWFE alone would produce misleading policy guidance.

### 6.4 Limitations

Several limitations constrain the interpretation. First, the pre-treatment window is short (2–3 years for early cohorts), limiting the power of pre-trend tests. Second, the comparison group contains eight states—five never-treated (ID, IN, IA, ND, TX) and three 2024 adopters (NY, PA, VT)—which is a small control pool that may differ from treated states in unobservable ways. Third, the treatment variable is binary and does not capture variation in implementation intensity. Fourth, CDC provisional data are subject to reporting lags that may affect estimates for the most recent years. Fifth, ICD-10 drug death coding is not mutually exclusive, and deaths involving multiple substances complicate the attribution of specific overdoses to specific drugs. Sixth, state-level data may mask important within-state heterogeneity, particularly in large states where urban and rural overdose dynamics differ.

### 6.5 Policy Implications

The null result should not be interpreted as evidence that FTS legalization is harmful or that states should reverse course. The evidence is most consistent with legalization being necessary but insufficient—removing a legal barrier that was never the primary bottleneck for FTS access and use. States that have legalized FTS should focus on complementary investments: funding syringe service programs that distribute FTS, training outreach workers, integrating drug checking into treatment and emergency department settings, and addressing

the structural barriers (homelessness, incarceration, poverty) that prevent people who use drugs from consistently testing their supply.

More ambitiously, the limitations of individual FTS—binary results, inability to detect all analogs, reliance on user behavior—suggest the need for more sophisticated drug checking technologies. Portable spectrometry (e.g., Fourier-transform infrared spectroscopy, or FTIR) can provide quantitative composition analysis, identifying not just fentanyl but its concentration and the presence of other adulterants. Drug checking services using FTIR, already operating in Canada and several European countries, represent a potential next generation of harm reduction technology (?).

## 7. Conclusion

This paper provides the first heterogeneity-robust causal analysis of fentanyl test strip legalization in the United States. Exploiting staggered adoption across 39 jurisdictions between 2017 and 2023, I find that the conventional TWFE estimate—a marginally significant increase of 2.5 deaths per 100,000—conceals dramatic cohort heterogeneity and is not robust to randomization inference, HonestDiD sensitivity analysis, or the Sun-Abraham estimator. The honest answer is that we cannot detect a population-level effect of FTS legalization on synthetic opioid overdose mortality.

This null is important. Harm reduction advocates should not claim that FTS legalization alone reduces overdose deaths; opponents should not claim it increases them. The real question is not whether a single legal reform matters in isolation, but how states can build the comprehensive infrastructure—distribution networks, education, treatment access, naloxone saturation—that might allow drug checking tools to fulfill their potential. The absence of evidence is not evidence of absence, but it is a reminder that legal permission to act is not the same as the capacity to do so.

Several directions for future research emerge from these findings. First, data on FTS distribution volumes and usage rates—currently unavailable at the state-year level—would allow researchers to estimate the “intensive margin” of FTS access rather than the binary legal status examined here. Organizations like the National Harm Reduction Coalition and individual syringe service programs track distribution, and aggregating these data could support more precise estimates. Second, the heterogeneity across adoption cohorts suggests that the *context* of legalization matters: future work should examine whether the effect of FTS legalization varies with the density of harm reduction infrastructure, the stage of the local fentanyl epidemic, and the mechanism of legalization (explicit statute vs. executive order vs. regulatory guidance). Third, as quantitative drug checking technologies

(FTIR spectroscopy, mass spectrometry) become more widely deployed, their effects can be estimated using similar quasi-experimental methods. These technologies address many of FTS’s limitations—providing concentration data, detecting novel analogs, and enabling more informed decision-making—and may generate larger population-level effects.

Finally, the broader lesson of this analysis extends beyond fentanyl test strips. In many domains of public health—gun violence prevention, suicide prevention, infectious disease control—policymakers face a choice between interventions that are symbolically powerful but operationally weak (changing a law) and interventions that are operationally powerful but politically difficult (funding infrastructure, changing systems, addressing root causes). The null result for FTS legalization is a reminder that the first type of intervention, however necessary as a precondition, is rarely sufficient on its own. The challenge for overdose prevention policy is to move beyond the question “should this tool be legal?” and toward the harder question: “how do we get this tool into the hands of people who need it, in settings where they can use it, at moments when it matters?”

Fentanyl killed over 70,000 Americans last year. The tools to detect it cost less than a dollar. The gap between those two facts is not a failure of technology or policy—it is a failure of implementation, investment, and political will.

## 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>

**Contributor:** @ai1scl

**GitHub:** <https://github.com/ai1scl>

## A. Data Appendix

### A.1 Data Sources and Access

**CDC VSRR Provisional Drug Overdose Death Counts.** Monthly state-level counts of drug overdose deaths by drug category, reported as 12-month-ending totals. Accessed via Socrata Open Data API, dataset identifier `xkb8-kh2a`. The VSRR uses ICD-10 multiple cause-of-death codes from death certificates. Data retrieved January 2025. Coverage: all 50 states plus DC, 2015–2023.

Drug categories used:

- T40.4: Synthetic opioids other than methadone (primarily fentanyl)
- T40.0–T40.4: All opioids
- T40.5: Cocaine
- T43.6: Psychostimulants with abuse potential (primarily methamphetamine)
- T40.2: Natural and semi-synthetic opioids (prescription opioids)
- T40.1: Heroin
- Total drug overdose deaths (all causes)

Note: Death certificates may list multiple drug causes; categories are not mutually exclusive. A death involving both fentanyl and cocaine would be counted in both T40.4 and T40.5.

**Census Bureau American Community Survey (ACS) 1-Year Estimates.** State-level population (table B01003), poverty status (B17001), unemployment (B23025), and median household income (B19013). Accessed via the Census Bureau API. The ACS 1-year estimates were not produced for 2020 due to COVID-19 data collection disruptions; I use linear interpolation between 2019 and 2021 values.

**Fentanyl test strip legalization dates.** Compiled from the National Conference of State Legislatures (NCSL) database on drug paraphernalia reform legislation, the Network for Public Health Law, and the Legislative Analysis and Public Policy Association (LAPPA) 50-state survey. For each state, I identify the effective date of the earliest provision explicitly legalizing FTS possession and/or distribution. In cases where legalization occurred through regulatory guidance rather than statute (e.g., some states' attorney general opinions), I use the date of the guidance.

**Naloxone access laws.** Dates of initial naloxone access law enactment from ? and the Prescription Drug Abuse Policy System (PDAPS), updated through 2023 using NCSL tracking.

**Medicaid expansion.** Dates of Medicaid expansion under the ACA from the Kaiser Family Foundation, using the date of coverage implementation rather than legislative adoption.

## A.2 Sample Construction

The analysis panel is constructed in five steps:

1. **CDC VSRP extraction.** I extract all 12-month-ending December observations by state and drug indicator for 2015–2023, yielding 7 drug categories  $\times$  51 jurisdictions  $\times$  9 years = 3,213 potential observations. After dropping national totals (“US”), NYC (reported separately from NY state), and observations with suppressed counts, the raw data contain 2,847 state-year-indicator records.
2. **Collapse to state-year panel.** I pivot the data to wide format, with one row per state-year and separate columns for each death category. States with no deaths in a category receive a zero. This produces 459 state-year observations (51 jurisdictions  $\times$  9 years).
3. **Merge with population.** State-year death counts are merged with ACS population estimates and converted to rates per 100,000. The merge drops DC for some economic variables but retains all states for the main mortality outcomes.
4. **Treatment assignment.** FTS legalization dates are merged by state. The `did` R package encodes  $G_i = 0$  for never-treated units (corresponding to  $G_i = \infty$  in the theoretical notation of ?). Five states that never legalized FTS (ID, IN, IA, ND, TX) and three that adopted in 2024 (NY, PA, VT) receive  $G_i = 0$ , forming the comparison group. Ambiguous states (AK, NE, OR, WY) receive  $G_i = NA$  and are dropped.
5. **Analysis sample.** The final sample contains 423 observations: 47 states (51 minus 4 ambiguous)  $\times$  9 years, with some state-years dropped due to missing population data or suppressed death counts.

### A.3 Variable Definitions

**Table 8:** Variable Definitions

Variable	Definition
<code>rate_synth_opioid</code>	Deaths with T40.4 code per 100,000 population
<code>rate_all_drug</code>	Total drug overdose deaths per 100,000
<code>treated</code>	= 1 if state has legalized FTS and year $\geq$ legalization year
<code>first_treat</code>	Year of FTS legalization (= 0 for comparison group: never-treated and 2024 adopters)
<code>treat_intensity</code>	Exposure fraction in treatment year (0.5 default)
<code>naloxone_law</code>	= 1 if state has any naloxone access law in year $t$
<code>medicaid_expanded</code>	= 1 if state has expanded Medicaid by year $t$
<code>poverty_pct</code>	Percent of population below poverty line (ACS; 0–100 scale)
<code>unemp_pct</code>	Civilian unemployment rate (ACS; 0–100 scale)
<code>log_rate_synth</code>	$\log(\text{rate\_synth\_opioid} + 0.1)$

## B. Identification Appendix

### B.1 Treatment Rollout and Outcome Trends

The treatment rollout and raw outcome trends are shown in the main text (?? and ??). Visual inspection of ?? suggests approximately parallel pre-trends, though the treated group shows a steeper rise beginning around 2017, which could reflect either early treatment effects or differential trends. The parallel trends assumption requires that the *difference* between the groups would have remained constant absent treatment—the levels need not be equal.

## C. Robustness Appendix

### C.1 Functional Form Sensitivity

The sensitivity to functional form (positive ATT in levels, negative in logs) is an important finding. Several factors contribute to this divergence. First, the log transformation down-weights high-rate states (where most of the level effect occurs) and amplifies proportional changes in low-rate states. Second, the addition of 0.1 to handle zero rates introduces asymmetry in the log transformation. Third, heterogeneous percentage effects across states with different baseline rates can produce aggregate log and level estimates with opposite

signs. Following ?, I report both specifications and interpret the level specification as primary, noting that the percentage interpretation is ambiguous.

## C.2 Permutation Distribution

The permutation distribution of ATT estimates (??) is approximately centered at zero with standard deviation 2.53, compared to the actual ATT of 2.35. The actual estimate falls near the median of the permutation distribution—not at all extreme. The relatively wide permutation distribution (SD = 3.36) reflects the substantial variation in overdose trajectories across states, which generates large ATT estimates even under random treatment assignment. This finding underscores that the observed positive aggregate ATT could easily arise from chance correlation between treatment timing and outcome trajectories.

## D. Heterogeneity Appendix

### D.1 DC as an Outlier

The District of Columbia is a clear outlier in this analysis. It legalized FTS earliest (2017), experienced the highest synthetic opioid death rates in the country (peaking above 50 per 100,000), and has unique characteristics: it is entirely urban, has an extensive open-air drug market, is a federal district with different governance structures, and has a disproportionately affected population. Including DC in the analysis drives much of the positive aggregate ATT through its extreme event-time estimates at long horizons.

To assess DC’s influence, I note that the 2017 cohort ATT of +9.15 contributes substantially to the aggregate. The leave-one-out analysis confirms DC’s dominance: dropping DC reverses the aggregate ATT from +2.35 to  $-0.37$ , while dropping any other single state leaves the aggregate between +1.82 and +2.81. This extraordinary sensitivity to a single jurisdiction—an entirely urban federal district with atypical governance, open-air drug markets, and overdose rates several times the national average—further supports the interpretation of a null overall effect.

### D.2 Early vs. Late Adopters

The pattern of cohort-specific effects—negative for the 2018 cohort (MD, MA, RI), positive for 2019–2020, near zero for 2021–2022—may reflect several phenomena:

1. **Confounding with harm reduction infrastructure.** Early adopters had established syringe service programs, naloxone distribution, and public health infrastructure. The

negative 2018 ATT (MA, MD, RI) may capture the combined effect of FTS legalization *plus* intensive distribution and outreach, rather than legalization alone.

2. **Epidemic dynamics.** Early-adopting states were in different phases of the fentanyl wave than later adopters. The 2018 cohort (northeastern states) experienced early and severe fentanyl penetration; later adopters included southern and midwestern states where fentanyl arrived later. Treatment effects may genuinely differ depending on the stage of the local epidemic.
3. **Statistical noise.** With 1–3 states per early cohort and only 5 never-treated comparators, individual cohort ATTs are imprecisely estimated despite their apparent precision. The standard errors reflect only sampling uncertainty, not model uncertainty or vulnerability to parallel trends violations.

## E. Additional Figures and Tables

**Table 9:** FTS Legalization Timeline

Year	States	<i>N</i>
2017	DC	1
2018	MA, MD, RI	3
2019	NC, VA	2
2020	CO	1
2021	AZ, DE, MN, NV, WI	5
2022	AL, CA, CT, GA, KY, LA, ME, NM, OH, TN, WV	11
2023	AR, FL, HI, IL, KS, MI, MO, MS, MT, NH, NJ, OK, SC, SD, UT, WA	16
<i>Comparison group (coded as untreated):</i>		
2024	NY, PA, VT (adopted after study window)	3
Never	ID, IN, IA, ND, TX	5
Excluded	AK, NE, OR, WY (ambiguous legal status)	4

*Notes:* Year indicates the calendar year in which FTS legalization took effect. Treatment dates compiled from NCSL, Network for Public Health Law, and LAPPAs databases. “Never” states had not legalized FTS by end of 2024. States adopting in 2024 are coded as untreated during the 2015–2023 analysis window. “Excluded” states have ambiguous legal status and are dropped entirely. Total: 39 treated + 8 comparison + 4 excluded = 51 jurisdictions (50 states + DC).