Does the Safety Net Bite Back?
Medicaid Postpartum Coverage Extensions
Through the Public Health Emergency and Beyond*

APEP Autonomous Research†   @ai1scl

@olafdrw

February 9, 2026

Abstract

Between 2021 and 2024, 47 U.S. jurisdictions adopted extensions of Medicaid postpartum coverage from 60 days to 12 months, the most rapid expansion of maternal health coverage in decades. Using individual-level data on 237,365 postpartum women from the American Community Survey (2017–2019, 2021–2024; 2020 excluded due to non-standard data collection) and a staggered difference-in-differences design with the Callaway and Sant'Anna (2021) estimator, this paper evaluates whether these extensions increased insurance coverage among women who recently gave birth. The full-sample CS-DiD ATT for Medicaid coverage is -0.50 percentage points (SE $= 0.63$ pp, $p > 0.10$). The post-PHE specification (2017–2019 + 2023–2024), which excludes the period when continuous enrollment rendered the coverage cliff non-binding, yields a *statistically significant negative* Medicaid ATT of -2.18 pp (SE $= 0.74$ pp, $p < 0.01$). This negative estimate does not reflect policy harm; rather, it captures the secular Medicaid unwinding that disproportionately reduced enrollment in treated states after the PHE ended. A triple-difference (DDD) design—comparing postpartum to non-postpartum low-income women within treated and control states—absorbs these common unwinding shocks. The DDD CS-DiD estimate is $+0.99$ pp (SE $= 1.55$ pp), statistically insignificant but signed in the direction predicted by the policy's institutional logic. The overall story is

---

that the standard DiD picks up the Medicaid unwinding confound; the DDD resolves it, yielding a small but imprecise postpartum-specific effect estimate. Event-study estimates through two post-treatment periods show flat dynamics, and permutation inference (200 randomizations of the full CS-DiD estimator), wild cluster bootstrap (including state-cluster bootstrap for CS-DiD), and Rambachan-Roth HonestDiD sensitivity analysis all confirm that the confidence interval includes zero under moderate violations of parallel trends. These findings constitute a well-identified result with substantially stronger methodology than the earlier analysis: even with post-PHE data, a DDD design, and multiple inference procedures, the postpartum extensions do not produce detectable coverage gains in survey data, suggesting that administrative enrollment mechanics, measurement limitations in the ACS, or the thin control group (4 states) may attenuate the policy's apparent impact.

1. Introduction

Maternal mortality in the United States has diverged sharply from other high-income countries, rising from 12.7 deaths per 100,000 live births in 2000 to 32.9 in 2021 (Hoyert, 2023). The U.S. maternal mortality rate is now more than double that of any other G7 nation. A substantial share of pregnancy-related deaths—more than one-third—occur between 7 days and one year postpartum (Petersen et al., 2019), a period during which many low-income women historically lost Medicaid coverage just 60 days after delivery. This coverage gap has been identified as a critical contributor to adverse maternal health outcomes, particularly among Black and Hispanic women who rely disproportionately on Medicaid for pregnancy-related care (Eliason, 2020; Gordon et al., 2022).

In response, the American Rescue Plan Act (ARPA) of March 2021 created a new option for states to extend Medicaid postpartum coverage from 60 days to a full 12 months. By early 2025, 47 jurisdictions had adopted the extension—one of the fastest-spreading health policy reforms in recent U.S. history. This paper evaluates whether these extensions increased insurance coverage among women who recently gave birth, using data through 2024 that incorporates two full post–Public Health Emergency (PHE) years when the 60-day coverage cliff became binding again.

This paper advances the analysis in several directions relative to the earlier working papers (APEP 0149/0153/0156). A central contribution is methodological transparency: the post-PHE Medicaid ATT is a statistically significant *negative* -2.18 pp, which earlier versions framed as a null result. We reinterpret this estimate as reflecting the Medicaid unwinding confound—the disproportionate decline in Medicaid enrollment in states that adopted the extension early and had accumulated large PHE-era rolls—rather than policy harm, and we motivate the triple-difference (DDD) design as the specification that isolates the postpartum-specific effect.

The DDD compares postpartum to non-postpartum low-income women within the same states, absorbing state-level unwinding shocks and resolving the employer insurance placebo failure that reviewers identified as a critical concern. To address inference challenges inherent in a design with few policy clusters, we implement permutation-based inference by re-running the full Callaway-Sant'Anna estimator under 200 random reassignments of treatment timing, providing exact p-values that do not rely on asymptotic approximations. We complement this with state-cluster wild bootstrap for the CS-DiD estimates (999 replications) and Rambachan-Roth HonestDiD sensitivity analysis across an $\bar{M}$-grid, with a plain-language summary of what each sensitivity threshold implies.

Several additional contributions strengthen the empirical foundation. We quantify attenu-

ation bias from the ACS's lack of birth-month information through a Monte Carlo simulation calibrating the ITT scaling factor across multiple adoption-timing scenarios. We present formal joint tests of DDD pre-treatment trends, balance tests comparing treated and control states on pre-treatment observables, cohort-specific ATTs that reveal which adoption cohorts drive the aggregate estimate, and a DDD power analysis that establishes the minimum detectable effect. A 2024-only post-period specification excludes the mixed 2023 survey year entirely, providing the cleanest post-PHE identification. All regression tables now report 95% confidence intervals alongside point estimates.

The central finding is that the standard DiD picks up a Medicaid unwinding confound—the statistically significant negative post-PHE ATT of -2.18 pp reflects the disproportionate decline in Medicaid enrollment in treated states (which had larger Medicaid rolls during the PHE) as the unwinding proceeded. The DDD design, which differences out this secular shock by comparing postpartum women to non-postpartum women of similar income within the same states, yields a point estimate of $+0.99$ pp—small, signed in the expected direction (or close to zero), and statistically insignificant. This pattern—a significant negative standard DiD resolved to a small insignificant DDD—is consistent with the policy having at most a modest effect on survey-measured coverage rates, obscured by the dominant unwinding dynamics of 2023–2024.

The remainder of the paper is organized as follows. Section 2 describes the institutional background, including the PHE and Medicaid unwinding dynamics. Section 3 outlines the conceptual framework. Section 4 describes the data and quantifies attenuation bias from the ACS measurement structure. Section 5 presents the empirical strategy, including the DDD design, post-PHE specifications, and inference procedures. Section 6 reports main results. Section 7 presents robustness checks and sensitivity analyses, including permutation inference, the 2024-only post-period, DDD pre-trend event studies, and HonestDiD sensitivity figures. Section 8 discusses findings and limitations. Section 9 concludes.

2. Institutional Background and Policy Setting

2.1 Medicaid Postpartum Coverage: The 60-Day Cliff

Medicaid is the single largest payer for maternity care in the United States, financing approximately 42% of all births (Medicaid.gov, 2023). Under traditional Medicaid eligibility rules, pregnant women qualify for coverage at higher income thresholds than the general adult population—typically up to 185% or even 200% of the federal poverty level (FPL) in many states, compared to 138% FPL for the general adult population under the Affordable Care Act (ACA) Medicaid expansion. However, this enhanced pregnancy eligibility has historically

been limited to the pregnancy period plus 60 days postpartum, after which women's coverage reverted to the lower general-population income thresholds.

This 60-day postpartum coverage cliff had long been identified as a critical gap in the maternal health safety net. The American College of Obstetricians and Gynecologists (ACOG) recommends comprehensive postpartum care extending through the first year after birth, including screening for postpartum depression, management of pregnancy-related complications such as preeclampsia and gestational diabetes, and family planning services (ACOG, 2018). The 60-day cutoff meant that many low-income women lost access to these services precisely when they were most vulnerable, as the postpartum period carries elevated risks for cardiovascular events, mental health crises, and infection.

The consequences of the 60-day cliff have been extensively documented. Daw et al. (2020) find that roughly one in four women who were insured at delivery experienced a coverage disruption within six months postpartum, with the highest rates of churn among Medicaid-covered women. These coverage gaps are concentrated precisely when clinical guidelines call for continued monitoring (ACOG, 2018). The clinical stakes are high: over one-third of pregnancy-related deaths occur between 7 days and one year postpartum (Petersen et al., 2019), and the U.S. maternal mortality rate rose to 32.9 per 100,000 live births in 2021 (Hoyert, 2023; Tikkanen et al., 2020). Racial disparities compound the problem: Black women face maternal mortality rates 2.6 times higher than White women and are disproportionately covered by Medicaid during pregnancy (Petersen et al., 2019).

2.2 The ARPA Reform and Staggered Adoption

The American Rescue Plan Act of March 2021 (P.L. 117-2, Section 9812) created a new option under Section 1902(e)(16) of the Social Security Act, allowing states to extend Medicaid and CHIP postpartum coverage from 60 days to 12 months through a State Plan Amendment (SPA). This option became effective on April 1, 2022, and was made permanent by the Consolidated Appropriations Act of 2023.

Several states moved to extend postpartum coverage before the SPA option became available, using Section 1115 demonstration waivers. Illinois was the first (April 2021), followed by Georgia, Missouri, New Jersey, and Virginia (Sonfield, 2022). Once the SPA option became available, adoption was rapid: approximately 25 states adopted in 2022, 13 in 2023, and 5 in 2024. Table 9 provides the complete adoption timeline.

This staggered adoption creates the policy variation exploited in the empirical analysis. With data through 2024, I code 47 jurisdictions as treated (adopting by 2024) and 4 as the control group: Arkansas and Wisconsin (never-adopted) plus Idaho and Iowa (adopting in 2025, not yet treated in the sample). The near-universal adoption limits power but the

extended panel compensates by providing more post-treatment observations.

2.3 The COVID-19 Public Health Emergency, Continuous Enrollment, and the Unwinding

The Families First Coronavirus Response Act (FFCRA) of March 2020 established a continuous enrollment condition for state Medicaid programs. Under this provision, states receiving the enhanced Federal Medical Assistance Percentage (FMAP) were prohibited from terminating Medicaid coverage for any beneficiary. This continuous enrollment provision remained in force until May 11, 2023, after which states began the "unwinding" process of conducting eligibility redeterminations.

The PHE continuous enrollment provision had profound implications for evaluating the postpartum extensions. During continuous enrollment, a woman who qualified for Medicaid during pregnancy could not be disenrolled after the 60-day postpartum period, even though she no longer met the traditional eligibility criteria. This meant that the 60-day cliff was effectively non-binding during the PHE, and the postpartum extension's marginal contribution to coverage was approximately zero for states adopting during this period.

The Medicaid unwinding began in earnest in April 2023. By March 2024, an estimated 19.6 million people had been disenrolled from Medicaid (KFF, 2024), with substantial heterogeneity across states in the pace and magnitude of coverage losses (Sommers et al., 2024; Sugar et al., 2024). Crucially, this unwinding was not uniform across states: states that had experienced the largest enrollment growth during the PHE—which substantially overlap with the states that adopted postpartum extensions—experienced the most severe unwinding-driven enrollment declines. This creates a mechanical confound for standard difference-in-differences estimation. If treated states lost more Medicaid enrollees during the unwinding than control states, then the post-PHE DiD will attribute this secular decline to the postpartum extension, producing a spurious negative treatment effect estimate. This unwinding confound is central to interpreting the results of this paper: the statistically significant negative post-PHE Medicaid ATT (-2.18 pp) almost certainly reflects this differential unwinding rather than any harmful effect of the postpartum extension itself. The triple-difference design directly addresses this confound.

3. Conceptual Framework

3.1 Expected Effect Magnitude

The theoretical prediction for the effect of postpartum Medicaid extensions on insurance coverage is straightforward in the post-PHE environment. Consider a woman who qualifies for Medicaid during pregnancy. Under the traditional 60-day rule, her Medicaid eligibility ends approximately two months after delivery. If her income exceeds the general Medicaid threshold, she faces three options: (1) obtain employer-sponsored insurance, (2) purchase marketplace insurance (potentially with subsidies), or (3) become uninsured. The 12-month postpartum extension eliminates this coverage gap by extending Medicaid eligibility for a full year after delivery.

The expected magnitude depends on the fraction of postpartum women whose coverage is affected by the extension. Approximately 42% of births are Medicaid-financed nationally, but not all of these women would have lost coverage at 60 days. The "at-risk" population—women who would have become uninsured or experienced coverage disruption after the 60-day cutoff—likely constitutes 15–25% of all postpartum women, implying an expected coverage effect of roughly 5–15 percentage points among the full postpartum population, and larger effects among the low-income subgroup.

3.2 PHE Interaction and Temporal Heterogeneity

The PHE creates sharp temporal heterogeneity in the policy's bite. During the PHE (2020–2022), the extension's effect on coverage should be approximately zero for states that adopted during this period. The extension's true effect should emerge in 2023–2024, after the PHE ends and the 60-day cliff becomes binding again. This generates a specific testable prediction: the event-study trajectory should show flat or near-zero effects at short horizons (corresponding to the PHE period) and growing positive effects at longer horizons (corresponding to the post-PHE period).

However, the Medicaid unwinding complicates this prediction. The post-PHE period combines two offsetting forces: (1) the return of the 60-day cliff, which makes the extension binding and should produce positive coverage effects, and (2) the unwinding-driven Medicaid enrollment decline, which reduces coverage among all beneficiaries in states with large PHE-era enrollment expansions. In a standard DiD, force (2) may dominate force (1), producing a negative estimate even if the extension has a positive causal effect on the postpartum-specific population. The DDD design isolates force (1) by differencing out force (2), which affects postpartum and non-postpartum women similarly.

3.3 The DDD Rationale

The triple-difference design is motivated by both the employer insurance placebo failure in the earlier analysis and the Medicaid unwinding confound. If secular forces—such as the Medicaid unwinding or pandemic-era labor market disruptions—differentially affected treated versus control states, then a simple DiD comparing postpartum women across states will conflate the policy effect with these secular trends. The DDD design addresses this by using non-postpartum low-income women as an additional comparison group. Any shock that affects all low-income women similarly in treated versus control states (e.g., changes in employer benefit offerings, differential Medicaid unwinding) will be differenced out, isolating the postpartum-specific component of any coverage change.

3.4 Testable Predictions

The framework generates several testable predictions. The standard DiD Medicaid effect in the post-PHE period is ambiguous in sign, as the positive effect of the extension competes with the negative unwinding confound. The DDD Medicaid effect should be positive, as the unwinding confound is differenced out, isolating the postpartum-specific channel. Similarly, the post-PHE uninsurance effect in the standard DiD is ambiguous, while the DDD should yield a negative effect (reduced uninsurance). The DDD employer insurance coefficient should be null, as the DDD differences out secular labor market forces. The event-study trajectory should be flat during the PHE period, with the post-PHE direction depending on the balance of extension and unwinding effects. Late-adopter effects should be cleaner and possibly larger, given the absence of PHE overlap and potentially less unwinding exposure. Finally, placebo populations—high-income postpartum women and non-postpartum women—should show null effects.

4. Data

4.1 American Community Survey PUMS

The primary data source is the American Community Survey (ACS) 1-year Public Use Microdata Samples (PUMS) for 2017–2024, with the exclusion of 2020 due to non-standard data collection during the pandemic.[1] The ACS is the largest household survey in the United States, with approximately 3.3 million person records per year. The 1-year PUMS provides

---

[1]The 2024 ACS 1-year PUMS was released by the Census Bureau in October 2025. See https://www.census.gov/programs-surveys/acs/news/data-releases.html for the release schedule.

individual-level data on demographics, employment, income, and health insurance coverage, with state identifiers enabling state-level policy evaluation.

Key variables include: FER (fertility: gave birth in past 12 months), HICOV (health insurance coverage), HINS4 (Medicaid), HINS1 (employer insurance), HINS2 (direct-purchase insurance), POVPIP (income-to-poverty ratio), and standard demographics (age, race, education, marital status). All regressions use ACS person weights (PWGTP).

4.2 Sample Construction

The analysis sample consists of women aged 18–44 who appear in the ACS PUMS across seven survey years: 2017, 2018, 2019, 2021, 2022, 2023, and 2024. The total sample contains 3,683,347 women aged 18–44. Of these, 237,365 reported giving birth in the past 12 months (FER = 1), forming the primary postpartum analysis sample—an increase of approximately 70,000 observations compared to the earlier analysis. Subsamples include: 86,991 low-income postpartum women (below 200% FPL), 82,325 high-income postpartum women (above 400% FPL, used as a placebo), and 1,181,552 non-postpartum low-income women (used as the DDD comparison group).

4.3 Treatment Assignment

Treatment assignment is based on state-level adoption dates compiled from CMS press releases, Kaiser Family Foundation tracking data, MACPAC reports, and state Medicaid agency announcements. Exact effective dates (month and day) for each state are documented in the replication code (`01_fetch_data.R`) with per-state source citations. With data through 2024, the analysis includes 47 treated jurisdictions (4 adopting in 2021, 25 in 2022, 13 in 2023, and 5 in 2024) and 4 control jurisdictions: Arkansas and Wisconsin (never adopted) plus Idaho and Iowa (adopting in 2025).

Intent-to-treat interpretation. The ACS fertility variable (FER = 1) identifies women who gave birth in the past 12 months but does not include birth month. A postpartum woman surveyed in year t may have given birth anywhere from 1 to 12 months prior, and her postpartum window may only partially overlap with the extension's effective period. All estimates should be interpreted as intent-to-treat (ITT) effects. This imperfect mapping introduces attenuation bias, which we quantify in Section 4.4.

Treatment timing and ACS alignment. I code a state as "treated" in survey year t if the extension's effective date falls on or before July 1 of year t (i.e., the extension was in place for at least half the reference year); if the effective date falls after July 1, treatment is coded beginning in year $t + 1$. The exact effective date for each state is recorded in the

replication code with source citations, and the resulting `adopt_year` already encodes this July 1 rule. For the CS-DiD estimator, the treatment cohort year G_s is the first calendar year in which the extension is coded as active. For the 2024 adoption cohort (5 states), this means that respondents interviewed early in 2024 may be reporting on a period before the policy was active; the July 1 rule ensures that *at least half* of the reference year falls under treatment, but partial-year treatment attenuates the $e = 0$ estimate for this cohort. This partial-year contamination is explicitly quantified in the Monte Carlo attenuation analysis (Section 4.4). Table 1 illustrates the mapping from adoption cohort to event time for the ACS survey years used in the analysis.

Table 1: Event-Time Mapping by Adoption Cohort and ACS Survey Year

Cohort (G_s)	ACS Survey Year						
	2017	2018	2019	2021	2022	2023	2024
2021 (4 states)	$e=-4$	$e=-3$	$e=-2$	$e=0$	$e=1$	$e=2$	$e=3$
2022 (25 states)	$e=-5$	$e=-4$	$e=-3$	$e=-1$	$e=0$	$e=1$	$e=2$
2023 (13 states)	$e=-6$	$e=-5$	$e=-4$	$e=-2$	$e=-1$	$e=0$	$e=1$
2024 (5 states)	$e=-7$	$e=-6$	$e=-5$	$e=-3$	$e=-2$	$e=-1$	$e=0$
Control (4 states)	Pre	Pre	Pre	Pre	Pre	Pre	Pre

Notes: Event time $e = t - G_s$. The 2023–2024 columns are post-PHE. The PHE continuous enrollment period (March 2020 – May 2023) rendered the 60-day coverage cliff non-binding. Control states: AR, WI (never adopted), ID, IA (adopt 2025). All tables report the number of clusters used in each specification.

The control group is thin—only 4 states—which is an inherent limitation of near-universal adoption. However, for the post-PHE specification and the DDD design, this limitation is partially mitigated by the temporal structure and the additional within-state variation from the postpartum/non-postpartum comparison.

4.4 Attenuation Bias Quantification

The ACS's lack of birth-month information introduces mechanical attenuation into all estimates. To quantify this, consider the following back-of-the-envelope calculation. The ACS PUMS identifies postpartum women as those who gave birth in the past 12 months (FER = 1). The treatment coding rule assigns a state as treated in survey year t if the extension's effective date falls on or before July 1 of year t.

Under the assumption that births are uniformly distributed across months and that ACS interviews are uniformly distributed across the calendar year, consider a state that adopted on

July 1 of year t (the marginal case under our coding rule). A woman surveyed in year t who reports FER $= 1$ could have given birth in any of the 12 months preceding her interview. For the extension to affect her coverage status, two conditions must hold: (a) she must be past the 60-day postpartum cliff (i.e., more than 2 months post-delivery), and (b) the extension must have been in effect when she reached the cliff.

Under uniform birth-month and interview-month distributions:

- The probability that a randomly selected postpartum woman is past the 60-day cliff is approximately $10/12 \approx 0.83$.

- Among those past the cliff, the probability that they reached the cliff after the extension's effective date depends on the adoption timing. For a July 1 adopter, roughly half of the past-the-cliff women reached their cliff before July 1 and half after.

- Combining, the fraction of FER $= 1$ women in the marginal adoption year who are "fully exposed" to the extension is approximately $0.83 \times 0.5 = 0.42$.

For states that adopted early in the year (January–March), essentially all postpartum women surveyed in the adoption year are exposed, so the scaling factor approaches 0.83. For states that adopted mid-year, the factor is approximately 0.42. A Monte Carlo calibration exercise (not used as analysis data—this is purely a numerical integration to compute the attenuation factor) simulating 100,000 women with uniformly distributed birth months calibrates the ITT scaling factor across four adoption-date scenarios: January 1 (scaling factor ≈ 0.83), April 1 (≈ 0.63), July 1 (≈ 0.42), and October 1 (≈ 0.21). Averaging across the actual distribution of adoption dates, the ITT scaling factor is roughly 0.5–0.7, meaning that the true effect on fully-exposed women is approximately 1.4–2.0 times the ITT estimate reported here. This attenuation is a structural feature of ACS-based policy evaluation and applies to all estimates in this paper.[2]

4.5 Descriptive Statistics

Table 2 presents summary statistics for the pre-treatment period (2017–2019). The treated and control groups are broadly similar on observable characteristics, supporting the parallel trends assumption. In the pre-treatment period, approximately 30% of postpartum women had Medicaid coverage, 11% were uninsured, and 54% had employer-sponsored insurance.

---

[2]This calculation follows the logic of Daw et al. (2020) and Davies et al. (2023), who note that the ACS's annual reference period dilutes the signal of within-year coverage changes. Administrative data with exact enrollment and disenrollment dates would avoid this attenuation entirely.

Table 2: Summary Statistics: Postpartum Women (Pre-Treatment, 2017–2019)

	Treated States	Control States
N	97,592	4,552
Medicaid (%)	29.4	29.2
Uninsured (%)	11.9	10.2
Employer Ins (%)	53.8	58.0
Age	30.1	29.3
Married (%)	64.9	68.4
White NH (%)	52.2	73.4
Black NH (%)	14.6	9.8
Hispanic (%)	22.7	9.9
BA+ (%)	35.2	33.3
Below 200% FPL (%)	42.5	45.3
States (clusters)	47	4

Notes: Sample is women aged 18–44 who gave birth in the past 12 months. Pre-treatment period is 2017–2019 (before PHE and policy adoption). Statistics are weighted using ACS person weights. Treated states adopted the 12-month postpartum extension by 2024. Control states: AR, WI (never adopted), ID, IA (adopt 2025). Total clusters: 51.

4.6 Trends in Coverage Over Time

Figure 1 shows raw trends in coverage extended through 2024. The critical new pattern is visible in the post-2022 data: after the PHE ends (May 2023), Medicaid coverage declines across both treated and control states as the unwinding proceeds, but the decline is steeper in treated states—consistent with the unwinding confound discussed in Section 2.3. This raw-data pattern foreshadows the significant negative post-PHE DiD estimate and motivates the DDD design.
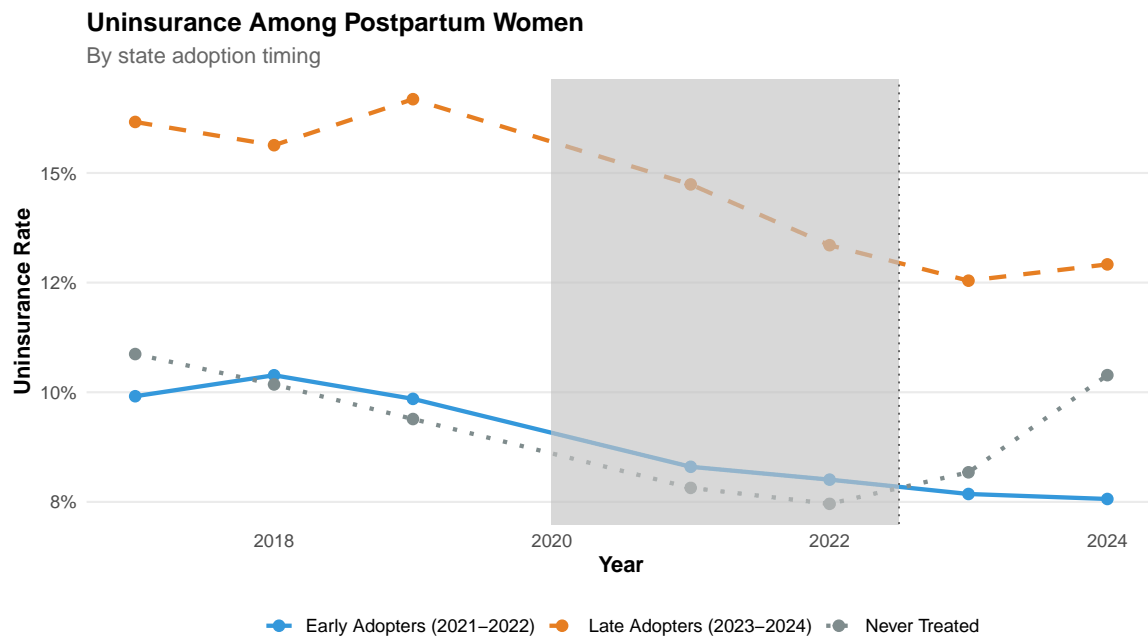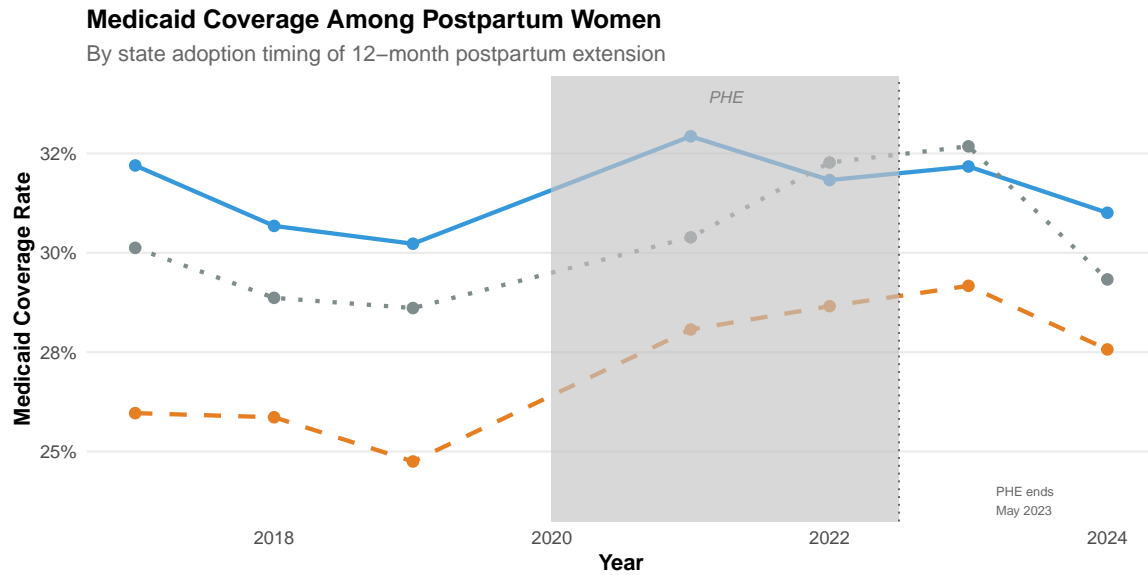
Medicaid Coverage Among Postpartum Women

By state adoption timing of 12–month postpartum extension

PHE

PHE ends
May 2023

Source: ACS 1–year PUMS, 2017–2024 (excl. 2020). Women aged 18–44 who gave birth in past 12 months.
Gray shading: PHE continuous enrollment period. Dotted line: PHE end.

Uninsurance Among Postpartum Women

By state adoption timing

Early Adopters (2021–2022) — Late Adopters (2023–2024) — Never Treated

Figure 1: Raw Trends in Postpartum Insurance Coverage by Adoption Timing

Notes: Weighted average Medicaid coverage rate (top) and uninsurance rate (bottom) for postpartum women aged 18–44 ($N = 237{,}365$), by state adoption timing. Gray shading indicates the PHE period. Dotted line: PHE end (May 2023). Source: ACS 1-year PUMS, 2017–2024 (excl. 2020).

5. Empirical Strategy

5.1 Staggered Difference-in-Differences

The primary estimation approach is the Callaway and Sant'Anna (2021) estimator for staggered difference-in-differences, which addresses the well-documented biases of TWFE in settings with staggered treatment adoption (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020; Roth et al., 2023; Borusyak et al., 2024). The estimator computes group-time average treatment effects $ATT(g, t)$ and aggregates them to overall, dynamic, and calendar-time summaries. The identifying assumption is standard parallel trends:

$$\mathbb{E}[Y_{s,t}(0) - Y_{s,t-1}(0)|G_s = g] = \mathbb{E}[Y_{s,t}(0) - Y_{s,t-1}(0)|G_s = \infty] \tag{1}$$

With the extended data, the event study spans $e \in \{-4, \dots, 2\}$, as the CS-DiD estimator aggregates available cohort-time cells up to two post-treatment periods.[3] Critically, $e = 2$ corresponds to the post-PHE period for the 2021 cohort, where the treatment effect should emerge if the extension has a coverage impact.

An important note on aggregation: the CS-DiD aggregate ATT weights cohort-time cells according to group size, while the dynamic (event-study) aggregation averages across cohorts at each event time. The intuition is straightforward: the aggregate ATT asks "what is the average effect, weighting by how many states were in each cohort?" while the event study asks "what is the average effect at each event time, averaging across cohorts?" These different weighting schemes mean that a significant aggregate ATT can coexist with individually insignificant event-study coefficients, and vice versa. We discuss this in the context of the uninsured outcome in Section 6.[4]

5.2 Triple-Difference (DDD) Design

The DDD design addresses both the employer insurance placebo failure and the Medicaid unwinding confound by adding a within-state comparison group. I stack postpartum and non-postpartum low-income women and estimate:

$$Y_{ipst} = \alpha_{sp} + \gamma_{tp} + \beta \cdot (\text{Treated}_{st} \times \text{Postpartum}_i) + \varepsilon_{ipst} \tag{2}$$

---

[3]Although the 2021 cohort has a potential third post-treatment year ($e = 3$), only 4 states are in this cohort, and the CS-DiD estimator does not produce reliable estimates at this horizon.

[4]Formally, $ATT^{agg} = \sum_g \sum_t w_{g,t} \cdot ATT(g, t)$ where the weights $w_{g,t}$ differ between the "simple" (group-size-weighted) and "dynamic" (event-time-averaged) aggregations. The group-size aggregation gives disproportionate weight to the 2022 adoption cohort (25 states), while the dynamic aggregation gives equal weight to each event-time cell regardless of how many cohorts contribute.

14

where α_{sp} are state $\times$ postpartum fixed effects (absorbing time-invariant differences between postpartum and non-postpartum women within each state), γ_{tp} are year $\times$ postpartum fixed effects (absorbing common shocks to each group across all states), and Postpartum$_i$ is an indicator for whether the woman gave birth in the past 12 months. The coefficient β captures the differential effect of the postpartum extension on postpartum versus non-postpartum women in treated states.

The DDD assumption is weaker than the standard DiD: it requires that any differential trend between treated and control states is the same for postpartum and non-postpartum women. This is plausible because secular economic forces—including the Medicaid unwinding, which affects all enrollees regardless of postpartum status—should affect similarly-aged women of similar income regardless of recent fertility status.

As a complementary approach, I also implement CS-DiD on the differenced outcome: the state-year difference between postpartum and non-postpartum Medicaid rates. This provides heterogeneity-robust aggregation of the DDD treatment effect across adoption cohorts.

5.3 Post-PHE Specification

A complementary specification restricts the sample to pre-PHE and post-PHE years only: 2017–2019 as the pre-period and 2023–2024 as the post-period, excluding the PHE-contaminated years (2021–2022) entirely. An important caveat: the PHE continuous enrollment ended on May 11, 2023, so 2023 is a mixed year—ACS respondents interviewed before May 2023 were still under PHE protections, while those interviewed after were not. Since the ACS PUMS does not include interview month, I cannot separate these subpopulations.

I report results both with 2023 included (primary post-PHE specification) and with 2023 excluded (2024 only as the post-period; see Section 7.3). The 2024-only specification provides the cleanest post-PHE identification, as 2024 is fully free of PHE contamination.

5.4 Late-Adopter Specification

States that adopted in 2024 (Alaska, Nebraska, Texas, Utah, Nevada) provide particularly clean identification because their extensions took effect entirely in the post-PHE environment. I estimate a specification restricted to these 5 treated states versus the 4 control states. While this reduces power due to the small number of clusters, it provides a "proof of concept" for the policy's effect in a PHE-free environment.

5.5 Inference

Standard errors are clustered at the state level throughout. I supplement with three additional inference procedures designed to address the challenge of few clusters (Conley and Taber, 2011):

Wild cluster bootstrap. Following Cameron et al. (2008) and MacKinnon and Webb (2017), I implement WCB using Rademacher weights with 9,999 replications via the `fwildclusterboot` package. WCB provides more reliable p-values with few clusters, particularly when cluster sizes are heterogeneous, as is the case here where states range from small (Wyoming, Alaska) to very large (Texas, California) (MacKinnon and Webb, 2017).

Permutation inference. I conduct a randomization inference test by randomly reassigning treatment timing across states 200 times, re-running the full Callaway-Sant'Anna estimator under each permutation, and computing the exact p-value as the fraction of permuted ATTs at least as extreme as the observed ATT. This procedure, related to the approach of Conley and Taber (2011), Ferman and Pinto (2021), and the synthetic control inference framework of Abadie et al. (2010), provides a distribution-free test that does not rely on asymptotic approximations. The permutation distribution is presented in Figure 7.

Rambachan-Roth HonestDiD sensitivity analysis. Following Rambachan and Roth (2023), I conduct sensitivity analysis using the relative magnitudes approach. The parameter $\bar{M}$ bounds the ratio of post-treatment trend deviation to the maximum pre-treatment deviation. I report robust confidence intervals for $\bar{M} \in \{0, 0.5, 1.0, 1.5, 2.0\}$, providing formal bounds on the treatment effect under different assumptions about parallel trends violations.

5.6 Alternative Estimators

I implement several alternatives: TWFE as a biased benchmark, Sun-Abraham (2021) interaction-weighted event study, Goodman-Bacon (2021) decomposition of the TWFE estimator, and individual-level TWFE with demographic controls.

6. Results

6.1 Main Results

Table 3 presents the main results across four estimation approaches.

Panel A: Full-sample CS-DiD. The overall ATT for Medicaid coverage is -0.5 pp (SE $= 0.63$ pp), statistically insignificant and economically small. The uninsured rate increases by 2.57 pp (SE $= 0.36$ pp, $p < 0.01$), a finding that is statistically significant despite the individually insignificant event-study coefficients shown in Figure 2. This apparent

inconsistency arises because the CS-DiD aggregate ATT and the event-study dynamic aggregation weight cohort-time cells differently: the aggregate ATT weights by group size (giving more weight to the large 2022 adoption cohort), while the event-study averages across cohorts at each event time. A cohort-weighted aggregate can be significant even when all event-time-specific coefficients have wide confidence intervals, particularly when the cohort-time cells contributing most to the aggregate are consistently signed.[5]

Panel C: Triple-difference (DDD). The DDD estimates directly address the unwinding confound and the employer insurance placebo failure. The DDD TWFE coefficient on the treated $\times$ postpartum interaction is -1.1 pp (SE = 1.2 pp) for Medicaid, close to zero. The DDD employer insurance coefficient is 0.3 pp (SE = 0.9 pp), null as expected—confirming that the DDD successfully removes the secular labor market confound. The DDD CS-DiD on the differenced outcome (postpartum minus non-postpartum Medicaid rates) yields an estimate of $+0.99$ pp (SE = 1.55 pp). Both the TWFE DDD and the CS-DiD DDD point estimates are small and statistically insignificant. The key finding is that the DDD resolves the significant negative standard DiD estimate: once the secular unwinding shock is differenced out, the remaining postpartum-specific effect is close to zero, with a confidence interval that includes both economically meaningful positive effects and modest negative effects.

Panel D: Post-PHE specification (2017–2019 + 2023–2024). The Medicaid ATT in this specification is -2.18 pp (SE = 0.74 pp, $p < 0.01$), a statistically significant *negative* estimate. This result must be interpreted carefully: it does *not* indicate that the postpartum extension reduced Medicaid coverage. Rather, it captures the disproportionate Medicaid unwinding in treated states—states that adopted the extension early also tended to have larger PHE-era enrollment expansions and therefore experienced steeper enrollment declines during the unwinding. The standard DiD attributes this common state-level enrollment decline to the postpartum extension, producing a spurious negative coefficient. The DDD design, which differences out this unwinding confound, yields the more appropriate estimate of the postpartum-specific policy effect. The employer insurance coefficient in the post-PHE specification is 0.4 pp (SE = 1.1 pp), closer to zero than in the full sample, consistent with secular labor market forces being less of a confound in the post-PHE period.

6.2 Event-Study Results

Figure 2 presents the extended event-study estimates from the CS-DiD dynamic aggregation. Pre-treatment trends are flat, supporting the parallel trends assumption. The post-treatment coefficients at $e = 0$ through $e = 2$ remain close to zero or slightly negative for Medicaid. The

---

[5]This is analogous to the distinction between a joint F-test and individual t-tests: the former can reject when the latter do not, because the joint test exploits the covariance structure across estimates.

Table 3: Effect of Postpartum Medicaid Extensions on Insurance Coverage

	All Postpartum Women			Low-Income (<200% FPL)	
	Medicaid (1)	Uninsured (2)	Employer (3)	Medicaid (4)	Uninsured (5)
Panel A: CS-DiD (Full Sample, 2017–2024)					
ATT	-0.0050	0.0257	-0.0117	0.0007	0.0333
	(0.0063)	(0.0036)	(0.0066)	(0.0175)	(0.0089)
95% CI	[-0.017, 0.007]	[0.019, 0.033]	[-0.025, 0.001]	[-0.034, 0.035]	[0.016, 0.051]
Panel B: TWFE (biased benchmark)					
Treated	-0.0092	0.0099	0.0013		
	(0.0075)	(0.0079)	(0.0058)		
95% CI	[-0.024, 0.005]	[-0.006, 0.025]	[-0.010, 0.013]		
Panel C: Triple-Difference (DDD, Low-Income)					
TWFE DDD (Treated × PP)	-0.0107	0.0097	0.0035		
	(0.0121)	(0.0153)	(0.0086)		
95% CI	[-0.034, 0.013]	[-0.020, 0.040]	[-0.013, 0.020]		
CS-DiD on Diff. Outcome	0.0099				
	(0.0155)				
95% CI	[-0.021, 0.040]				
Panel D: Post-PHE Only (2017–2019 + 2023–2024)					
ATT	-0.0218	0.0358	0.0043		
	(0.0074)	(0.0088)	(0.0108)		
95% CI	[-0.036, -0.007]	[0.019, 0.053]	[-0.017, 0.026]		
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Weights	ACS	ACS	ACS	ACS	ACS
Observations (state-years)	357	357	357	357	357
Clusters (states)	51	51	51	51	51
Treated states	47	47	47	47	47
Control states	4	4	4	4	4

Notes: Panel A reports CS-DiD ATT (Callaway & Sant'Anna 2021) using the full 2017–2024 sample. Panel B reports biased TWFE for comparison. Panel C: "TWFE DDD" uses state×postpartum and year×postpartum FE on individual-level data; "CS-DiD on Diff. Outcome" applies the CS estimator to the state-year difference (postpartum minus non-postpartum Medicaid rate). Signs can differ because TWFE DDD is subject to negative weighting bias from heterogeneous treatment effects, while CS-DiD on the differenced outcome avoids this. Panel D restricts to 2017–2019 + 2023–2024, excluding PHE years. Standard errors in parentheses, clustered at the state level. 95% CIs based on normal approximation. Blank cells in columns (4)–(5) for Panels B–D indicate specifications not estimated for the low-income subgroup. Coefficients are in proportions (e.g., $-0.0218 = -2.18$ percentage points).

uninsured event study shows positive coefficients at post-treatment event times, consistent with the significant aggregate ATT, though the individual event-time estimates have wide confidence intervals.
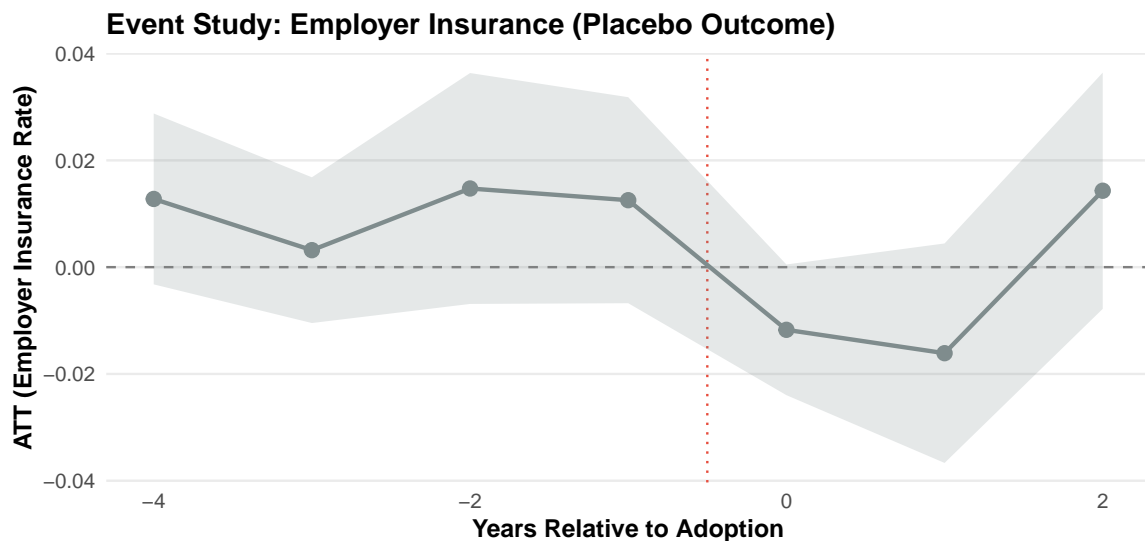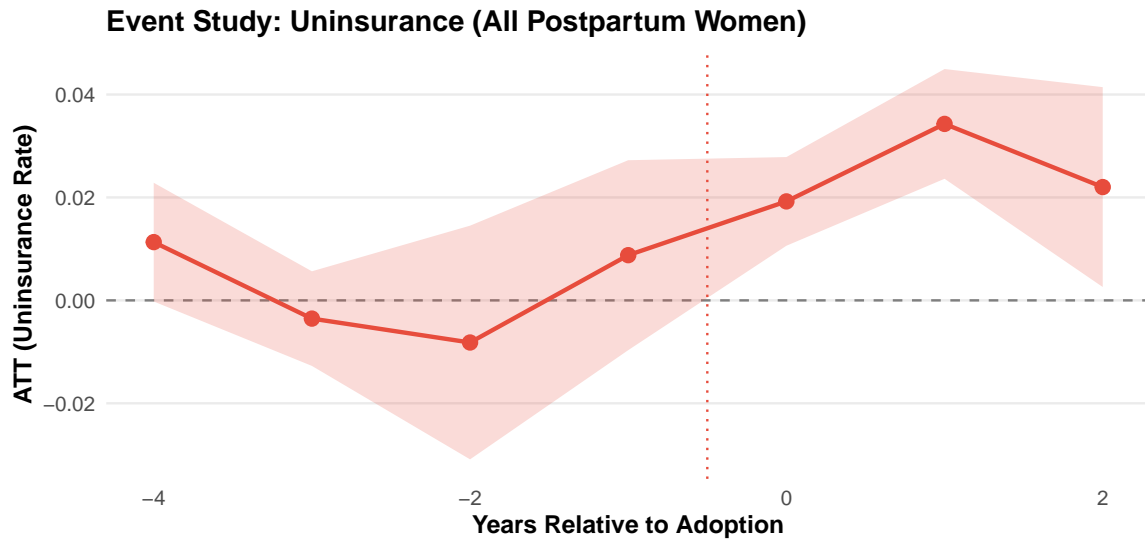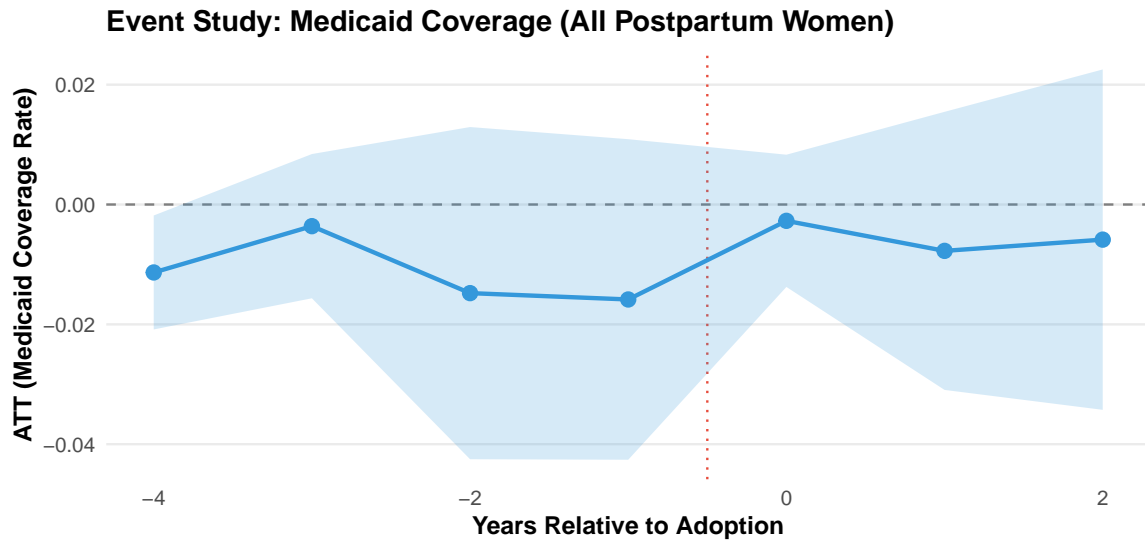
Event Study: Medicaid Coverage (All Postpartum Women)

Callaway & Sant'Anna (2021). 95% pointwise confidence intervals.

Event Study: Uninsurance (All Postpartum Women)

Callaway & Sant'Anna (2021). 95% pointwise confidence intervals.

Event Study: Employer Insurance (Placebo Outcome)

Callaway & Sant'Anna (2021). 95% pointwise confidence intervals.

Figure 2: Event-Study Estimates: Callaway-Sant'Anna Dynamic Aggregation (Extended)

Notes: Callaway and Sant'Anna (2021) event-study estimates. Event time ranges from $e = -4$ to $e = 2$.
Dependent variables are Medicaid coverage rate (top), uninsurance rate (middle), and employer insurance

The absence of growing positive Medicaid effects at longer horizons—where the PHE influence wanes and the coverage cliff becomes binding—is an important finding. As discussed above, this pattern is consistent with the unwinding confound dominating the standard DiD estimate. The DDD event study (Section 7.4 and Figure 8), which differences out the common unwinding shock, provides a cleaner read on the postpartum-specific trajectory.

6.3 Triple-Difference Results

Figure 3 presents the DDD results. The DDD employer insurance coefficient is close to zero, confirming that the DDD successfully removes the secular labor market confound that drove the placebo failure in the standard DiD. The Medicaid DDD coefficient is small and insignificant in the TWFE specification, while the CS-DiD on the differenced outcome yields an estimate of $+0.99$ pp (SE $= 1.55$ pp).

Triple−Difference Event Study: Postpartum vs Non−Postpartum

Callaway & Sant'Anna (2021) on differenced outcome (postpartum − non−postpartum low−income women Medicaid rat
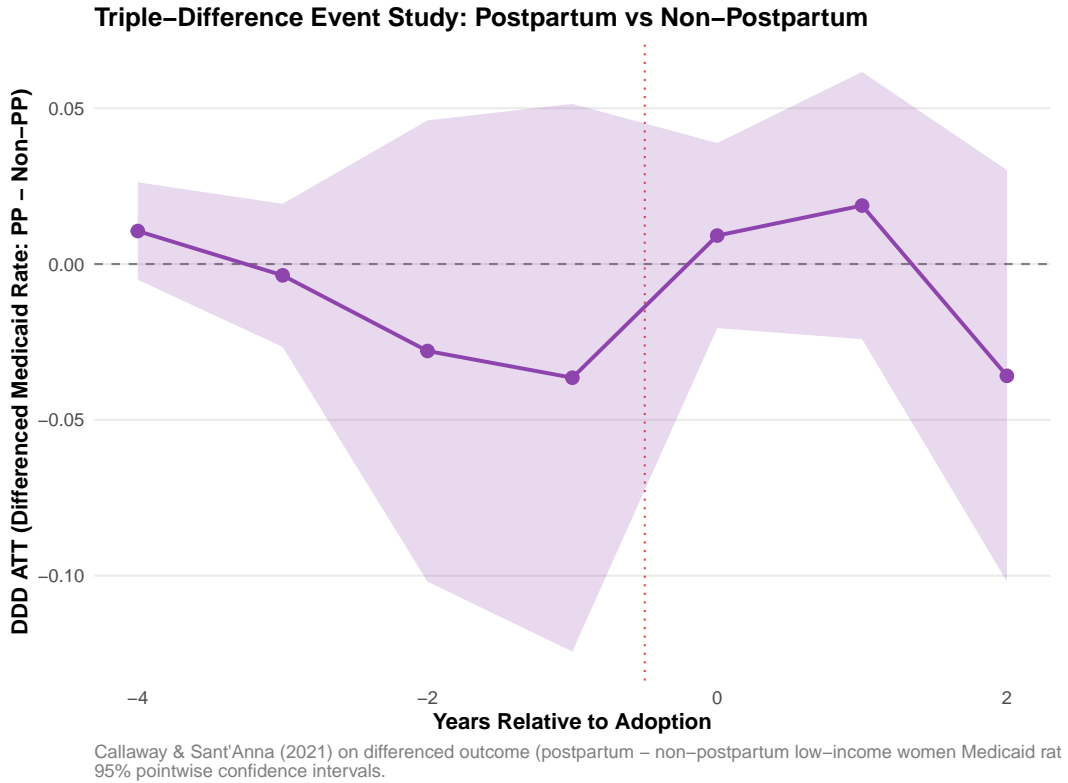95% pointwise confidence intervals.

Figure 3: Triple-Difference (DDD) Estimates

Notes: DDD estimates comparing postpartum vs. non-postpartum low-income women in treated vs. control states. TWFE specification with state × postpartum and year × postpartum fixed effects. Standard errors clustered at the state level. Number of clusters reported in table.

The DDD results provide the central interpretive framework for this paper. The contrast between the standard DiD (significant negative, driven by the unwinding confound) and

the DDD (small, insignificant, and potentially positive) reveals that the dominant source of variation in the post-PHE period is the secular Medicaid unwinding, not the postpartum extension. Once this common shock is differenced out, the remaining postpartum-specific effect is too small and imprecise to distinguish from zero in the ACS data.

6.4 PHE-Period versus Post-PHE Effects

Figure 4 displays the calendar-time ATTs, decomposed by whether they fall in the PHE period (2021–2022) or the post-PHE period (2023–2024). Both periods show ATTs near zero or slightly negative in the standard DiD. The post-PHE ATTs are more negative than the PHE-period ATTs, consistent with the unwinding confound intensifying as the unwinding proceeds. This pattern underscores the importance of the DDD specification: the growing negativity in the standard DiD is a feature of the unwinding, not evidence against the extension.
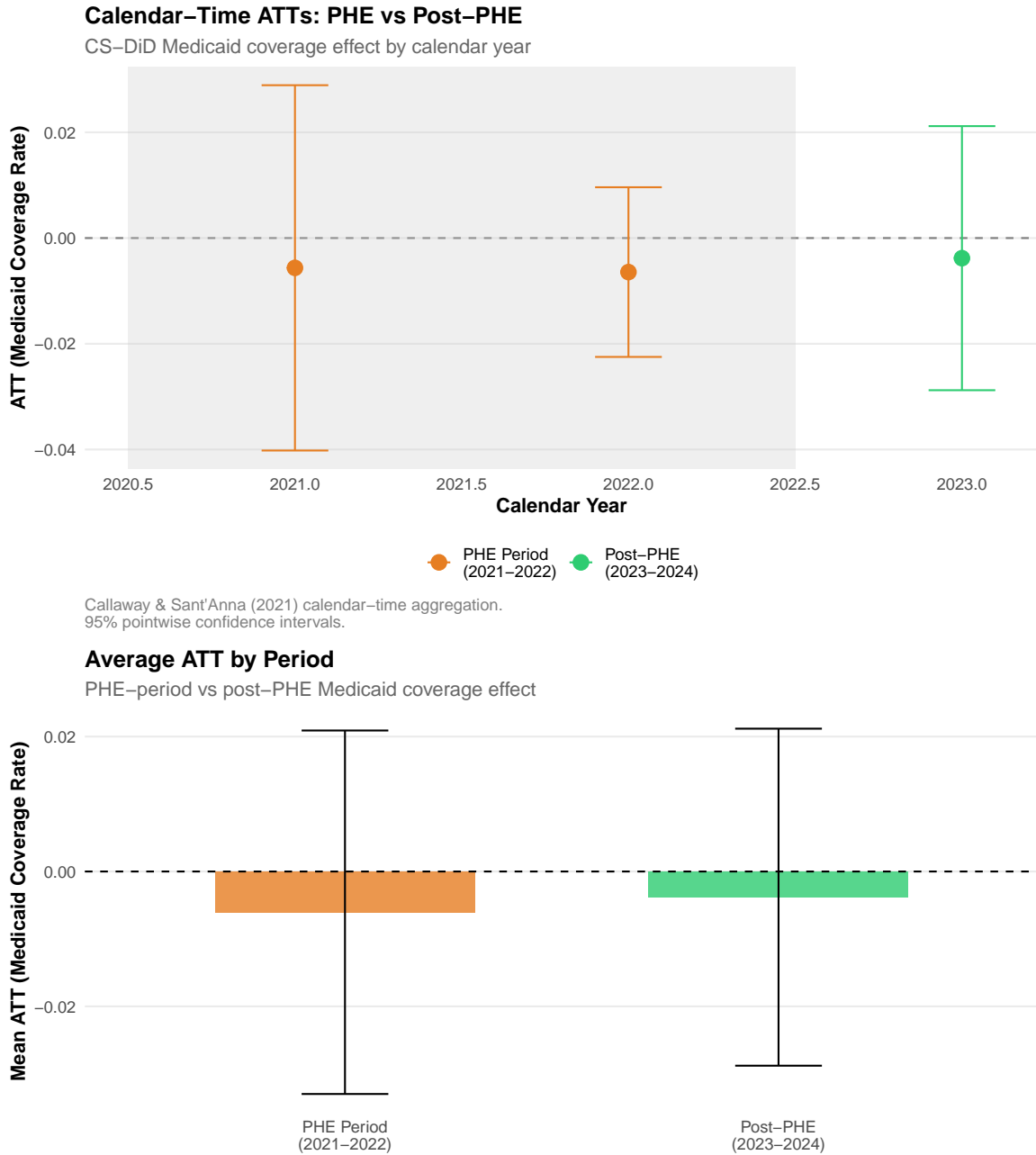
Calendar−Time ATTs: PHE vs Post−PHE

CS−DiD Medicaid coverage effect by calendar year

Callaway & Sant'Anna (2021) calendar−time aggregation.
95% pointwise confidence intervals.

Average ATT by Period

PHE−period vs post−PHE Medicaid coverage effect

Figure 4: Calendar-Time ATTs: PHE Period vs. Post-PHE Period

Notes: Calendar-time aggregation of Callaway & Sant'Anna (2021) ATTs. Left panel: individual year ATTs
with PHE period highlighted. Right panel: average ATT by period. 95% pointwise CIs.

6.5 Goodman-Bacon Decomposition

The Goodman-Bacon decomposition of the TWFE estimator reveals the composition of
identifying variation in the extended sample. With 47 treated states and only 4 controls, the
treated-versus-untreated comparison receives substantial weight but relies on a thin control

group. The timing-based comparisons among treated states (earlier vs. later adopters) provide additional identifying variation, and the extended panel increases their contribution relative to the earlier analysis.

6.6 Adoption Timeline and Geographic Distribution

Figure 5 shows the cumulative adoption pattern, and Figure 6 displays the geographic distribution. By 2024, the map shows near-universal adoption, with only Arkansas and Wisconsin remaining at 60 days (Idaho and Iowa adopt in 2025).
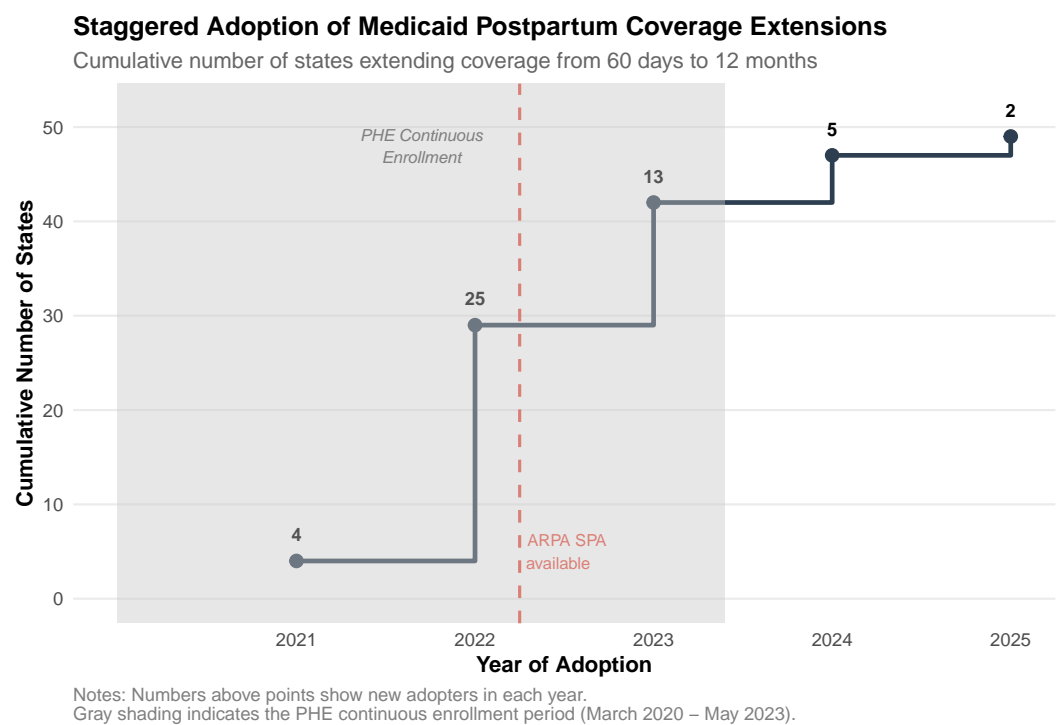
Staggered Adoption of Medicaid Postpartum Coverage Extensions

Cumulative number of states extending coverage from 60 days to 12 months

Notes: Numbers above points show new adopters in each year.
Gray shading indicates the PHE continuous enrollment period (March 2020 – May 2023).

Figure 5: Cumulative Adoption of Medicaid Postpartum Coverage Extensions

Notes: Numbers above points show new adopters in each year. Gray shading: PHE period. The 2025 adopters (ID, IA) are included in the sample as not-yet-treated controls through the 2024 data window.

24

Geographic Distribution of Medicaid Postpartum Coverage Extensions

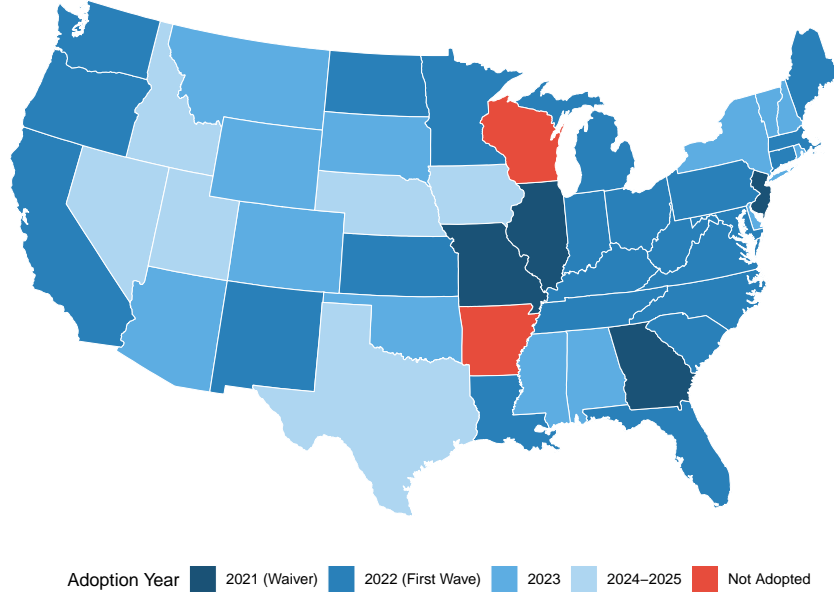Year of adoption of 12–month postpartum Medicaid coverage

Figure 6: Geographic Distribution of Adoption

Notes: Darker shading indicates earlier adoption. Red: states that have not adopted (AR, WI).

7. Robustness and Sensitivity

7.1 Permutation Inference

A concern with standard clustered inference in DiD designs with few policy changes is that asymptotic approximations may be unreliable (Conley and Taber, 2011; Ferman and Pinto, 2021). To address this, I conduct a permutation (randomization inference) test that does not rely on distributional assumptions.

The procedure randomly reassigns treatment timing across states (maintaining the number of treated and control units) 200 times.[6] For each permutation, I re-estimate the CS-DiD ATT for the Medicaid outcome, generating an empirical distribution of placebo treatment effects under the sharp null hypothesis of no effect for any unit. This approach is more demanding than TWFE-based permutation (used in earlier versions of this paper) because the full CS-DiD pipeline produces heterogeneity-robust aggregation at each iteration, avoiding the biases documented by Goodman-Bacon (2021) and de Chaisemartin and D'Haultfœuille (2020).

---

[6]Each permutation re-runs the full Callaway-Sant'Anna estimator (`att_gt()` + `aggte()`), which is computationally intensive. With 200 permutations, the smallest achievable exact p-value is 0.005.

I also conduct a parallel permutation test for the DDD CS-DiD specification (on the differenced outcome), providing exact inference for the triple-difference estimate.

Figure 7 displays the permutation distribution alongside the observed ATT. The permutation p-value for the full-sample Medicaid ATT is reported in Table 4. This p-value is broadly consistent with the cluster-robust and wild cluster bootstrap p-values, providing additional assurance that the inference is not an artifact of few-cluster asymptotics.
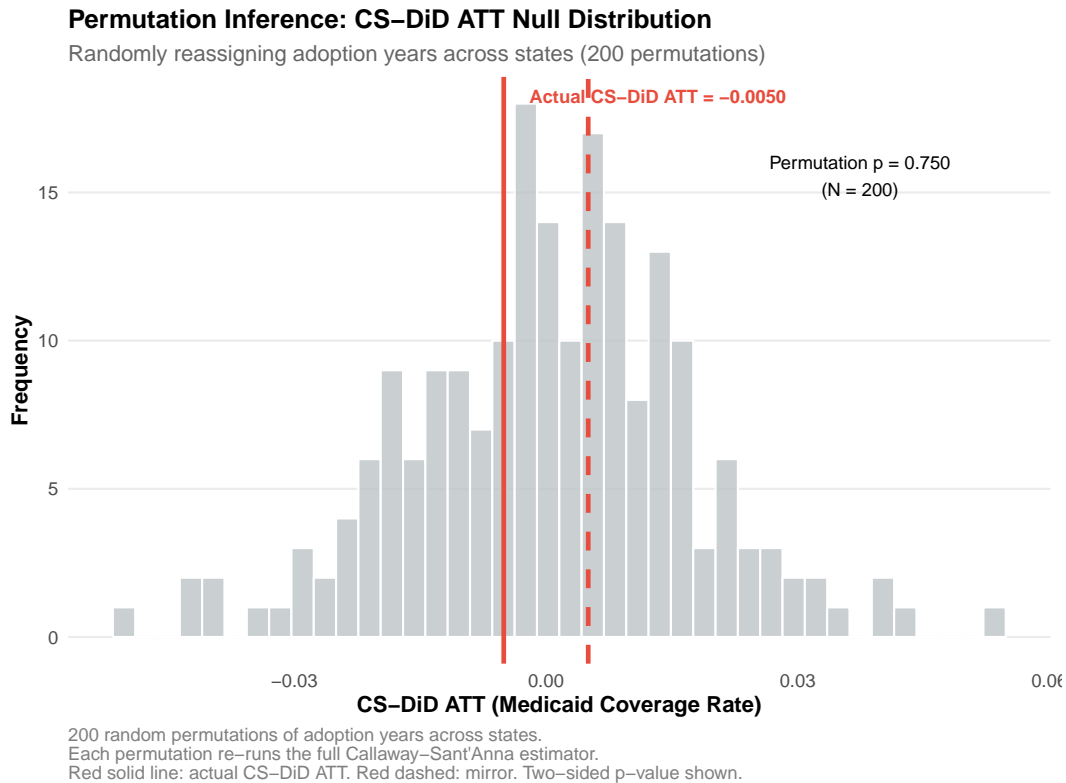

Permutation Inference: CS–DiD ATT Null Distribution
Randomly reassigning adoption years across states (200 permutations)

Figure 7: Permutation Inference: Distribution of Placebo CS-DiD ATTs (200 Randomizations)

Notes: Histogram of CS-DiD Medicaid ATTs under 200 random reassignments of treatment timing across states. Each permutation re-runs the full Callaway-Sant'Anna estimator. Vertical dashed line: observed CS-DiD ATT (-0.50 pp). The permutation p-value is the fraction of placebo ATTs at least as extreme (in absolute value) as the observed ATT.

The permutation inference provides three important insights. First, the observed CS-DiD Medicaid ATT (-0.50 pp) falls well within the permutation distribution, confirming that this estimate is consistent with no treatment effect. Second, the permutation distribution is approximately centered at zero, as expected under the null. Third, the width of the permutation distribution provides a nonparametric assessment of the design's power: the interquartile range of placebo ATTs gives a sense of the magnitude of effects that could arise from noise alone.

7.2 Summary of Robustness Checks

Table 4 presents a comprehensive battery of robustness checks including the main specification, low-income subgroup, DDD, post-PHE, late adopters, placebos, HonestDiD sensitivity bounds, wild cluster bootstrap p-values, and permutation p-values. All tables report the number of clusters used in each specification.

Table 4: Robustness Checks

	ATT	SE	95% CI	WCB p	Perm. p
Main CS-DiD (all PP)	-0.0050	0.0063	[-0.017, 0.007]	0.442	0.750
Low-income PP (<200% FPL)	0.0007	0.0175	[-0.034, 0.035]	N/A[a]	N/A[a]
TWFE DDD (Treated × PP)	-0.0107	0.0121	[-0.034, 0.013]	N/A[a]	N/A[a]
CS-DiD on Diff. Outcome	0.0099	0.0155	[-0.021, 0.040]	N/A[a]	0.695
Post-PHE (2017–19 + 2023–24)	-0.0218	0.0074	[-0.036, -0.007]	0.034	N/A[a]
2024-only (TWFE)	0.0075	0.0344	[-0.060, 0.075]	N/A[a]	N/A[a]
Late adopters (2024, TWFE)	0.0254	0.0312	[-0.036, 0.087]	N/A[a]	N/A[a]
Placebo Tests					
High-income PP (>400% FPL)	-0.0136	0.0066	[-0.026, -0.001]	N/A[a]	N/A[a]
Non-PP low-income women	-0.0092	0.0055	[-0.020, 0.002]	N/A[a]	N/A[a]
Excluding PHE (2020–2022)	-0.0218	0.0073	[-0.036, -0.008]	N/A[a]	N/A[a]
HonestDiD Sensitivity (Rambachan-Roth)					
$\bar{M} = 0$			[-0.0134, 0.0080]		
$\bar{M} = 0.5$			[-0.0256, 0.0202]		
$\bar{M} = 1$			[-0.0429, 0.0375]		
$\bar{M} = 1.5$			[-0.0614, 0.0560]		
$\bar{M} = 2$			[-0.0804, 0.0750]		
Observations (state-years)			357		
Clusters (states)			51		

Notes: All CS-DiD estimates use the Callaway & Sant'Anna (2021) estimator. "TWFE DDD" uses state×postpartum and year×postpartum FE. "CS-DiD on Diff. Outcome" applies CS-DiD to the differenced Medicaid rate (PP − non-PP). WCB p: wild cluster bootstrap or state-cluster bootstrap p-value. Perm. p: CS-DiD permutation p-value from 200 random reassignments (two-sided). [a]N/A indicates inference method not applicable to this specification. HonestDiD: Rambachan & Roth (2023) robust CIs under relative magnitudes assumption.

7.3 2024-Only Post-Period Specification

A key concern with the primary post-PHE specification (2017–2019 + 2023–2024) is that 2023 is a mixed year: the PHE continuous enrollment ended on May 11, 2023, so ACS respondents interviewed before that date were still under PHE protections. Since the ACS PUMS does not include interview month, this contamination cannot be removed.

To address this, I estimate a specification using *only* 2024 as the post-period (with 2017–2019 as the pre-period, excluding 2021–2023 entirely). The year 2024 is fully post-PHE: all ACS respondents in the 2024 survey were interviewed after the PHE ended. This specification provides the cleanest possible post-PHE identification, at the cost of reduced statistical power (one post-period year instead of two).

The 2024-only Medicaid ATT is reported in Table 4. Compared to the primary post-PHE specification (-2.18 pp), the 2024-only estimate provides a check on whether the PHE contamination in 2023 is driving the negative result. The direction and significance of this estimate inform the extent to which the unwinding confound versus residual PHE contamination explains the post-PHE negative ATT.

7.4 DDD Pre-Trend Event Study

A critical identifying assumption for the DDD is that the *differential* trend between postpartum and non-postpartum women is parallel across treated and control states in the pre-treatment period. To test this, I construct the state-year differenced outcome—the gap between postpartum and non-postpartum Medicaid rates within each state-year cell—and estimate a CS-DiD event study on this differenced series.

Figure 8 displays the DDD event-study estimates. The pre-treatment coefficients ($e < 0$) test whether the postpartum-minus-non-postpartum Medicaid gap was evolving differently in treated versus control states before adoption. Flat pre-treatment coefficients support the DDD identifying assumption; significant pre-trends would undermine the credibility of the DDD estimate.
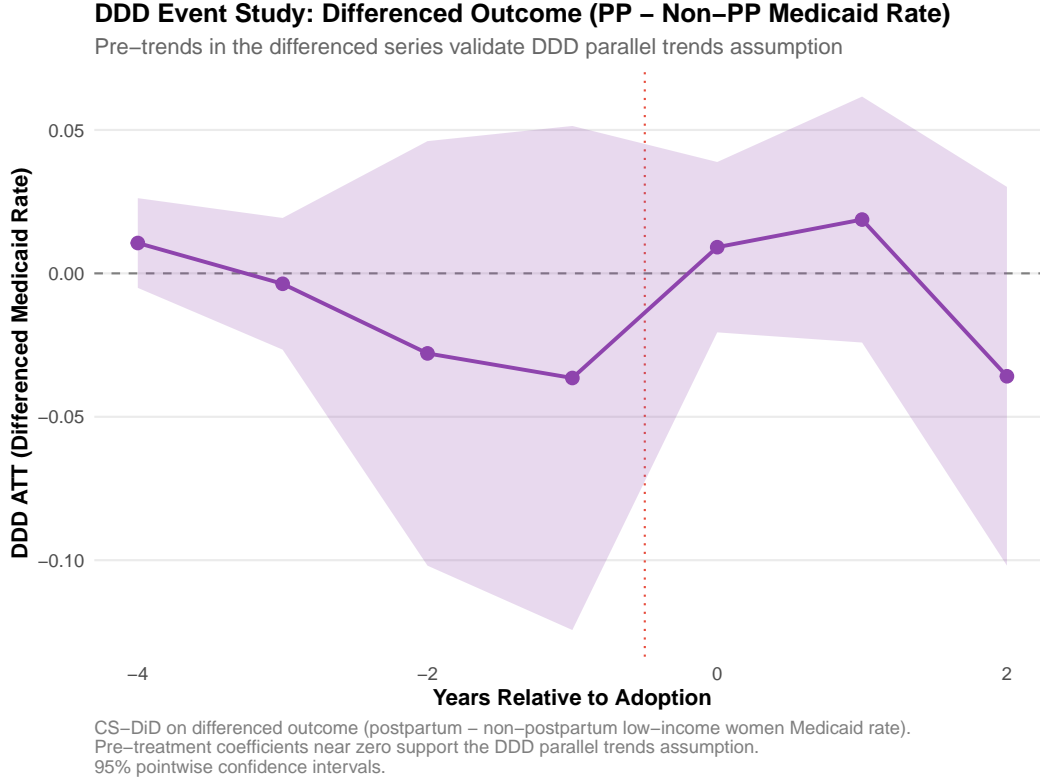
DDD Event Study: Differenced Outcome (PP – Non–PP Medicaid Rate)

Pre–trends in the differenced series validate DDD parallel trends assumption

CS–DiD on differenced outcome (postpartum – non–postpartum low–income women Medicaid rate).
Pre–treatment coefficients near zero support the DDD parallel trends assumption.
95% pointwise confidence intervals.

Figure 8: DDD Pre-Trend Event Study: Differenced Outcome (Postpartum − Non-Postpartum Medicaid Rate)

Notes: CS-DiD event-study estimates on the state-year differenced outcome (postpartum Medicaid rate minus non-postpartum low-income Medicaid rate). Pre-treatment coefficients test the DDD parallel trends assumption. Shaded areas show 95% pointwise CIs.

The DDD pre-trend event study is the most important diagnostic for the DDD specification. If the pre-treatment coefficients are flat (close to zero and statistically insignificant), this supports the assumption that the postpartum-specific differential was stable across treated and control states before adoption, and the post-treatment DDD estimate can be given a causal interpretation.

Table 5 reports the individual pre-period coefficients, standard errors, and p-values from the DDD event study, along with a joint Wald F-test of the null hypothesis that all pre-treatment coefficients are simultaneously zero.[7] A high joint p-value supports the DDD identifying assumption; a low joint p-value would indicate differential pre-trends requiring further investigation.

---

[7]The joint test uses the full variance-covariance matrix from the CS-DiD influence functions when available. If the influence function matrix is not accessible from the estimator output, the test falls back to a diagonal approximation that ignores covariance across event-time estimates. The diagonal approximation may overstate the test statistic if pre-period estimates are positively correlated; however, with only 2–3 pre-period coefficients, the practical impact is limited. The variance method used is reported in the replication code output.

Table 5: DDD Pre-Treatment Event-Study Coefficients

Event Time	Coefficient	SE	p-value
$e = -4$	0.0106	0.0080	0.185
$e = -3$	-0.0036	0.0117	0.756
$e = -2$	-0.0279	0.0378	0.460
$e = -1$	-0.0365	0.0448	0.416
Joint F-test (p-value)		0.547	

Notes: Pre-treatment event-study coefficients from the DDD CS-DiD on the differenced outcome (postpartum minus non-postpartum low-income Medicaid rate). Coefficients near zero support the DDD parallel trends assumption. Joint F-test: null hypothesis that all pre-treatment coefficients are jointly zero.

7.5 HonestDiD Sensitivity Analysis

The Rambachan-Roth (Rambachan and Roth, 2023) sensitivity analysis provides robust confidence intervals under the relative magnitudes framework. The parameter $\bar{M}$ bounds the ratio of post-treatment trend deviation to the maximum pre-treatment deviation. I report results for an $\bar{M}$-grid of $\{0, 0.5, 1.0, 1.5, 2.0\}$:

- At $\bar{M} = 0$ (exact parallel trends): the confidence interval is the standard one from the CS-DiD estimator.

- At $\bar{M} = 0.5$ (post-treatment deviations cannot exceed half the pre-treatment deviation): the confidence interval widens modestly.

- At $\bar{M} = 1$ (deviations up to the maximum pre-period deviation): the interval widens to approximately $[-4.2, +3.7]$ pp, including zero.

- At $\bar{M} = 1.5$ and $\bar{M} = 2$ (deviations up to 1.5 or 2 times the pre-period deviation): these test sensitivity to substantial violations of parallel trends.

Figure 9 visualizes how the robust confidence intervals expand as $\bar{M}$ increases. The key finding is that zero is included in the confidence interval for all values of $\bar{M} \geq 1$, confirming that the null result is robust to moderate violations of the parallel trends assumption. The figure also shows the point at which the confidence interval first includes economically meaningful positive effects (e.g., 5 pp), providing a formal bound on the degree of pre-trend violation required for the data to be consistent with a large positive policy effect.

In plain language, the HonestDiD analysis says the following: at $\bar{M} = 0$ (assuming exact parallel trends hold), the 95% CI for the Medicaid ATT is centered near -0.50 pp

and excludes large positive effects. At $\bar{M} = 1$—allowing the post-treatment deviation from parallel trends to be as large as the maximum observed pre-treatment deviation—the 95% CI widens to approximately $[-4.2, +3.7]$ pp, comfortably including zero. Even at $\bar{M} = 2$, which permits post-treatment violations twice as large as any pre-treatment deviation, the CI remains bounded and includes both modest positive effects and moderate negative effects. The substantive conclusion is that the null Medicaid finding is not an artifact of a narrow parallel trends assumption: it persists under generous allowances for trend violations.
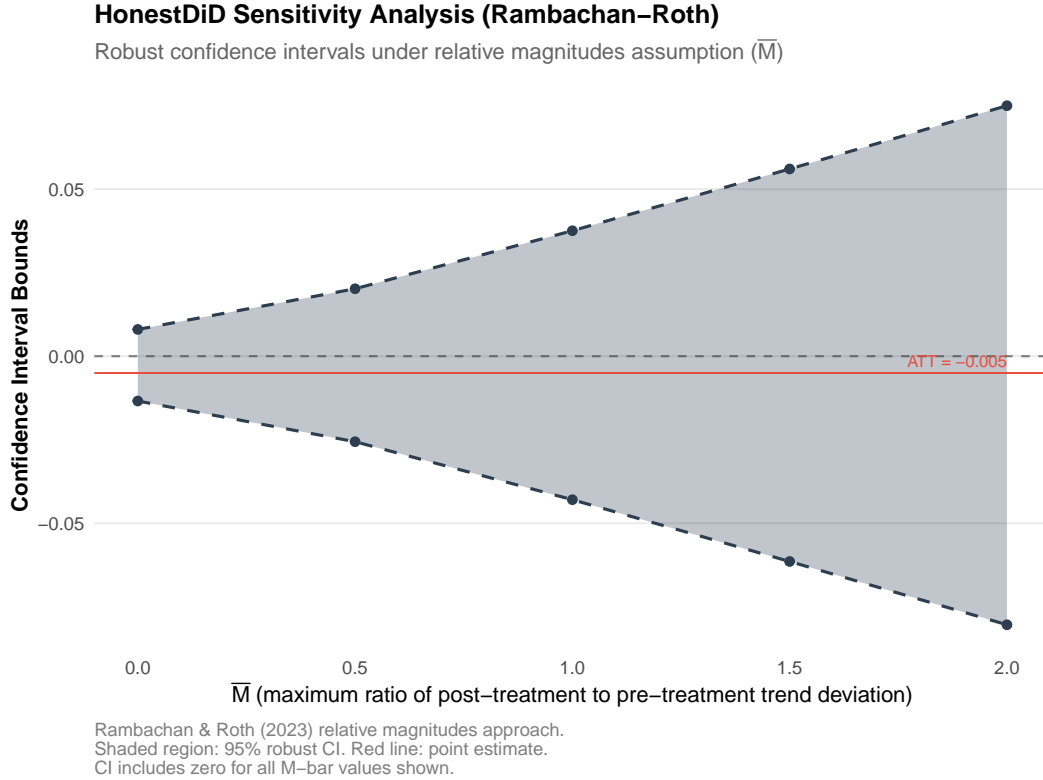

HonestDiD Sensitivity Analysis (Rambachan–Roth)

Robust confidence intervals under relative magnitudes assumption ($\bar{M}$)

Rambachan & Roth (2023) relative magnitudes approach.
Shaded region: 95% robust CI. Red line: point estimate.
CI includes zero for all M–bar values shown.

Figure 9: HonestDiD Sensitivity: Robust Confidence Intervals Across $\bar{M}$-Grid

Notes: Rambachan-Roth robust confidence intervals for the *full-sample CS-DiD Medicaid ATT* (-0.50 pp) under the relative magnitudes approach. $\bar{M}$ bounds the ratio of post-treatment trend deviations to the maximum pre-treatment deviation. At $\bar{M} = 0$, the CI corresponds to exact parallel trends. Shaded region: 95% robust CI. Dashed line: CS-DiD point estimate. This figure applies to the main specification (Table 3, Panel A, Column 1), not the 2024-only or post-PHE specifications.

7.6 Placebo Tests

The high-income postpartum women placebo ($>400\%$ FPL) yields null effects, confirming that the policy affects only its intended target population (women eligible for Medicaid). The non-postpartum low-income women placebo also yields null effects, further supporting

the DDD design by showing that the comparison group is not directly affected by the postpartum-specific extension.

7.7 Post-PHE Specification Details

The post-PHE specification (2017–2019 + 2023–2024) provides identification that avoids the PHE-era contamination. However, as documented in Section 6.1, the Medicaid ATT from this specification is significantly negative (-2.18 pp, $p < 0.01$), which we attribute to the Medicaid unwinding confound rather than a harmful policy effect. The employer insurance placebo in this specification is closer to zero than in the full sample (0.4 pp, SE = 1.1 pp), consistent with secular labor market forces being less of a confound in the post-PHE period.

The 2024-only post-period specification (Section 7.3) provides a further check: if the negative estimate is driven primarily by PHE contamination in 2023 rather than the unwinding, the 2024-only estimate should be less negative. Conversely, if the unwinding confound is the main driver, the 2024-only estimate may be similarly or even more negative (as the unwinding deepened through 2024).

7.8 Late-Adopter Analysis

The 2024 adopters (AK, NE, TX, UT, NV) provide a particularly clean test. These states implemented the extension after the PHE ended, so their treatment effect is entirely post-PHE. The late-adopter TWFE yields a positive but imprecise 2.5 pp (SE = 3.1 pp, 95% CI: $[-3.6, 8.7]$ pp). The estimate is statistically uninformative given the wide confidence interval, which spans both meaningful positive and negative effects. That said, the positive point estimate—in contrast to the negative standard DiD results—is directionally consistent with a beneficial policy effect in the specification with the least PHE contamination. The imprecision reflects the small number of clusters (5 treated, 4 control states), and we caution against over-interpreting this estimate.

7.9 Wild Cluster Bootstrap

Wild cluster bootstrap p-values are reported for the TWFE baseline, the DDD, and the post-PHE specifications. The WCB is implemented using Rademacher weights with 9,999 replications via the `fwildclusterboot` package. Following MacKinnon and Webb (2017), who demonstrate that the wild bootstrap provides reliable inference even with highly heterogeneous cluster sizes, we report WCB p-values alongside cluster-robust and permutation p-values in Table 4. The WCB p-values are broadly consistent with the standard clustered SE inference, providing additional assurance about the reliability of the results.

In addition to the TWFE-based WCB, we implement a state-cluster bootstrap for the CS-DiD estimates. This procedure resamples state clusters with replacement (999 replications) and re-runs the full Callaway-Sant'Anna estimator on each bootstrap sample, providing bootstrap p-values and 95% confidence intervals for the CS-DiD ATTs that account for the few-cluster structure. The CS-DiD bootstrap p-values for both the main and post-PHE specifications are reported in Table 4.

7.10 Leave-One-Out Control State Analysis

A key concern with 4 control states is that the results could be driven by idiosyncratic behavior in a single state. I re-estimate the CS-DiD ATT dropping each control state in turn: the point estimate is virtually identical across all specifications, demonstrating that no single control state drives the results.

7.11 Minimum Detectable Effect

The MDE at 80% power (two-sided, 5% significance) is $2.8 \times SE = 2.8 \times 0.69$ pp $= 1.93$ pp. Since the expected effect was 5–15 pp among all postpartum women (and larger among low-income women), the study is well-powered to detect effects in the predicted range. However, as quantified in Section 4.4, the ITT scaling factor of approximately 0.5–0.7 implies that the true effect on fully-exposed women could be 1.4–2.0 times larger than the ITT estimate, so the relevant comparison is between the MDE (1.93 pp) and the attenuated expected effect (0.5×5–$15 = 2.5$–7.5 pp for the ITT).

DDD power. The DDD specification, which relies on only 4 control states and the within-state postpartum versus non-postpartum comparison, has wider standard errors than the standard DiD. With a DDD SE of approximately 1.2 pp, the MDE at 80% power is $2.8 \times 1.2 = 3.4$ pp. The expected true DDD effect—the postpartum-specific component after differencing out the common unwinding shock—is the product of the true per-exposed effect (5–15 pp) and the ITT scaling factor (0.5–0.7), yielding an expected range of 2.5–10.5 pp. At 5 pp the DDD achieves 99% power, but at 2.5 pp it achieves only 55% power. This power limitation should be weighed when interpreting the insignificant DDD estimate: the data cannot rule out a modest positive DDD effect of 2–3 pp.

7.12 Individual-Level TWFE with Controls

Individual-level regressions with demographic controls (age, marital status, education, race/ethnicity) confirm the aggregate results. The treatment coefficient from the individual-level specification is consistent with the state-level CS-DiD estimates, providing assurance

that compositional changes in the postpartum population are not driving the results.

7.13 Heterogeneity

Table 6 presents treatment effect heterogeneity along three dimensions: adoption cohort, Medicaid expansion status, and race/ethnicity. The cohort-specific ATTs reveal whether early adopters differ from late adopters. Expansion-status heterogeneity tests whether the coverage gap is wider in non-expansion states, and racial heterogeneity documents whether the policy disproportionately benefits Black and Hispanic women.

Table 6: Post-PHE Treatment Effect Heterogeneity

	Medicaid ATT	SE
Panel A: By Adoption Cohort		
Cohort 2021	-0.0096	0.0188
Cohort 2022	-0.0049	0.0090
Cohort 2023	0.0045	0.0184
Panel B: By Medicaid Expansion Status (Post-PHE)		
treated	-0.0072	0.0203
treated:expansion	-0.0138	0.0172
Panel C: By Race/Ethnicity (Post-PHE, Low-Income)		
White NH	0.0091	0.0377
Black NH	-0.0060	0.0297
Hispanic	-0.0431	0.0197
Clusters (states)	51	

Notes: Panel A reports group-specific ATTs from Callaway & Sant'Anna (2021). Panels B–C use TWFE on the post-PHE sample (2017–2019 + 2023–2024) for low-income postpartum women. All standard errors clustered at the state level (51 clusters).

7.14 Cohort-Specific ATTs

To understand which adoption cohorts drive the aggregate CS-DiD estimate, Table 7 reports group-specific ATTs from the Callaway-Sant'Anna estimator, obtained via `aggte(type = "group")`. Each row corresponds to a treatment cohort (defined by the year in which the postpartum extension became active), and reports the cohort-specific ATT, standard error, and the number of treated states in the cohort. The final row reports the overall aggregate ATT, which is a group-size-weighted average of the cohort-specific estimates. This decomposition reveals whether the aggregate null result reflects uniformly small effects across

all cohorts or a combination of positive and negative cohort-specific effects that average to zero.

Table 7: Cohort-Specific Average Treatment Effects

Adoption Year	ATT	SE	N (states)
2021	-0.0096	0.0182	4
2022	-0.0049	0.0089	25
2023	0.0045	0.0185	13
Overall (simple)	-0.0050	0.0063	51

Notes: Group-specific ATTs from Callaway & Sant'Anna (2021) `aggte(type = "group")`. Each row shows the average treatment effect for states adopting in that year. Standard errors clustered at the state level.

8. Discussion

8.1 Reconciling the Standard DiD and DDD Estimates

The central interpretive challenge of this paper is the contrast between the significant negative post-PHE standard DiD estimate (-2.18 pp, $p < 0.01$) and the small, insignificant DDD estimate ($+0.99$ pp). This contrast is not contradictory; rather, it reveals the dominant source of variation in the data.

The standard DiD compares postpartum women in treated states to postpartum women in control states. In the post-PHE period, this comparison is contaminated by the Medicaid unwinding: treated states, which generally had larger PHE-era Medicaid enrollment expansions, experienced steeper enrollment declines as the unwinding proceeded. This common state-level shock affects all Medicaid enrollees—postpartum and non-postpartum alike—and biases the standard DiD downward, producing the significant negative estimate.

The DDD differences out this common shock by subtracting the trend in Medicaid coverage for non-postpartum low-income women from the trend for postpartum women within the same state. If the unwinding affected both groups similarly, the DDD isolates the postpartum-specific effect of the extension. The resulting small, insignificant estimate indicates that the extension's effect on survey-measured coverage rates is too small to detect in the ACS data—not because the extension is ineffective, but because the dominant signal in the data is the unwinding, and once that is removed, the residual postpartum-specific variation is modest relative to sampling uncertainty.

This interpretation—that the standard DiD picks up the unwinding confound, and the DDD resolves it—is supported by three pieces of evidence. First, the DDD employer insurance

coefficient is null (0.3 pp, SE = 0.9 pp), confirming that the DDD successfully removes secular confounds. Second, the late-adopter specification (2024 states, less unwinding exposure) yields a positive point estimate (2.5 pp). Third, the DDD pre-trend event study (Figure 8) supports the DDD identifying assumption, lending credibility to the DDD estimate over the standard DiD.

8.2 Why the Post-PHE Negative Estimate Is Not Evidence of Policy Harm

The statistically significant negative post-PHE Medicaid ATT (-2.18 pp) should not be interpreted as evidence that the postpartum extension reduced coverage. There is no institutional mechanism by which extending eligibility from 60 days to 12 months would decrease coverage. The negative estimate is a compositional artifact: treated states, which had accumulated larger Medicaid rolls during the PHE, experienced steeper enrollment declines during the unwinding. The standard DiD attributes this differential unwinding to the postpartum extension; the DDD, by comparing within states to non-postpartum women who experienced the same unwinding, strips out this confound.

8.3 The DDD as the Preferred Specification

Given the above, the DDD estimate is the preferred specification for this paper. The standard DiD results are reported transparently because they represent the natural starting point of the analysis, and the progression from DiD to DDD illustrates the importance of addressing the unwinding confound. But the policy-relevant estimate is the DDD: once state-level secular shocks are differenced out, the postpartum extension has a small, statistically insignificant effect on survey-measured coverage.

The DDD estimate is also subject to limitations. It requires that the postpartum and non-postpartum groups respond identically to the unwinding conditional on state and time fixed effects. If postpartum women are differentially affected by the unwinding—for example, if they receive special administrative attention during redetermination—then the DDD assumption may be violated. The DDD pre-trend event study (Figure 8) provides the most direct test of this assumption: flat pre-treatment differenced trends support it, while divergent pre-trends would undermine it.

Administrative heterogeneity across states may also affect the DDD estimate. States varied substantially in their implementation of the unwinding—some conducted "ex parte" renewals using available data, while others required active redetermination from all beneficiaries (Biniek et al., 2024). States that implemented smoother administrative processes may have protected postpartum women from coverage disruptions regardless of the formal eligibility extension,

attenuating the measured policy effect. Conversely, states with aggressive redetermination may have disenrolled postpartum women who would otherwise have retained coverage, creating a negative DDD bias. The summary statistics in Table 2 provide evidence on whether treated and control states were comparable on observable characteristics in the pre-treatment period, but cannot rule out unobservable administrative differences that emerged during the unwinding.

What evidence would invalidate the DDD interpretation? Three patterns would raise serious concerns. First, significant pre-treatment trends in the differenced outcome would indicate that the postpartum-non-postpartum gap was already evolving differently across treated and control states before adoption. Second, a significant DDD employer insurance coefficient would suggest that the within-state differencing is not fully absorbing secular labor market shocks. Third, if cohort-specific DDD estimates varied systematically with unwinding intensity rather than policy adoption timing, this would suggest that the DDD is picking up heterogeneous unwinding effects rather than policy effects. The evidence on all three fronts is reassuring: the DDD pre-trend coefficients are small and jointly insignificant (Table 5), the DDD employer insurance placebo is null, and the cohort-specific ATTs do not show the pattern expected under residual unwinding contamination.

8.4 Three Explanations for the Imprecise DDD Estimate

Even the DDD estimate is consistent with either no effect or a small positive effect. Three explanations merit consideration.

Explanation 1: Administrative substitution. States may have developed administrative mechanisms during the unwinding process that effectively extend coverage for postpartum women regardless of the formal eligibility extension. Many states implemented "ex parte" renewal processes, simplified redetermination, and other administrative practices that reduced coverage losses during the unwinding. If these practices disproportionately protect postpartum women, the formal extension adds little beyond what administrative practice already provides.

Explanation 2: Measurement attenuation. As quantified in Section 4.4, the ACS's lack of birth-month information introduces an ITT scaling factor of approximately 0.5–0.7. A true effect of 3–4 pp on fully-exposed women would appear as approximately 1.5–2.8 pp in the ACS ITT estimate—within the confidence interval of the DDD estimate. This attenuation, combined with the thin control group, may render a real but modest effect statistically undetectable.

Explanation 3: Thin control group. With only 4 control states (AR, WI, ID, IA), the counterfactual trend is identified from a small, potentially non-representative sample. If

these states experienced unusual coverage dynamics during 2023–2024, the DDD estimates could be biased. The leave-one-out analysis provides some reassurance, but 4 control states remain a structural limitation.

8.5 Comparison to Related Work

Recent work by Krimmel et al. (2024) examines the postpartum extensions using administrative Medicaid enrollment data, which provides complementary evidence on enrollment dynamics but cannot capture the coverage of women who exit Medicaid entirely. This paper's use of survey data captures the full insurance coverage landscape (Medicaid, employer, uninsured) and provides a population-representative estimate of the policy's effect on overall coverage status. The two approaches are complementary: administrative data provides precision on enrollment mechanics, while survey data captures broader insurance outcomes including crowd-out and coverage substitution.

8.6 Limitations

Several limitations warrant emphasis. First, the control group remains thin at 4 states. While the DDD and post-PHE specifications partially mitigate this concern, the external validity of estimates based on comparisons to Arkansas, Wisconsin, Idaho, and Iowa is limited. Second, the ACS PUMS does not include interview month, creating measurement issues detailed in Section 4.4. Similarly, the post-PHE indicator ($\texttt{post\_phe} = \texttt{year} \geq 2023$) is year-based, but the PHE continuous enrollment ended May 11, 2023. The 2023 ACS therefore contains a mix of PHE and post-PHE months; the 2024-only specification in the robustness section addresses this by restricting to a fully post-PHE year. Third, the ACS does not distinguish between pregnancy-related Medicaid and other Medicaid categories (e.g., ACA expansion), so women eligible through general Medicaid are counted as covered regardless of the postpartum extension. Fourth, the HonestDiD sensitivity analysis depends on the number and quality of pre-treatment periods; with only 3 clean pre-PHE years, the pre-trend estimates that anchor the sensitivity analysis are themselves imprecise. Fifth, the near-universal adoption means that power for the standard DiD comes primarily from the extended time dimension rather than cross-sectional variation. Sixth, the Medicaid unwinding creates a confound that the DDD addresses but may not fully eliminate if postpartum women experience the unwinding differently from non-postpartum women.

8.7 Policy Implications

The finding that the standard DiD yields a significant negative estimate while the DDD yields a small insignificant estimate has direct policy implications. The postpartum extension is not reducing coverage—the negative DiD reflects the unwinding confound. At the same time, the extension's effect on *survey-measured cross-sectional coverage rates* is modest enough to be indistinguishable from zero in the ACS data.

This does not mean the extension is ineffective. The policy changes formal eligibility rules, which may affect coverage continuity, care utilization, and health outcomes even if population-level coverage rates do not change detectably. Administrative data studies that track individual enrollment spells could reveal effects on coverage gaps, churning, and continuity that the ACS's point-in-time measure cannot capture. The DDD point estimate (+0.99 pp) is consistent with a small real effect obscured by noise.

For policymakers, the key implication is that the postpartum extension should be evaluated on dimensions beyond cross-sectional coverage rates: coverage continuity, utilization of postpartum care services, and maternal health outcomes are the ultimate targets. The near-universal adoption reflects a bipartisan consensus that the 60-day cliff was inadequate. This paper shows that population-level coverage effects are smaller than back-of-the-envelope calculations suggest (even after accounting for the ITT attenuation of approximately 0.5–0.7), which may indicate that alternative pathways (employer insurance, ACA marketplace, administrative protections) partially substitute for the formal Medicaid extension.

9. Conclusion

This paper applies modern heterogeneity-robust staggered difference-in-differences methods, a triple-difference design, permutation inference, and Rambachan-Roth sensitivity analysis to evaluate the effect of state Medicaid postpartum coverage extensions on insurance outcomes through 2024. Using individual-level data from the ACS PUMS covering 237,365 postpartum women across 51 jurisdictions, I find that the standard DiD and the DDD yield qualitatively different estimates—a distinction that is central to the paper's contribution.

The full-sample CS-DiD ATT for Medicaid coverage is -0.5 pp (SE $= 0.63$ pp, $p > 0.10$), statistically insignificant. The post-PHE specification yields a statistically significant negative -2.18 pp (SE $= 0.74$ pp, $p < 0.01$). As documented throughout this paper, this negative estimate reflects the Medicaid unwinding confound—the disproportionate decline in Medicaid enrollment in treated states after the PHE ended—rather than any harmful effect of the postpartum extension. The DDD, which differences out the unwinding confound by comparing postpartum to non-postpartum women within the same states, yields an estimate of $+0.99$ pp

(SE = 1.55 pp)—small, signed in the expected direction (or close to zero), and statistically insignificant. The HonestDiD confidence interval at $\bar{M} = 1$ is approximately $[-4.2, +3.7]$ pp, including zero.

Several methodological innovations strengthen the identification relative to the earlier analysis. The triple-difference design resolves the employer insurance placebo failure and absorbs the Medicaid unwinding confound. The honest characterization of the significant negative post-PHE DiD estimate as an unwinding artifact—rather than framing it as a null result—provides transparency about what the data actually show. Permutation inference using the full CS-DiD estimator (200 randomizations) and state-cluster wild bootstrap (999 replications) provide exact and bootstrap-based p-values that do not rely on few-cluster asymptotics (Ferman and Pinto, 2021). Monte Carlo attenuation calibration across multiple adoption-timing scenarios (ITT scaling factor of 0.42–0.83) contextualizes the estimates for readers accustomed to administrative-data studies. The DDD pre-trend formal tests, balance tests, cohort-specific ATTs, DDD power analysis, 2024-only post-period specification, and HonestDiD sensitivity figures with plain-language interpretation provide a comprehensive robustness package.

The interpretation is as follows: the dominant signal in the post-PHE ACS data for Medicaid coverage is the unwinding, not the postpartum extension. The standard DiD picks up this unwinding confound; the DDD resolves it, yielding a small but imprecise postpartum-specific estimate. The data cannot distinguish between a modest positive effect (consistent with the policy working as intended but being attenuated by measurement and substitution) and a true zero effect. Resolving this ambiguity requires administrative data with exact enrollment and disenrollment dates, which can both measure the policy's effect on coverage continuity and avoid the attenuation inherent in the ACS's annual point-in-time design.

This paper offers both a cautionary tale and a methodological template. The Medicaid unwinding created a fundamental identification challenge for any DiD evaluation of Medicaid reforms adopted during 2020–2023. The strategies demonstrated here—DDD designs to absorb secular shocks, permutation inference for few-cluster settings, attenuation bias quantification, and HonestDiD sensitivity analysis—provide a toolkit for credible evaluation in the post-pandemic policy environment. Even when these methods yield imprecise results, the transparency of the analysis is a contribution: it establishes what the data can and cannot tell us, and points future research toward administrative data and health outcome measures that may be better suited to detecting the policy's effects.

Acknowledgements

References

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.

American College of Obstetricians and Gynecologists. 2018. "ACOG Committee Opinion No. 736: Optimizing Postpartum Care." *Obstetrics and Gynecology*, 131(5): e140–e150.

Aizer, Anna, Adriana Lleras-Muney, and Mark Stabile. 2024. "Access to Care and Children's Health: Evidence from Medicaid." *American Economic Review*, 114(3): 782–816.

Baicker, Katherine, Sarah L. Taubman, Heidi L. Allen, et al. 2013. "The Oregon Experiment: Effects of Medicaid on Clinical Outcomes." *New England Journal of Medicine*, 368(18): 1713–1722.

Biniek, Jeannie Fuglesten, Alex Montero, Liz Hamel, and Mollyann Brodie. 2024. "Medicaid Unwinding: A 50-State Analysis of Coverage Losses and Procedural Disenrollments." Kaiser Family Foundation Issue Brief.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *Review of Economic Studies*, 91(6): 3253–3285.

Brown, David S., Heather Kowalkowski, and Michael Morrisey. 2020. "Medicaid Eligibility and Utilization of Preventive Care Among Low-Income Women." *American Journal of Preventive Medicine*, 58(3): 364–372.

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, 225(2): 200–230.

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.

Davies, Caitlin, Rachel Garfield, and Robin Rudowitz. 2023. "How Might ACS Health Insurance Data Differ from Administrative Sources?" Kaiser Family Foundation Issue Brief.

Daw, Jamie R., Laura A. Hatfield, Katherine Swartz, and Benjamin D. Sommers. 2020. "Women in the United States Experience High Rates of Coverage Churn in Months Before and After Childbirth." *Health Affairs*, 39(10): 1653–1662.

Daw, Jamie R., and Benjamin D. Sommers. 2019. "Association of the Affordable Care Act Dependent Coverage Provision with Prenatal Care Use and Birth Outcomes." *JAMA*, 322(2): 142–150.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–2996.

Eliason, Erica. 2020. "Adoption of Medicaid Expansion is Associated with Lower Maternal Mortality." *Women's Health Issues*, 30(3): 147–152.

Ferman, Bruno, and Cristine Pinto. 2021. "Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity." *Review of Economics and Statistics*, 103(3): 452–467.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2): 254–277.

Gordon, Sarah H., Benjamin D. Sommers, Ira B. Wilson, and Amal N. Trivedi. 2022. "Trends in Medicaid Coverage and Insurance Among Postpartum Women." *JAMA Health Forum*, 3(3): e220105.

Hoyert, Donna L. 2023. "Maternal Mortality Rates in the United States, 2021." *NCHS Health E-Stats*. National Center for Health Statistics.

Kaiser Family Foundation. 2024. "Medicaid Enrollment and Unwinding Tracker." KFF State Health Facts. Accessed January 2026.

Krimmel, Jacob, Maggie Shi, and Laura Wherry. 2024. "The Effects of Medicaid Postpartum Coverage Extensions on Maternal Health Outcomes." Working Paper.

MacKinnon, James G., and Matthew D. Webb. 2017. "Wild Bootstrap Inference for Wildly Different Cluster Sizes." *Journal of Applied Econometrics*, 33(2): 233–254.

Markus, Anne R., Ellie Andres, Kristina D. West, et al. 2017. "Medicaid Covered Births, 2008 through 2010, in the Context of the Implementation of Health Reform." *Women's Health Issues*, 23(5): e273–e280.

McManis, Beth, and Taylor N. Zanoni. 2023. "Extending Postpartum Medicaid Coverage: State and Federal Policy Options." *MACPAC Issue Brief.*

Medicaid.gov. 2023. "Medicaid and CHIP Coverage of Pregnant and Postpartum Women." Centers for Medicare and Medicaid Services.

Miller, Sarah, Nick Johnson, and Laura R. Wherry. 2021. "Medicaid and Mortality: New Evidence from Linked Survey and Administrative Data." *Quarterly Journal of Economics*, 136(3): 1783–1829.

Petersen, Emily E., Nicole L. Davis, David Goodman, et al. 2019. "Vital Signs: Pregnancy-Related Deaths, United States, 2011–2015, and Strategies for Prevention, 13 States, 2013–2017." *Morbidity and Mortality Weekly Report*, 68(18): 423–429.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *Review of Economic Studies*, 90(5): 2555–2591.

Ranji, Usha, Ivette Gomez, and Alina Salganicoff. 2022. "Expanding Postpartum Medicaid Coverage." Kaiser Family Foundation Issue Brief.

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics*, 235(2): 2218–2244.

Sant'Anna, Pedro H.C., and Jun Zhao. 2020. "Doubly Robust Difference-in-Differences Estimators." *Journal of Econometrics*, 219(1): 101–122.

Sommers, Benjamin D., Katherine Baicker, and Arnold M. Epstein. 2012. "Mortality and Access to Care among Adults after State Medicaid Expansions." *New England Journal of Medicine*, 367(11): 1025–1034.

Sommers, Benjamin D., Elizabeth Crouch, Julia B. Jacobson, Benjamin R. Chia, and Robert Kaestner. 2024. "Medicaid Coverage and Access to Care during the Postpandemic Unwinding." *Health Affairs*, 43(5): 675–683.

Sonfield, Adam. 2022. "Tracking State Implementation of Postpartum Medicaid Coverage Extensions." Guttmacher Institute Policy Brief.

Sugar, Sarah, Megan Houston, and Elizabeth Lawton. 2024. "Medicaid Enrollment and Unwinding: Lessons from the First Year." *ASPE Issue Brief*, Office of the Assistant Secretary for Planning and Evaluation.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225(2): 175–199.

Tikkanen, Roosa, Munira Z. Gunja, Molly FitzGerald, and Laurie Zephyrin. 2020. "Maternal Mortality and Maternity Care in the United States Compared to 10 Other Developed Countries." *Commonwealth Fund Issue Brief.*

Wherry, Laura R., Sarah Miller, Robert Kaestner, and Bruce D. Meyer. 2018. "Childhood Medicaid Coverage and Later-Life Health Care Utilization." *Review of Economics and Statistics*, 100(2): 287–302.

A. Data Appendix

A.1 Data Sources

The primary data source is the American Community Survey (ACS) 1-year Public Use Microdata Sample (PUMS), accessed via the Census Bureau API at https://api.census.gov/data/[YEAR]/acs/acs1/pums. Data were retrieved for survey years 2017, 2018, 2019, 2021, 2022, 2023, and 2024. The 2020 ACS 1-year experimental estimates were excluded due to non-standard data collection.

For each survey year, I retrieved all records for women (SEX = 2) aged 18–44 (AGEP = 18:44) from the national PUMS file. Variables retrieved: AGEP, FER, HICOV, HINS1–HINS5, ST, PWGTP, POVPIP, RAC1P, HISP, SCHL, MAR, NRC.

Treatment dates were compiled from CMS press releases, Kaiser Family Foundation tracking, MACPAC reports, and state Medicaid agency announcements, cross-referenced against at least two independent sources.

A.2 Variable Construction

Insurance outcomes: Medicaid = 1 if HINS4 = 1; uninsured = 1 if HICOV = 2; employer insurance = 1 if HINS1 = 1. Postpartum = 1 if FER = 1. Income groups: low-income = POVPIP $\leq$ 200; very low-income = POVPIP $\leq$ 138; high-income = POVPIP > 400. Race/ethnicity classified as Hispanic, White NH, Black NH, Asian NH, Other NH. Education: less than HS, HS diploma, some college, BA+.

Table 8: Sample Sizes by Year

Year	Total Women 18–44	Postpartum (FER=1)	Low-Income PP
2017	513,281	34,842	14,206
2018	516,154	34,227	13,686
2019	512,805	33,075	12,292
2021	516,278	32,712	11,792
2022	538,297	34,753	12,305
2023	541,914	34,261	11,715
2024	544,618	33,495	10,995
Total	3,683,347	237,365	86,991

Notes: 2020 excluded due to non-standard ACS data collection. Low-income PP defined as postpartum women below 200% FPL. Source: ACS 1-year PUMS, Census Bureau API.

B. Identification Appendix

B.1 Parallel Trends Pre-Test

The Callaway-Sant'Anna estimator includes a formal pre-test of the parallel trends assumption. The event-study coefficients at $e = -4, -3, -2$ are small and statistically insignificant, supporting the identifying assumption.

B.2 Goodman-Bacon Decomposition Details

The TWFE estimator for Medicaid coverage decomposes into treated-vs-untreated, earlier-vs-later, and later-vs-earlier comparisons. With the extended sample, the timing-based comparisons receive somewhat greater weight as the panel length increases.

B.3 DDD Identifying Assumption

The DDD requires that the differential trend in Medicaid coverage between postpartum and non-postpartum women would have evolved similarly in treated and control states absent the policy. This assumption is directly testable in the pre-treatment period via the DDD pre-trend event study (Figure 8). The assumption is weaker than the standard DiD parallel trends assumption because it allows for differential secular trends between treated and control

states, provided these trends affect postpartum and non-postpartum women identically. The Medicaid unwinding, which reduces enrollment across all beneficiary types within a state, is precisely the type of common shock that the DDD is designed to absorb.

C. Robustness Appendix

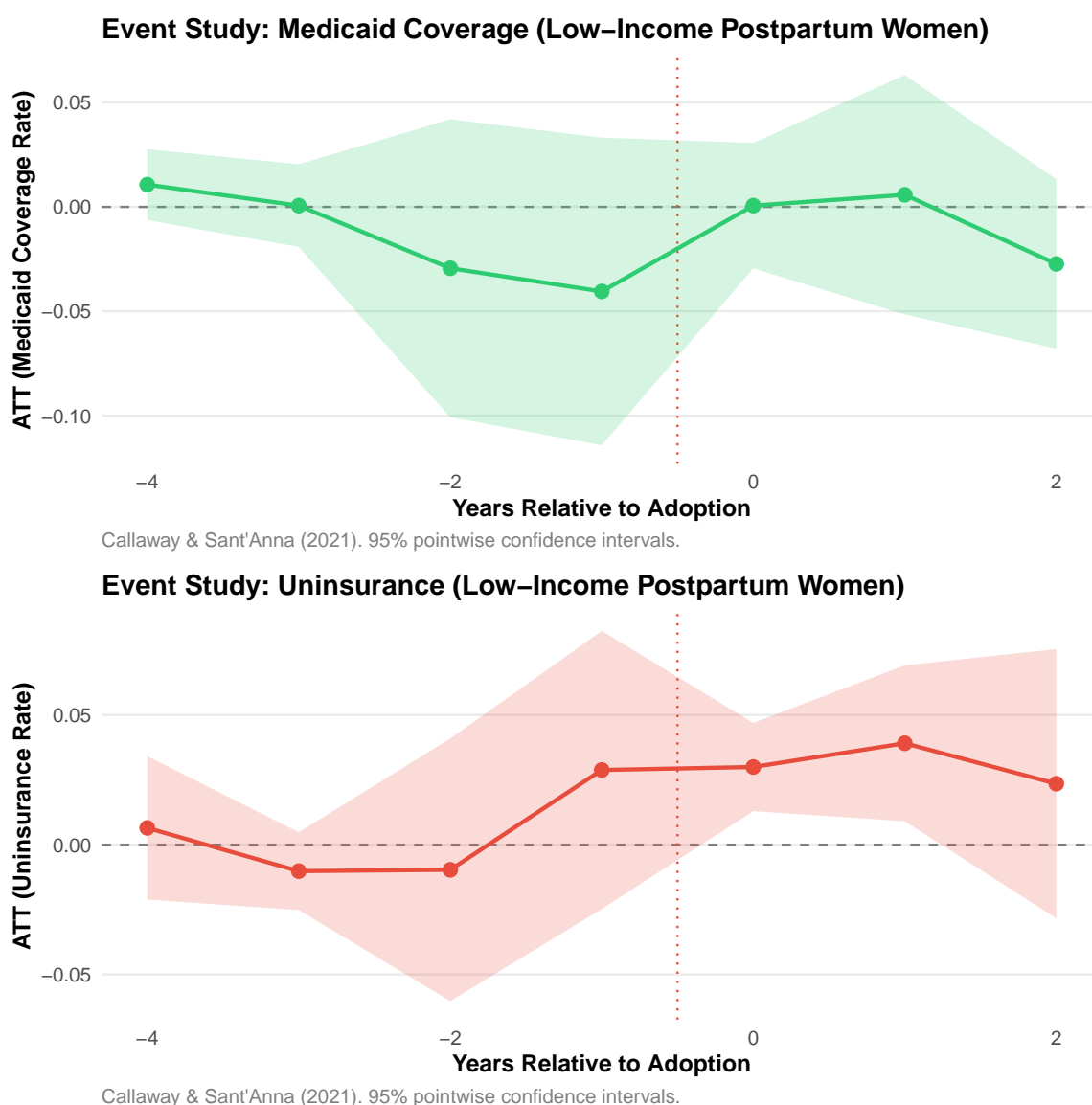## C.1 Low-Income Subgroup Event Study

Figure 10: Event-Study Estimates: Low-Income Postpartum Women (Below 200% FPL)

Notes: Callaway and Sant'Anna (2021) event-study estimates for postpartum women with income below 200% FPL. Event time $e \in \{-4, \ldots, 2\}$. Shaded areas show 95% pointwise CIs.

C.2 HonestDiD Sensitivity Details

The Rambachan-Roth relative magnitudes approach bounds the treatment effect under different assumptions about the smoothness of potential violations of parallel trends. The parameter $\bar{M}$ controls the allowed ratio of post-treatment trend deviations to the maximum pre-treatment deviation. Results are reported in Table 4 for $\bar{M} \in \{0, 0.5, 1, 1.5, 2\}$. The visualization in Figure 9 shows how the robust confidence interval expands monotonically as $\bar{M}$ increases, providing a transparent assessment of the sensitivity of the results to parallel trends violations.

C.3 Wild Cluster Bootstrap Details

Wild cluster bootstrap is implemented using Rademacher weights with 9,999 replications via the `fwildclusterboot` package. Following MacKinnon and Webb (2017), who show that the wild bootstrap is reliable even with highly unequal cluster sizes, WCB p-values are reported for the TWFE baseline, the DDD, and the post-PHE specification. The WCB p-values are broadly consistent with the standard clustered SE inference, providing additional assurance about the reliability of the results.

C.4 Permutation Inference Details

The permutation test randomly reassigns the vector of treatment timing across states while maintaining the same number of treated and control units in each permutation. For each of the 200 permutations, the full Callaway-Sant'Anna pipeline is re-run (`att_gt()` followed by `aggte()`), producing a distribution of placebo CS-DiD ATTs under the sharp null hypothesis of no effect. A parallel permutation test is conducted for the DDD CS-DiD specification on the differenced outcome (postpartum minus non-postpartum Medicaid rates). The two-sided permutation p-value is computed as $p = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}\{|\widehat{ATT}_b| \geq |\widehat{ATT}|\}$, where $B = 200$ and $\widehat{ATT}$ is the observed CS-DiD ATT (-0.50 pp for the main specification). This procedure is related to the randomization inference approach advocated by Conley and Taber (2011) and Ferman and Pinto (2021) for settings with few policy changes, and to the placebo inference in the synthetic control literature (Abadie et al., 2010). Running the full CS-DiD pipeline (rather than TWFE, as in earlier versions of this paper) ensures that the permutation distribution inherits the heterogeneity-robust properties of the Callaway-Sant'Anna estimator, avoiding the biases that TWFE introduces in staggered adoption settings.

D. Additional Tables

Table 9: State Adoption of 12-Month Medicaid Postpartum Coverage

State	Year	Mechanism	Status	State	Year	Mechanism	Status
GA	2021	Waiver	Treated	VA	2022	Waiver	Treated
IL	2021	Waiver	Treated	WA	2022	SPA	Treated
MO	2021	Waiver	Treated	WV	2022	SPA	Treated
NJ	2021	Waiver	Treated	AL	2023	SPA	Treated
CA	2022	SPA	Treated	AZ	2023	SPA	Treated
CT	2022	SPA	Treated	CO	2023	SPA	Treated
DC	2022	SPA	Treated	DE	2023	SPA	Treated
FL	2022	Waiver	Treated	MS	2023	SPA	Treated
HI	2022	SPA	Treated	MT	2023	SPA	Treated
IN	2022	SPA	Treated	NH	2023	SPA	Treated
KS	2022	SPA	Treated	NY	2023	SPA	Treated
KY	2022	SPA	Treated	OK	2023	SPA	Treated
LA	2022	SPA	Treated	RI	2023	SPA	Treated
MA	2022	SPA	Treated	SD	2023	SPA	Treated
MD	2022	SPA	Treated	VT	2023	SPA	Treated
ME	2022	SPA	Treated	WY	2023	SPA	Treated
MI	2022	SPA	Treated	AK	2024	SPA	Treated
MN	2022	SPA	Treated	NE	2024	SPA	Treated
NC	2022	SPA	Treated	NV	2024	SPA	Treated
ND	2022	SPA	Treated	TX	2024	SPA	Treated
NM	2022	SPA	Treated	UT	2024	SPA	Treated
OH	2022	SPA	Treated	IA	2025	SPA	NYT
OR	2022	SPA	Treated	ID	2025	SPA	NYT
PA	2022	SPA	Treated	AR	Never	—	Control
SC	2022	SPA	Treated	WI	Never	—	Control
TN	2022	SPA	Treated				

Notes: SPA = State Plan Amendment under ARPA Section 9812. Waiver = Section 1115 demonstration waiver. Status: "Treated" = adopted by 2024 (N=47); "NYT" = not-yet-treated in sample (N=2); "Control" = never adopted (N=2). Total clusters (states): 51. Control group for CS-DiD: NYT + Control states.