Paid Family Leave and Female Labor Market Outcomes: Evidence from the United States

Abraham Puotrer Sansah

Universidad de Alicante, Spain

asp223@alu.ua.es

September, 2026

Abstract

The United States remains one of the few high-income countries without a national paid family leave (PFL) policy, although several states have implemented their own programs. This paper examines whether state PFL policies improve maternal labor-market attachment following childbirth. Using 2001–2024 American Community Survey (ACS) microdata on women aged 25–44 who reported giving birth during the preceding 12 months, the analysis employs the Callaway and Sant'Anna (2021) difference-in-differences estimator to exploit staggered state policy adoption. PFL exposure increased sustained full-time employment by 1.49 percentage points, equivalent to a 3.86% increase relative to the weighted control-group mean of 38.60%. The estimated effect on labor force participation is positive but not statistically significant, while the estimated effect on usual weekly hours worked is close to zero and also not statistically significant. Subgroup point estimates vary by education and race/ethnicity, with a larger sustained full-time employment estimate among highly educated women and larger labor force participation and usual weekly hours estimates among women with low education. Overall, the findings support a positive short-run effect of state PFL availability on sustained full-time employment among recent mothers.

Keywords: paid family leave, labor force participation, sustained full-time employment, usual weekly hours worked, staggered difference-in-differences, United States

1. Introduction

Despite substantial evidence on the importance of paid family leave (PFL) for women's labor-market outcomes (Baum & Ruhm, 2016; Bailey et al., 2025), the United States remains one of the few high-income countries without a national PFL policy. Workers who need time away from work to care for a newborn child rely largely on the federal Family and Medical Leave Act (FMLA), which provides up to twelve weeks of unpaid, job-protected leave to eligible workers. Eligibility is limited by requirements relating to employer size, job tenure, and hours worked (U.S. Department of Labor, 2025). Moreover, access to employer-sponsored paid leave remains uneven and concentrated among higher-paid workers (Boesch, 2019).

In the absence of a national PFL program, a number of states have established paid family leave programs, beginning with California, where benefits became payable in 2004. These programs generally provide partial wage replacement during eligible periods of leave and are financed through payroll contributions from employees, employers, or both (Bipartisan Policy Center, 2025). By 2024, thirteen states and the District of Columbia had enacted mandatory programs, although not all had begun paying benefits by the end of the American Community Survey (ACS) sample period in 2024 (Bipartisan Policy Center, 2025).

Existing evidence on the labor-market consequences of PFL is mixed. Early studies of California's program found substantial increases in leave-taking and, among mothers who returned to work, modest gains in hours worked and wages, although they found no clear effect on employment itself (Rossin-Slater et al., 2013; Baum & Ruhm, 2016). By contrast, Das and Polachek (2015) find that California's program increased both the unemployment rate and the duration of unemployment. Using administrative tax records, Bailey et al. (2025) found no overall gains in employment or earnings and identified persistent negative effects among first-time mothers.

At the same time, evidence points to benefits beyond labor-market outcomes, alongside potential costs. Using a staggered-adoption design across six states, Wells et al. (2026) find that PFL increased breastfeeding duration and reduced postpartum depressive symptoms but was also associated with adverse effects on certain birth outcomes. Relatedly, Nguyen et al. (2023) found substantial increases in paid-leave use following the implementation of New York's PFL program. These findings suggest that the effects of PFL may vary according to policy design, the population examined, the time horizon considered, the outcome measured, and the empirical method employed.

Despite this growing literature, evidence on maternal labor-market outcomes remains concentrated in evaluations of individual states, particularly California. Studies examining leave-taking, labor force attachment, unemployment, and longer-run career outcomes relied on a single state as the primary source of treatment variation, as in the studies discussed above. Although Wells et al. (2026) demonstrate the value of exploiting PFL implementation across multiple states using a staggered-adoption estimator, their analysis focused primarily on maternal health rather than labor-market outcomes. Comparable multi-state evidence on maternal labor-market attachment therefore remains limited. This paper addresses this gap by examining the effect of state PFL availability on labor force participation, sustained full-time employment, and usual weekly work hours among employed women who recently gave birth across multiple adopting states.

The broader motivation for this study is the labor-market disadvantage associated with motherhood, commonly described as the "motherhood penalty" or "child penalty" (Kleven et al., 2019). Using individual-level microdata from the 2001–2024 American Community Survey (ACS), obtained through IPUMS USA (Ruggles et al., 2025), this paper examines labor-market outcomes among women aged 25–44 who reported giving birth during the preceding 12 months. Because the ACS is a repeated cross-sectional survey and does not follow the same women before and after childbirth, the analysis does not estimate the child penalty in its canonical sense. Instead, it estimates how the availability of a state PFL program affects maternal labor-market outcomes around childbirth.

The staggered introduction of state PFL programs creates substantial variation in policy exposure across states and over time, which this paper exploits for identification. However, staggered policy adoption also creates an important methodological challenge. Conventional two-way fixed effects (TWFE) estimators can produce misleading estimates when treatment timing and treatment effects vary across groups. This problem arises because already-treated units may serve as controls for later-treated units, and some comparisons may receive problematic weights (de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021).

To address these concerns, this paper employs a staggered difference-in-differences (DiD) design based on the group-time average treatment effect framework developed by Callaway and Sant'Anna

(2021). This estimator accommodates variation in treatment timing and heterogeneous treatment effects while avoiding inappropriate comparisons between earlier and later-treated states. Identification is based on variation in the timing of PFL benefit availability across states, with treated cohorts compared with appropriate not-yet-treated or never-treated states. Pooling multiple state policy implementations within this framework also makes it possible to assess whether the estimated labor-market effects extend beyond any single state program.

Importantly, the analysis measures exposure to the availability of a state PFL program rather than individual participation in that program. The ACS does not report whether a woman was eligible for, applied for, or received PFL benefits. Consequently, the estimates are interpreted as intention-to-treat effects of state-level policy availability rather than as the effects of actual benefit receipt. This distinction is important because program take-up is not universal. Bailey et al. (2025), for example, report that approximately 16% of California women giving birth used paid leave following the introduction of the state program. The effect of actually receiving PFL benefits may therefore differ from the policy availability effect estimated in this paper.

The preferred weighted estimate indicates that PFL exposure increased sustained full-time employment among recent mothers by 1.49 percentage points, equivalent to a 3.86% increase relative to the weighted control-group mean of 38.60%. The estimated effect on labor force participation is positive but not statistically significant, while the estimated effect on usual weekly hours among employed recent mothers is close to zero and also not statistically significant. Subgroup estimates indicate heterogeneous responses by education and race/ethnicity. Sustained full-time employment increases are larger among highly educated women, whereas the larger labor force participation and weekly-hours estimates occur among women with low education.

This paper makes three main contributions. First, it provides multi-state evidence on the effects of PFL on maternal labor-market outcomes, rather than relying on a single state as the primary source of treatment variation. Second, it applies a modern staggered difference-in-differences framework that accommodates variation in treatment timing and heterogeneous treatment effects across adoption cohorts and over time. Third, by extending the analysis through 2024, it incorporates more recent state PFL implementations and provides updated evidence on whether PFL strengthens women's labor-market attachment following childbirth.

The remainder of the paper proceeds as follows. Section 2 presents the institutional background. Section 3 describes the data and measures, and Section 4 outlines the empirical strategy. Section 5 presents the main results. Section 6 reports robustness checks and identification diagnostics, followed by the heterogeneity analysis in Section 7. Section 8 discusses the findings and concludes.

2. Institutional Background

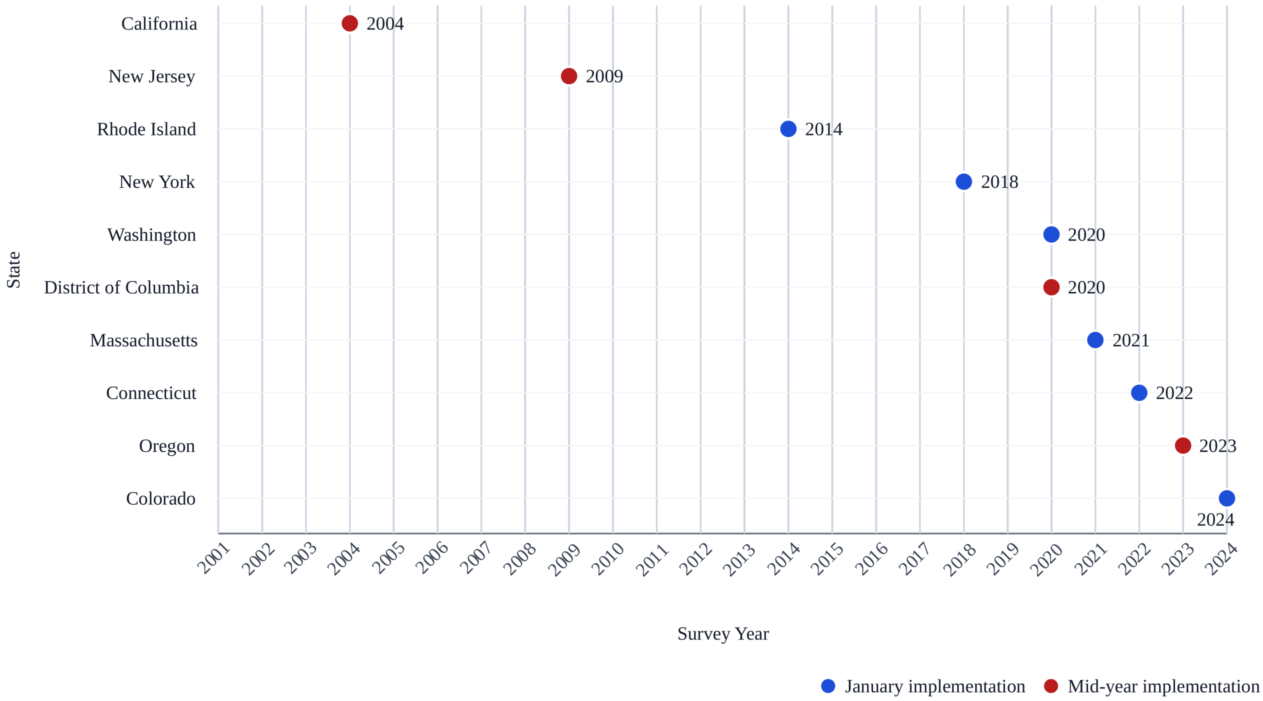
Paid leave policy in the United States developed in three broad phases (Bailey et al., 2025). The first phase followed the Pregnancy Discrimination Act of 1978, which prohibited state Temporary Disability Insurance (TDI) programs from excluding disabilities related to pregnancy and childbirth. This change extended access to partially paid maternity leave through existing TDI programs in California, Hawaii, New Jersey, New York, Rhode Island, and Puerto Rico. Stearns (2015) shows that this early TDI-based expansion produced measurable effects on maternal and infant outcomes. These pre-existing programs therefore form part of the institutional background against which subsequent PFL policies were introduced. In states that already operated TDI programs, the estimated effect of PFL should be understood as the effect of expanding the existing paid-leave system to cover family caregiving and bonding, rather than as a transition from no paid leave.

The second phase involved the expansion of unpaid, job-protected leave and culminated in the federal Family and Medical Leave Act (FMLA) of 1993. The FMLA provides eligible employees with up to 12 weeks of unpaid, job-protected leave. However, eligibility depends on firm size, job tenure, and hours worked. Consequently, FMLA coverage is not universal.

The third phase, which is the focus of this paper, began in 2004 when California's Paid Family Leave Act took effect. At its introduction, the program provided up to six weeks of partially paid leave, replacing approximately 55% of previous earnings subject to a maximum benefit. Benefits could be used to bond with a new child or care for a seriously ill family member. The program was financed through employee payroll contributions and did not impose a firm size eligibility restriction. Other states subsequently introduced their own PFL programs with different implementation dates and program characteristics.

Figure 1 presents the timing of PFL benefit implementation across the adopting jurisdictions included in the analysis. The markers indicate the calendar year in which each jurisdiction began paying benefits. Because the ACS identifies the survey year but not the precise relationship between the month of childbirth and the month of policy implementation, programs introduced partway through a calendar year are assigned to the following year as their analytic treatment cohort. The preferred treatment year is therefore the first full ACS calendar year in which benefits were available. For example, California began paying benefits in July 2004 and is assigned to the 2005 treatment cohort; New Jersey began paying benefits in July 2009 and is assigned to the 2010 cohort; the District of Columbia began paying benefits in July 2020 and is assigned to the 2021 cohort; and Oregon began paying benefits in September 2023 and is assigned to the 2024 cohort. Appendix Table A.1 reports both the actual (statutory) implementation year and the analytic treatment cohort assigned to each treated state.

Figure 1: State Paid Family Leave Implementation and Analytic Treatment Cohorts



Notes: States are shown on the vertical axis, and ACS survey years from 2001 to 2024 are shown on the horizontal axis. Markers indicate the calendar year in which each jurisdiction began paying PFL wage-replacement benefits. Blue markers identify programs implemented in January, while red markers identify programs implemented later in the calendar year. For programs implemented partway through a calendar year, the preferred analytic treatment cohort is the following year, representing the first full ACS calendar year of benefit exposure. The adopting jurisdictions are California, New Jersey, Rhode Island, New York, Washington, the District of Columbia, Massachusetts, Connecticut, Oregon, and Colorado. States that did not begin paying PFL benefits during the study period serve as never-treated controls. Hawaii is excluded because its longstanding TDI program provided paid maternity benefits before the analysis period, making it unsuitable as a never-treated control. Appendix Table A.1 reports the actual implementation and analytic treatment cohort years.

State PFL programs differ substantially in generosity and institutional design. Wage-replacement rates generally ranged from approximately 50% to 90%, while the maximum duration of benefits has varied across programs and over time. Eligibility requirements, benefit caps, financing arrangements, and job-protection provisions also differ. These institutional differences, together with variation in state labor markets and demographic composition, create the possibility that the effects of PFL vary across states, adoption cohorts, and population groups.

These policy differences could, in principle, be used to examine whether more generous or longer programs produce larger effects. However, the relatively small number of adopting jurisdictions in the sample prevents a sufficiently precise analysis of individual program characteristics. Dividing the treated jurisdictions into categories based on wage-replacement rates, benefit duration, eligibility rules, or job protection would produce small comparison groups. In addition, several program features changed over time, making it difficult to isolate the effect of any single characteristic. The Callaway and Sant'Anna (2021) group-time estimator nevertheless allows treatment effects to vary across adoption cohorts and time since implementation. This paper therefore examines heterogeneity across demographic groups, adoption cohorts, and event time, while leaving a formal analysis of heterogeneity by specific program characteristics for future research.

3. Data and Measures

3.1 Data Source and Sample

The analysis uses repeated cross-sectional microdata from the 2001–2024 American Community Survey (ACS), obtained through IPUMS USA (Ruggles et al., 2025), at the individual level. Because the ACS is not a panel, effects are identified from changes across state-year cohorts as states introduced PFL at different times, following the repeated cross-sectional framework of Callaway and Sant'Anna (2021). The analysis includes all U.S. states except Hawaii. Hawaii is excluded because its pre-existing Temporary Disability Insurance program predates the analysis period, making it unsuitable as either a newly treated or never-treated state. The analytic sample consists of women aged 25–44 who reported giving birth during the preceding 12 months.

Because the ACS does not report the exact month of birth, some women observed in a state's initial implementation year could be classified as exposed even if they gave birth before PFL benefits became available. Under nondifferential measurement error, such timing misclassification would tend to attenuate the estimated effects toward zero, although the direction of bias cannot be guaranteed if birth timing is systematically related to labor-market outcomes. The preferred treatment cohort is therefore defined as the first full ACS calendar year of benefit exposure. As a timing robustness check, the analysis excludes the partially exposed implementation state-years for California (2004), New Jersey (2009), the District of Columbia (2020), and Oregon (2023), while retaining the preferred first-full-year treatment definition. The corresponding estimates are discussed in Section 6.3 and reported in Appendix Table A.7.

3.2 Outcome Measures and Covariates

The two primary outcomes are labor force participation and sustained full-time employment. Labor force participation equals one if a woman is employed or unemployed and in the labor force, and zero if she is not in the labor force. It is defined for the full analytic sample.

Sustained full-time employment equals one if a woman is employed, usually works at least 35 hours per week, and worked approximately 40 or more weeks during the preceding year. It equals zero for women who do not satisfy these conditions. This outcome is also defined for the full analytic sample.

Usual weekly hours worked is examined as a secondary outcome and is measured using the number of hours an employed respondent usually works per week. Because this outcome is observed only among employed women, it captures the intensive margin of labor supply rather than employment participation.

The preferred specification controls for age, age squared, educational attainment, and race/ethnicity. Educational attainment is grouped into low, medium, and high education. Race/ethnicity is classified into four mutually exclusive categories: White non-Hispanic, Black non-Hispanic, Hispanic, and other non-Hispanic. The race/ethnicity measure combines the ACS race and Hispanic-origin variables so that women of Hispanic origin are classified as Hispanic regardless of race, while the remaining categories refer to non-Hispanic respondents.

Occupation and industry are excluded because PFL may affect whether mothers retain their pre-birth jobs or change jobs to accommodate caregiving responsibilities. By reducing the income loss during leave, PFL may help preserve an existing employment relationship, a mechanism consistent with the job-continuity explanation discussed by Baum and Ruhm (2016). Retaining or changing employers may also affect the occupation or industry in which a mother works. Occupation and industry observed after policy implementation are therefore potential post-treatment variables; controlling for them could absorb part of the policy effect the analysis seeks to estimate. Appendix Table A.2 reports the ACS/IPUMS variables and coding rules used to construct the sample, outcomes, treatment variables, and covariates.

3.3 Pre-reform Descriptive Characteristics

Table 1 compares the characteristics and labor-market outcomes of women in future-treated states before PFL implementation with those of women in states that remained untreated throughout the sample period. Restricting the comparison to pre-reform observations ensures that women in future-treated states are observed before PFL benefits became available. Appendix Table A.3 provides the corresponding covariate-balance and common-support diagnostics used to assess whether the treated and comparison groups overlap on the observed controls.

Table 1 reveals several pre-reform differences between future-treated and never-treated states. Women in future-treated states were, on average, older and more likely to be married, and the two groups also differed in their educational and racial/ethnic composition. Pre-reform labor-market outcomes were lower in future-treated states. Labor force participation was 60.84% in future-treated states compared with 64.50% in never-treated states; the reported difference is −3.65 percentage points before rounding the displayed means. Sustained full-time employment was 33.62% compared with 39.72%, a difference of −6.10 percentage points. Usual weekly hours among employed women were approximately 0.71 hours lower.

These differences show that the two groups were not identical in their observed pre-reform levels. However, differences in baseline levels do not necessarily invalidate the difference-in-differences design, which requires comparable trends in untreated potential outcomes rather than identical levels. The event study estimates and joint pre-trend tests presented later provide a more direct assessment of this assumption.

Table 1: Pre-reform characteristics and outcomes: Future-treated versus never-treated states

Characteristic	Never-treated	Future-treated	Difference (Future treated −Never treated)	SE	p-value
Panel A. Sample characteristics					
Respondent age, mean	31.51	32.21	0.70	0.14	0.0001
Age group, %					
Age 25–34	74.52	68.99	-5.53	1.02	0.0001
Age 35–44	25.48	31.01	5.53	1.02	0.0001
Educational attainment, %					
Low education	8.45	9.41	0.96	1.55	0.5391
Medium education	52.33	47.56	-4.78	1.40	0.0013
High education	39.21	43.03	3.82	2.49	0.1311
Race / ethnicity, %					
White, non-Hispanic	61.51	58.45	-3.06	5.31	0.5671
Black, non-Hispanic	13.78	8.56	-5.22	2.28	0.0266
Hispanic	16.83	20.88	4.05	5.11	0.4312
Other, non-Hispanic	7.89	12.11	4.22	1.09	0.0003
Married, %	74.62	78.13	3.51	1.33	0.0111
Panel B. Labor-market outcomes					
Labor force participation, %	64.50	60.84	-3.65	1.98	0.0713
Sustained full-time employment, %	39.72	33.62	-6.10	1.76	0.0011
Usual weekly hours among employed, mean	36.81	36.10	-0.71	0.30	0.0206
Observations	403,588	81,290			

Notes: The sample includes women aged 25–44 in the 2001–2024 ACS who reported giving birth in the preceding 12 months. Future-treated observations are women in states that later adopted PFL, observed before the first full year of benefit availability; never-treated observations are women in states without PFL during the sample period. Means are weighted using ACS person weights. The Difference column reports future-treated minus never-treated estimates from weighted regressions, with state-clustered standard errors. Binary variables are reported in percentages and percentage points; age and weekly hours are reported in years and hours. The observations row reports pre-reform sample sizes; weekly hours are based on 247,266 never-treated and 48,405 future-treated employed women.

4. Empirical Strategy

4.1 Treatment Timing and Policy Exposure

The empirical design exploits variation in the timing of state PFL benefit implementation. Let G_s denote the analytic treatment cohort assigned to state s, where $G_s = 0$ for states that did not provide PFL benefits during the sample period. A woman observed in state s and year t is classified as exposed to PFL when the state has reached its assigned treatment year:

$$D_{st} = 1(G_s > 0 \text{ and } t \geq G_s) \tag{1}$$

For programs introduced partway through a calendar year, the analytic treatment cohort is defined as the first full ACS calendar year in which benefits were available. Figure 1 presents the implementation timing, while Appendix Table A.1 reports the actual implementation year and the corresponding analytic treatment year for each jurisdiction. States that did not adopt PFL during the analysis period serve as never-treated comparisons. States that adopted PFL later contribute to the not-yet-treated comparison group before reaching their assigned treatment year. Once a state reaches its treatment year, it remains treated for the remainder of the sample period.

Treatment is measured at the state-year level and represents PFL policy availability rather than individual participation. Because the ACS does not observe individual eligibility, leave-taking, applications, or benefit receipt, the estimates are interpreted as intent-to-treat effects of state PFL availability.

4.2 Group-Time Treatment Effects and Identification

The analysis uses the group-time average treatment effect framework developed by Callaway and Sant'Anna (2021). Let $Y_{ist}(1)$ denote the potential labor-market outcome for woman *i* in state *s* and year *t* under exposure to an active PFL policy, and let $Y_{ist}(0)$ denote the corresponding potential outcome in the absence of exposure. For states first treated in year *g*, the group-time average treatment effect on the treated is:

$$\boldsymbol{ATT(g,t) = \mathbb{E}[Y_{ist}(1) - Y_{ist}(0) \mid G_s = g]\,, \qquad t \geq g} \tag{2}$$

The untreated potential outcome$Y_{ist}(0)$ is not observed for women in treated states after PFL implementation. The Callaway-Sant'Anna estimator constructs this counterfactual by comparing the change in outcomes for a treated cohort with the corresponding change among states that are either not yet treated or never treated at time *t*. Already-treated states are not used as controls for later-treated states.

The central identifying assumption is the conditional parallel trend. Conditional on observed covariatesX_{ist}, the average change in untreated potential outcomes for cohort *g* is assumed to equal the corresponding change for the valid comparison group. For a pre-treatment reference period $g-1$, this condition can be expressed as:

$$\mathbb{E}[\boldsymbol{Y}_{ist}(0) - \boldsymbol{Y}_{is,g-1}(0) \mid \boldsymbol{G}_s = \boldsymbol{g}, \boldsymbol{X}_{ist}] = \mathbb{E}[\boldsymbol{Y}_{ist}(0) - \boldsymbol{Y}_{is,g-1}(0) \mid \boldsymbol{C}_t = 1, \boldsymbol{X}_{ist}] \tag{3}$$

where C_t identifies states that provide valid comparisons in year t, consisting of never-treated and not-yet-treated states. The covariate vector contains age, age squared, educational attainment, and race/ethnicity. These characteristics are strongly associated with maternal labor-market outcomes and are plausibly determined before contemporaneous PFL exposure.

Conditional parallel trends do not require future-treated and never-treated states to have identical pre-reform outcome levels. The descriptive differences reported in Table 1 are therefore not, by themselves, evidence against the research design. The relevant condition is that, after conditioning on the included covariates, labor-market outcomes in the treated and comparison states would have followed comparable trends in the absence of PFL. The paper evaluates this assumption using the event study evidence in Figures 2-4, the restricted joint pre-trend tests in Appendix Table A.4, and the event specific lead estimates in Appendix Table A.5.

The design also relies on three additional conditions. First, treatment is absorbing: once a state begins providing PFL benefits, it remains treated in subsequent years. Second, there is no anticipation: outcomes before the assigned treatment year are not affected by future PFL availability. Third, the design assumes that PFL implementation does not induce differential compositional changes between treated and comparison states. Because interstate migration could affect this condition, migration-related composition is examined as an additional diagnostic.

4.3 Estimation, Aggregation, and Event Time Effects

The preferred specification uses the doubly robust inverse-probability-weighted estimator (Sant'Anna & Zhao, 2020) developed for staggered-adoption settings by Callaway and Sant'Anna (2021). The estimator combines an outcome-regression model with a propensity-score weighting model. Subject to the identifying assumptions, it remains consistent if either the outcome-regression model or the propensity-score model is correctly specified.

The estimator is appropriate in this setting because states adopted PFL at different times and the effects may vary across adoption cohorts and periods since implementation. Conventional two-way fixed-effects estimators can produce misleading results under such heterogeneity because already-treated states may be used as controls for later-treated states and some treatment comparisons may receive problematic weights. The Callaway-Sant'Anna framework avoids these inappropriate comparisons by estimating separate treatment effects for each eligible cohort and calendar year.

The overall average treatment effect is obtained by aggregating the estimated group-time effects:

$$\boldsymbol{ATT^o = \sum_{g}\sum_{t\geq g} w_{g,t} ATT(g,t), \qquad \sum_{g}\sum_{t\geq g} w_{g,t} = 1} \tag{4}$$

where $W_{g,t}$ denotes the weight assigned to cohort g in calendar year t. The overall ATT is therefore a weighted average of the cohort-time effects that can be identified within the sample. It summarizes the average effect of PFL exposure among treated state-year cohorts rather than imposing a common effect across all states and periods. Dynamic effects are examined by reorganizing the group-time estimates according to event time. Event time is defined as:

$$\boldsymbol{e = t - g} \tag{5}$$

where negative values denote years before treatment, e=0 is the treatment year, and positive values denote years after treatment. Event-time effects are obtained by aggregating the group-time effects at each relative period:

$$\boldsymbol{\theta_e = \sum_{g:g+e\in\mathcal{T}} w_{g,e} ATT(g, g+e)} \tag{6}$$

The event-study window covers five years before and five years after treatment, e=−5,…,+5. The pre-treatment coefficients are used to assess whether treated and comparison states followed similar trends before PFL implementation. A joint Wald test is therefore used to assess whether the five pre-treatment coefficients are jointly equal to zero:

$$\boldsymbol{H_0 : \theta_{-5} = \theta_{-4} = \theta_{-3} = \theta_{-2} = \theta_{-1} = 0} \tag{7}$$

The test is restricted to the five pre-treatment event times shown in the event-study window. It uses the estimated pre-treatment coefficients and their covariance matrix to test whether they are jointly equal to zero. Rejection of the null indicates evidence of differential pre-treatment trends over this period.

4.4 Weighting and Statistical Inference

The preferred estimates use ACS person weights, while unweighted estimates are reported as a robustness check. Standard errors are clustered at the state level because PFL treatment varies at the state level, and statistical inference is based on 999 bootstrap replications.

4.5 Identification Diagnostics and Robustness Checks

The empirical design includes several robustness analyses and identification diagnostics. The robustness analyses assess whether the estimated effects are sensitive to influential treatment cohorts, unusual survey years, survey weighting, and treatment-timing decisions. One specification excludes California because it adopted PFL near the beginning of the sample period and therefore contributes disproportionately to longer post-treatment horizons. A second excludes the 2020 ACS because the COVID-19 pandemic affected both ACS data collection and labor-market conditions.

Unweighted estimates are also reported to assess sensitivity to the use of ACS person weights. Because weighted and unweighted specifications may estimate different population averages when treatment effects are heterogeneous (Solon et al., 2015), differences between them are interpreted as evidence that the estimated effects depend on the weighting scheme and population represented, rather than as a mechanical test of model validity.

Alternative treatment-timing specifications address uncertainty arising from the absence of exact childbirth dates. The preferred specification assigns treatment beginning in the first full ACS calendar year of PFL benefit exposure. An additional specification excludes observations from the partial implementation years in California (2004), New Jersey (2009), the District of Columbia (2020), and Oregon (2023), while retaining the preferred first-full-year treatment definition. These checks assess whether the estimates are sensitive to the classification of women observed around the policy implementation date.

The identification diagnostics include placebo adoption-date tests, joint tests of the pre-treatment coefficients over event times -5 to -1, and a migration-composition diagnostic. The migration diagnostic uses whether a woman was born in her current state as an indirect proxy for migration-related compositional changes. Together, these analyses assess the plausibility of the identifying assumptions and clarify the limitations of the estimates.

4.6 Heterogeneity Analysis

The analysis examines whether the effects of PFL differ by educational attainment and race/ethnicity. Educational attainment is divided into low, medium, and high education groups, while race/ethnicity is classified into four mutually exclusive categories: White non-Hispanic, Black non-Hispanic, Hispanic, and other non-Hispanic.

The Callaway–Sant'Anna estimator is re-estimated separately for each subgroup using the same treatment definition, comparison group, weighting procedure, and state-clustered inference as in the preferred specification. Education controls are omitted from the education-specific models, while race/ethnicity controls are omitted from the race/ethnicity-specific models. Supplementary pooled interaction tests are used to assess whether estimated effects differ across categories. Because these pooled tests are not direct comparisons of the separately estimated subgroup-specific Callaway–Sant'Anna effects, they are interpreted as supplementary evidence of heterogeneity.

5. Results

5.1 Main ATT Estimates

Table 2 reports the aggregated group-time average treatment effects of PFL exposure on the three labor-market outcomes. For interpretation, the table also reports descriptive control-group means, calculated using ACS person weights across never-treated and not-yet-treated observations in the analysis sample, with each observation counted once. These means provide a benchmark for interpreting the magnitude of the estimates.

PFL exposure increased sustained full-time employment among recent mothers by an estimated 1.49 percentage points. This corresponds to a 3.86% increase relative to the weighted descriptive control-group mean of 38.60%. The result indicates that PFL availability increased sustained full-time employment among women who recently gave birth.

The estimated effect on labor force participation is 1.51 percentage points, equivalent to 2.37% of the weighted descriptive control-group mean of 63.83%. However, the estimate is not statistically significant.

For hours worked, the estimated effect is 0.0180 hours, equivalent to approximately 0.05% of the descriptive control-group mean of 36.68 hours. The estimate is close to zero and not statistically significant, providing no clear evidence that PFL exposure changes usual weekly hours among employed recent mothers. Because usual weekly hours are measured only among employed women, this outcome captures the intensive margin of labor supply.

The sustained full-time employment result is consistent with shorter-run evidence that PFL can support maternal labor-market attachment (Rossin-Slater et al., 2013; Baum and Ruhm, 2016). It is more favorable than Bailey et al. (2025), who report negative longer-run effects on cumulative employment and earnings among first-time mothers in California. The findings are not necessarily contradictory because the studies examine different populations, outcomes, time horizons, and treatment parameters. Bailey et al. focus on first-time mothers in California using administrative data, whereas this paper examines women aged 25–44 across multiple adopting states using repeated ACS cross-sections.

Table 2: Callaway-Sant'Anna ATT: Paid Family Leave Effects on Sustained Full-Time Employment, Labor Force Participation, and Usual weekly hours worked

	Sustained full-time employment	Labor force participation	Usual weekly hours worked
ATT	0.0149***	0.0151	0.0180
	(0.0046)	(0.0092)	(0.2360)
Control mean	0.3860	0.6383	36.6827
Observations	586,649	586,649	357,972

Notes: Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimates and standard errors are reported to four decimal places. Sustained full-time employment and labor force participation are reported in decimal units; multiplying a coefficient by 100 gives the percentage-point effect. Usual weekly hours worked is reported in hours among employed recent mothers with valid UHRSWORK values. All three models use the Callaway and Sant'Anna group-time difference-in-differences estimator with ACS person weights and standard errors clustered by state. The models control for age, age squared, education category, and race/ethnicity. Control means are descriptive ACS-person-weighted means for observations with $Gs = 0$ or $t < Gs$; each observation is counted once. Binary-outcome means are reported as proportions. These benchmarks are not ATT-weighted counterfactual means. The overall ATT should be read as an aggregation of cohort-time effects across treated states and survey years, rather than as the effect of any single state program.

5.2 Placebo Tests

To assess whether changes resembling the main PFL effects were already present before actual policy implementation, placebo lead tests assign treated states hypothetical adoption dates before their actual treatment dates and exclude observations from actual implementation onward. A statistically significant placebo estimate would indicate possible pre-existing trends or anticipation, thereby weakening the causal interpretation of the corresponding main estimate.

Table 3 reports three and five-year placebo leads. Using two horizons reduces reliance on a single hypothetical date (Roth et al., 2023). The three-year design retains California by assigning it a 2002 placebo date, leaving 2001 as its pre-placebo year. California is excluded from the five-year design because its hypothetical 2000 adoption date precedes the ACS sample. The five-year lead also corresponds to the longest pre-treatment period in the main event-study window. All specifications otherwise follow the estimator, weighting, controls, and state-clustered inference used in Table 2.

Table 3: Callaway-Sant'Anna ATT Placebo Estimates Using Hypothetical Pre-Treatment Adoption Dates

Placebo design	N	Sustained full time employment	Labor force participation	N, hours	Usual weekly hours worked
3-year lead	464,865	0.0054 (0.0076)	0.0247*** (0.0063)	282,186	0.0140 (0.2430)
5-year lead, excluding California	460,298	0.0021 (0.0069)	0.0016 (0.0072)	280,014	0.0910 (0.2370)

Notes: Binary-outcome estimates are reported in decimal units, while weekly hours are reported in hours. State-clustered standard errors are in parentheses. * p<0.10, ** p<0.05, and *** p<0.01. All models use the doubly robust Callaway-Sant'Anna estimator, ACS person weights, and controls for age, age squared, education, and race/ethnicity.

For sustained full-time employment, neither placebo estimate is statistically significant, providing support for the positive main ATT. For labor force participation, the three-year placebo estimate is positive and statistically significant at 2.47 percentage points, indicating some pre-policy differences, while the five-year placebo estimate is small and statistically insignificant. The weekly-hours placebo estimates are also statistically insignificant.

5.3 Event-Study Evidence and Pre-Treatment Trends

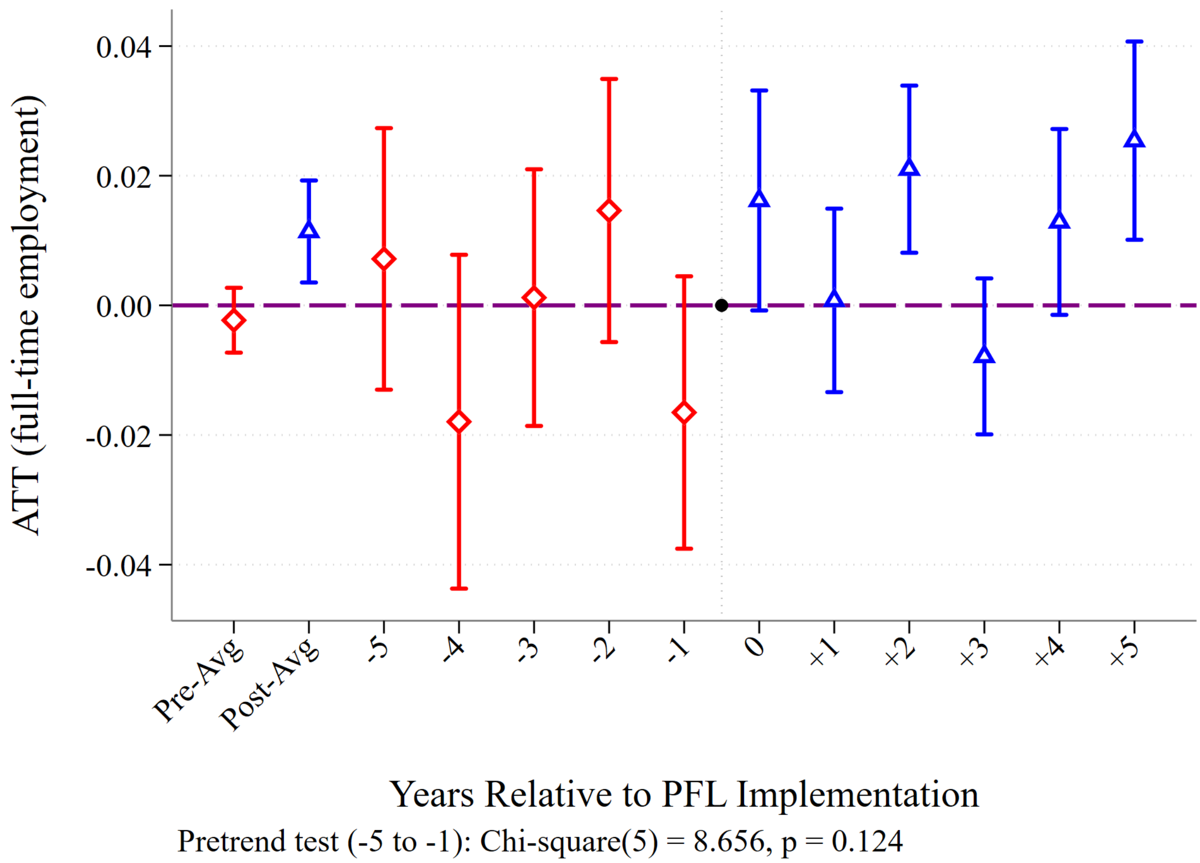
Figures 2–4 present event-study estimates for sustained full-time employment, labor force participation, and usual weekly hours worked. The pre-treatment coefficients over event times −5 to −1 are used to assess whether treated and comparison states followed similar trends before PFL implementation. The joint Wald test described in Section 4.3 evaluates whether these five pre-treatment coefficients are jointly equal to zero. Appendix Tables A.4 and A.5 report the joint tests and the underlying lead estimates.

Figure 2 presents the event-study estimates for sustained full-time employment. All five pre-treatment coefficients are statistically indistinguishable from zero, and the joint test does not reject the null that they are jointly equal to zero ($\chi2$ (5) =8.656, p=0.124). The pre-treatment evidence is therefore consistent with parallel trends for sustained full-time employment over the observed five-year window.

Figure 3 presents the event-study estimates for labor force participation. Most pre-treatment coefficients are statistically indistinguishable from zero, but the estimate at event time −3 is positive and statistically significant. The joint test rejects the null that the five pre-treatment coefficients are jointly equal to zero ($\chi2$ (5) =38.441, p<0.001), indicating differential pre-policy trends in labor force participation.

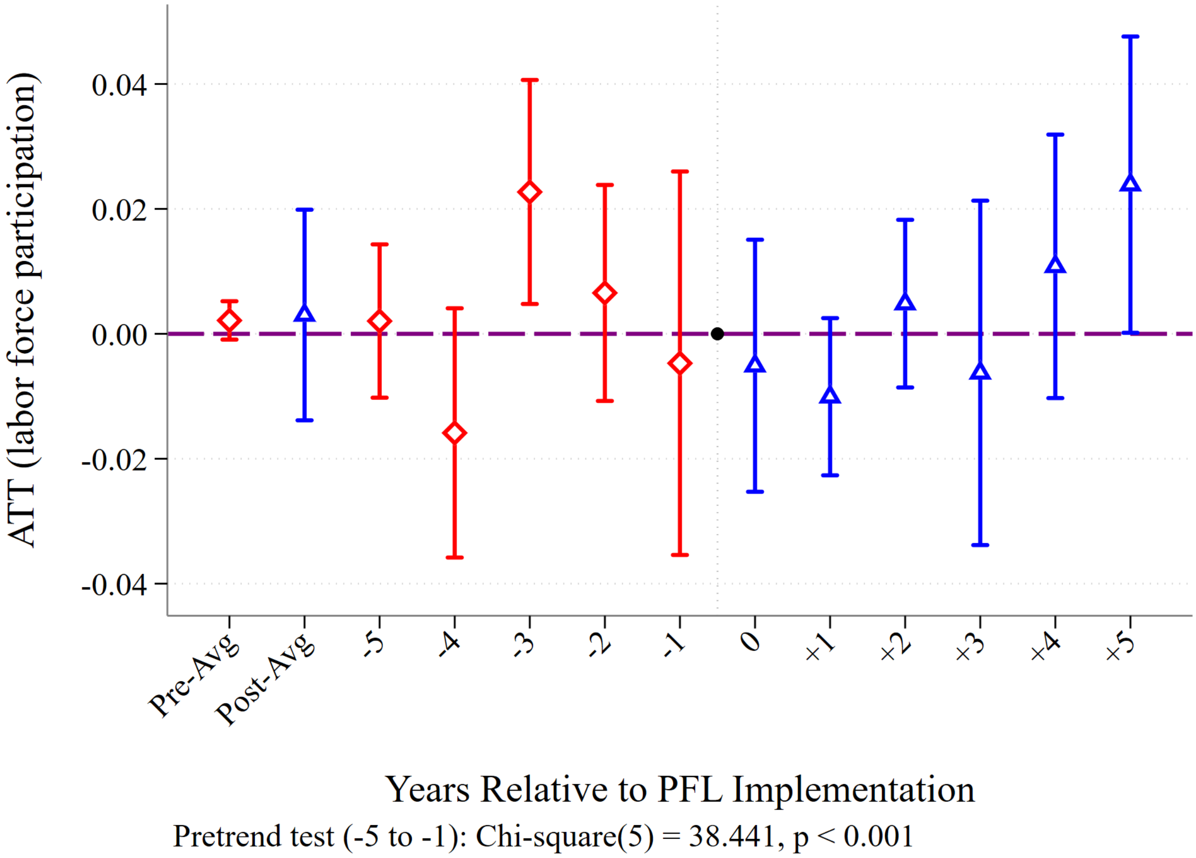
Figure 4 presents the event-study estimates for usual weekly hours worked. All five pre-treatment coefficients are statistically indistinguishable from zero, and the joint test fails to reject the null that they are jointly equal to zero at the 5 percent significance level, $\chi^2(5) = 10.281$, $p = 0.068$. The pre-treatment evidence is therefore consistent with parallel trends for usual weekly hours worked over the observed five-year window.

.Figure 2: Event Study Estimates for Sustained Full-Time Employment



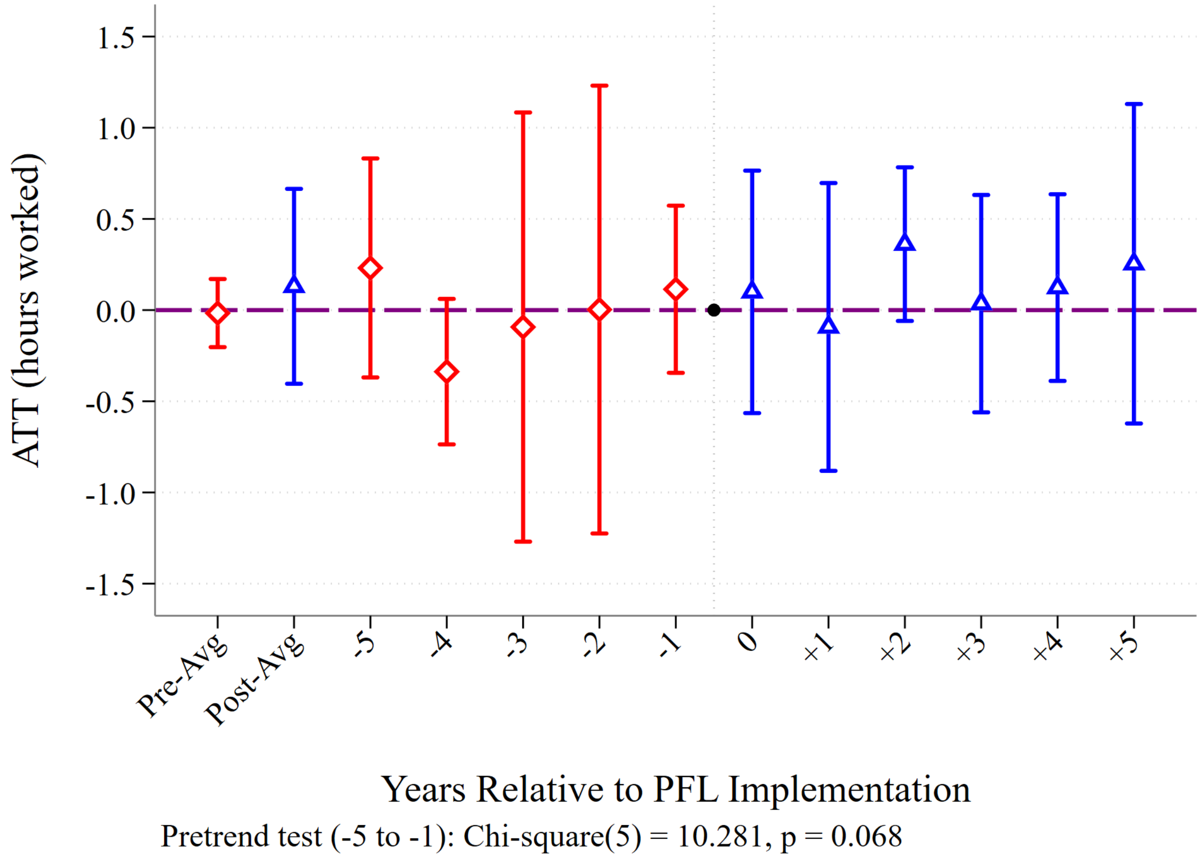
Notes: Red diamonds and blue triangles denote pre- and post-treatment estimates, respectively. The red circle and blue square report the average pre- and post-treatment estimates. Vertical bars show 95 percent confidence intervals, and the dashed horizontal line marks zero.

Figure 3: Event Study Estimates for Labor Force Participation



Notes: Red diamonds and blue triangles denote pre- and post-treatment estimates, respectively. The red circle and blue square report the average pre- and post-treatment estimates. Vertical bars show 95 percent confidence intervals, and the dashed horizontal line marks zero.

Figure 4: Event Study Estimates for Usual Weekly Hours Worked



Notes: Red diamonds and blue triangles denote pre- and post-treatment estimates, respectively. The red circle and blue square report the average pre- and post-treatment estimates. Vertical bars show 95 percent confidence intervals, and the dashed horizontal line marks zero.

Overall, the joint pre-treatment tests fail to reject the null hypothesis that the pre-treatment coefficients are jointly equal to zero for sustained full-time employment and usual weekly hours worked at the 5 percent significance level, providing support for the parallel-trends assumption for these outcomes. In contrast, the joint test rejects the null for labor force participation, indicating evidence of differential pre-treatment trends for this outcome. Therefore, the estimated effect on labor force participation should be interpreted cautiously and cannot be given the same causal interpretation as the sustained full-time employment estimate.

6. Robustness and Identification Diagnostics

6.1 Sample-Exclusion Sensitivity Analysis

Table 4 reports two sample-exclusion sensitivity checks: excluding California and excluding the 2020 ACS wave. In both specifications, the sustained full-time employment estimate remains positive. The estimates are equivalent to 2.67% and 4.09% of the respective descriptive control-group means of 38.95% and 38.36% (1.04 and 1.57 percentage points). These benchmarks apply the control-mean definition in Section 5.1 to each restricted analysis sample. The labor force participation estimate is small when California is excluded and positive when the 2020 ACS wave is excluded. The usual-weekly-hours estimates remain close to zero. Overall, the sample-exclusion checks are consistent with the positive sustained full-time employment result.

Table 4: Callaway-Sant'Anna ATT Estimates: Sample-Exclusion Sensitivity Analysis

	Sustained full-time employment	Labor force participation	Usual weekly hours worked
Exclude California	0.0104**	0.0037	0.1610
	(0.0047)	(0.0113)	(0.2580)
Observations	513,317	513,317	316,305
Exclude 2020 ACS	0.0157***	0.0163*	0.0120
	(0.0047)	(0.0097)	(0.2420)
Observations	562,577	562,577	341,888

Notes: Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Estimates and standard errors are reported to four decimal places. Sustained full-time employment and labor force participation are reported in decimal units; usual weekly hours worked is reported in hours among employed recent mothers with valid UHRSWORK values. The estimator, comparison group, weights, controls, and clustering are the same as in Table 2.

6.2 Sensitivity to ACS Person Weighting

Section 4.4 explains that weighted and unweighted specifications may estimate different population averages when treatment effects are heterogeneous. Table 5 therefore compares the preferred ACS-weighted specification with an otherwise identical unweighted specification for labor force participation, sustained full-time employment, and usual weekly hours. The estimator, comparison groups, sample, treatment timing, covariates, and state clustered inference are held constant, so the only difference is the use of ACS person weights.

The sustained full-time employment estimate remains positive and statistically significant in both the weighted and unweighted specifications. The estimate for usual weekly hours remains close to zero in both cases. The labor force participation estimate is larger in the unweighted specification, indicating some sensitivity to the use of ACS person weights. The weighted specification remains preferred because it represents the ACS target population.

Table 5: Callaway-Sant'Anna ATT: Sensitivity to ACS Person Weighting

Specification	Sustained full-time employment	Labor force participation	Usual weekly hours worked
Weighted	0.0149***	0.0151	0.0180
	(0.0046)	(0.0092)	(0.2360)
Unweighted	0.0225***	0.0307**	0.0970
	(0.0061)	(0.0124)	(0.2360)
Observations	586,649	586,649	357,972

Notes: Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Estimates and standard errors are reported to four decimal places. Sustained full-time employment and labor force participation are reported in decimal units; usual weekly hours worked is reported in hours among employed recent mothers with valid UHRSWORK values. All models use the doubly robust Callaway-Sant'Anna estimator and control for age, age squared, education, and race/ethnicity. The weighted models apply ACS person weights.

6.3 Additional Identification and Validity Checks

Appendix Table A.6 reports the estimated PFL effect on whether a U.S.-born woman resides in her state of birth. The estimate is equivalent to 4.44% of the weighted descriptive control-group mean of 63.25% (2.81 percentage points) and is not statistically significant. It provides no clear evidence of a change in birthplace–residence composition. However, residence in one's birth state is only an indirect proxy for interstate migration.

Appendix Table A.7 excludes state-years in which PFL benefits were available for only part of the calendar year. The sustained full-time employment estimate remains positive but is no longer statistically significant: 1.07 percentage points, equivalent to 2.77% of the restricted analysis sample's descriptive control-group mean of 38.70%. The labor force participation estimate is −1.06 percentage points, equivalent to −1.66% of its control-group mean of 63.90%. The weekly-hours estimate is 0.377 hours, or 1.03% of its control-group mean of 36.69 hours. These results indicate sensitivity to the treatment of partially exposed implementation years.

7. Heterogeneity Analysis by Educational Attainment and Race/Ethnicity

Table 6 presents subgroup-specific estimates by educational attainment and race/ethnicity. Low education includes women with less than a high-school education, medium education includes women who completed grade 12 or up to two years of college, and high education includes women with four or more years of college. Race/ethnicity is classified into four mutually exclusive groups: White non-Hispanic, Black non-Hispanic, Hispanic, and other non-Hispanic.

For sustained full-time employment, the estimated effect for highly educated women is 2.22 percentage points and is statistically significant at the 5% level. Relative to their weighted descriptive control-group mean of 50.35%, this represents a 4.41% increase. The estimates for women with low and medium education are not statistically significant. Across race/ethnicity groups, the estimate for White non-Hispanic women is 2.44 percentage points and is statistically significant at the 1% level, equivalent to 6.01% of their weighted descriptive control-group mean of 40.58%. The corresponding estimate for Black non-Hispanic women is also positive but not statistically significant.

For labor force participation, the estimated effect for women with low education is 6.44 percentage points and is statistically significant at the 5% level. Relative to their weighted descriptive control-group mean of 38.26%, this represents a 16.83% increase. The estimates for women with medium and high education are not statistically significant. Across race/ethnicity groups, the estimate for White non-Hispanic women is 1.32 percentage points and is statistically significant at the 5% level, equivalent to 2.01% of their weighted descriptive control-group mean of 65.57%. The Black non-Hispanic point estimate is positive but not statistically significant.

For usual weekly hours, the estimated effect for employed women with low education is 2.18 hours per week and is statistically significant at the 5% level. Relative to their weighted descriptive control-group mean of 34.55 hours, this represents a 6.31% increase. The estimates for women with medium and high education are close to zero and not statistically significant. Across race/ethnicity groups, the estimate for employed Black non-Hispanic women is −1.43 hours per week, equivalent to −3.82% of their weighted descriptive control-group mean of 37.43 hours, and is statistically significant at the 10% level. The estimates for the remaining race/ethnicity groups are not statistically significant.

Table 6: Callaway–Sant'Anna Subgroup ATT Estimates by Educational Attainment and Race/Ethnicity

	Sustained full-time employment	Labor force participation	Usual weekly hours worked
Panel A: Education			
Low education	-0.0088	0.0644**	2.1810**
	(0.0152)	(0.0262)	(0.8560)
Observations	42,605	42,605	12,619
Medium education	0.0127	-0.0017	-0.0500
	(0.0097)	(0.0128)	(0.4060)
Observations	281,626	281,626	155,732
High education	0.0222**	0.0143	-0.0670
	(0.0090)	(0.0120)	(0.2490)
Observations	262,418	262,418	189,620
Panel B: Race/ethnicity			
White non-Hispanic	0.0244***	0.0132**	0.1290
	(0.0078)	(0.0054)	(0.3120)
Observations	368,567	368,567	236,266
Black non-Hispanic	0.0209	0.0322	-1.4280*
	(0.0181)	(0.0239)	(0.7670)
Observations	50,834	50,834	31,991
Hispanic	0.0079	0.0080	0.3600
	(0.0110)	(0.0157)	(0.5510)
Observations	104,312	104,312	53,115
Other non-Hispanic	0.0043	0.0159	-0.3209
	(0.0151)	(0.0196)	(0.6788)
Observations	62,936	62,936	36,600

Notes: Standard errors are reported in parentheses. One, two, and three asterisks indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Estimates and standard errors are reported to four decimal places. Sustained full-time employment and labor force participation are reported in decimal units. Usual weekly hours worked is reported in hours among employed recent mothers with valid UHRSWORK values. Models stratified by education omit education controls. Models stratified by race/ethnicity omit race/ethnicity controls but retain education controls. All specifications use ACS person weights and state-clustered standard errors.

One possible interpretation of the education-specific pattern is that PFL may support employment continuity differently across workers with different labor-market circumstances. The relatively larger sustained full-time employment estimate among highly educated women may reflect stronger attachment to established full-time jobs, whereas the larger labor-force-participation and usual-hours estimates among women with low education may reflect adjustment along different employment margins. However, the ACS does not observe individual leave-taking, benefit receipt, job protection, or employer characteristics, so these mechanisms cannot be identified directly.

Overall, the subgroup estimates provide suggestive evidence of differences in the pattern and magnitude of labor-market responses to PFL across education and race/ethnicity groups. The sustained full-time employment estimate is more pronounced among highly educated women, while the labor-force-participation and usual weekly hours estimates are relatively larger among women with low education. Estimates also vary across racial/ethnic groups, although differences in sample size and precision warrant caution in interpreting these patterns. White non-Hispanic women constitute the largest race/ethnicity subgroup, representing 62.83% of the analytic sample.

8. Discussion and Conclusion

This paper provides multi-state evidence on the short-run effects of state paid family leave (PFL) policies on maternal labor-market outcomes following childbirth. Under the identifying assumptions, the preferred weighted estimate indicates that PFL exposure increased sustained full-time employment among recent mothers by 1.49 percentage points. The estimate is statistically significant at the 1% level. Relative to the weighted descriptive control-group mean of 38.60%, calculated using never-treated and not-yet-treated observations, this represents a 3.86% relative increase. It is equivalent to approximately 15 additional women in sustained full-time employment per 1,000 recent mothers.

The identification evidence is strongest for sustained full-time employment. The joint pre-treatment test fails to reject the null that the five pre-treatment coefficients are jointly equal to zero at the 5% significance level. The observed pre-treatment evidence is therefore consistent with the parallel-trends assumption for this outcome. The three-year and five-year placebo estimates are also statistically insignificant. In addition, the sustained full-time employment estimate remains positive when California or the 2020 ACS wave is excluded and when ACS person weights are omitted.

The labor force participation estimate is positive but statistically insignificant. For this outcome, the joint pre-treatment test rejects the null that the five pre-treatment coefficients are jointly equal to zero at the 1% significance level. The three-year placebo estimate for labor force participation is also statistically significant at the 1% level. The labor force participation estimate therefore does not receive the same support from the identification diagnostics as the sustained full-time employment estimate. The estimated effect on usual weekly hours among employed recent mothers is close to zero and statistically insignificant. For this outcome, the joint pre-treatment test fails to reject the null that the five pre-treatment coefficients are jointly equal to zero at the 5% significance level. These findings describe immediate labor-market outcomes around childbirth. The analytic sample consists of women who reported giving birth during the preceding 12 months. The results therefore capture short-run maternal employment around a recent birth rather than longer-term employment trajectories

The pattern across outcomes is consistent with PFL supporting employment continuity following childbirth rather than substantially changing the hours worked by employed mothers. Partial wage replacement may reduce the financial cost of taking leave and help mothers maintain their connection to employment, thereby facilitating a return to sustained full-time work. The positive sustained full-time employment estimate and the near-zero usual-weekly-hours estimate are consistent with adjustment occurring through continued employment rather than through changes in the intensive margin of labor supply. The ACS does not observe individual leave-taking, benefit receipt, or continuity with the same employer, so this mechanism cannot be tested directly.

The education-specific results suggest that the estimated responses occur along different labor-market margins across educational groups. Among highly educated women, the estimated effect on sustained full-time employment is 2.22 percentage points, statistically significant at the 5% level and equivalent to 4.41% of their weighted descriptive control-group mean of 50.35%. One possible interpretation is that highly educated women are more likely to hold established full-time positions to which they can return following leave. Among women with low education, the larger estimated responses occur in labor force participation and usual weekly hours, and both estimates are statistically significant at the 5% level. For these women, partial wage replacement may be more relevant to remaining economically active or returning to employment following childbirth. The aggregate pre-treatment evidence for labor force participation should be considered when interpreting the low-education participation estimate.

Race/ethnicity-specific estimates differ in magnitude and precision. The sustained full-time employment estimate for White non-Hispanic women is 2.44 percentage points and statistically significant at the 1% level. The corresponding estimate for Black non-Hispanic women is also positive and similar in magnitude, at 2.09 percentage points, but is less precisely estimated. In the binary-outcome analytic sample, White non-Hispanic women constitute the largest subgroup, with 368,567 observations, compared with 50,834 Black non-Hispanic women, 104,312 Hispanic women, and 62,936 other non-Hispanic women. The greater precision of the White non-Hispanic estimates is consistent with their larger sample, although precision also depends on state-level treatment variation and clustering. Differences in statistical significance should not be interpreted as evidence that the effects are confined to White non-Hispanic women.

Educational composition also differs across race/ethnicity groups. Using ACS person weights, the high-education share in the full analytic sample is 48.79% among White non-Hispanic women, compared with 24.25% among Black non-Hispanic women and 18.52% among Hispanic women. The corresponding share is higher among other non-Hispanic women, at 55.39%. Because the sustained full-time employment estimate is also larger among highly educated women, differences in educational composition may help explain some of the descriptive overlap between the education- and race/ethnicity-specific patterns. However, the race/ethnicity-specific models retain education controls, so educational composition alone does not explain the race/ethnicity-specific estimates.

One important limitation concerns potential fertility-related selection. Because the main sample is restricted to women reporting a birth during the preceding 12 months, PFL could affect the composition of the analytic sample if policy availability influences whether or when women have children. For example, if PFL changes fertility among women with different levels of education, earnings potential, or prior labor-market attachment, part of the observed change in maternal employment could reflect a change in who enters the recent-mother sample rather than only a change in employment following childbirth. Because the analysis does not directly test whether PFL affects the probability of giving birth during the preceding 12 months, fertility-related selection cannot be ruled out.

Overall, the evidence is most supportive of a positive short-run effect of state PFL exposure on sustained full-time employment among recent mothers, while the evidence for labor force participation and usual weekly hours is less conclusive. The heterogeneity results further suggest that estimated maternal labor-market responses to PFL vary across demographic groups and labor-market margins. By examining multiple adopting states, several labor-market outcomes, and maternal subgroups, this study contributes multi-state evidence on how state PFL policies may support mothers' short-run employment following childbirth.

References

Bailey, M. J., Byker, T. S., Patel, E., & Ramnath, S. (2025). The long-run effects of California's Paid Family Leave Act on women's careers and childbearing: New evidence from a regression discontinuity design and U.S. tax data. *American Economic Journal: Economic Policy, 17*(1), 401–431. https://doi.org/10.1257/pol.20200277

Baum, C. L., & Ruhm, C. J. (2016). The effects of paid family leave in California on labor market outcomes. *Journal of Policy Analysis and Management, 35*(2), 333–356.

Bipartisan Policy Center. (2025). *State paid family leave laws across the U.S.*

Boesch, D. (2019). The urgent case for permanent paid leave. Center for American Progress.

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

Das, T., & Polachek, S. W. (2015). Unanticipated effects of California's paid family leave program. Contemporary Economic Policy, 33(4), 619-635.

De Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review, 110*(9), 2964–2996.

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

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). Child penalties across countries: Evidence and explanations. *AEA Papers and Proceedings, 109*, 122–126.

Nguyen, T., Dennison, B. A., Radigan, A., FitzPatrick, E., Zhang, W., & Ncube, B. (2023). New York State's Paid Family Leave Program is associated with more equitable and increased use of paid leave following childbirth. *Maternal and Child Health Journal, 27*(3), 516–526.

Rossin-Slater, M., Ruhm, C. J., & Waldfogel, J. (2013). The effects of California's paid family leave program on mothers' leave-taking and subsequent labor market outcomes. *Journal of Policy Analysis and Management, 32*(2), 224–245.

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

Ruggles, S., Flood, S., Sobek, M., Backman, D., Cooper, G., Rivera Drew, J. A., Richards, S., Rogers, R., Schroeder, J., & Williams, K. C. W. (2025). *IPUMS USA: Version 16.0* [Data set]. IPUMS. https://doi.org/10.18128/D010.V16.0

Sant'Anna, P. H. C., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics, 219*(1), 101–122.

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources, 50*(2), 301–316.

Stearns, J. (2015). The effects of paid maternity leave: Evidence from Temporary Disability Insurance. *Journal of Health Economics, 43*, 85–102.

U.S. Department of Labor, Wage and Hour Division. (2025). *Fact Sheet #28: The Family and Medical Leave Act*. https://www.dol.gov/agencies/whd/fact-sheets/28-fmla

Wells, W. M., White, J. S., Collin, D. F., Wang, G., Modrek, S., & Hamad, R. (2026). Effects of US state paid family leave policies on perinatal and postpartum health: A quasi-experimental analysis. *American Journal of Epidemiology, 195*(1), 246–254.

Appendix

This appendix provides supporting information on treatment timing, variable construction, covariate balance, pre-treatment diagnostics, robustness, and heterogeneity. Tables A.1–A.3 report treatment cohorts, variable coding, and covariate balance. Tables A.4 and A.5 report the joint pre-treatment tests and event-specific estimates. Tables A.6 and A.7 report the migration diagnostic and partial-implementation-year sensitivity analysis. Table A.8 presents supplementary pooled heterogeneity tests, and Table A.9 reports education distributions by race/ethnicity.

Appendix Table A.1: State Paid Family Leave Treatment Cohorts

State	**Actual PFL year**	**Analytic treatment cohort year**
California	2004	2005
New Jersey	2009	2010
Rhode Island	2014	2014
New York	2018	2018
Washington	2020	2020
District of Columbia	2020	2021
Massachusetts	2021	2021
Connecticut	2022	2022
Oregon	2023	2024
Colorado	2024	2024

Notes: Actual implementation year is the calendar year in which each state began paying PFL wage-replacement benefits (California, July 2004; New Jersey, July,2009; Rhode Island, January 2014; New York, January 2018; Washington, January 2020; District of Columbia, July 2020; Massachusetts, January 2021; Connecticut, January 2022; Oregon, September 2023; Colorado, January 2024). The analytic treatment cohort year is the first full ACS calendar year of policy exposure; for states with mid-year implementation (California, New Jersey, District of Columbia, and Oregon), the cohort year is coded as the following calendar year, consistent with the treatment-timing rule described in Section 4.1

Appendix Table A.2: Variable Coding and Sample Construction Rules

Component	ACS variable(s)	Codes or rule used	Use in analysis
Sample years	year	2001–2024	ACS calendar years used in the paper sample
Female sample	sex	2	Female respondents only
Recent mother	fertyr	2	Child born within the last 12 months
Age restriction	age	25–44	Women aged 25 to 44
Household population	gq	1, 2, 5	Household/noninstitutional population restriction used in clean file
Treatment cohort	gvar	CA 2005; NJ 2010; RI 2014; NY 2018; WA 2020; DC 2021; MA 2021; CT 2022; OR 2024; CO 2024	First full ACS calendar year of paid PFL benefit exposure
Labor force participation	empstat	1 or 2 = 1; 3 = 0	In labor force if employed or unemployed
Employment	empstat	1 = 1; 2 or 3 = 0	Currently employed
Sustained full-time employment	empstat uhrswork	empstat 1 and uhrswork 35-98 and wkswork2 4-6	Employed 35+ usual weekly hours and 40+ weeks
Education category	educ	0-5 low; 6-8 medium; 9-11 high	Education controls and timing diagnostic
Race/ethnicity	race hispan	Hispanic if hispan = 1–4; otherwise White if race = 1; Black if race = 2; Other non-Hispanic for all remaining non-Hispanic race codes	Race/ethnicity controls
Migration proxy	bpl statefip	bpl equals statefip for US-born state/DC birthplaces	Proxy for born in current state; not direct migration

Notes: This table summarizes the ACS/IPUMS coding rules used to construct the analytic sample, treatment timing, outcomes, controls, and migration proxy. The treatment cohort is the first full ACS calendar year of paid PFL benefit exposure.

Appendix Table A.3: Covariate Balance and Common Support

Covariate	Active PFL mean	Raw comparison mean	IPW comparison mean	Raw SMD	IPW SMD
Age	32.6350	31.6380	32.6710	0.2160	-0.0080
Low education	0.1000	0.0860	0.0990	0.0480	0.0030
Medium education	0.4720	0.5150	0.4690	-0.0860	0.0040
High education	0.4280	0.3990	0.4310	0.0590	-0.0060
White non-Hispanic	0.3820	0.6090	0.3800	-0.4670	0.0040
Black non-Hispanic	0.0670	0.1280	0.0670	-0.2060	0.0010
Hispanic	0.3610	0.1760	0.3640	0.4270	-0.0070
Other non-Hispanic	0.1900	0.0870	0.1880	0.3020	0.0030

Notes: Active PFL refers to women observed in state-years with active paid-family-leave benefits. The IPW comparison mean reweights comparison observations by p/(1-p), where p is the estimated probability of active-PFL exposure from a logit model using age, age squared, education category, and race/ethnicity. SMD denotes standardized mean difference. Propensity scores range from 0.051 to 0.413 in the common support; 0.0% of weighted observations fall outside this range; Marital status is included as an out-of-model diagnostic because it is not in the preferred control set.

Appendix Table A.4: Joint Wald Test of Pre-Treatment Coefficients

Outcome	N	Chi-square	df	p-value	Decision at 5%
Labor force participation	586,649	38.4410	5	<0.0010	Reject joint-zero null
Sustained full-time employment	586,649	8.6560	5	0.1240	Fail to reject joint-zero null
Usual weekly hours worked	357,972	10.2810	5	0.0680	Fail to reject joint-zero null

Notes: The null hypothesis is that the pre-treatment event-time coefficients for e=−5… −1 are jointly equal to zero. Decisions are based on a 5 percent significance level. Rejection of the null indicates evidence of differential pre-treatment trends over the specified window.

Appendix Table A.5: Event-Specific Pre-Treatment Lead Estimates

Event time	Estimate	SE	p-value	95% CI
Labor force participation				
−5	0.0020	0.0060	0.7430	[-0.0100, 0.0140]
−4	-0.0160	0.0100	0.1190	[-0.0360, 0.0040]
−3	0.0230	0.0090	0.0130	[0.0050, 0.0410]
−2	0.0070	0.0090	0.4580	[-0.0110, 0.0240]
−1	-0.0050	0.0160	0.7640	[-0.0350, 0.0260]
Sustained full-time employment				
−5	0.0070	0.0100	0.4860	[-0.0130, 0.0270]
−4	-0.0180	0.0130	0.1720	[-0.0440, 0.0080]
−3	0.0010	0.0100	0.9060	[-0.0190, 0.0210]
−2	0.0150	0.0100	0.1580	[-0.0060, 0.0350]
−1	-0.0170	0.0110	0.1230	[-0.0380, 0.0040]
Usual weekly hours worked				
−5	0.2310	0.3060	0.4510	[-0.3690, 0.8310]
−4	-0.3380	0.2040	0.0970	[-0.7360, 0.0610]
−3	-0.0930	0.6000	0.8770	[-1.2700, 1.0840]
−2	0.0030	0.6270	0.9960	[-1.2250, 1.2310]
−1	0.1140	0.2340	0.6250	[-0.3440, 0.5730]

Notes: Event time is measured relative to the analytic treatment cohort year. These estimates provide the event-specific components behind the joint anticipation test.

Appendix Table A.6: Migration Composition Diagnostic

Outcome	N	ATT	SE	p-value
Born in current state	461,009	0.0281	0.0177	0.1115

Notes: The proxy equals one when a US-born respondent is observed in her birth state. The estimate is obtained using the Callaway-Sant'Anna group-time difference-in-differences estimator with ACS person weights, controls for age, age squared, education, and race/ethnicity, and standard errors clustered by state. The proxy is not a direct migration measure, but it checks for estimated effects on composition in birthplace-current-state attachment.

Appendix Table A.7: Sensitivity to Excluding Partial Implementation Years

	Sustained full-time employment	Labor force participation	Usual weekly hours worked
Preferred: first full ACS year	0.0149***	0.0151	0.0180
	(0.0046)	(0.0092)	(0.2360)
	N = 586,649	N = 586,649	N = 357,972
Exclude partial implementation years	0.0107	-0.0106*	0.3770*
	(0.0071)	(0.0057)	(0.2200)
	N = 503,187	N = 503,187	N = 308,785

Notes: Standard errors are in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. Estimates and standard errors are reported to four decimal places. Sustained full-time employment and labor force participation are reported in decimal units; usual weekly hours worked is reported in hours among employed women with a recent birth and valid UHRSWORK values. The preferred specification codes each mid-year-adopting state's treatment cohort as the following full ACS calendar year. The partial-year exclusion retains that preferred cohort definition while removing California 2004, New Jersey 2009, District of Columbia 2020, and Oregon 2023 from the estimation sample. The estimator, comparison group, weights, and controls are otherwise the same as in Table 2.

Appendix Table A.8: Formal Tests of Heterogeneity

Outcome	Subgroup dimension	N	F statistic	p-value	Conclusion
Labor force participation	Education	586,649	7.9400	0.0010	Reject equality at 1 percent
Labor force participation	Race/ethnicity	586,649	14.3690	<0.0010	Reject equality at 1 percent
Sustained full-time employment	Education	586,649	20.1560	<0.0010	Reject equality at 1 percent
Sustained full-time employment	Race/ethnicity	586,649	5.0280	0.0041	Reject equality at 1 percent
Usual weekly hours worked	Education	357,972	12.1090	<0.0010	Reject equality at 1 percent
Usual weekly hours worked	Race/ethnicity	357,972	2.9340	0.0425	Reject equality at 5 percent

Notes: Tests are weighted pooled interaction models with state and year fixed effects, ACS person weights, and standard errors clustered by state. The null hypothesis is equality of the PFL coefficient across subgroups within the listed dimension. These tests are supplementary diagnostics; subgroup-specific estimates in the main text remain based on the stratified CS-DID models.