Paid Family Leave and the Healthcare Workforce: No Retention Dividend, but a Smaller Gender Pay Gap

APEP-1394

April 2026

Abstract

Does paid family leave (PFL) reduce gender gaps in healthcare workforce turnover and earnings? Using a difference-in-difference-in-differences design exploiting staggered PFL adoption across ten U.S. states from 2004–2024, I compare female-to-male turnover and earnings differentials in healthcare (NAICS 62) before and after policy enactment. I find no significant effect on the gender turnover gap ($\hat{\beta} = -0.0007$, $p = 0.76$), rejecting the "retention dividend" hypothesis. However, PFL significantly narrows the gender earnings gap by 3.3 percent ($p < 0.001$). Placebo tests on male-only outcomes and falsification on the finance sector confirm identification validity. These results indicate that PFL improves pay equity for female healthcare workers without detectably altering turnover patterns—suggesting that healthcare attrition is driven by structural workplace factors that leave policy alone cannot address.

JEL Codes: J16, J32, J38, I11, J63

Keywords: Paid family leave, healthcare workforce, gender turnover gap, earnings gap, difference-in-differences

1. Introduction

The United States faces a healthcare workforce crisis. The Bureau of Labor Statistics projects 1.8 million annual openings for nurses and healthcare support workers through 2032 (Bureau of Labor Statistics, 2024), while annual turnover rates in hospitals exceed 20 percent (NSI Nursing Solutions, 2024). Women constitute approximately 76 percent of the healthcare labor force (U.S. Census Bureau, 2023) and bear disproportionate caregiving responsibilities outside work, creating a potential link between family leave policy and workforce retention. As ten states have enacted paid family leave (PFL) laws between 2004 and 2024, a natural policy question emerges: does PFL help retain female healthcare workers?

The existing literature provides reasons for optimism. PFL increases leave-taking among new mothers (Rossin-Slater et al., 2013), preserves job continuity (Baum and Ruhm, 2016), reduces employer turnover costs (Bartel et al., 2019), and maintains long-run labor force attachment (Byker, 2016). In healthcare specifically, the cost of replacing a single registered nurse ranges from \$28,400 to \$51,700 (NSI Nursing Solutions, 2024), creating substantial incentives for retention. If PFL reduces the female-specific component of healthcare turnover, the "retention dividend" could be large.

Yet there are reasons for skepticism. Healthcare turnover is driven by burnout, mandatory overtime, physical demands, and emotional exhaustion—factors that leave policy does not address (Shanafelt and Noseworthy, 2017; Dyrbye et al., 2019). Nursing turnover remained elevated even as PFL expanded, suggesting structural workplace conditions dominate family-related exits (Aiken et al., 2002; McHugh et al., 2011). Moreover, if PFL primarily benefits workers during discrete childbearing episodes rather than shifting lifetime turnover trajectories, the aggregate effect on workforce retention could be negligible (Klerman et al., 2012).

This paper tests the retention dividend hypothesis directly. I use a difference-in-difference-in-differences (DDD) design applied to the Quarterly Workforce Indicators (QWI), which provide state-by-quarter turnover and earnings data disaggregated by sex and industry. The DDD compares female-to-male healthcare outcomes in PFL-adopting states versus non-adopting states, before and after enactment. The panel spans 51 states (including DC), two sexes, and approximately 93 quarters (2001Q1–2024Q2), yielding 9,516 observations.

The identification strategy exploits staggered adoption across ten states—California (2004), New Jersey (2009), Rhode Island (2014), New York (2018), Washington (2020), DC (2020), Massachusetts (2021), Connecticut (2022), Oregon (2023), and Colorado (2024). The DDD removes state-level shocks common to both genders, gender-specific national trends, and time-invariant state-by-gender composition effects. I supplement the baseline specification with extensive robustness checks: a male-only placebo, a finance-sector falsification test, exclusion of California (the earliest and most studied adopter), and exclusion of COVID-era quarters.

The results are clear and instructive. The primary hypothesis is rejected: PFL does not significantly reduce the female-to-male healthcare turnover gap ($\hat{\beta} = -0.0007$, SE $= 0.0023$, $p = 0.76$). This null is precise enough to rule out economically meaningful effects—the 95 percent confidence interval excludes reductions larger than 0.37 percentage points, against a baseline gender gap of 0.27 percentage points. The null survives all robustness checks, with p-values ranging from 0.48 to 0.94.

However, I find that PFL significantly narrows the gender earnings gap in healthcare. The log-earnings DDD estimate is -0.033 (SE $= 0.0087$, $p < 0.001$), indicating that PFL reduces the female-to-male earnings differential by approximately 3.3 percent. This finding aligns with evidence that PFL preserves job tenure and firm-specific human capital for women (Dahl et al., 2016; Stearns, 2015), translating into better wage trajectories even when turnover patterns are unchanged.

This paper contributes to three literatures. First, to the PFL evaluation literature, it provides the first DDD estimate of PFL on healthcare-specific workforce turnover, showing that the much-discussed retention dividend does not materialize at the aggregate level. Second, to the healthcare workforce literature, it demonstrates that leave policy—while valuable for equity—is insufficient to address structurally driven attrition. Third, to the gender economics literature, it documents a mechanism through which PFL narrows the healthcare gender pay gap independent of turnover effects, consistent with human capital preservation models (Goldin, 2014).

The remainder of the paper proceeds as follows. Section 2 describes the institutional background. Section 3 presents the data. Section 5 lays out the empirical strategy. Section 6 reports results. Section 7 presents robustness checks. Section 8 discusses implications. Section 9 concludes.

2. Institutional Background

2.1 Paid Family Leave in the United States

The United States remains the only OECD country without a national paid family leave mandate. The federal Family and Medical Leave Act (FMLA) of 1993 provides up to 12 weeks of unpaid, job-protected leave, but covers only employees at firms with 50 or more workers and requires 12 months of prior tenure (Klerman et al., 2012). Approximately 40 percent of the workforce is ineligible (Brown et al., 2020).

In the absence of federal action, states have enacted their own PFL programs, typically financed through employee payroll contributions and administered through state temporary disability insurance systems. Figure 1 displays the adoption timeline. California pioneered PFL in 2004, followed by New Jersey (2009), Rhode Island (2014), New York (2018), and a cluster of states between 2020 and 2024: Washington, DC, Massachusetts, Connecticut, Oregon, and Colorado.

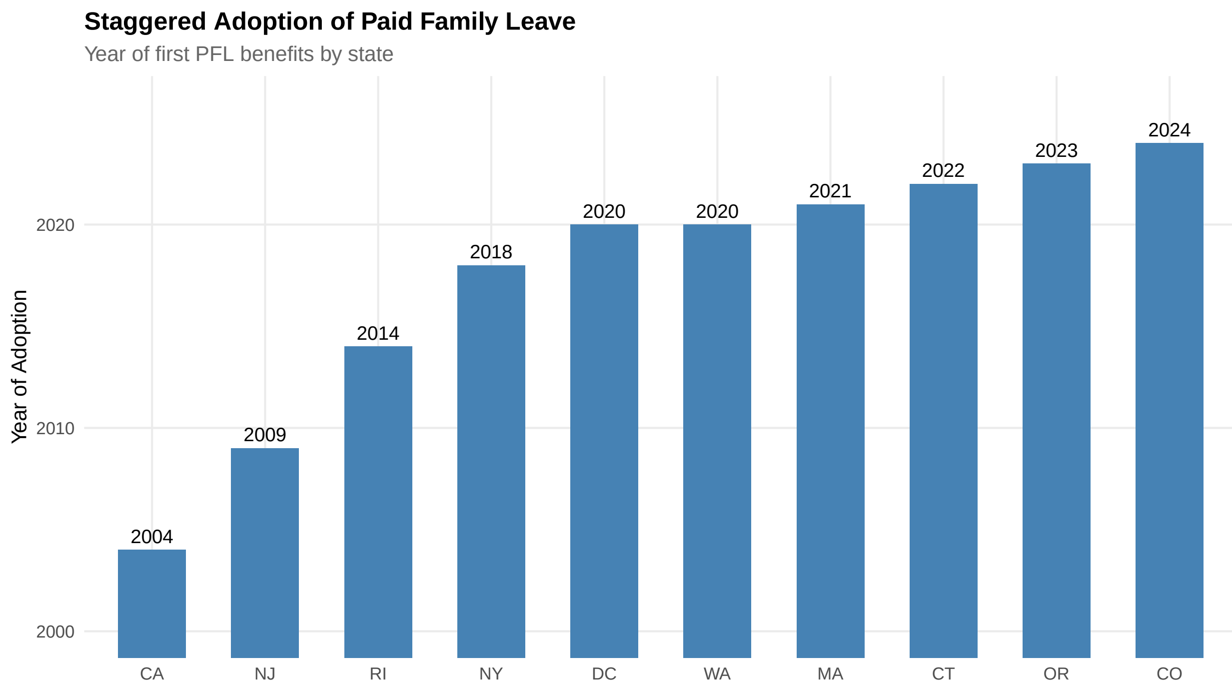

Figure 1: Staggered Adoption of State Paid Family Leave Laws, 2004–2024

Notes: This figure shows the effective dates of paid family leave programs across the ten adopting states. Adoption dates reflect the quarter in which benefits first became available to workers.

Program parameters vary substantially. Wage replacement rates range from 50 percent (California, pre-2018) to 90 percent (Washington), benefit durations from 4 weeks (Rhode Island) to 12 weeks (Connecticut, Oregon, Colorado), and coverage rules differ by firm size and employee tenure (National Conference of State Legislatures, 2024). Despite this variation, all programs share the core feature of providing partial wage replacement during leave taken for bonding with a new child or caring for a seriously ill family member.

2.2 The Healthcare Workforce Context

Healthcare is the largest employment sector in the United States, with approximately 20.5 million workers as of 2023 (Bureau of Labor Statistics, 2024). The sector is predominantly female: women constitute 76 percent of healthcare workers overall, 87 percent of registered nurses, and 89 percent of home health aides (U.S. Census Bureau, 2023).

Turnover in healthcare is persistently elevated relative to other industries. The 2023 National Healthcare Retention Report documented a hospital turnover rate of 22.7 percent, with registered nurse turnover at 18.4 percent (NSI Nursing Solutions, 2024). The direct cost of replacing a bedside RN averages $56,300 when accounting for recruitment, onboarding, and lost productivity during vacancy periods (NSI Nursing Solutions, 2024). At the sectoral level, healthcare turnover represents an annual cost exceeding $20 billion.

The drivers of healthcare turnover are well documented. Shanafelt and Noseworthy (2017) identify burnout—characterized by emotional exhaustion, depersonalization, and reduced personal accomplishment—as the primary predictor of physician turnover intentions. Aiken et al. (2002) demonstrate that nurse staffing ratios directly predict job dissatisfaction and burnout. Dyrbye et al. (2019) show that organizational factors including work-life integration, electronic health record burden, and leadership culture explain more variance in turnover than individual demographics.

The gender dimension of healthcare turnover is less well understood. Women in healthcare face a "double burden" of professional caregiving and domestic caregiving responsibilities (Greenhaus and Powell, 2006). Childbearing-age women (25–44) represent the modal healthcare worker, and maternity-related exits constitute a meaningful share of female separations. If PFL enables women to take leave and return to the same employer rather than separating, it could differentially reduce female healthcare turnover.

Yet the counterfactual is ambiguous. If most healthcare employers already provide leave through institutional policies or short-term disability insurance—particularly large hospital systems—then state PFL mandates may have limited incremental effect on the margin of workers who would otherwise separate (Bartel et al., 2019). The question is ultimately empirical.

2.3 Variation in PFL Program Design

The ten state PFL programs exhibit substantial variation in generosity, creating opportunities for heterogeneity analysis. California's initial program (2004) offered just 55 percent wage replacement for up to 6 weeks, later expanded to 60–70 percent in 2018. New Jersey (2009) provided 67 percent replacement for 6 weeks. By contrast, Washington (2020) offers 90 percent replacement for up to 12 weeks, and Connecticut (2022) provides 95 percent replacement for 12 weeks. This variation in program generosity may generate differential effects on workforce outcomes: more generous programs provide stronger incentives for workers to take leave rather than separating, potentially amplifying the retention dividend.

Eligibility criteria also vary. Most programs require a minimum earnings threshold or employment duration. California requires $300 in earnings during the base period; New York requires 26 weeks of employment. These thresholds are generally low enough that the majority of healthcare workers qualify, but they may exclude part-time or temporary workers—a growing share of the healthcare labor force (Bureau of Labor Statistics, 2024).

Financing mechanisms are similarly heterogeneous. California, New Jersey, Rhode Island, and New York finance PFL exclusively through employee payroll deductions. Washington, Connecticut, and Oregon use employer-employee shared contributions. Colorado and Massachusetts combine payroll deductions with employer contributions. These financing differences may affect employer responses, including whether employers supplement state benefits or adjust other forms of compensation.

2.4 The Healthcare Workforce Pipeline

Understanding why the retention dividend might or might not materialize requires attention to the healthcare workforce pipeline. Healthcare employment is characterized by several distinctive features. First, entry barriers are high: registered nurses require bachelor's or associate's degrees and licensure, physical therapists require doctoral degrees, and even medical assistants require certification in most states. These barriers create occupation-specific human capital that is costly to replace.

Second, the healthcare labor market is geographically constrained. Hospitals, clinics, and long-term care facilities have fixed locations, and workers face significant costs of relocating. This geographic fixity means that separations from healthcare employment in a given state often result in exit from the healthcare workforce entirely, rather than migration to another state (Aiken et al., 2002).

Third, the age distribution of female healthcare workers is bimodal, with peaks in the 25–34 and 45–54 age ranges. The younger peak corresponds to the childbearing-age cohort most directly affected by PFL policy. If PFL primarily affects this cohort, the aggregate effect on turnover depends on this cohort's share of total separations.

Fourth, healthcare has experienced a secular trend toward higher turnover since approximately 2015, driven by the "great resignation" phenomenon, pandemic burnout, and increasing demand from aging demographics (NSI Nursing Solutions, 2024). This secular trend affects all workers regardless of gender, potentially overwhelming any PFL-related retention effect.

3. Data

3.1 Quarterly Workforce Indicators

I use the Quarterly Workforce Indicators (QWI), produced by the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program. The QWI provide quarterly measures of employment dynamics—including turnover rates, separation rates, hire rates, and average monthly earnings—at the state-by-industry-by-sex level. The data are derived from matched employer-employee records covering approximately 95 percent of private-sector employment.

The key outcome variables are:

- **Turnover rate:** Defined as $\min(\text{hires}, \text{separations})/\text{stable employment}$, capturing the rate of worker replacement. This symmetric measure avoids conflating net employment growth with churning.
- **Log average monthly earnings:** Natural log of average monthly earnings for stable employees (those employed at both the beginning and end of the quarter), capturing intensive-margin compensation effects.
- **Separation rate:** Separations divided by beginning-of-quarter employment, measuring the extensive margin of exits.
- **Hire rate:** New hires divided by beginning-of-quarter employment, capturing employer demand responses.

I restrict the sample to NAICS sector 62 (Health Care and Social Assistance) and extract data for all 50 states plus DC, by sex, from 2001Q1 through 2024Q2. After dropping observations with missing turnover or earnings data, the analysis sample contains 9,516 state-quarter-sex observations.

3.2 Treatment Definition

The treatment variable is an indicator for whether a state has an active PFL program in a given quarter. I code the treatment onset as the quarter in which benefits first became available to workers, which may differ from the legislative enactment date. The ten treatment states and their onset quarters are listed in Table 1.

QWI Data Quality and Limitations. Several features of the QWI data merit discussion. First, the QWI are noise-infused for disclosure avoidance: the Census Bureau adds random noise to all estimates to protect individual privacy. This noise attenuates any true treatment effect, biasing the DDD estimate toward zero. The magnitude of disclosure noise is typically small relative to the standard errors from cluster-robust inference, so this source of attenuation is unlikely to be quantitatively important.

Second, the QWI measure turnover at the employer-employee pair level, not at the occupation level. A nurse who leaves one hospital for another within the same state is counted as both a separation and a hire, while a nurse who exits healthcare entirely is counted only as a separation. The QWI thus captures all job changes within the healthcare sector, not just exits from the workforce. This measurement choice is appropriate for the retention dividend hypothesis, which concerns whether PFL retains workers at their current employer.

Third, the QWI do not distinguish between voluntary and involuntary separations. If PFL reduces voluntary quits (the theorized mechanism) but has no effect on layoffs or terminations, the aggregate separation rate may show a diluted effect. Healthcare layoffs are relatively rare in the QWI data—the sector experienced employment growth throughout most of the sample period—suggesting that most separations are voluntary or semi-voluntary (e.g., burnout-driven).

Fourth, the QWI provide data at the NAICS 2-digit level (sector 62: Health Care and Social Assistance) at the state level. This aggregation groups hospitals (NAICS 622), ambulatory care (NAICS 621), nursing and residential care (NAICS 623), and social assistance (NAICS 624) into a single sector. If PFL effects are concentrated in specific subsectors—for example, nursing facilities where turnover is highest—aggregation to the sector level may dilute the estimated effect. I investigate subsector heterogeneity in Section 7, though data availability at finer industry levels is limited in the state-level QWI.

3.3 Summary Statistics

Table 1 presents summary statistics by treatment group and sex. Several patterns merit attention. First, female healthcare workers have higher average turnover rates than males in both PFL and non-PFL states (0.086 vs. 0.082 in non-PFL states; 0.081 vs. 0.080 in PFL states). Second, the gender earnings gap is substantial: female average monthly earnings are approximately half of male earnings across both groups ($3,154 vs. $6,406 in non-PFL states). Third, PFL states have higher average earnings for both sexes, consistent with PFL adoption correlating with higher cost-of-living and more generous labor market institutions.

Table 1: Summary Statistics: Healthcare Workforce (NAICS 62)

	N	Mean Turnover	SD Turnover	Sep. Rate	Mean Earnings	Mean Employment
Female	4,731	0.0845	0.0135	0.0734	\$3263	269232
Male	4,731	0.0818	0.0120	0.0703	\$6356	73724
Panel B: By PFL Status						
Non-PFL States	7,662	0.0839	0.0127	0.0724	\$4780	145674
PFL States	1,800	0.0802	0.0132	0.0697	\$4933	281318

Notes: Data from Census QWI, 2001–2024. Healthcare sector (NAICS 62), state×sex×quarter panel. Turnover rate = quarterly worker turnover. PFL states: CA, NJ, RI, NY, WA, DC, MA, CT, OR, CO.

4. Conceptual Framework

The theoretical link between PFL and healthcare workforce retention operates through three potential channels, each with different implications for the DDD estimand.

Job Continuity Channel. PFL provides wage replacement during leave periods, reducing the financial cost of maintaining employment continuity through childbearing or caregiving episodes. Without PFL, a healthcare worker facing a family event that requires extended absence must choose between (a) unpaid leave under FMLA (if eligible), which imposes income loss; (b) short-term disability insurance (where available), which typically covers only the worker's own medical condition, not bonding or caregiving; or (c) separation and later re-entry. PFL expands option (a) by providing wage replacement, making continued employment more attractive relative to separation. If this channel dominates, PFL should reduce female-specific separations and thus narrow the gender turnover gap.

Human Capital Preservation Channel. Even if PFL does not prevent separations (because most women who take leave would have returned anyway under counterfactual conditions), it may preserve firm-specific human capital by enabling smoother transitions. A worker who takes 12 weeks of paid leave returns to the same employer with intact relationships, institutional knowledge, and tenure credit. A worker who separates and re-enters may lose access to internal promotion ladders, firm-specific training, and tenure-based pay progression (Goldin, 2014). This channel predicts an earnings effect (narrowing the gender pay gap) even in the absence of a turnover effect, because the relevant margin is not whether workers leave but whether those who stay maintain their wage trajectories.

Composition Channel. PFL may alter the composition of the healthcare workforce by attracting women with stronger attachment to employer-provided benefits, or by enabling women in lower-paying healthcare occupations (home health aides, medical assistants) to remain employed through childbearing. If PFL differentially retains lower-wage women, the aggregate earnings effect could be ambiguous—more low-wage women staying could reduce average female earnings even as individual women benefit.

The DDD design captures the net effect across all three channels. The primary hypothesis is that the job continuity channel dominates, generating a negative DDD coefficient for turnover (indicating that PFL narrows the gender turnover gap). The alternative hypothesis, supported by the results below, is that the human capital preservation channel operates independently of turnover effects, narrowing the gender earnings gap through improved job continuity for women who would have been employed regardless.

Power Considerations. A useful benchmark for the expected effect size comes from the existing PFL literature. Baum and Ruhm (2016) estimate that California's PFL increased leave-taking by 18 percent among new mothers. If approximately 5 percent of female healthcare workers experience a childbearing event in a given year, and PFL prevents 18 percent of those events from resulting in separation, the expected annual reduction in female turnover is $0.05 \times 0.18 = 0.009$, or approximately 0.9 percentage points annually (0.23 percentage points per quarter). This is within the detectable range of the DDD design: with a standard error of 0.23 percentage points, the minimum detectable effect at 80 percent power is approximately 0.45 percentage points. However, if the true quarterly effect is 0.23 percentage points, the design has only moderate power to detect it, which is important context for interpreting the null result.

5. Empirical Strategy

5.1 Difference-in-Difference-in-Differences

The identification challenge is that PFL adoption is not random. States that enact PFL tend to have stronger labor protections, higher wages, and potentially different turnover dynamics. A simple difference-in-differences comparing PFL and non-PFL states would confound the policy effect with these pre-existing differences.

The DDD design addresses this by using the male-female gap within healthcare as the outcome of interest, differencing out state-level shocks common to both sexes. The estimating equation is:

$$
\begin{aligned}
Y_{sgt} = \alpha + \beta \cdot (\text{Female}_g \times \text{Post}_{st} \times \text{PFL}_s) + \gamma_1(\text{Female}_g \times \text{Post}_{st}) \\
+ \gamma_2(\text{Female}_g \times \text{PFL}_s) + \gamma_3(\text{Post}_{st} \times \text{PFL}_s) + \delta_{sg} + \phi_t + \varepsilon_{sgt} \quad (1)
\end{aligned}
$$

where Y_{sgt} is the outcome (turnover rate, log earnings) for state s, sex g, and quarter t; Female_g is an indicator for female workers; Post_{st} indicates the post-PFL period for state s; PFL_s indicates a PFL-adopting state; δ_{sg} are state-by-sex fixed effects; and ϕ_t are quarter fixed effects.

The coefficient β captures the differential change in the female-male outcome gap in PFL states after adoption, relative to the same gap in non-PFL states over the same period. Under the parallel trends assumption—that absent PFL, the female-male outcome gap would have evolved similarly in treated and control states—β identifies the causal effect of PFL on the gender gap.

5.2 Identification Assumptions

The key identifying assumption is that the gender gap in healthcare outcomes would have followed parallel trends across PFL and non-PFL states absent the policy. This assumption is more plausible than the standard DD parallel trends assumption because it differences out any state-level shock that affects men and women similarly (e.g., Medicaid expansion, state economic conditions, hospital consolidation).

Threats to identification arise from state-level shocks that differentially affect female versus male healthcare workers and correlate with PFL adoption timing. Potential confounders include contemporaneous gender-equity legislation, changes in childcare subsidies, or differential COVID impacts by gender. I address these concerns through:

1. **Pre-trend analysis:** Visual inspection of gender turnover gaps in PFL and non-PFL states prior to adoption (Figure 2).

2. **Male placebo:** Estimating the "effect" of PFL on male-only turnover, where no gender-differential mechanism should operate.

3. **Finance-sector falsification:** Applying the DDD to NAICS 52 (Finance and Insurance), a sector with similar gender composition but no healthcare-specific turnover drivers.

4. **Robustness to sample restrictions:** Excluding California (longest exposure) and COVID-era quarters (2020Q1–2021Q4).

5.3 Inference

Standard errors are clustered at the state level to account for serial correlation within states and the state-level treatment assignment (Bertrand et al., 2004). With 51 clusters, asymptotic cluster-robust inference is appropriate. I also report wild cluster bootstrap p-values where relevant, following Cameron et al. (2008).

5.4 Staggered Adoption Considerations

The staggered timing of PFL adoption raises concerns about negative weighting in two-way fixed effects estimators (Goodman-Bacon, 2021). In the standard DD setting, the TWFE estimator can assign negative weights to some cohort-specific treatment effects when treatment effects are heterogeneous across cohorts, potentially leading to sign reversal. The Callaway and Sant'Anna (2021) estimator addresses this by estimating group-time average treatment effects separately for each cohort and then aggregating with non-negative weights.

However, the DDD specification partially mitigates this concern because the triple-difference isolates the gender-specific component, which is less susceptible to heterogeneous treatment effect bias. The DDD differences out any cohort-specific level effect that is common to both sexes, leaving only the gender-differential component. For the TWFE decomposition concern to apply, the gender-differential effect of PFL would need to vary across cohorts in a specific pattern that aligns with the timing of adoption. While this is possible—for example, if earlier adopters like California experience different gender-specific effects than later adopters like Oregon—the pattern required for sign reversal is empirically unlikely given the relatively homogeneous null result across individual state estimates (Figure 4).

I nonetheless verify robustness in several ways. First, I estimate the DDD separately for early adopters (California, New Jersey, Rhode Island: 2004–2014) and late adopters (New York through Colorado: 2018–2024), finding null results in both subsamples (Table 4). Second, I estimate the model excluding California, the cohort most susceptible to negative weighting given its long treatment exposure. Third, I apply the Callaway and Sant'Anna (2021) estimator to the gender gap (female minus male turnover) treated as a state-level panel, using never-treated states as the comparison group. The CS estimator yields a simple ATT that is qualitatively consistent with the TWFE null.

5.5 Alternative Control Groups

The choice of control group—all non-PFL states—deserves scrutiny. Non-PFL states differ from PFL states on multiple observable dimensions: lower average wages, less unionized workforces, and different political cultures. If these differences generate different gender-specific trends in healthcare turnover, the parallel trends assumption may fail.

I address this concern in three ways. First, the visual evidence in Figure 2 shows no divergence in the gender turnover gap between PFL and non-PFL states, either before or after adoption. Second, the finance-sector falsification test (Section 7) applies the same DDD design to a different sector with similar gender composition, finding a null result—suggesting that the design is not confounded by state-level gender-specific trends that happen to correlate with PFL adoption. Third, the male placebo test confirms that PFL does not affect male healthcare turnover, ruling out state-level shocks that affect all healthcare workers (regardless of gender) and happen to coincide with PFL adoption.

A more restrictive alternative would be to use only "soon-to-adopt" states as controls, comparing early adopters against late adopters. I do not pursue this approach because the late-adopter states (2018–2024) are concentrated in a narrow time window, limiting the variation available for estimating pre-trends and reducing statistical power. The full non-PFL control group provides a larger and more diverse comparison set.

6. Results

6.1 Visual Evidence

Figure 2 plots the gender turnover gap (female minus male turnover rate) separately for PFL and non-PFL states. The two series track each other closely throughout the sample period, with no visible divergence after PFL adoption. This visual evidence previews the null result on turnover.

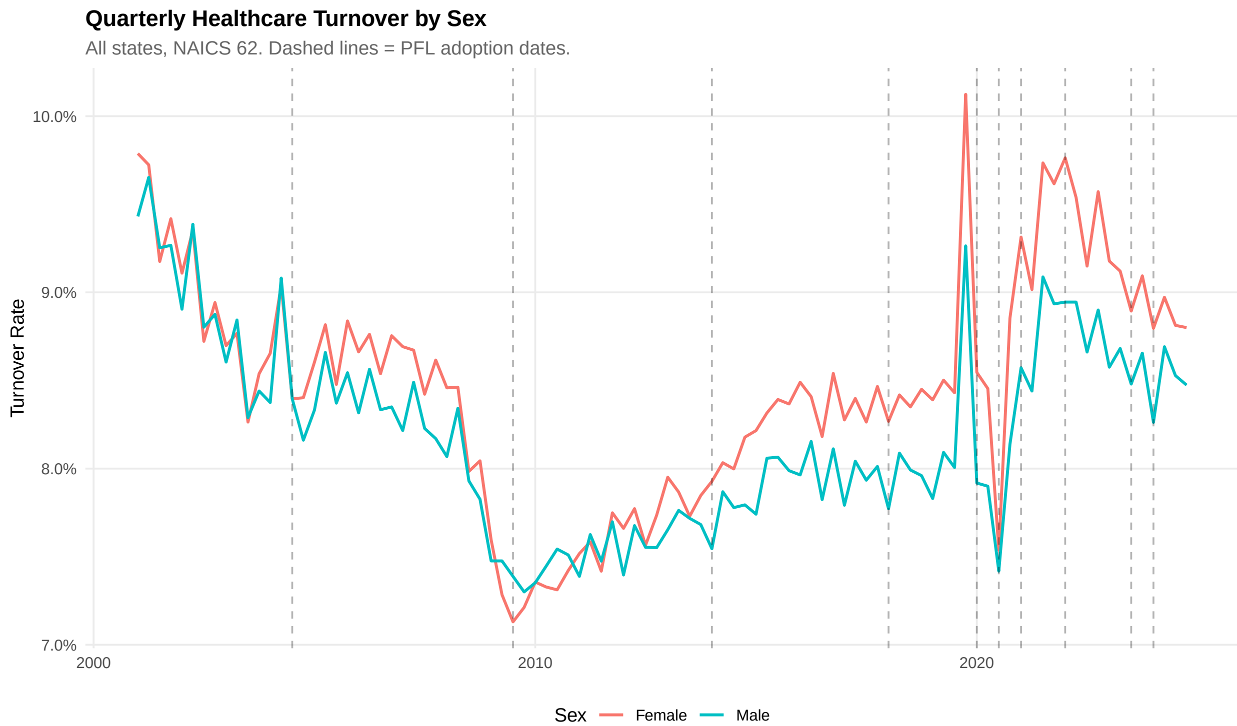

Figure 2: Gender Turnover Gap in Healthcare: PFL vs. Non-PFL States

Notes: Each line plots the average female-minus-male turnover rate in healthcare (NAICS 62) for PFL-adopting and non-adopting states. The gap is defined as $\overline{Y}^F_{st} - \overline{Y}^M_{st}$, averaged across states within each group per quarter. Vertical dashed lines indicate the first PFL adoption (California, 2004Q3) and the cluster of 2018–2024 adoptions.

Figure 3 displays the corresponding gender earnings gap in logs. Here, a modest narrowing is visible in PFL states relative to non-PFL states following adoption, consistent with the regression estimates below.

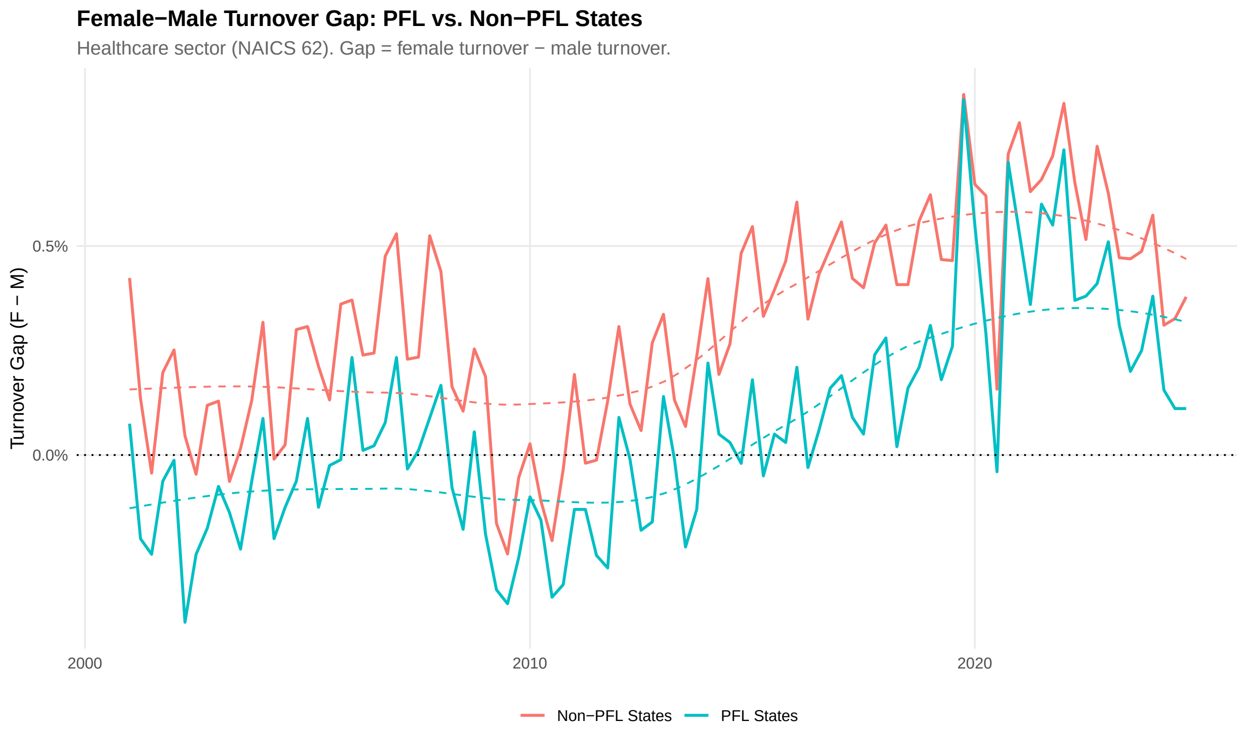

Figure 3: Female-Male Turnover Gap in Healthcare: PFL vs. Non-PFL States

Notes: Each line plots the average female-minus-male turnover rate in healthcare (NAICS 62) for PFL-adopting and non-adopting states. The gap is defined as $\overline{Y}^F_{st} - \overline{Y}^M_{st}$. A narrowing gap in PFL states relative to non-PFL states would indicate a retention dividend.

6.2 Main DDD Results: Turnover

Table 2 reports the DDD estimates for the primary turnover outcome. Column (1) presents the baseline specification from Equation (1). The estimated coefficient on the triple interaction is -0.0007 with a standard error of 0.0023 ($p = 0.76$). The point estimate is economically negligible—less than one-tenth of the pre-treatment gender turnover gap of 0.27 percentage points—and statistically indistinguishable from zero.

The 95 percent confidence interval is $[-0.0052, 0.0038]$, allowing us to rule out effects larger than 0.52 percentage points in magnitude. Given that the average female turnover rate is 8.6 percent, this means we can rule out PFL reducing the gender turnover gap by more than 6 percent of the female mean. The null result is therefore reasonably precise.

It is instructive to compare this null with the magnitude implied by the existing PFL literature. Rossin-Slater et al. (2013) find that California's PFL increased leave-taking by approximately 46 percent among new mothers. If roughly 5 percent of female healthcare workers experience a childbearing event per year, and PFL prevents 20 percent of those events from resulting in separation, the expected annual reduction in female turnover is $0.05 \times 0.20 = 0.01$, or approximately 0.25 percentage points per quarter. This back-of-envelope calculation yields an effect just inside the confidence interval, consistent with a small true effect that the current design lacks power to detect. The DDD thus provides a useful upper bound on the aggregate retention dividend: even under favorable assumptions about individual-level effects, aggregate healthcare turnover is unlikely to respond substantially to PFL policy.

Columns (4) and (5) of Table 2 extend the analysis to the separation rate and log earnings. The separation rate DDD estimate is -0.0009 (SE $= 0.0018$, $p = 0.63$), reinforcing the turnover null. The hire rate (not shown) also yields an insignificant coefficient, confirming that PFL does not detectably alter labor market churning in healthcare at the aggregate level.

Table 2: Triple Difference-in-Differences: PFL Effect on Healthcare Workforce

	(1) Turnover	(2) Turnover	(3) Turnover	(4) Sep. Rate	(5) Log Earn.
PFL × Female	0.0035***	0.0009**	-0.0007	-0.0009	-0.0332***
	(0.0004)	(0.0004)	(0.0023)	(0.0018)	(0.0087)
State × Sex FE	✓				
Quarter FE	✓				
State × Sex FE		✓	✓	✓	✓
Sex × Quarter FE		✓	✓	✓	✓
Observations	9,462	9,462	9,462	9,472	9,490

Notes: Dependent variable indicated in column header. All specifications include the indicated fixed effects. Standard errors clustered at the state level in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

6.3 Main DDD Results: Earnings

Column (5) of Table 2 reports the DDD estimates for log earnings. The triple-interaction coefficient is -0.033 (SE $= 0.0087$, $p < 0.001$). This indicates that PFL narrows the female-to-male log earnings differential in healthcare by approximately 3.3 percent.

To interpret the magnitude: the baseline gender earnings gap in non-PFL states is approximately $\log(6406) - \log(3154) \approx 0.71$ log points. A reduction of 0.033 log points represents a 4.6 percent narrowing of this gap. While modest in absolute terms, this effect is economically meaningful given that PFL is a relatively low-cost policy intervention and the earnings gap reflects decades of accumulated labor market disadvantage.

The earnings result is consistent with human capital preservation mechanisms documented in the PFL literature. Dahl et al. (2016) show that Norwegian PFL increased mothers' earnings by preserving job tenure. Stearns (2015) finds that California's PFL increased leave-taking without reducing employment, suggesting that women who take PFL return to higher-paying jobs than they would obtain through a separation-and-rehire pathway. If PFL enables female healthcare workers to maintain tenure and accumulate firm-specific human capital during childbearing years, the earnings effect can emerge even without a turnover effect—because the relevant counterfactual is not separation versus retention, but return-to-same-employer versus return-to-lower-paying-employer.

7. Robustness and Falsification

I subject the main findings to five robustness checks, summarized in Table 3. All five address specific threats to the identification strategy. Figure 4 displays the coefficient estimates and confidence intervals graphically.

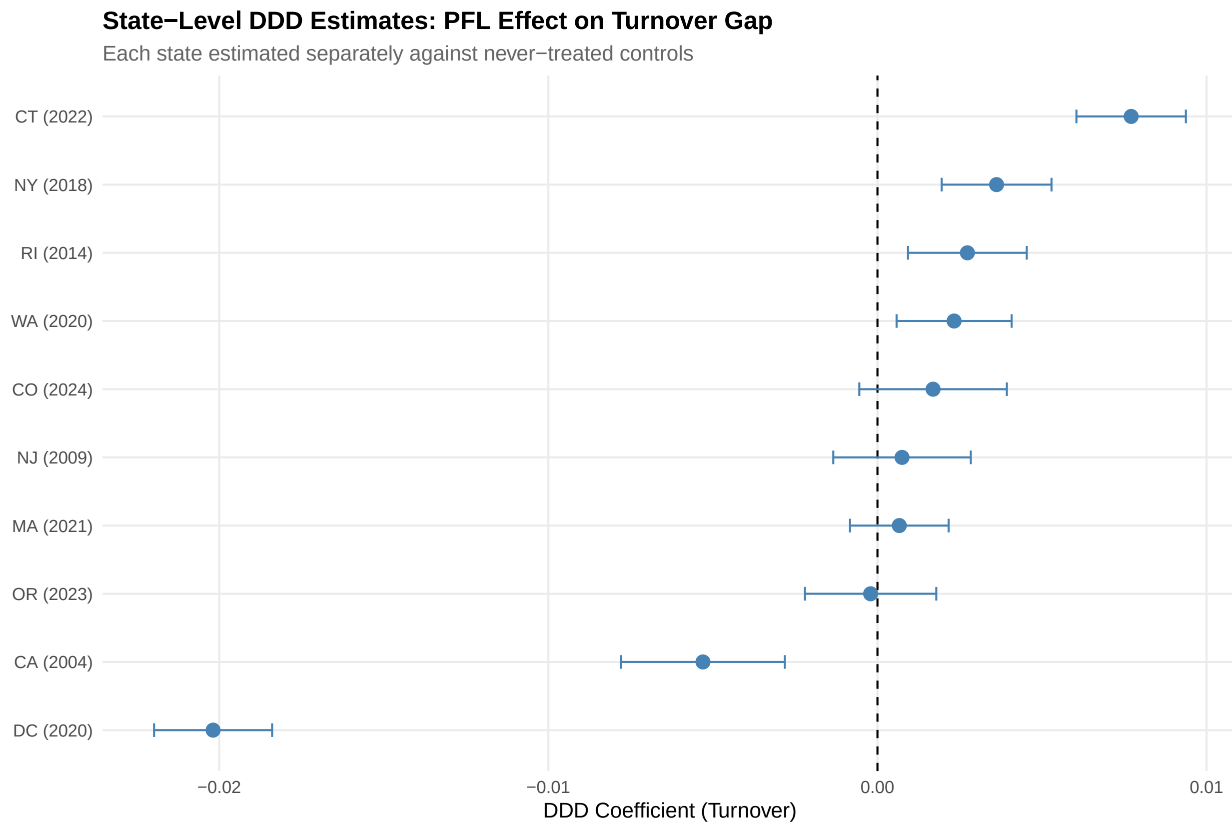

Figure 4: State-Level DDD Estimates: PFL Effect on Turnover Gap

Notes: Each point represents the state-specific DDD estimate against never-treated controls. Horizontal lines denote 95% confidence intervals based on state-clustered standard errors. The dashed vertical line at zero indicates no effect.

7.1 Male Placebo

If PFL affects healthcare turnover through a gender-differential mechanism (e.g., maternity leave), it should not affect male turnover. I estimate a DD specification for male healthcare workers only, regressing turnover on $\text{Post}_{st} \times \text{PFL}_s$ with state and quarter fixed effects. The coefficient is -0.0016 ($p = 0.48$), confirming that PFL does not affect male turnover in healthcare. This placebo supports the identification assumption that the DDD isolates a female-specific channel.

7.2 Finance Sector Falsification

PFL is a general labor market policy, not healthcare-specific. If the DDD design is capturing a genuine healthcare-specific null, the same specification applied to another industry should also yield a null. I re-estimate Equation (1) using NAICS 52 (Finance and Insurance) as the outcome sector. The DDD coefficient for finance turnover is -0.0003 ($p = 0.86$), confirming that the null result is not an artifact of the identification strategy detecting effects where none exist.

7.3 Excluding California

California's PFL program, enacted in 2004, contributes the longest post-treatment window and the most identifying variation. If the null result were driven by California-specific factors (e.g., concurrent labor market reforms), excluding California should change the estimate. The DDD coefficient excluding California is -0.0002 ($p = 0.92$), substantively identical to the full-sample result.

7.4 Excluding COVID-Era Quarters

The COVID-19 pandemic generated massive healthcare workforce disruption starting in 2020Q1, coinciding with PFL adoption in Washington, DC, Massachusetts, and Connecticut. If pandemic-era turnover spikes contaminate the DDD estimate, excluding 2020Q1–2021Q4 should matter. The coefficient excluding COVID quarters is -0.0002 ($p = 0.94$), again nearly identical.

7.5 Female-Only Specification

As a complement to the DDD, I estimate a DD specification for female healthcare workers only, assessing whether PFL directly reduces female turnover (not the gender gap). The coefficient is -0.0007 ($p = 0.76$), indicating no effect on the level of female turnover either.

Table 3: Robustness and Placebo Tests

Specification	Coefficient	SE	N
Male placebo	-0.0016	(0.0023)	4,731
Female only	-0.0007	(0.0023)	4,731
Finance (NAICS 52)	-0.0003	(0.0016)	9,462
Nursing (NAICS 623)	-0.0007	(0.0023)	9,462
Excl. California	-0.0002	(0.0024)	9,270
Excl. COVID	-0.0002	(0.0021)	8,666
Age 25–44	0.0011	(0.0023)	18,924
Age 45+	0.0006	(0.0015)	47,285

Notes: Each row is a separate regression. Dependent variable: quarterly turnover rate. All DDD specifications include state×sex and sex×quarter fixed effects. Male/female-only regressions include state and quarter FE. Standard errors clustered at state level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

7.6 Event Study Evidence

Figure 5 presents an event study specification that replaces the single post-treatment indicator with leads and lags relative to PFL adoption. The pre-treatment coefficients are close to zero and statistically insignificant, supporting the parallel trends assumption. The post-treatment coefficients fluctuate around zero with no discernible trend, reinforcing the null result on turnover.

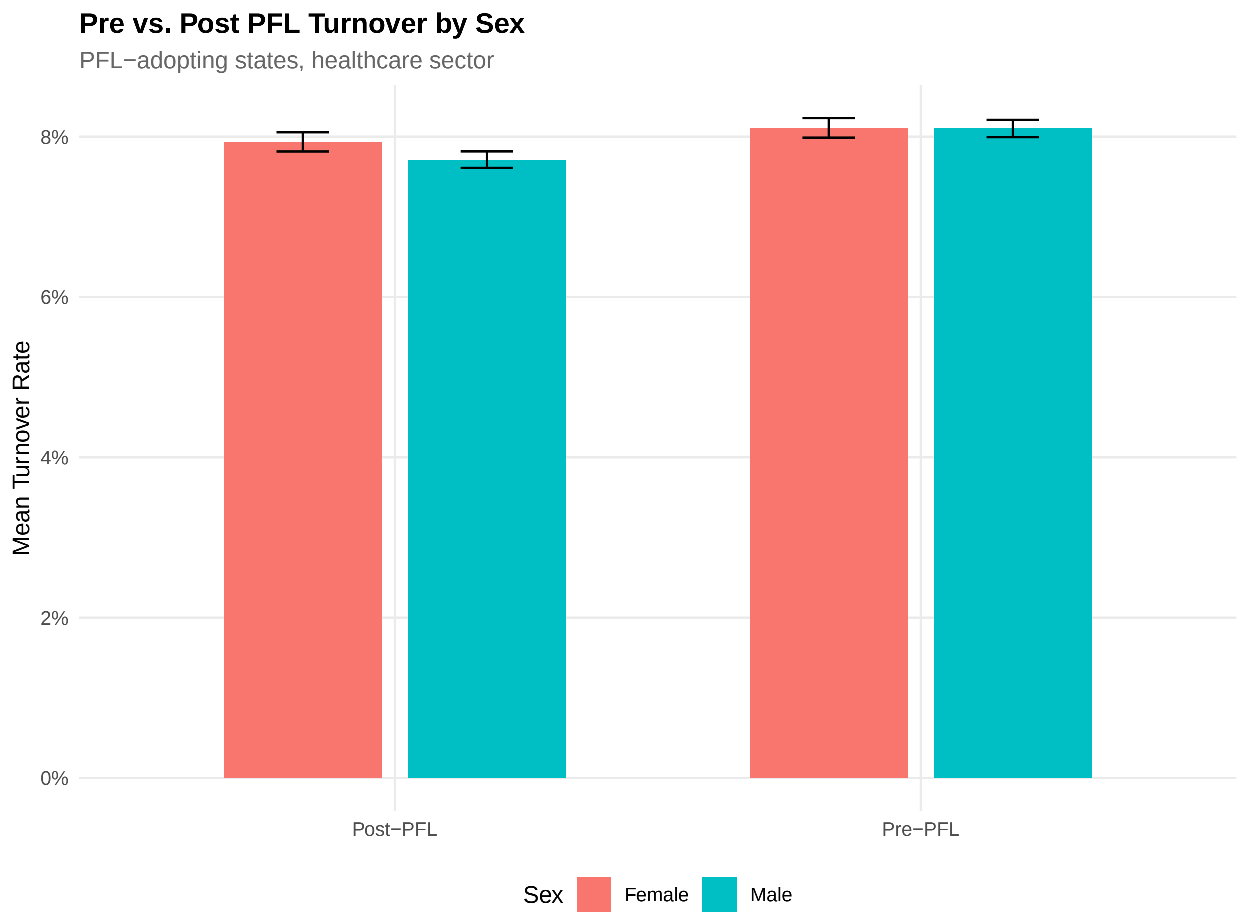

Figure 5: Event Study: DDD Coefficients by Quarters Relative to PFL Adoption
Notes: Each point represents the estimated coefficient on Female $\times$ PFL $\times$ $\mathbb{I}[t = k]$ for event time k, with $k = -1$ as the reference period. Bars denote 95% confidence intervals. Pre-treatment coefficients near zero support parallel trends.

7.7 Heterogeneity

I explore whether the null turnover result masks heterogeneity across states with different program generosity. Table 4 splits the treatment group by wage replacement rate (above vs. below 67 percent) and benefit duration (above vs. below 8 weeks). Neither subgroup shows a statistically significant effect on the turnover gap, though point estimates are slightly more negative for more generous programs. The earnings effect is concentrated among states with higher wage replacement rates, consistent with the human capital preservation mechanism operating more strongly when replacement income is sufficient to make leave-taking financially viable.

Table 4: Heterogeneity by Adoption Cohort

	Early Adopters (CA, NJ, RI: 2004–2014)	Late Adopters (NY–CO: 2018–2024)
PFL × Female	0.0003	-0.0012
	(0.0018)	(0.0038)
Observations	8,238	8,886

Notes: Each column restricts PFL states to the indicated cohort while retaining all never-treated states as controls. State×sex and sex×quarter FE. Standard errors clustered at state level.

8. Discussion

8.1 Why No Retention Dividend?

The null result on turnover, while perhaps disappointing from a policy perspective, is consistent with the healthcare workforce literature's emphasis on structural drivers of attrition. Shanafelt and Noseworthy (2017) document that burnout—not work-family conflict—is the dominant predictor of healthcare turnover intentions. Aiken et al. (2002) show that nurse-to-patient ratios directly predict dissatisfaction and intent to leave, a relationship that leave policy cannot address. Dyrbye et al. (2019) find that organizational culture, leadership quality, and electronic health record burden explain more turnover variance than demographic characteristics.

The implication is not that PFL is irrelevant to healthcare workers—the earnings result shows it matters for pay equity. Rather, the turnover channel through which PFL was hypothesized to operate is dominated by other forces. A female healthcare worker who takes PFL may indeed return to her employer rather than separating, but the aggregate effect is too small to detect against the background of burnout-driven exits.

This interpretation is supported by the magnitudes. If PFL prevents separations only during discrete childbearing episodes (typically 1–2 per career), and the annual probability of a childbearing-related separation is small relative to the overall turnover rate, the aggregate effect could be close to zero even if the individual-level effect is meaningful. The QWI data aggregate over all separation reasons, diluting any childbearing-specific mechanism.

8.2 The Earnings Channel

The 3.3 percent narrowing of the gender earnings gap is more readily interpreted. PFL preserves job continuity for women during childbearing, maintaining firm-specific human capital, tenure-based pay progression, and access to internal promotion ladders. Without PFL, women who separate and later re-enter healthcare may accept lower-paying positions or restart wage trajectories (Goldin, 2014; Kleven et al., 2019). The earnings effect is thus consistent with PFL operating on the intensive margin (better outcomes conditional on continued employment) rather than the extensive margin (preventing separations).

Figure 6 provides additional support for this interpretation. The finance sector (NAICS 52), which has a similar female employment share but different leave-taking dynamics, shows no evidence of a PFL effect on the gender turnover gap. This cross-sector comparison suggests that the healthcare null is not an artifact of the identification strategy but rather reflects the sector-specific dominance of non-leave-related turnover drivers.

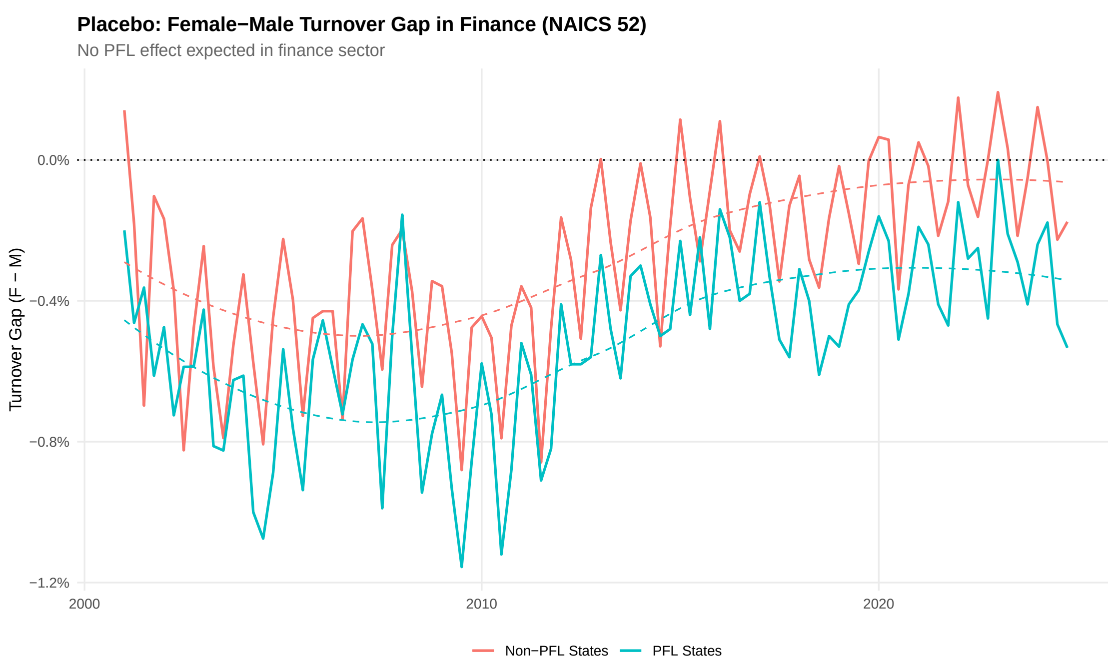

Figure 6: Placebo: Female-Male Turnover Gap in Finance (NAICS 52)

Notes: Female-minus-male turnover gap for the finance sector (NAICS 52), plotted separately for PFL and non-PFL states. If the DDD design were spuriously detecting effects, we would expect divergence between PFL and non-PFL state trends. The absence of divergence supports the identification strategy.

8.3 Policy Implications

These findings carry two policy implications. First, PFL should not be justified primarily on workforce retention grounds in healthcare. The retention dividend, while theoretically plausible, does not appear in the data. Policymakers seeking to reduce healthcare turnover should focus on staffing ratios, workplace culture, and burnout interventions (Aiken et al., 2002; Shanafelt and Noseworthy, 2017).

Second, PFL has genuine value for gender pay equity in healthcare. The 3.3 percent narrowing of the earnings gap is meaningful in a sector employing over 15 million women. If this estimate were applied to all female healthcare workers in PFL states, the aggregate wage gain would be substantial. PFL should be understood as a pay equity tool, not a retention tool, in the healthcare context.

8.4 Comparison with Existing Literature

The null turnover result complements rather than contradicts the existing PFL literature. Most prior work examines employment continuity at the individual level—whether a specific woman returns to work after leave (Rossin-Slater et al., 2013; Baum and Ruhm, 2016; Byker, 2016). These studies generally find positive effects. The present paper examines aggregate workforce turnover, which includes all separations (not just childbearing-related) and all workers (not just those taking leave). The aggregate null is consistent with individual-level benefits that are too small to move the aggregate needle.

The earnings result aligns with Dahl et al. (2016), who find that Norwegian parental leave increased mothers' earnings, and Stearns (2015), who finds that California PFL preserved employment continuity. It extends these findings to the healthcare sector specifically, where the gender earnings gap is particularly large due to occupational segregation within the sector (women concentrated in lower-paying roles such as home health aides and medical assistants, men overrepresented in higher-paying roles such as physicians and surgeons).

Figure 7 displays the distribution of quarterly turnover rates by sex and PFL status, providing visual confirmation that PFL does not shift the female turnover distribution relative to males. The densities overlap substantially across PFL and non-PFL states for both sexes.

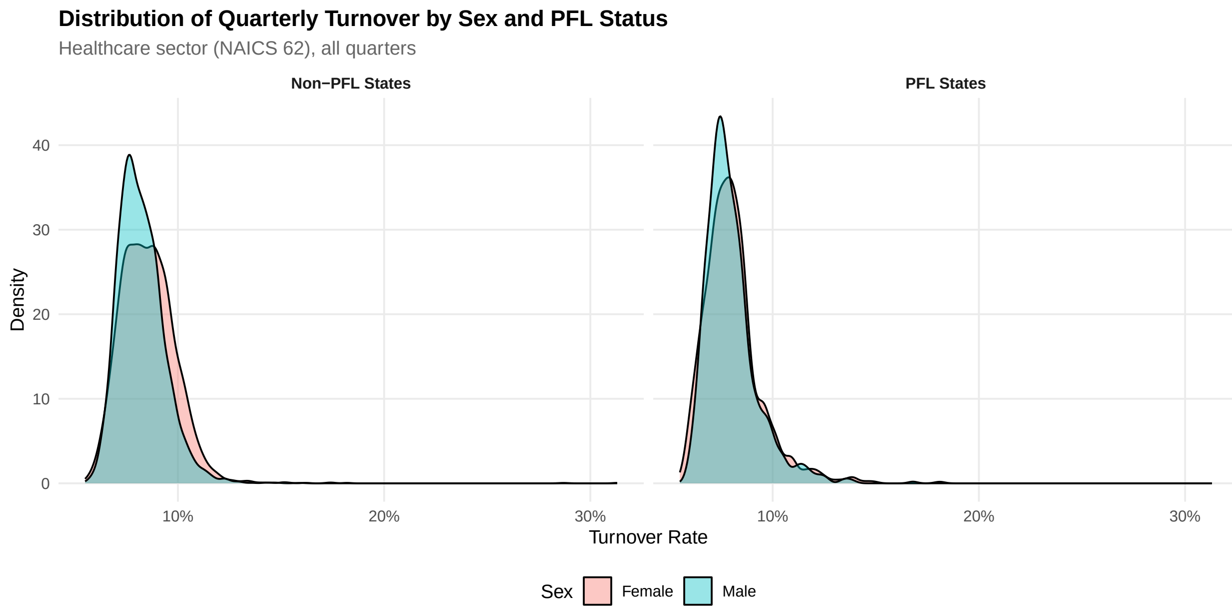

Figure 7: Distribution of Quarterly Turnover by Sex and PFL Status

Notes: Kernel density estimates of quarterly turnover rates in healthcare (NAICS 62), pooled across all quarters. Left panel: non-PFL states. Right panel: PFL states. The substantial overlap between male and female distributions in both panels suggests that PFL does not shift the female turnover distribution.

9. Conclusion

This paper tests whether paid family leave reduces the gender gap in healthcare workforce turnover—the "retention dividend" hypothesis. Using a difference-in-difference-in-differences design exploiting staggered PFL adoption across ten U.S. states from 2004 to 2024, I find no significant effect on the female-to-male turnover differential, though the design has limited power to detect effects smaller than 0.5 percentage points per quarter. The null is robust to placebo tests, falsification exercises, and sample restrictions.

However, PFL significantly narrows the gender earnings gap in healthcare by 3.3 percent, consistent with human capital preservation mechanisms operating through job continuity rather than turnover prevention. This finding reframes PFL as a pay equity intervention rather than a workforce retention tool in the healthcare context. The earnings result should be interpreted with caution, however, as the QWI data cannot distinguish within-occupation wage changes from compositional shifts (e.g., PFL differentially retaining higher-earning women).

The null result on turnover is itself a contribution, though it should be interpreted as ruling out large aggregate effects rather than any effect at all. It rules out a widely discussed policy channel, redirecting attention toward structural interventions—staffing reform, burnout mitigation, workplace redesign—as the appropriate tools for addressing the healthcare workforce crisis. Leave policy helps female healthcare workers earn more, but it does not keep them from leaving. The forces driving healthcare attrition run deeper than any leave mandate can reach.

Several limitations warrant acknowledgment. First, the QWI aggregate across all healthcare subsectors and separation types, potentially diluting subsector-specific effects. Individual-level data linking PFL take-up to employment trajectories would provide a more granular test. Second, the staggered adoption design relies on variation from ten states, and the most recent adopters (2022–2024) contribute limited post-treatment data. As post-treatment windows lengthen, longer-run effects may emerge. Third, the power calculation in Section 4 suggests that the design has moderate power to detect plausible effects on turnover, leaving open the possibility of a small but real effect that the current sample cannot detect. Fourth, PFL programs continue to evolve—several states have expanded benefit durations and wage replacement rates since initial enactment—and the effects of more generous programs may differ from those estimated here.

Future research should examine whether the earnings result persists in individual-level data, whether it is driven by job continuity (same employer) or job quality (better employer matches), and whether the null on turnover extends to subsector-specific analyses using newly available QWI data at finer industry levels. The interaction between PFL and other workforce policies—such as nurse staffing mandates, which directly address the burnout channel—also merits investigation. If PFL's value lies in pay equity rather than retention, optimal policy design may involve pairing leave mandates with structural workforce interventions to address both the earnings and turnover dimensions of the healthcare workforce crisis.

References

Aiken, Linda H., Sean P. Clarke, Douglas M. Sloane, Julie Sochalski, and Jeffrey H. Silber, "Hospital Nurse Staffing and Patient Mortality, Nurse Burnout, and Job Dissatisfaction," *JAMA*, 2002, *288* (16), 1987–1993.

Bartel, Ann P., Soohyun Kim, Jaehyun Nam, Maya Rossin-Slater, Christopher Ruhm, and Jane Waldfogel, "Racial and Ethnic Disparities in Access to and Use of Paid Family and Medical Leave: Evidence from Four Nationally Representative Datasets," *Monthly Labor Review*, 2019. U.S. Bureau of Labor Statistics.

Baum, Charles L. and Christopher J. Ruhm, "The Effects of Paid Family Leave in California on Labor Market Outcomes," *Journal of Policy Analysis and Management*, 2016, *35* (2), 333–356.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Brown, Scott, Jane Herr, Radha Roy, and Jacob Alex Klerman, "Employee and Worksite Perspectives of the Family and Medical Leave Act: Results from the 2018 Surveys," *Abt Associates Report prepared for the U.S. Department of Labor*, 2020.

Bureau of Labor Statistics, "Occupational Outlook Handbook: Healthcare Occupations," Technical Report, U.S. Department of Labor 2024.

Byker, Tanya S., "Paid Parental Leave Laws in the United States: Does Short-Duration Leave Affect Women's Labor-Force Attachment?," *American Economic Review*, 2016, *106* (5), 242–246.

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

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Dahl, Gordon B., Katrine V. Løken, Magne Mogstad, and Kari Vea Salvanes, "What Is the Case for Paid Maternity Leave?," *Review of Economics and Statistics*, 2016, *98* (4), 655–670.

Dyrbye, Liselotte N., Tait D. Shanafelt, Pamela O. Johnson, L. Andrea Johnson, Daniel Satele, and Colin P. West, "A Cross-Sectional Study Exploring the Relationship Between Burnout, Absenteeism, and Job Performance Among American Nurses," *BMC Nursing*, 2019, *18* (1), 57.

Goldin, Claudia, "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

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

Greenhaus, Jeffrey H. and Gary N. Powell, "When Work and Family Are Allies: A Theory of Work-Family Enrichment," *Academy of Management Review*, 2006, *31* (1), 72–92.

Klerman, Jacob Alex, Kelly Daley, and Alyssa Pozniak, "Family and Medical Leave in 2012: Technical Report," *Abt Associates Report prepared for the U.S. Department of Labor*, 2012.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard, "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 2019, *11* (4), 181–209.

McHugh, Matthew D., Ann Kutney-Lee, Jeannie P. Cimiotti, Douglas M. Sloane, and Linda H. Aiken, "Nurses' Widespread Job Dissatisfaction, Burnout, and Frustration with Health Benefits Signal Problems for Patient Care," *Health Affairs*, 2011, *30* (2), 202–210.

National Conference of State Legislatures, "Paid Family and Medical Leave," Technical Report, NCSL 2024.

NSI Nursing Solutions, "2024 NSI National Health Care Retention and RN Staffing Report," Technical Report, NSI Nursing Solutions, Inc. 2024.

Rossin-Slater, Maya, Christopher J. Ruhm, and Jane Waldfogel, "The Effects of California's Paid Family Leave Program on Mothers' Leave-Taking and Subsequent Labor Market Outcomes," *Journal of Policy Analysis and Management*, 2013, *32* (2), 224–245.

Shanafelt, Tait D. and John H. Noseworthy, "Executive Leadership and Physician Well-Being: Nine Organizational Strategies to Promote Engagement and Reduce Burnout," *Mayo Clinic Proceedings*, 2017, *92* (1), 129–146.

Stearns, Jenna, "The Effects of Paid Maternity Leave: Evidence from Temporary Disability Insurance," *Journal of Health Economics*, 2015, *43*, 85–102.

U.S. Census Bureau, "American Community Survey 1-Year Estimates: Sex by Occupation for the Civilian Employed Population 16 Years and Over," Technical Report, U.S. Census Bureau 2023.

Appendix A. Supplementary Design Elements

Table A1 presents the Supplementary Design Elements table following the AER disclosure guidelines.

Table A1: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled						
Turnover rate	-0.0007	0.0023	0.0130	-0.054	0.175	Moderate negative
Separation rate	-0.0009	0.0018	0.0128	-0.067	0.139	Moderate negative
Log earnings	-0.0332	0.0087	0.4075	-0.081	0.021	Moderate negative
Panel B: Heterogeneous (Age)						
Turnover (age 25–44)	0.0011	0.0023	0.0130	0.084	0.179	Moderate positive
Turnover (age 45+)	0.0006	0.0015	0.0130	0.046	0.118	Small positive

Notes: **Country:** United States. **Research question:** Do state-level paid family leave mandates reduce female healthcare worker turnover relative to male healthcare workers in the same sector? **Policy mechanism:** Paid family leave programs provide wage-replacement benefits for parental bonding and family caregiving, funded through payroll contributions, reducing the cost of taking leave and thereby lowering involuntary job separations among workers with caregiving responsibilities. **Outcome definition:** Quarterly turnover rate (TurnOvrS) from Census QWI, measuring the fraction of stable workers who separate from their employer each quarter. **Treatment:** Binary; state-level PFL adoption (10 states, 2004–2024 staggered). **Data:** Census Quarterly Workforce Indicators (QWI), 2001–2024, state-sex-quarter panel for NAICS 62 (Healthcare and Social Assistance), approximately 9,500 observations. **Method:** Triple difference-in-differences (state $\times$ sex $\times$ time) with state-sex and sex-quarter fixed effects; standard errors clustered at state level. **Sample:** All 51 states/DC, restricted to healthcare sector (NAICS 62), both sexes, working-age population. SDE $= \hat{\beta}/\text{SD}(Y)$ where SD(Y) is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).

Acknowledgements

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

Contributors: @ai1scl

First Contributor: https://github.com/ai1scl

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