Who Bears the Burden? Heterogeneous Labor Market Penalties of Child and Eldercare*

Shinnosuke Kikuchi†
UCSD

February 26, 2026

Abstract

This paper investigates the labor market effects of family caregiving, focusing on childcare and eldercare. Using large panel data in Japan and an event-study design that accounts for staggered treatment timing, I find sizable and persistent employment penalties for females after childbirth. Mothers' employment decreases by 36 percentage points one year after childbirth and stays lower five years later by approximately 19 percentage points. These effects vary by job characteristics, contract type, and co-residence status, highlighting substantial heterogeneity. In contrast, eldercare has smaller, at most 5 percentage points, and often statistically insignificant average effects on employment. However, eldercare penalties are larger and statistically significant, reaching up to 10 percentage points, for females with certain pre-event characteristics: those in low-teleworkability jobs, those in high physical proximity jobs, those on non-regular contracts, and those employed in small firms.

*I thank Minamo Mikoshiba and seminar participants at Recruit Works Institute and Recruit Co., Ltd for valuable comments. I also thank the anonymous referees for their helpful comments that improved the quality of the manuscript. The data for this analysis, Japanese Panel Study of Employment Dynamics (JPSED, Recruit Works Institute), was provided by the Social Science Japan Data Archive, Center for Social Research and Data Archives, Institute of Social Science, The University of Tokyo. The data for this analysis, Japan Household Panel Survey (JHPS/KHPS), was provided by the Panel Data Research Center, Institute for Economic Studies, Keio University. I thank Daiji Kawaguchi and Sagiri Kitao for their help with data access. The views expressed in this paper are my own and do not represent the views of organizations with which I am affiliated. All remaining errors and omissions are my own.

†Email: shkikuchi@ucsd.edu

1 Introduction

In developed countries, workforces are graying, fertility rates are below the replacement level, and female labor force participation has reached record levels (Acemoglu et al., 2022; Organisation for Economic Co-operation and Development (OECD), 2024; International Labour Organization, 2018). These changes highlight two career challenges: caring for new children and caring for aging parents. Child penalties are well mapped–the atlas of Kleven et al. (2024) shows earnings losses for mothers in every country it covers using event study designs. Evidence on eldercare penalties is sparse.

This paper studies the impact of family caregiving responsibilities, specifically related to childcare and eldercare, on labor market outcomes using event study designs following Callaway and Sant'Anna (2021). This study's empirical setting is Japan, one of the most rapidly aging societies. I use data from the Japanese Panel Study of Employment Dynamics (JPSED), which tracks roughly 50,000 adults per year between 2016 and 2024, and offers a richer sample than most national longitudinal surveys by recording jobs, earnings, occupations, and family care. I supplement the JPSED data with another longitudinal dataset from the Japan Household Panel Survey (JHPS), which has longer panels (2004 to 2021) with smaller cross-sectional samples, which are about 2,000 per year. I also use data on task contents by occupation from the Japan O-NET data. These data let us follow labor-market outcomes around both care events with fine detail.

First, the arrival of a first child consistently imposes significant and persistent penalties on females' labor market outcomes, while males experience no statistically detectable effects. This is consistent with the previous literature for many countries, including Japan (Kleven et al., 2024). Female employment rates decline sharply, with the JPSED data showing a peak drop of approximately 0.36 points in year +1, and the JHPS data indicating an even larger peak decline of about 0.76 points at year +1. Females' annual earnings also fall substantially, with losses of approximately 94 (in JPY 10K units) in JPSED starting from year +1, and a larger reduction of around 140 (in JPY 10K units) in JHPS from year +1. Weekly hours worked for females decrease by approximately 5 hours in JPSED. The reduction persists for several years post-birth. In JHPS, the estimates are less precise, but the decline is clear from year 2 onwards. Heterogeneous analyses reveal that this childcare penalty is not uniform: mothers in high-teleworkability jobs prior to childbirth do not experience declines in employment rates beyond year 0. Mothers with regular employment contracts prior to childbirth experience a substantially reduced penalty, with non-regular workers showing markedly larger employment declines. Co-residence with parents mitigates longer-term employment loss.

Second, I find that the average effects of entering eldercare on employment, earnings, and hours are close to zero for both genders. For both males and females, employment changes remain within ± 0.05 and are never statistically significant in either dataset. Their earnings and hours losses are not statistically significant in the JPSED samples. Post-event females' declines in earnings and hours worked are statistically significant in the JHPS sample for some later periods, but the sizes are mostly modest. Heterogeneous analyses indicate that the eldercare penalty on female employment is not uniform. Females in high-teleworkability jobs, those with high intensity in face-to-face discussion, those with low physical proximity requirements, those with regular contracts, and those in large firms do not experience statistically significant decreases in employment rates.

Comparing these two caregiving responsibilities, the childcare penalty on female labor market outcomes is substantially larger and more consistently significant than the eldercare penalty. The employment declines for mothers after childbirth are several times greater (e.g., 0.37 to 0.78 points) than the maximum observed employment dip for elder care (0.05 points). Similarly, earning losses for mothers after childbirth (about 100 to 150 in JPY 10K units) far exceed the small earning losses observed for females in elder care (at most 30 in JPY 10K units).

Literature. This paper contributes to two strands of the literature. First, a rich body of research documents large and persistent "child penalties"—the adverse labor market effects of having children on mothers' careers. Across many countries, females' employment and earnings trajectories diverge sharply from males' after the birth of a first child, while fathers' outcomes remain essentially unchanged. Using Danish administrative data, Kleven et al. (2019) show that mothers' earnings drop by about 20% relative to fathers and never fully recover over the subsequent decade. Swedish evidence from a within-couple event-study design finds a comparable pattern: 15 years after childbirth, mothers' annual earnings are 32% lower than their partners' on average, whereas fathers' earnings are virtually unaffected (Angelov et al., 2016). In fact, extensive cross-country evidence confirms that sharp, long-lasting female employment losses after childbirth are the norm across high-, middle-, and low-income economies (Kleven et al., 2024).

Recent studies have investigated the mechanisms behind these motherhood penalties. In particular, employer practices and career dynamics within firms can amplify the long-term gender gap. For instance, Okuyama et al. (2025) use personnel records from a large Japanese firm to decompose the motherhood pay gap. They find that immediately after childbirth, mothers' earnings losses are driven mainly by reductions in time-based pay (e.g., lower base salary due to shorter hours), but over time, a widening gap in promotion-based pay becomes the dominant factor.

Other papers also document a large and persistent child (motherhood) penalty in Japan. Using administrative tax records and large-scale panel surveys, studies such as Fukai and Kondo (2025), Hsu (2021), Mugiyama (2024), and Komura (2022) find substantial declines in women's labor market outcomes around first birth, including employment, earnings/income, and (to a lesser extent) hours worked.[1] My estimates for childbirth in the JPSED and JHPS are consistent with this evidence: employment and income fall sharply after the first birth, and the adjustment is concentrated on the extensive margin, with comparatively smaller changes in hours conditional on employment.

This paper contributes to this literature by replicating the overall childcare penalty pattern in Japan using an event-study approach and by showing substantial heterogeneity across groups of individuals with different pre-birth characteristics (such as job task content, co-residence with parents, or employment contract type).

Second, this paper contributes to the literature on the labor market impacts of eldercare. In contrast to the extensive research on new mothers, the labor market impacts of caring for aging parents (i.e. unpaid *eldercare*) have been less thoroughly quantified, and the evidence to date is more mixed. Notably, virtually no prior work has directly compared childcare and eldercare penalties within the same empirical framework—a gap this paper addresses by examining both types of caregiving side-by-side.

Many earlier studies using panel data found at most modest average effects of informal eldercare on employment. For example, Bolin et al. (2008) analyze adults aged 50+ in several European countries and estimate that providing care to an elderly parent is associated with only a very small reduction in the caregiver's employment probability (on the order of a few percentage points). Similarly, Oshio and Usui (2018) report that when Japanese females begin providing care to an older parent, their labor force participation drops by roughly 3% on average, with no significant change in hours worked among those who remain employed.[2]

More recent evidence suggests that eldercare responsibilities can indeed hinder females' careers under certain conditions. Using U.S. administrative data, Maestas et al. (2024) document divergent trajectories by gender around a parent-care event: when daugh-

[1]Much of this literature implements pooled event-study designs in the spirit of the Kleven et al. (2019) approach. While informative, conventional pooled TWFE event studies can embed problematic comparisons under staggered timing when treatment effects are heterogeneous. For this reason, I use the Callaway and Sant'Anna (2021) estimator with not-yet-treated comparisons and show in Appendix E that the estimates are sensitive to estimator choice.

[2]For other studies using Japanese panel data on eldercare and labor outcomes, see, for example, Fukahori et al. (2015); Kikuzawa and Uemura (2021); Oshio and Usui (2017). See also Kohara and Ohtake (2011); Yamada and Shimizutani (2015).

ters start providing care to an elderly parent, their employment and earnings tend to drop only at year +1, whereas sons show little change at caregiving onset (often because sons step in only after their own careers have been interrupted for other reasons). However, emerging econometric research cautions that the estimators used in earlier studies may be biased when treatment timing is staggered and effects are heterogeneous. In particular, conventional difference-in-difference designs can produce biased estimates in such settings (Baker et al., 2022; Borusyak et al., 2024; Callaway and Sant'Anna, 2021; De Chaisemartin and d'Haultfoeuille, 2023; Roth et al., 2023).[3]

This paper's contribution is to apply an event-study design that accounts for staggered treatment timing to the context of eldercare, providing a more credible estimate of causal effects. Following Callaway and Sant'Anna (2021), I implement an event-study approach for caregiving onset and find that the average effect of an elderly parent's care need on females' employment is close to zero and often not statistically significant.[4] Importantly, this null average effect masks considerable heterogeneity: the analysis reveals that caregiving penalties vary markedly by individual and job characteristics. In particular, females in less flexible jobs (e.g., roles with low teleworkability or high physical proximity), those on non-regular contracts, those employed in smaller firms, and those not cohabiting with their parents experience significantly larger employment declines when taking on parent-care duties. By contrast, females with more flexible work arrangements or additional support show relatively negligible labor impacts. In sum, different types of family care impose different constraints on workers, underscoring the novel insight that childcare and eldercare may require distinct policy responses.

2 Context and Data

2.1 Institutional Context: Japan's LTCI Certification System

Japan introduced public long-term care insurance (LTCI) in 2000 to respond to rapid population ageing and to reduce the burden of family-based eldercare (Tamiya et al., 2011). The system is mandatory and nationally regulated, but municipalities administer eligibility and service delivery. A key feature for this paper is that eligibility is determined through a formal care-need certification process rather than informal family reporting

[3] Notably, Oshio and Usui (2017) avoid this staggered-timing bias by using only two survey waves. Their IV and fixed-effects estimates of caregiving's impact on employment are statistically insignificant, consistent with the present study's findings.

[4] This negligible average eldercare effect is in line with the findings of Oshio and Usui (2017), who also report no significant impact on females' employment.

alone.

In practice, certification is decided through a standardized multi-step assessment. Applicants are first evaluated using a detailed questionnaire on physical and cognitive functioning, then assigned a preliminary category by algorithm, and finally reviewed by a local expert committee (Tamiya et al., 2011). Certified individuals are classified into seven levels, commonly grouped as lighter support-need levels (Yoshien 1–2) and more intensive care-need levels (Yokaigo 1–5). Higher levels imply a higher monthly ceiling of publicly covered in-kind services and, in turn, greater expected time demands for care coordination and family involvement.

This institutional structure helps interpret my treatment definition. In both datasets, event time zero for eldercare corresponds to the first year a parent or parent-in-law is reported as newly certified or newly needing care. Because the first certification can occur at either light or severe levels, onset does not mechanically represent the same caregiving burden across households. For this reason, I complement the baseline onset analysis with evidence on severity at onset and by event time, using the richer care-level information available in the JHPS.

2.2 Japanese Panel Study of Employment Dynamics

My primary data source is the Japanese Panel Study of Employment Dynamics (JPSED), an annual survey of working-age adults conducted each January. I use the 2016–2024 waves; each wave interviews about 50,000 respondents. I restrict samples of individuals aged between 20 and 59, not in schools.

JPSED records age, gender, education, marital status, annual earnings, weekly hours, employment status, employer size, and a three-digit occupation code with more than 200 categories. The large panel sharpens precision relative to most national datasets; its yearly sample is about five times the size of the U.S. Panel Study of Income Dynamics (PSID). Detailed occupation codes let me link jobs to task measures and examine heterogeneity by work content.

For childcare analysis, I use children's ages to identify the first year a respondent has a child; that year is event time zero for childbirth. For eldercare analysis, I use a question that asks whether a parent or parent-in-law received long-term-care certification in the previous year. I set event time zero for elder care in the first year, where this indicator equals one.

Table 1 reports weighted means for the JPSED sample in 2016 and 2024. Male employment rate is high—93.2% in 2016 and 93.2% in 2024—while the female rate rises from

74.1% to 80.1%, shrinking but not closing the gender gap. Regular-worker shares are constant at around 87% for males and at around 47 to 49% for females. College attainment increases for both genders, from 36.2% to 39.5% among males and from 19.4% to 26.6% among females. Care events remain infrequent: roughly 1 to 2% of respondents experience a first birth each year, and new elder-care responsibilities affect about 1.5% of males and decline from 3 to 2% among females. Average earnings and hours worked have been stable for both genders.[5]

Table 1: Summary Statistics (JPSED)

	Males		Females	
	2016	2024	2016	2024
Age	42.7	43.6	42.7	44.0
Employment Rate (%)	93.2	93.2	74.1	80.1
Regular Worker Share (%)	86.4	87.5	47.1	49.2
College Educated (%)	36.2	39.5	19.4	26.6
Start Childcare (%)	1.7	2.1	1.9	2.3
Start Eldercare (%)	1.6	1.6	3.1	2.3
Annual Earnings (JPY 10K)	470.4	463.4	219.7	236.8
Weekly Hours Worked	44.0	42.1	33.2	32.6
Num. of Obs.	14404	15450	14987	14083

Note: This table shows weighted means of employment status, contract type, education, caregiving indicators, annual earnings (JPY 10K, 2020 price), and weekly hours for respondents aged 20–59 in the JPSED data. The weight is a survey weight.

Table 2 reports pre-event characteristics for the event-study samples, separately for childbirth and eldercare. For each event type, I restrict attention to respondents who eventually experience the event during the sample period (ever-treated), keep only pre-event observations ($t \leq -1$), and split them into "Not-yet-treated" years ($t \leq -2$) and the "Just treated" reference period ($t = -1$). Overall, the means of non-labor market, demographic variables are very similar across the not-yet-treated and just-treated groups for both childbirth and eldercare, including ages, and college education. The main exception is partner status in the childbirth sample: the share with a partner is substantially higher at $t = -1$ than in earlier pre-event years, which is consistent with family formation and marriage often occurring around the time of first birth. In the event-study specifications, I therefore control for partner status (along with age and age squared), so differences in partnership rates are absorbed by the covariates rather than confounding the estimated

[5] Annual earnings are in real terms in 2020's price, using the consumer price index for all goods. The data source is Statistics Bureau of Japan (2025).

dynamic treatment effects. In addition, labor market variables, such as employment rates, annual earnings, and weekly hours worked, are not different across not-yet-treated and just-treated groups.

Table 2: Summary Statistics (JPSED)

	Childcare		Eldercare	
	Not-yet-treated	Just treated	Not-yet-treated	Just treated
Age	31.8	33.0	48.5	48.9
College Educated (%)	47.2	45.2	29.6	28.7
Share of Having Partners (%)	72.5	95.1	67.7	68.4
Regular Worker Share (%)	83.5	82.9	61.7	61.1
Employment Rate (%)	89.5	85.7	81.8	81.3
Annual Earnings (JPY 10K)	374.9	394.3	369.1	354.1
Weekly Hours Worked	40.8	39.7	37.7	37.1
N individuals	4193	5111	3822	3878
N person-years	10131	5111	11152	3878

Note: This table shows means of demographic variables for those aged 20–59 in the JPSED data. The samples are restricted to respondents who eventually experience the corresponding event (childbirth or eldercare onset) during the sample period (i.e., ever-treated). For each event type, I further restrict the sample to pre-event observations only ($t \leq -1$), and split them into two groups: "Not-yet-treated" observations with $t \leq -2$, and "Just treated" observations at $t = -1$. N individuals report the number of unique respondents in each group, and N person-years report the number of respondent-year observations.

2.3 Japan Household Panel Survey

My supplementary data source is the Japan Household Panel Survey (JHPS), an annual household panel that began in 2004. I use the 2004–2021 waves; each wave interviews about 7,000 households. Since the variables for eldercare have been available since 2009. Thus, I use the 2009-2021 wave when I analyze the eldercare penalty. I restrict samples of individuals aged between 20 and 59, not in schools.

JHPS records age, gender, education, marital status, children's years of birth, earnings, hours, employment status, and caregiving details. The 18-year span lets me track earnings and care events over nearly two decades. Its caregiving module provides information not available in JPSED.

For childcare analysis, I use children's years of birth to set event time zero at the first year a respondent has a newborn. For eldercare analysis, I use the item "Family member needing care." and the item to identify which family member newly needs LTC. I construct a dummy that equals one in the first year when a parent or parent-in-law is reported as needing care; that year is event time zero for eldercare.

Table 3 reports comparable statistics for the JHPS sample. In the overlapping year 2016, employment rates are almost the same as the JPSED samples: 95.0% for males (93.2% in JPSED) and 74.8% for females (74.1% in JPSED). The share of regular workers is lower than in the JPSED sample: 73.8% for males (86.4% in JPSED) and 26.0% for females (47.1% in JPSED). The fraction of college-educated respondents is higher than in the JPSED sample: 48.6% for males (36.2% in JPSED) and 33.3% for females (19.4% in JPSED). Annual earnings are higher in the JHPS sample for males, while hours worked are almost identical to those in the JPSED samples.

Table 3: Summary Statistics (JHPS)

	Males			Females		
	2004	2016	2021	2004	2016	2021
Age	35.3	42.8	42.5	32.9	41.2	41.4
Employment Rate (%)	92.9	95.0	93.7	63.2	74.8	78.4
Regular Worker Share (%)	63.9	73.8	73.5	27.6	26.0	32.1
College Educated (%)	41.7	48.6	55.5	26.1	33.3	39.6
Start Childcare (%)	3.4	1.0	0.3	2.3	0.8	0.5
Start Eldercare (%)	.	3.3	1.3	.	3.9	1.6
Annual Earnings (JPY 10K)	426.3	535.2	527.3	215.6	212.3	230.9
Weekly Hours Worked	47.3	45.0	42.2	32.2	31.1	29.5
Num. of Obs.	593	988	1148	555	861	1118

Note: This table shows weighted means of employment status, contract type, education, caregiving indicators, annual earnings (JPY 10K, 2020 price), and weekly hours for respondents aged 20–59 in the JHPS data. The weight is a survey weight. In 2004, the eldercare question was not asked.

Table 4 reports a comparable table to Table 2 for JHPS samples. Once again, all pre-event, non-labor market characteristics are similar between not-yet-treated and just-treated samples, except for the partner status.[6]

2.4 Japan O-NET

Finally, I also use Japan O-NET, an occupational information network maintained by the Ministry of Health, Labour and Welfare and the Japan Institute for Labour Policy and

[6] The share of having partners in the JHPS sample, 61.6%, is smaller than expected. Nevertheless, the average of the share of having partners at t=0 for childcare exceeds 87%, so that the difference might come from sampling frames, not from the noisy measurement.

Table 4: Summary Statistics (JHPS)

	Childcare		Eldercare	
	Not-yet-treated	Just treated	Not-yet-treated	Just treated
Age	29.1	30.9	40.8	41.9
College Educated (%)	48.8	53.9	42.3	45.0
Share of Having Partners (%)	23.3	61.6	69.0	64.8
Regular Worker Share (%)	69.7	67.3	46.0	49.9
Employment Rate (%)	95.0	92.9	80.7	85.8
Annual Earnings (JPY 10K)	349.8	368.6	428.1	398.9
Weekly Hours Worked	43.6	40.6	40.5	38.9
N individuals	227	297	652	725
N person-years	913	297	4360	725

Note: This table shows means of employment status, contract type, education, annual earnings (JPY 10,000, 2020 price), weekly hours, age, and whether having partners for respondents aged 20–59 in the JHPS data. The samples are restricted to respondents who eventually experience the corresponding event (childbirth or eldercare onset) during the sample period (i.e., ever-treated). For each event type, I further restrict the sample to pre-event observations only ($t \leq -1$), and split them into two groups: "Not-yet-treated" observations with $t \leq -2$, and "Just treated" observations at $t = -1$. N individuals report the number of unique respondents in each group, and N person-years report the number of respondent-year observations.

Training (Japan Institute for Labour Policy and Training (JILPT), 2025). The database assigns 1–5 scores to roughly 500 three-digit occupations for many task descriptors.

I focus on two: "physical proximity to others" and "face-to-face discussion." The proximity score marks jobs that require workers to stand close to coworkers or customers; Mongey et al. (2021) show that such jobs were least adaptable to social-distancing rules and suffered larger employment losses.

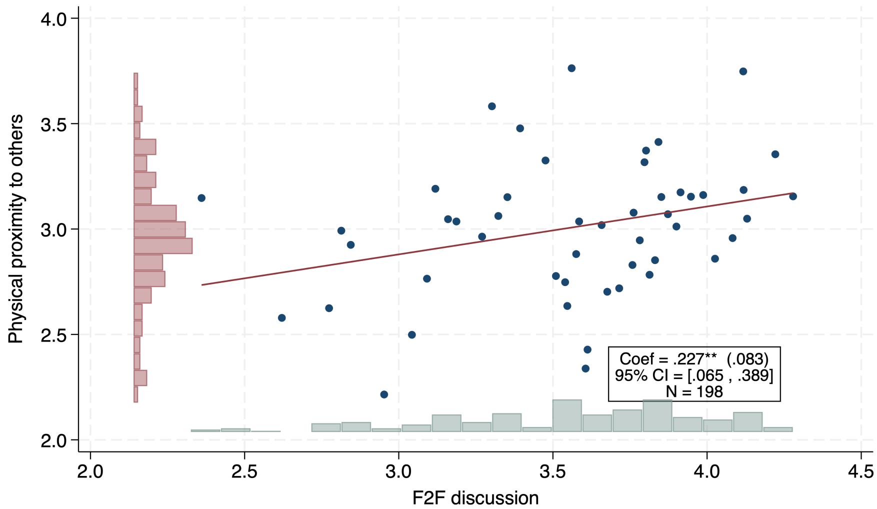
The face-to-face discussion score tracks the importance of in-person interaction; Dingel and Neiman (2020) use similar items to label occupations that can be done from home.

Figure 1 shows the bin-scattered plot for the correlation between the proximity score and the face-to-face discussion score across occupations and their histograms. The two scores are correlated and indicate that occupations that require more physical proximity to others at work also require more face-to-face discussion.

3 Childcare Penalty

In this section, I show the average treatment effects of childcare on labor market outcomes using the event study design following Callaway and Sant'Anna (2021). I use the not-

Figure 1: Physical proximity to others and F2F Discussion



Note: This figure shows the bin-scattered plot for the correlation between the proximity score and the face-to-face discussion score across occupations and their histograms.

yet-treated as control groups and implement a doubly robust DiD estimator based on IPW (inverse probability weighting) and OLS. All the regressions are weighted by sample weights. As covariates, I include age, age squared, and a dummy variable of having a partner.

Because the identification strategy relies on not-yet-treated comparisons, an important requirement is that treated and not-yet-treated observations have sufficient overlap in pre-event characteristics. To make this transparent, I report event-study-aligned descriptive statistics that compare not-yet-treated observations ($t \leq -2$) to the reference-period observations at $t = -1$ ("just to be treated") for both childbirth and eldercare (Tables 2 and 4). These tables show that the two groups are very similar in baseline employment, contract type, education, earnings, hours, and age; the main exception is partner status around childbirth, which is mechanically expected given family formation timing. Partner status is included as a baseline covariate in all specifications, so this imbalance is absorbed by the covariate adjustment in the doubly robust procedure.

Finally, I use a parsimonious covariate set. In practice, including many additional background variables leads to numerical instability and non-convergence in my application, reflecting limited overlap/support in some cohort-by-event-time cells. I therefore prioritize covariates that are consistently available in both surveys and yield stable estimation.

3.1 Childcare Main results

Figure 2 presents the event-study coefficients detailing the impact of a first child's birth on key labor market outcomes—employment, annual earnings, and weekly hours worked—for both males and females.

Panels (a) and (b) illustrate the effects on the employment rate. For males, in both JPSED (Panel a) and JHPS (Panel b), the coefficients representing the average treatment effect are close to zero across all observed periods, from five years before to five years after the child's birth. The 95% confidence intervals for these male coefficients consistently overlap with zero, indicating no statistically significant change in their employment rates attributable to becoming fathers.

In contrast, females experience a pronounced decline in employment. In the JPSED data (Panel a), female employment begins to drop around the event year (year 0). The reduction becomes statistically significant and substantial from year 0, falling by approximately 0.22 points. The peak effect is observed at year +1, with a reduction of about 0.36 points in the employment rate. This negative impact persists, with employment remaining approximately 0.26 points below the baseline through year +4 and still around -0.19 at year +5.

The JHPS data (Panel b) corroborates this finding, though the magnitude of the decline is larger. Female employment in JHPS shows a drop starting around year 0, reaching a peak decline of approximately 0.8 points at year +1. This effect remains statistically significant and substantial, with the employment rate still more than about 0.4 points lower at year +5. The qualitative pattern of a persistent drop in female employment post-childbirth is thus identical across both datasets.

Panels (c) and (d) shift the focus to annual earnings, measured in 10,000 JPY. Similar to employment, males' earnings show no decline. For females, however, the birth of a child coincides with a significant reduction in earnings. According to JPSED estimates (Panel c), females' annual earnings start to decline around year 0, and from year +1 onwards, they experience a loss of approximately 94 (in JPY 10K units), which corresponds to 46% of the average income of females at t=-1 in the JPSED sample.[7] This substantial earnings penalty persists through year +5. The JHPS data (Panel d) shows a similar, though larger, earnings decline for females, with annual earnings dropping by approximately 140 (in JPY 10K units) from year +1 to +2 and remaining around 100 (in JPY 10K units) of reduction through year +5. This decline is about half of the average income of females at t=-1 in the JHPS sample.

[7] See Figure B.1 in Appendix B for the specifications using log income as the outcome. The reduction is around 0.5 log points, which is broadly consistent with the numbers in the percentage.

Finally, Panels (e) and (f) examine the impact on weekly hours worked. Once again, male weekly hours worked are statistically unchanged in response to childbirth in both JPSED (Panel e) and JHPS (Panel f), with coefficients hovering near zero. Conversely, females' weekly hours worked contract significantly. In the JPSED data (Panel e), females' weekly hours decrease by approximately 5 hours from year +1 onwards (about 15% of the average hours worked of females at t=-1), but this reduction becomes statistically insignificant and not persistent through year +5. The JHPS data (Panel f) mirrors this trend, showing a reduction in female weekly hours of about 10 hours from year +3, which also remains persistent. Note that hours worked are measured conditional on employment (i.e., among employed respondents). Thus, the hours profiles should be interpreted as intensive-margin responses among those who remain employed, while the employment profiles capture extensive-margin adjustments. The fact that employment changes substantially, whereas conditional hours change little, indicates that the main adjustment occurs on the extensive margin rather than via hours reductions among the employed.

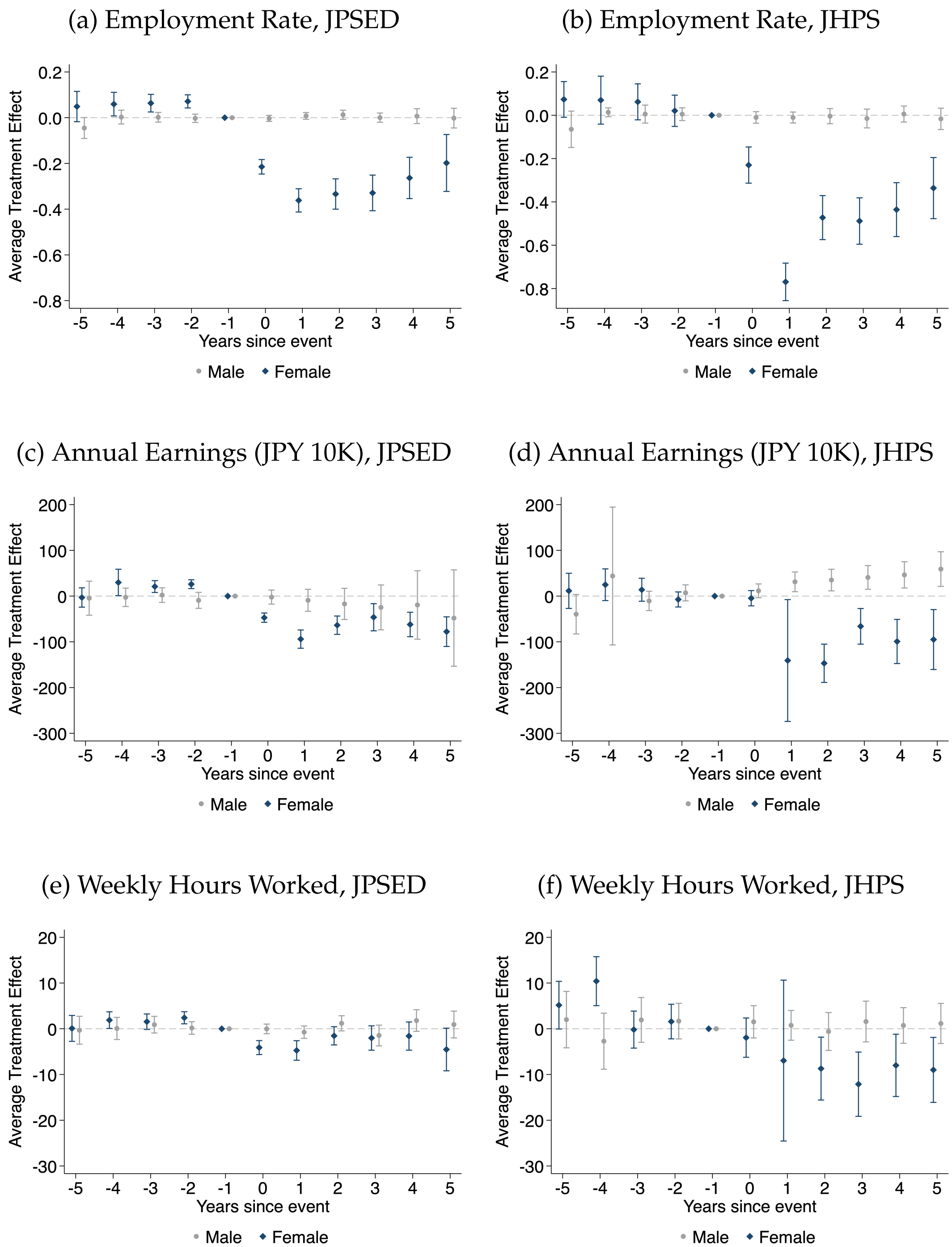
In summary, the event-study analyses from both JPSED and JHPS consistently demonstrate that the arrival of a first child imposes significant and persistent penalties on females' labor market outcomes across employment rates, annual earnings, and weekly hours worked. Males, on the other hand, experience no statistically detectable effects on these dimensions. While the precise magnitudes of these penalties vary between the two datasets, the qualitative message is unequivocal: the childcare penalty is predominantly, if not exclusively, borne by mothers in Japan.

3.2 Childcare Penalty: Heterogeneous effects

This subsection examines the heterogeneous impact of childbirth on mothers' employment rates, using the more granular occupational data available in JPSED. Because some heterogeneity dimensions (e.g., teleworkability and face-to-face/physical interaction measures) are plausibly correlated with COVID-era labor market shocks, I interpret these measures as broader job attributes related to flexibility and workplace interaction and explicitly assess sensitivity to the COVID period. In particular, Appendix Section D replicates the heterogeneity results restricting the sample to the pre-COVID period (until 2019), and the qualitative patterns are unchanged.

Individuals are pre-assigned to binary groups based on characteristics measured before the childbirth event. Specifically, I assign individuals to groups based on their modal characteristics between year -4 and year -1. These characteristics include occupation-level features such as teleworkability (computed from JPSED as the share of workers who tele-

Figure 2: Childcare Penalty



Note: This figure shows event-study coefficients for employment, annual earnings (in JPY 10K), and weekly hours worked for males and females around the birth of their first child. The x-axis represents years since the event, with year 0 being the year of birth. Estimates are derived from JPSED and JHPS data. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. The control groups are the not-yet-treated. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals.

work), physical proximity score from the O-NET data, and face-to-face discussion score from the O-NET data, as well as individual-level factors like regular-contract status, employer firm size (300 or more employees versus smaller firms), and co-residence with parents. I then compare the event-study coefficients for the employment rate of females within each group. Figure 3 visually presents these comparisons.

Panel (a) explores how teleworkability affects mothers' employment. Post-birth, females who were in low-teleworkability jobs before the birth see their employment fall sharply. Their employment rate drops by a peak of nearly 0.40 points in year +3. In contrast, females who were in high-teleworkability jobs before the birth face a smaller penalty. While their employment drop is around 0.15 points in year 0, their responses from year +1 onwards are not statistically significant, and the 95% confidence intervals often include zero. This suggests that teleworkability helps workers return to work after childbirth.

Panels (b) and (c) investigate the impact of occupations' task content. Panel (b) studies the importance of *face-to-face discussion* requirements. Both groups show negative employment responses, and their negative responses of about 0.30 to 0.35 points at years +2 and +3 are similar. The differences are not statistically significant. This suggests that the importance of face-to-face discussion in an occupation might not significantly correlate with the magnitude of the childcare penalty. Panel (c) focuses on *physical proximity* requirements at work. Similar to Panel (b), the results are similar across the two groups, which indicates that the importance of physical proximity does not explain significant heterogeneity in the childcare penalty among mothers.

Panel (d) differentiates based on pre-birth *regular worker* status. Mothers who held regular contracts before childbirth experience a smaller employment exit, with a maximum drop of less than 0.23 points in year +1. In contrast, those on non-regular contracts see a larger decline, around 0.64 points in year +3 post-birth.

Panel (e) explores differences by employer *firm size*. One might expect that large firms might offer better job retention, possibly due to more structured leave policies or greater resources for accommodating working mothers. However, there is no statistically significant difference.

Panel (f) considers the effect of *cohabitation with parents*. While the initial employment responses are similar, females cohabiting with their parents do not have a statistically significant employment penalty in year 5. This suggests that intra-household childcare support from grandparents may not act as an immediate buffer against initial employment exits after childbirth, but it can be a crucial buffer against persistent labor market exit.

Collectively, these heterogeneous analyses underscore that the childcare penalty on female employment is not uniform. Specifically, pre-birth job characteristics such as high teleworkability and holding a regular employment contract appear to substantially reduce the penalty, with non-regular workers experiencing markedly larger employment declines. Furthermore, while co-residence with parents emerges as a factor mitigating longer-term employment loss, the specific task content of occupations, like requirements for face-to-face discussion, did not clearly differentiate the extent of the childcare penalty in this analysis.

4 Eldercare Penalty

In this section, I estimate the average treatment effects of eldercare on labor market outcomes using an event-study design following Callaway and Sant'Anna (2021). I use the not-yet-treated as the control group and implement a doubly robust DiD estimator that combines inverse probability weighting (IPW) and outcome regression (OLS). All regressions are weighted by sample weights. As covariates, I include age, age squared, and an indicator for having a partner.

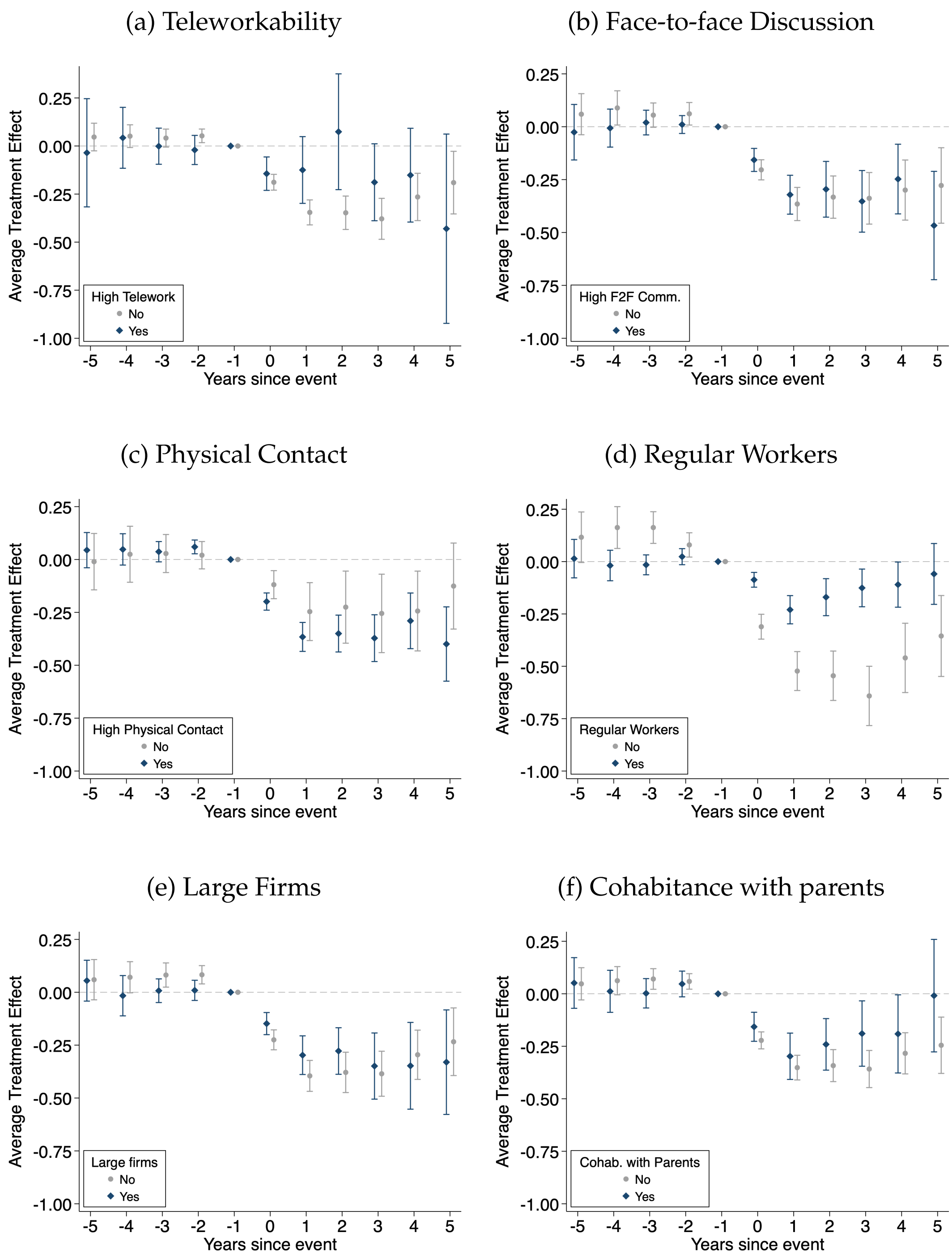
4.1 Event Definition and Severity

Definition The key treatment event is the onset of long-term care (LTC) needs in the respondent's family. In the JPSED, the event is defined as the first time a parent or parent-in-law receives an LTC certificate. In the JHPS, the event is defined as the first time a parent or parent-in-law needs care. The JHPS data also records the level of care needed. In contrast, the JPSED does not record the severity categories in the same way. For comparability across datasets, my baseline specifications therefore use the common event definition available in both surveys: the first year in which the survey indicates that a parent or parent-in-law begins requiring long-term care. Using the JHPS information, I additionally document the distribution of care-need levels at the onset of care (Table A.1) and assess whether the estimated effects differ by severity (Figure A.4).

4.2 Eldercare Penalty: Main results

Figure 4 presents event-study coefficients for the impact of eldercare onset on employment, annual earnings, and weekly hours for both genders, using JPSED and JHPS data.

Figure 3: Childcare Penalty Heterogeneity (Female Employment, JPSED)



Note: This figure shows event-study coefficients for employment rates for females around the birth of their first child. The x-axis represents years since the event, with year 0 being the year of birth. Estimates are derived from JPSED data. The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals.

Panel (a) reports employment-rate effects in JPSED. Coefficients for both males and females are not statistically significant, with 95% confidence intervals always overlapping zero. Panel (b) shows JHPS estimates: males' employment changes remain within ± 0.03, never significant; females' employment changes are also not statistically significant. Thus, the employment penalty from eldercare is at most 0.05 for females and indistinguishable from zero.

Panels (c) and (d) report annual earnings effects (in 10K JPY). In JPSED (panel c), changes in both males' and females' earnings are not statistically significant. In JHPS (panel d), the female drop peaks at -30 in year +3 (statistically significant), which corresponds to 15% of the average income of females at t=-1 in the JHPS sample. Male coefficients remain insignificant. Compared to the childcare penalty, which can decrease by more than 100 at their peaks, these eldercare earnings losses are smaller.

Panels (e) and (f) show weekly hours worked. The declines are only statistically significant for females from year 3 beyond in the JHPS samples.

In summary, the analyses using both JPSED and JHPS confirm that the average eldercare penalty on employment, earnings, and hours is negligible relative to the childcare penalty. Male labor market outcomes remain near zero throughout, while female employment dips by no more than 0.05 in JHPS. Earnings and hours worked remain mostly flat with some delayed effects from years 2 or 3.

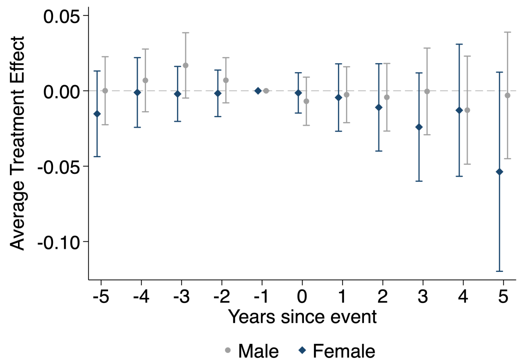
Interpreting dynamic effects A central interpretational issue is that the onset of care needs may correspond to relatively light needs for some households, while caregiving demands can intensify as health conditions progress, as shown in Figure A.3. This feature makes delayed labor-market responses plausible even if immediate employment effects are modest. The event-study design is well-suited to capturing such delayed adjustments because it traces outcomes over multiple years after the onset of care needs.[8]

Severity heterogeneity As discussed when defining the treatment event, the onset of LTC needs contains various levels of severity. In Appendix A.2, to assess whether average effects mask stronger impacts in more difficult situations, I re-estimate the main event-study specifications separately by severity using the JHPS care-need categories and report

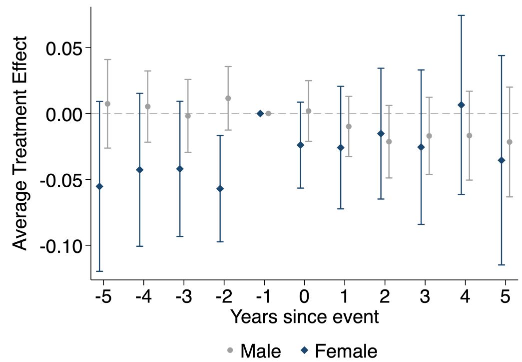
[8] More generally, childcare and eldercare differ fundamentally in their time-cost dynamics: childcare typically becomes less time-intensive as children age, whereas eldercare can become more demanding over time. This asymmetry implies that the labor-market impact of eldercare may emerge with a delay as care needs intensify rather than immediately at onset. Consistent with this interpretation, Nishimura and Oikawa (2025) show that the expansion of nursing home capacity affects female employment, suggesting that constraints associated with eldercare can intensify at later stages rather than immediately upon certification.

Figure 4: Eldercare Penalty

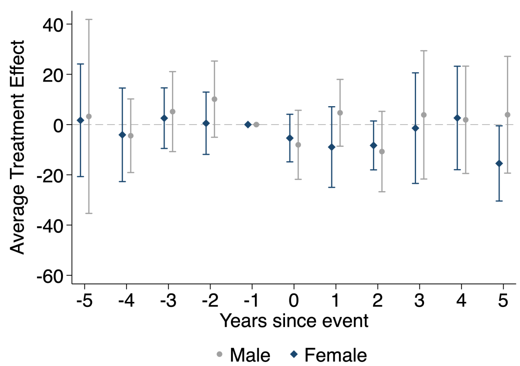
(a) Employment Rate, JPSED



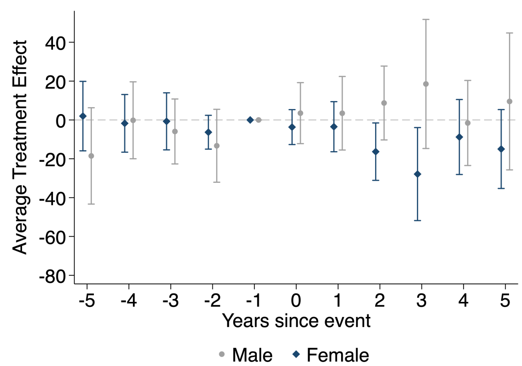
(b) Employment Rate, JHPS



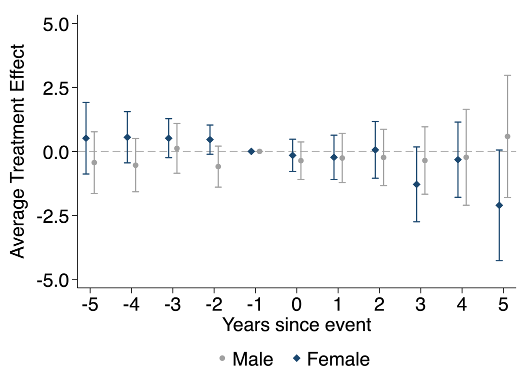
(c) Annual Earnings (JPY 10K), JPSED



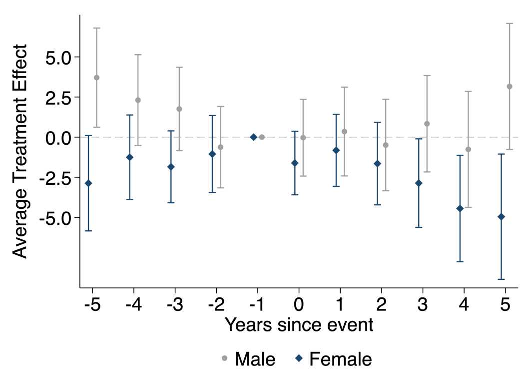
(d) Annual Earnings (JPY 10K), JHPS



(e) Weekly Hours Worked, JPSED



(f) Weekly Hours Worked, JHPS



Note: This figure shows event-study coefficients for employment, annual earnings (in JPY 10,000s), and weekly hours worked for males and females around the onset of caregiving needs. The x-axis represents years since the event, with year 0 being the year of the onset. Estimates are derived from JPSED and JHPS data. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. The control groups are the not-yet-treated. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals.

these results in the Appendix. The qualitative patterns are similar across severity groups, and the main conclusions are unchanged.[9]

Caregiving measures Finally, both surveys contain a direct measure of whether the respondent provides informal care. In Appendix A.1, I use this caregiving indicator as an additional outcome to verify that the onset-of-care event coincides with informal caregiving behavior.

4.3 Eldercare Penalty: Heterogeneous effects

Figure 4 shows that the average effects of entering elder care on employment, earnings, and hours are close to zero, but these averages mask important heterogeneity. In the remainder of this subsection, I split the sample along the same pre-defined characteristics used for the childcare analysis—occupation-based task content (teleworkability, physical proximity, face-to-face discussion), contract type, firm size, and co-residence with parents—and compare treated and not-yet-treated workers within each group. Groups are fixed using pre-event information for treated individuals (their modal value before care begins), so post-event sorting cannot bias the estimates.

Figure 5 shows the results by different subgroups. Panel (a) explores how teleworkability affects females' employment. Post-event, females who were in low-teleworkability jobs before the event see their employment fall sharply. Their employment rate drops by a peak of nearly 0.07 points in year +3. In contrast, females who were in high-teleworkability jobs before the event face a smaller penalty. Their responses are not statistically significant, and the 95% confidence intervals often include zero.[10]

Panel (b) investigates the impact of *face-to-face discussion* requirements. Females who were in occupations with more frequent face-to-face discussion did not see a statistically significant decline, while the other experienced statistically significant declines in years 0 to 3 and 5. The differences of the estimates are not statistically significant except for year 3.

Panel (c) focuses on *physical proximity* requirements at work. High physical contact jobs are associated with statistically significant employment declines for females. Females in jobs with low physical proximity requirements do not experience an eldercare penalty

[9] The JHPS contains information on the severity of care needs at the onset of care. I use this information to split the sample into light and heavy care needs and re-estimate the main specifications. I find no statistically significant differences in treatment effects by severity. See Figure A.4.

[10] Note that the differences are not statistically significant except for years 0 and 3.

until year 5, with a large confidence interval. The estimates for the different groups are statistically not different.

Panel (d) differentiates based on pre-event *regular worker* status. Females who held regular contracts before the LTC event experience a statistically insignificant employment exit. Those on non-regular contracts see a statistically significant decline, around 0.03 to 0.10 points in the initial years post-event.[11]

Panel (e) explores differences by employer *firm size*. Penalties are notably larger for females working in small firms. Their employment rate drops by a significant 0.03 points from year +1 and persists at larger levels, which are close to 0.1 points. For females in large firms ("Large firms - Yes", meaning $\geq$ 300 employees), the effect is not statistically significant. This suggests that large firms might offer better job retention, possibly due to more structured leave policies or greater resources for accommodating working females.

Panel (f) considers the effect of *cohabitation with parents*. There is no clear difference.

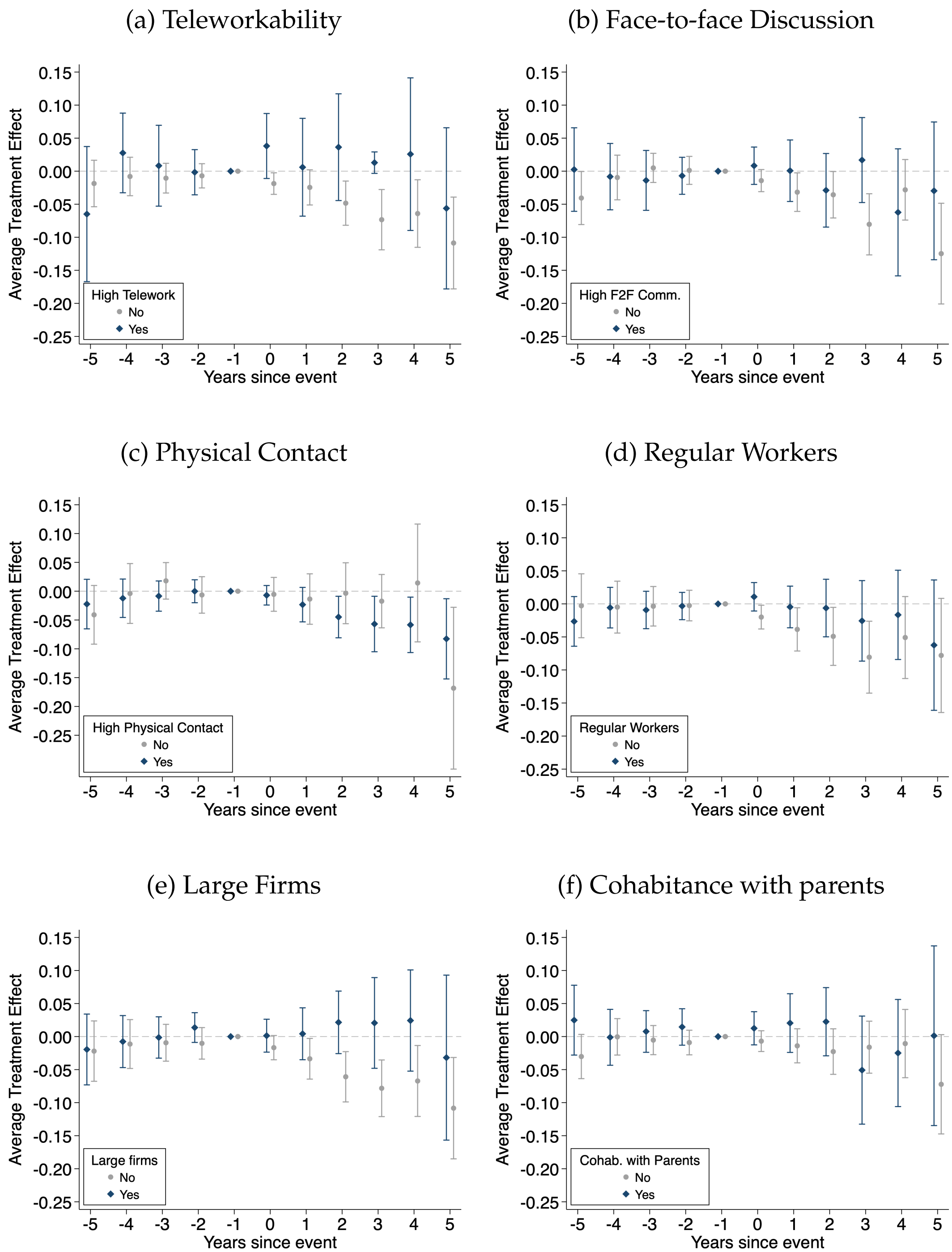
Collectively, these heterogeneous analyses underscore that the eldercare penalty on female employment is not uniform. The effects observed vary across different subgroups of females. Specifically, the penalty is less pronounced or statistically insignificant for females in high-teleworkability jobs, those with low physical proximity requirements, those with regular contracts, and those in larger firms. These findings indicate that specific job characteristics are associated with differing magnitudes of the eldercare penalty on female employment.

5 Conclusion

In conclusion, I find that motherhood imposes a substantial and persistent *childcare penalty* on females' employment in Japan. This employment loss after childbirth is large on average and heterogeneous across individual and job characteristics, indicating that some females face much steeper career costs of motherhood than others. By contrast, the labor market consequence of family eldercare, an *eldercare penalty*, is much smaller. My estimates suggest that, on average, the impact of an elderly parent's care needs on a female's employment is close to zero and often not statistically significant. However, this modest average masks considerable heterogeneity: certain subgroups of females experience non-trivial employment declines when taking on caregiving responsibilities for elderly family members. Importantly, my event-study approach—accounting for the staggered timing of caregiving onset—was crucial in detecting these nuanced effects and avoiding biases that could arise with conventional estimators.

[11] The differences of the estimates are not statistically different, though.

Figure 5: Eldercare Penalty, Heterogeneity (Female Employment, JPSED)



Note: This figure shows event-study coefficients for employment rates for females around the onset of caregiving needs. The x-axis represents years since the event, with year 0 being the year of the onset. Estimates are derived from JPSED data. The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals.

Looking ahead, my findings open two avenues for further research and policy analysis. First, the absence of a large average employment effect for eldercare should not be interpreted as zero welfare cost. Many females likely continue working while caregiving by increasing effort or sacrificing leisure and well-being, implying hidden burdens not captured by employment rates. Developing a structural model of females' labor supply and caregiving choices would help quantify these welfare implications, illuminating how caregiving affects utility, productivity, or health even when employment is maintained.

Second, my results have implications for the design of social policies to support working caregivers. Prior evidence shows that formal long-term care support can mitigate the burden of informal caregiving on labor supply. For instance, Fu et al. (2017) find that the introduction of Japan's public Long-Term Care Insurance in 2000 significantly increased labor force participation among family caregivers. Similarly, Mikoshiba (2025) uses a structural model to demonstrate that a generous universal LTC insurance program can enhance caregiver welfare while sustaining labor supply. Building on these insights, future work should explore how targeted policies, such as improved caregiver leave provisions, flexible work arrangements, and expanded access to formal care services, could reduce the remaining labor market penalties of child-rearing and eldercare in Japan.

References

Acemoglu, D., Mühlbach, N. S., and Scott, A. J. (2022). The rise of age-friendly jobs. *The Journal of the Economics of Ageing*, 23:100416.

Angelov, N., Johansson, P., and Lindahl, E. (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics*, 34(3):545–579.

Baker, A. C., Larcker, D. F., and Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395.

Bolin, K., Lindgren, B., and Lundborg, P. (2008). Your next of kin or your own career? caring and working among the 50+ of europe. *Journal of Health Economics*, 27(3):718–738.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285.

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

Crespo, L. and Mira, P. (2014). Caregiving to elderly parents and employment status of european mature women. *Review of Economics and Statistics*, 96(4):693–709.

De Chaisemartin, C. and d'Haultfoeuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, 26(3):C1–C30.

Dingel, J. I. and Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, 189:104235.

Fu, R., Noguchi, H., Kawamura, A., Takahashi, H., and Tamiya, N. (2017). Spillover effect of japanese long-term care insurance as an employment promotion policy for family caregivers. *Journal of Health Economics*, 56:103–112.

Fukahori, R., Sakai, T., and Sato, K. (2015). The effects of incidence of care needs in households on employment, subjective health, and life satisfaction among middle-aged family members. *Scottish Journal of Political Economy*, 62(5):518–545.

Fukai, T. and Kondo, A. (2025). Parental earnings trajectories around childbirth in japan: Evidence from local tax records. Technical report, Research Institute of Economy, Trade and Industry (RIETI).

Hsu, C.-H. (2021). Parity-specific motherhood penalties: Long-term impacts of childbirth on women's earnings in japan. *Advances in Life Course Research*, 50:100435.

International Labour Organization (2018). Ilo labour force estimates and projections (lfep) 2018: Key trends. Statistical brief, International Labour Organization (ILO). Accessed: 2025-06-07.

Japan Institute for Labour Policy and Training (JILPT) (2025). Occupational information database: Simple numerical download data, version 3.0. Downloaded from the Job Tag website on 3 June 2025 and processed by the author. URL: https://shigoto.mhlw.go.jp/User/download.

Kikuzawa, S. and Uemura, R. (2021). Parental caregiving and employment among midlife women in japan. *Research on Aging*, 43(2):107–118.

Kleven, H., Landais, C., and Leite-Mariante, G. (2024). The child penalty atlas. *Review of Economic Studies*, page rdae104.

Kleven, H., Landais, C., and Søgaard, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, 11(4):181–209.

Kohara, M. and Ohtake, F. (2011). Altruism and the care of elderly parents: evidence from japanese families. *Japanese Economy*, 38(2):3–18.

Komura, N. (2022). Child penalty and the gender gap. In *Report of the Study Group on Work, Work Styles, and Wages: Toward Realizing a Society Where Each Individual Can Fully Demonstrate Their Abilities*, chapter 3, pages 42–56. Policy Research Institute, Ministry of Finance, Japan, Tokyo, Japan. In Japanese.

Maestas, N., Messel, M., and Truskinovsky, Y. (2024). Caregiving and labor supply: New evidence from administrative data. *Journal of Labor Economics*, 42(S1):S183–S217.

Mikoshiba, M. (2025). Universal insurance with in-kind transfers: The welfare effects of long-term care insurance in japan. Discussion Paper 25-E-030, Research Institute of Economy, Trade and Industry (RIETI). March 2025.

Mongey, S., Pilossoph, L., and Weinberg, A. (2021). Which workers bear the burden of social distancing policies? *Journal of Economic Inequality*, 19(3):509–526.

Mugiyama, R. (2024). Educational differences in the motherhood penalty on wage trajectories in japan: The role of loss of experience. *CSRDA Discussion Paper*, 87.

Nishimura, Y. and Oikawa, M. (2025). Who has benefited from nursing home expansion in japan?: The effects of government supply-side intervention. *The Journal of the Economics of Ageing*, page 100604.

Okuyama, Y., Murooka, T., and Yamaguchi, S. (2025). Unpacking the child penalty using personnel data: How promotion practices widen the gender pay gap. *IZA Discussion Paper*, (17673).

Organisation for Economic Co-operation and Development (OECD) (2024). *Society at a Glance 2024: OECD Social Indicators*. OECD Publishing, Paris, 10th edition. Published 20 June 2024.

Oshio, T. and Usui, E. (2017). Informal parental care and female labour supply in japan. *Applied Economics Letters*, 24(9):635–638.

Oshio, T. and Usui, E. (2018). How does informal caregiving affect daughters' employment and mental health in japan? *Journal of the Japanese and International Economies*, 49:1–7.

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

Statistics Bureau of Japan (2025). Consumer price index for all goods. https://www.e-stat.go.jp/dbview?sid=0003427113. e-Stat database; Accessed: June 12, 2025.

Tamiya, N., Noguchi, H., Nishi, A., Reich, M. R., Ikegami, N., Hashimoto, H., Shibuya, K., Kawachi, I., and Campbell, J. C. (2011). Population ageing and wellbeing: Lessons from japan's long-term care insurance policy. *The Lancet*, 378(9797):1183–1192.

Yamada, H. and Shimizutani, S. (2015). Labor market outcomes of informal care provision in japan. *The Journal of the Economics of ageing*, 6:79–88.

Appendix

Contents

A Details of Eldercare Giving

A.1 Actual Caregiving

Actual Caregiving as Event (JPSED) A natural alternative to defining the event as the onset of a care need is to use the respondent's *actual informal caregiving* as the event. This addresses the concern that the onset of a care need may not always translate one-to-one into the respondent providing care (e.g., because other family members are primary caregivers, formal services are used, or caregiving responsibilities are delayed). At the same time, two caveats are important for interpretation. First, the "actual caregiving" measure is available only in a limited subset of survey waves (in and after 2017 for the JHPS and in and after 2018 for the JPSED) and observations, which substantially shortens the available pre-event window and reduces precision.[12] Second, actual caregiving is itself a behavioral choice that may respond to labor market conditions and household arrangements, whereas LTC needs is more plausibly external to the respondent's short-run labor supply decisions. For these reasons, I treat caregiving-based definitions as informative complements rather than as a universally superior substitute for the LTC-need event.

With these caveats in mind, I use the JPSED measure of whether the respondent starts providing informal care and re-estimate the event-study specification, treating the onset of caregiving as the event. Figure A.1 reports the corresponding labor market responses by gender.

As discussed, the sample periods only span from 2018 to 2022, which ends up with noisy and imprecise estimates.

Actual Caregiving as Outcome (JPSED) A complementary check is to treat informal caregiving as an *outcome* and examine whether the onset of a care need in the family is associated with an increase in the respondent's caregiving. I therefore estimate an event-study where the event is the onset of a care need (as in the baseline definition) and the outcome is whether the respondent provides informal care.

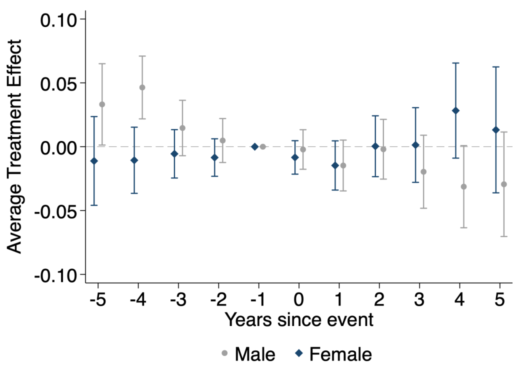
Figure A.2 shows the corresponding caregiving responses. This exercise helps assess the empirical linkage between the onset-of-care measure and caregiving behavior, while recognizing that the mapping need not be mechanical for all households. As above, this analysis is based on a caregiving variable observed only in a limited subset of survey years and respondents, so it should be interpreted as supplementary evidence rather than as a definitive validation test. At the onset of the LTC needs for parents (or in-laws), both

[12] For this reason, we can only rely on the JPSED data, because the JHPS data have smaller samples and the most of the coefficients are not estimated.

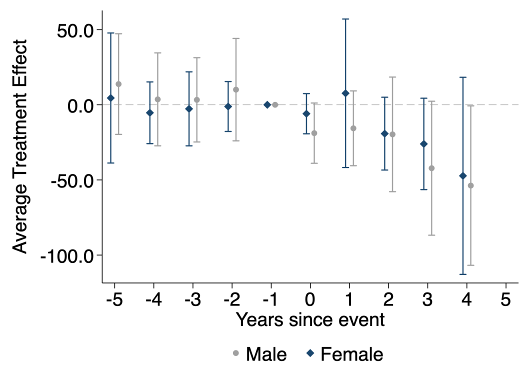
males and females start to provide informal care by 16% pt more in Year 0. Most of the other estimates are imprecise because we have very limited samples.

Figure A.1: Caregiving as Event (JPSED)

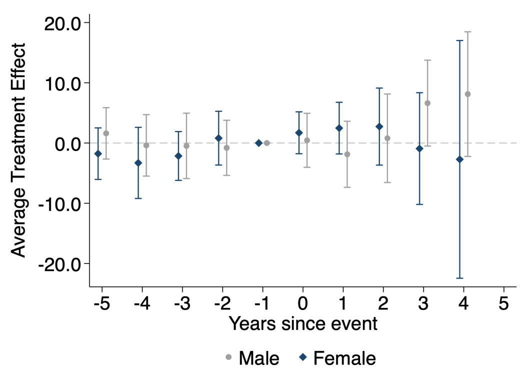
(a) Employment Rate



(b) Income

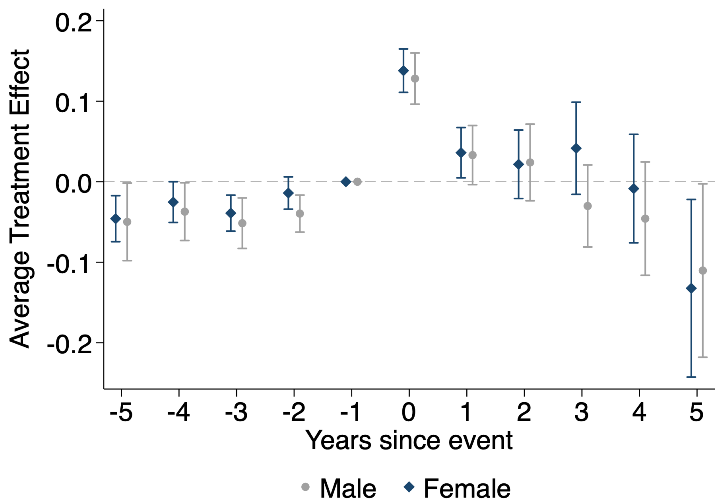


(c) Hours



Note: This figure reports event-study estimates of the effect of the onset of informal caregiving on the respondent's employment rate, income, and hours worked in the JPSED. The event is defined as the first year in which the respondent reports providing informal care. This variable is available only for a limited set of survey years and respondents, which shortens the pre-event window and reduces statistical power. Estimates use a not-yet-treated control group and a doubly robust DiD estimator following Callaway and Sant'Anna (2021), with covariates age, age squared, and an indicator for having a partner, weighted by survey weights. The omitted period is event time -1. Points show estimated average treatment effects by event time, and vertical bars show 95% confidence intervals.

Figure A.2: Caregiving Responses: Caregiving as Outcome (JPSED)



Note: This figure reports event-study estimates of the effect of the onset of a care need in the family on the respondent's informal caregiving behavior in the JPSED. The event is defined as the first year in which a parent (or parent-in-law) is reported to begin requiring long-term care, and the outcome is an indicator of whether the respondent provides informal care. The caregiving outcome is available only for a limited set of survey years and respondents, which shortens the pre-event window and reduces statistical power. Estimates use a not-yet-treated control group and a doubly robust DiD estimator following Callaway and Sant'Anna (2021), with covariates age, age squared, and an indicator for having a partner, weighted by survey weights. The omitted period is event time -1. Points show estimated average treatment effects by event time, and vertical bars show 95% confidence intervals.

A.2 Levels of Severity

Levels of Severity at the Onset Table A.1 reports the distribution of long-term care (LTC) severity at the onset of the event in the JHPS, where the event is defined as the first LTC certification of a parent or parent-in-law.[13] As the table makes clear, the event includes both relatively light categories (Jiritsu/Not-applied, Keika, Yoshien 1–2, 37%) and more intensive care-need categories (Yokaigo 1–5, 63%). This directly addresses the concern that the empirical "event" might capture only minor cases: in the JHPS, first certification spans the full range of severity levels.

Table A.1: Level of LTC Severity at the Onset

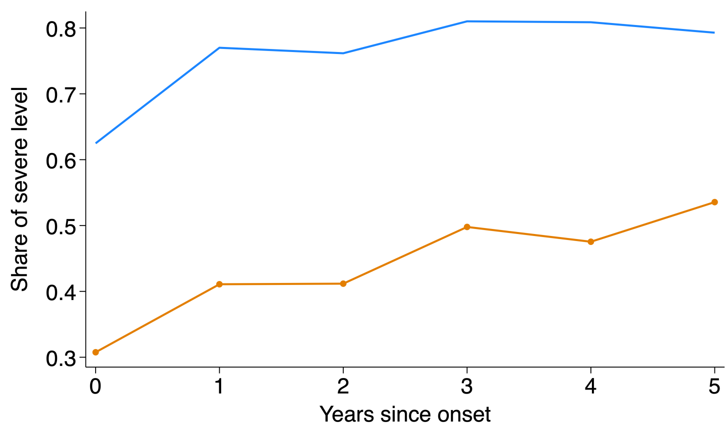
	N	Share
Level		
Jiritsu/Not-applied	90	12.41
Keika	25	3.45
Yoshien 1	84	11.59
Yoshien 2	73	10.07
Yokaigo 1	111	15.31
Yokaigo 2	119	16.41
Yokaigo 3	94	12.97
Yokaigo 4	64	8.83
Yokaigo 5	65	8.97
Total	725	100.00

Note: This table shows the distribution of long-term care (LTC) certification levels at the onset of the event in the JHPS, defined as the first year in which a parent (or parent-in-law) receives LTC certification. Jiritsu, Not-applied, Keika, and Yoshien 1–2 correspond to relatively light care needs, while Yokaigo 1–5 correspond to more severe care needs.

Levels of Severity over Time To further investigate whether the labor-market impact of eldercare may emerge with a delay if care needs intensify over time, I also document how the severity composition evolves following the onset of long-term care (LTC). Figure A.3 plots, by event time, the share of severe LTC cases (Yokaigo) among all LTC cases (Yoshien + Yokaigo) and the share of more severe LTC cases (Yokaigo Level 3 and above) among all LTC cases (Yoshien + Yokaigo). The figure shows that the severity mix shifts toward more severe categories in the years after onset, consistent with the interpretation that time costs and constraints associated with eldercare may increase as conditions progress. This provides complementary evidence for interpreting dynamic treatment effects with some delays in the event-study framework.

[13] The JHPS data has such information, while the JPSED data does not.

Figure A.3: Share of Severe LTC over Time since Onset



Note: This figure shows the share of severe LTC cases (Yokaigo) in the blue line and the share of more severe LTC cases (Yokaigo Level 3 and above) in the orange line with dots among all LTC cases (Yoshien + Yokaigo). They are plotted by event time (years since the onset of LTC). The onset is defined as the first year in which a parent (or parent-in-law) receives LTC certification in the JHPS. Each point corresponds to the mean of an indicator for Yokaigo within an event-time bin.

Eldercare Penalty by Severity Figure A.4 shows the eldercare penalty by severity of the LTC needs at the onset of the event in JHPS data.

Panels (a) and (b) show the responses of employment rates. Even with severe cases (Panel a), females' employment declines are not statistically significant. Notably, the decline for males is statistically significant for severe needs of LTC, but the magnitudes are at most 0.04pt. Panels (c) and (d) show the responses of income. While they are not statistically different from each other, females' declines in income in year 3 are statistically significant for the low case. Panels (e) and (f) show the responses of weekly hours worked. In both cases, the declines are present for females in later periods, but the differences between high and low cases are not statistically significant.

Figure A.5 repeats the same exercise with a different cutoff. The new cutoff is whether the level is Yokaigo Level 3 or above. The qualitative patterns are similar, but the differences are less visible because the sample sizes are smaller.

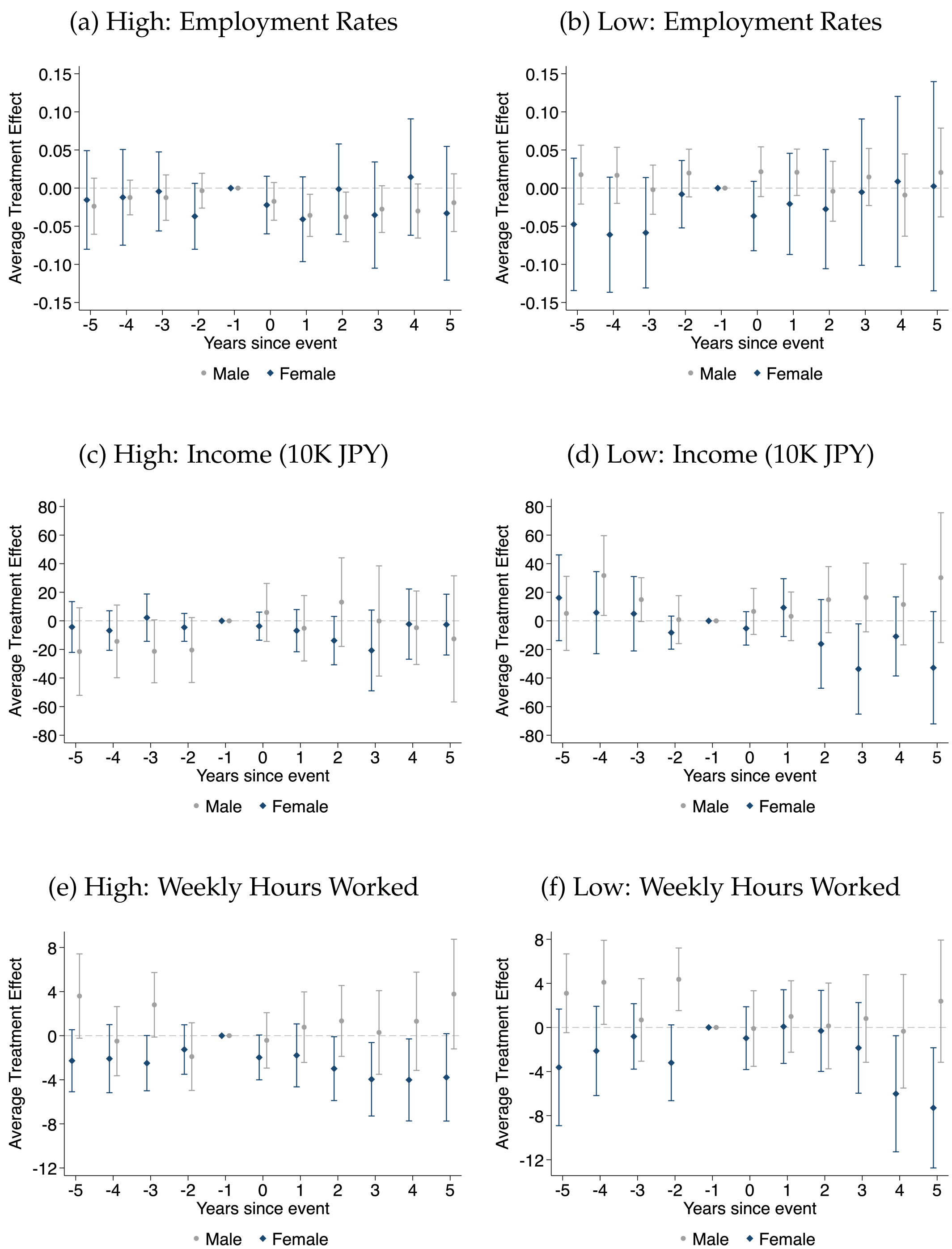
B Results on Log Income

This appendix reports event-study results using log income as the outcome. The motivation is interpretability: coefficients in log specifications can be read approximately as percentage changes, which complements the main analysis that reports income in levels (yen) and discusses magnitudes in percentage terms by scaling level coefficients by the treated group's mean pre-event income at event time $k = -1$.

A key limitation of the log specification is sample selection. Taking logs requires restricting attention to observations with positive income, which implicitly conditions on being employed and having positive earnings. As a result, the log-income estimates should be interpreted as evidence on income changes *among those with positive earnings*, rather than unconditional income effects that combine extensive- and intensive-margin adjustments.

With this caveat in mind, Figure B.1 shows that the childcare penalty on income for females is clearly present in the JPSED data (persistent declines over 0.5 log points), consistent with the main results using income in levels. In contrast, the eldercare penalty on income is imprecisely estimated in the log specification, which mirrors the broader finding that average eldercare effects are modest and less precisely estimated than childcare effects in the JPSED.

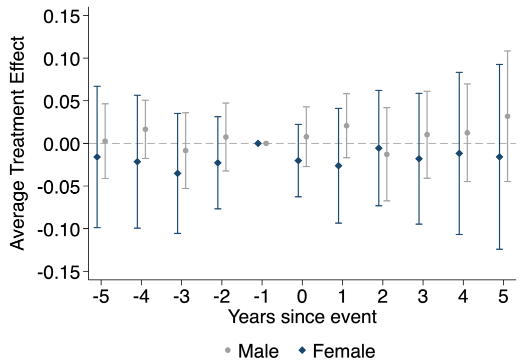
Figure A.4: Eldercare Penalty by Severity (Yokaigo)



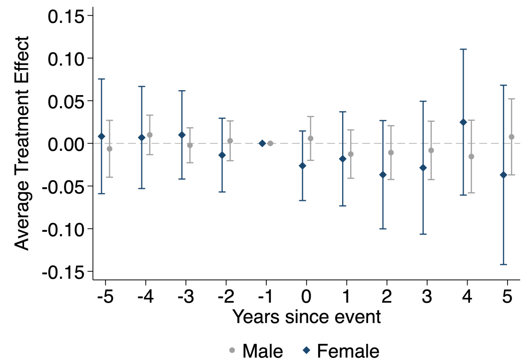
Note: This figure shows event-study coefficients for employment rates, income, and hours worked around the onset of caregiving needs. The x-axis represents years since the event, with year 0 being the year of the onset. Estimates are derived from JHPS data. The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals.

Figure A.5: Eldercare Penalty by Severity (Yokaigo Levle 3 and Above)

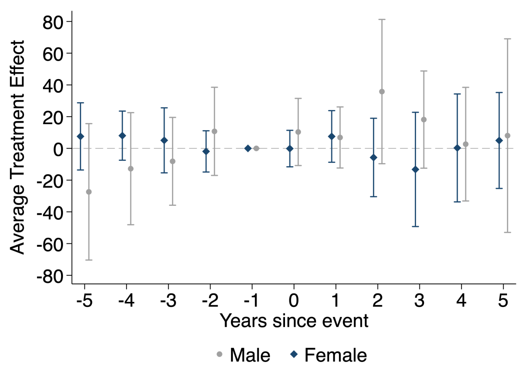
(a) High: Employment Rates



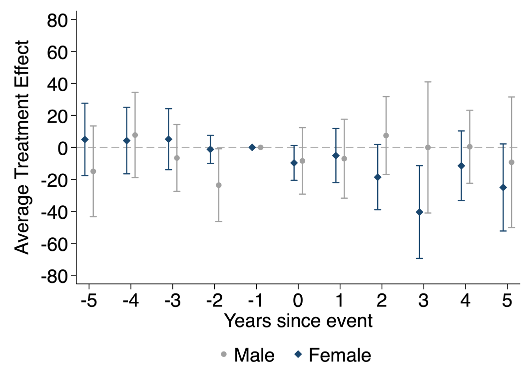
(b) Low: Employment Rates



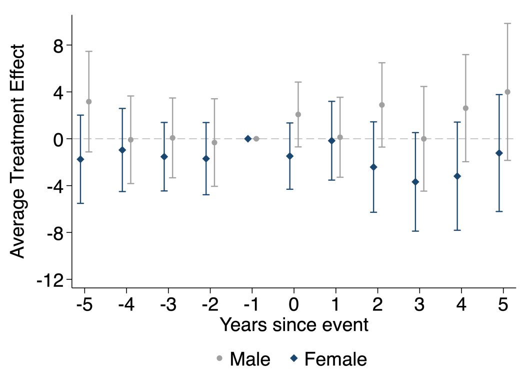
(c) High: Income (10K JPY)



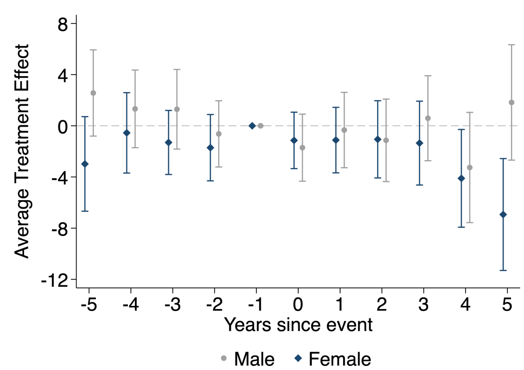
(d) Low: Income (10K JPY)



(e) High: Weekly Hours Worked



(f) Low: Weekly Hours Worked

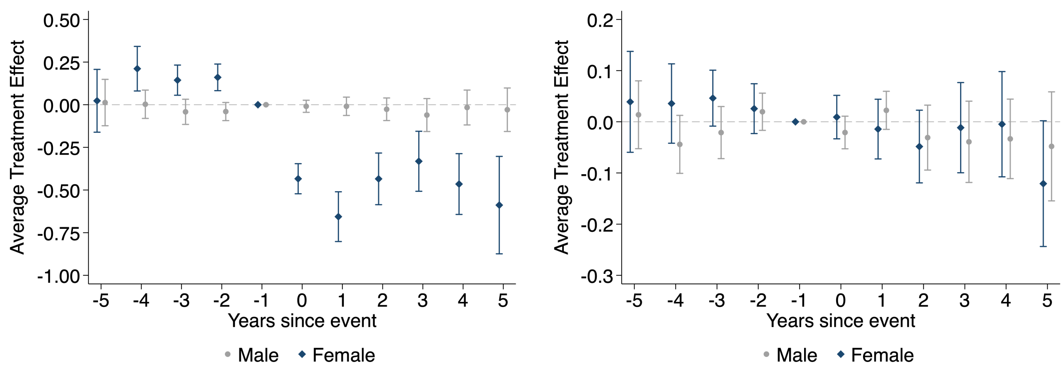


Note: This figure shows event-study coefficients for employment rates, income, and hours worked around the onset of caregiving needs. The x-axis represents years since the event, with year 0 being the year of the onset. Estimates are derived from JHPS data. The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals.

Figure B.1: Childcare and Eldercare Penalty on Log Income (JPSED)

(a) Childcare Penalty

(b) Eldercare Penalty



Note: This figure compares baseline event-study estimates of the childcare and eldercare penalty on log income using the JPSED data. The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals.

C Sample Restrictions

C.1 Restricting samples to those with parents alive

A common approach in the eldercare-penalty literature is to restrict attention to respondents who are *at risk* of providing care, typically operationalized as having at least one living parent (and/or parent-in-law) at baseline; see, for example, Crespo and Mira (2014). This restriction helps interpret estimated labor-market responses as arising from changes in caregiving needs among families that could plausibly face such shocks, rather than from compositional differences driven by respondents who are never exposed to eldercare risk.

Motivated by this concern, I conduct the following robustness check using the JHPS. I restrict the sample to respondents for whom the survey indicates that at least one *own parent* is alive. I then re-estimate the baseline event-study specification using the not-yet-treated as the control group and the doubly robust DiD estimator following Callaway and Sant'Anna (2021).

I focus on the employment-rate outcome, because the parental-alive indicator is missing for a large share of observations (about 75%), which substantially reduces sample size and precision for other outcomes such as hours and income.

Figure C.1 reports the labor market responses in this restricted sample.

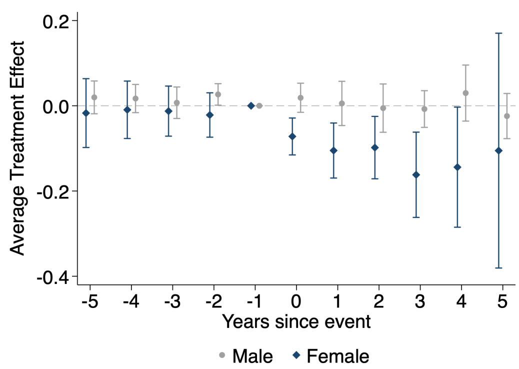
Panel (a) shows the employment responses. The dynamic pattern is qualitatively similar to the baseline results, but the estimates are more precise.

Panels (b) and (c) show the responses of income and hours worked. The estimates are imprecise but qualitatively similar to the baseline results.

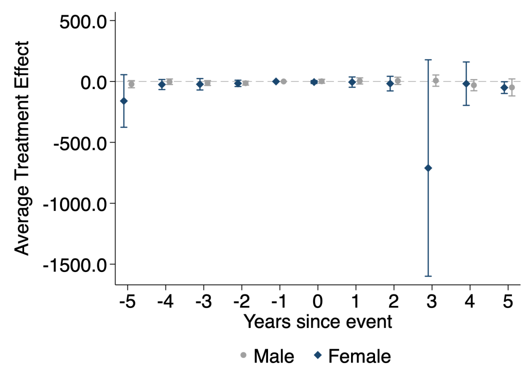
Three limitations are important for interpreting this exercise. First, the JPSED does not include a parental-survival measure that can be mapped to the JHPS question, so the restriction is feasible only in the JHPS and cannot be implemented symmetrically across the two datasets. Second, the baseline eldercare event is defined using both parents *and parents-in-law*, whereas the only available survival question in the JHPS pertains to the respondent's *own parents* only. As a result, the restriction cannot rule out cases in which the observed event is driven by a parent-in-law whose survival status is unobserved, implying that the "at-risk" sample remains contaminated by individuals who may not be at risk with respect to the relevant parent type that triggers the event. Third, the JHPS survival dummy is observed for only 22.1% of the sample (77.9% missing), which sharply reduces effective sample size and thus statistical power, making the resulting estimates materially less precise and potentially less representative. In addition, conditional on being observed, the measure has very limited variation: 98% of respondents report that at least one of their own parents is alive. This implies that the restriction is weak in practice and should be interpreted as a coarse, one-sided screen rather than a stringent "at-risk" definition. Taken together, these constraints mean that the exercise provides, at best, a partial approximation to the ideal at-risk sample and should not be interpreted as a fully harmonized restriction for both parents and parents-in-law.

Figure C.1: Restricting the sample to respondents with at least one parent alive (JHPS)

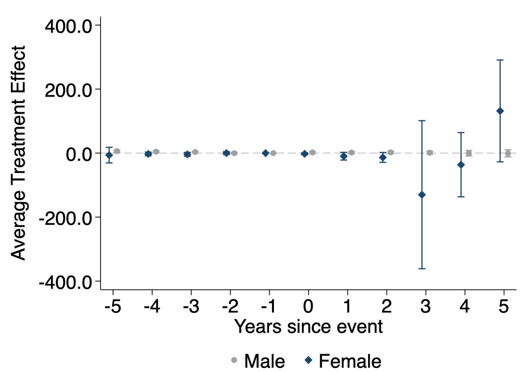
(a) Employment Rates



(b) Income



(c) Hours Worked



Note: This figure reports event-study estimates of the effect of the onset of a parent requiring long-term care on the respondent's employment rate, income, and hours worked in the JHPS, restricting the sample to respondents for whom the survey indicates that at least one *own parent* is alive. The event is defined as the first year in which a parent is reported to begin requiring long-term care. The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals.

C.2 Restricting samples to complete Subgroup-Defining variables

Several heterogeneity analyses define subgroups using characteristics measured before the event (e.g., teleworkability, face-to-face intensity, and regular employment status). In these analyses, subgroup membership is defined using the pre-event period, and in particular, event time -1 is the omitted baseline period. When the subgroup-defining variable is missing at event time -1, subgroup assignment is not well-defined. I therefore do not impute subgroup status and instead restrict the estimation sample to observations for which the relevant subgroup-defining variable is observed at event time -1.

This restriction implies that subgroup figures can differ from the full-sample event-study results not only because they compare effects across groups, but also because they are estimated on a reduced "non-missing" sample. To assess the quantitative importance of this sample-composition channel, I conduct the following diagnostic exercise for each subgroup scenario:

(i) Estimate the baseline (non-heterogeneous) event-study specification on the full eligible sample (as in Figure 3).

(ii) Re-estimate the same baseline specification on the restricted sample that retains only observations with non-missing subgroup-defining variables at event time -1 (the estimation sample used for the corresponding subgroup figure).

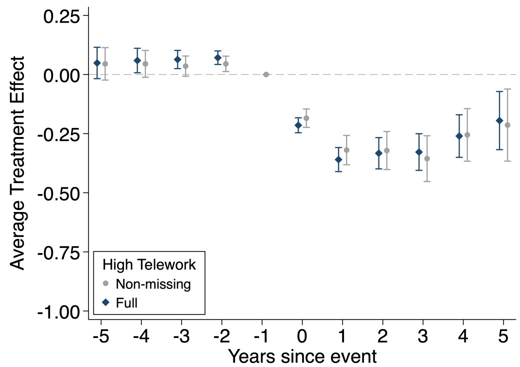
(iii) Compare the resulting average treatment effects (ATEs) and dynamic paths.

The main conclusion is that the childcare estimates are essentially unchanged when moving from the full sample to the corresponding non-missing samples. For eldercare, some estimates differ: in most cases, restricting to non-missing subgroup-defining variables yields more precise estimates, and the ATEs become more negative. This pattern is consistent with the idea that missingness is not random and that the restricted samples are more selected and, mechanically, smaller, which affects precision and can affect point estimates. I therefore interpret the subgroup results with this sample restriction in mind.

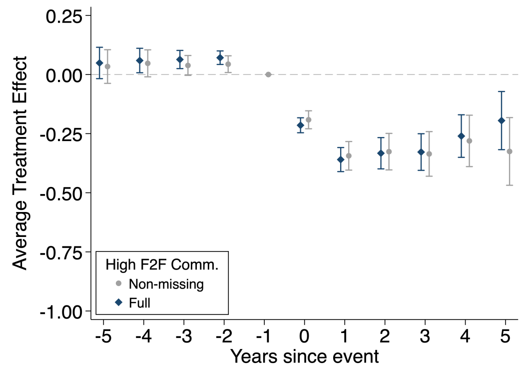
For each heterogeneity dimension Z, I define the restricted sample as observations satisfying $Z_{i,t=-1}$ observed, where $t = -1$ is event time relative to the focal care event (childbirth or onset of care need). I then re-estimate the baseline Callaway–Sant'Anna event study using the not-yet-treated as the control group and the same covariates and weights as in the main specifications.

Figure C.2: Baseline childcare estimates: full sample vs non-missing subgroup-defining sample (JPSED)

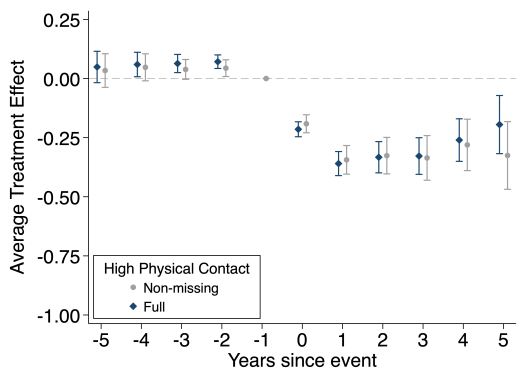
(a) Teleworkability



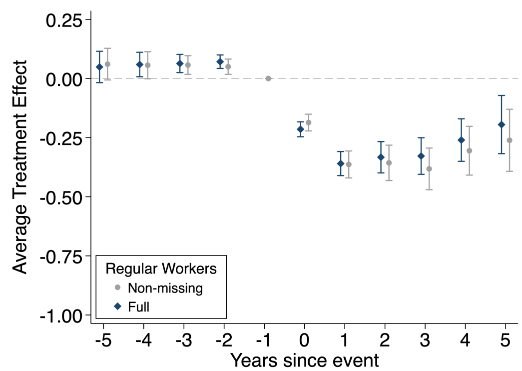
(b) Face-to-face Discussion



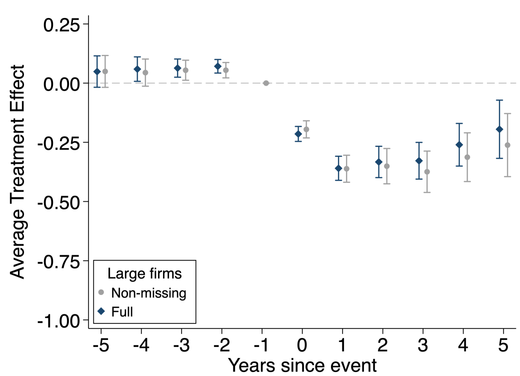
(c) Physical Contact



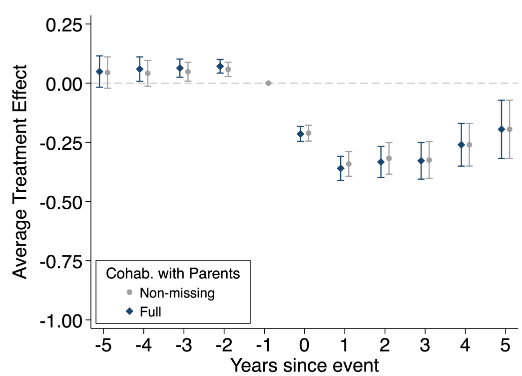
(d) Regular Workers



(e) Large Firms



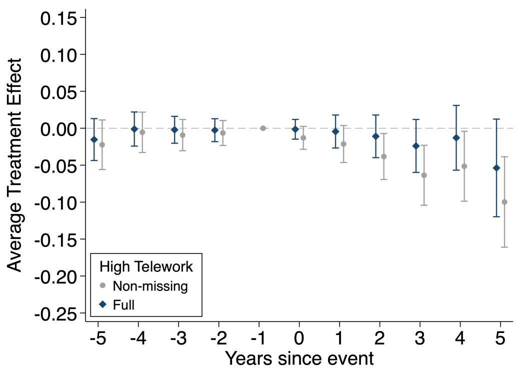
(f) Cohabitance with parents



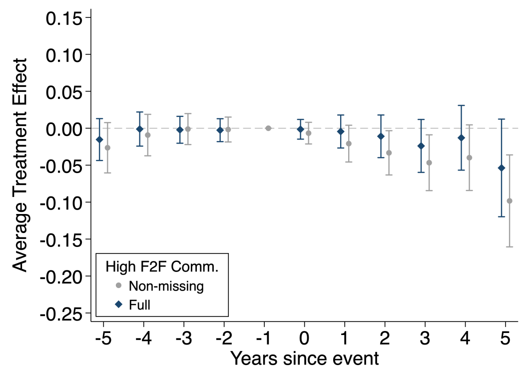
Note: This figure compares baseline event-study estimates of the childcare penalty on female employment using (i) the full eligible sample (as in Figure 3) and (ii) the restricted sample requiring non-missing subgroup-defining variables at event time -1 (as in the subgroup analyses). The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals.

Figure C.3: Baseline eldercare estimates: full sample vs non-missing subgroup-defining sample (JPSED)

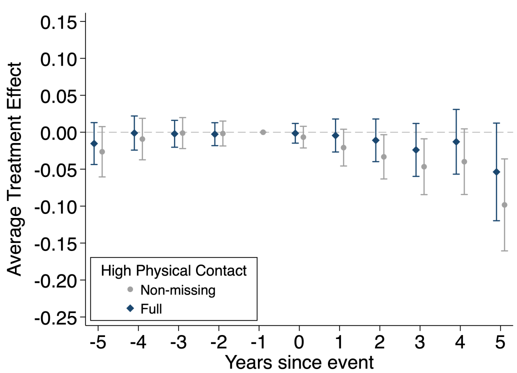
(a) Teleworkability



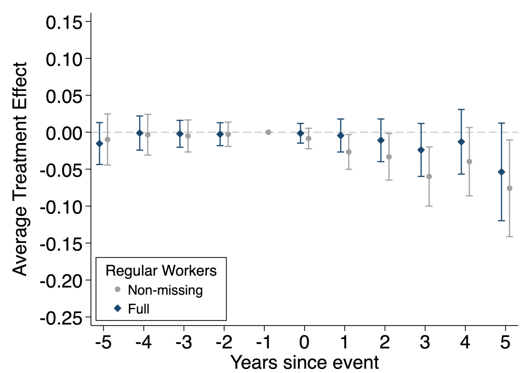
(b) Face-to-face Discussion



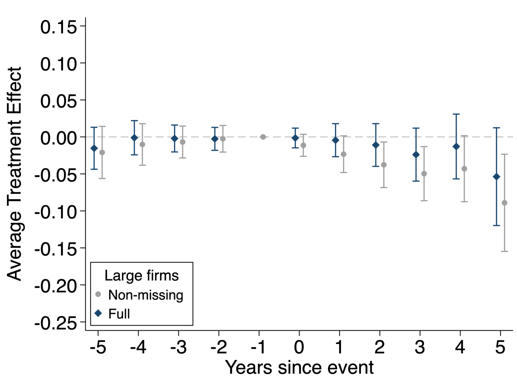
(c) Physical Contact



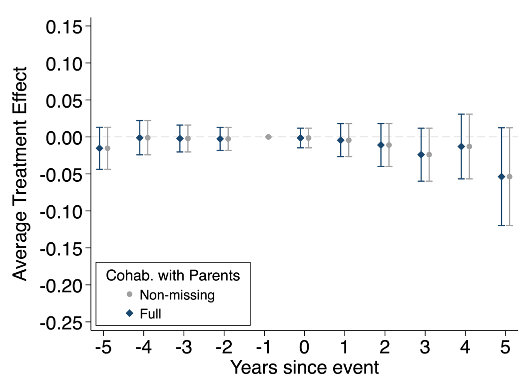
(d) Regular Workers



(e) Large Firms



(f) Cohabitance with parents



Note: This figure compares baseline event-study estimates of the eldercare penalty on female employment using (i) the full eligible sample and (ii) the restricted sample requiring non-missing subgroup-defining variables at event time -1 (as in the subgroup analyses). The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals.

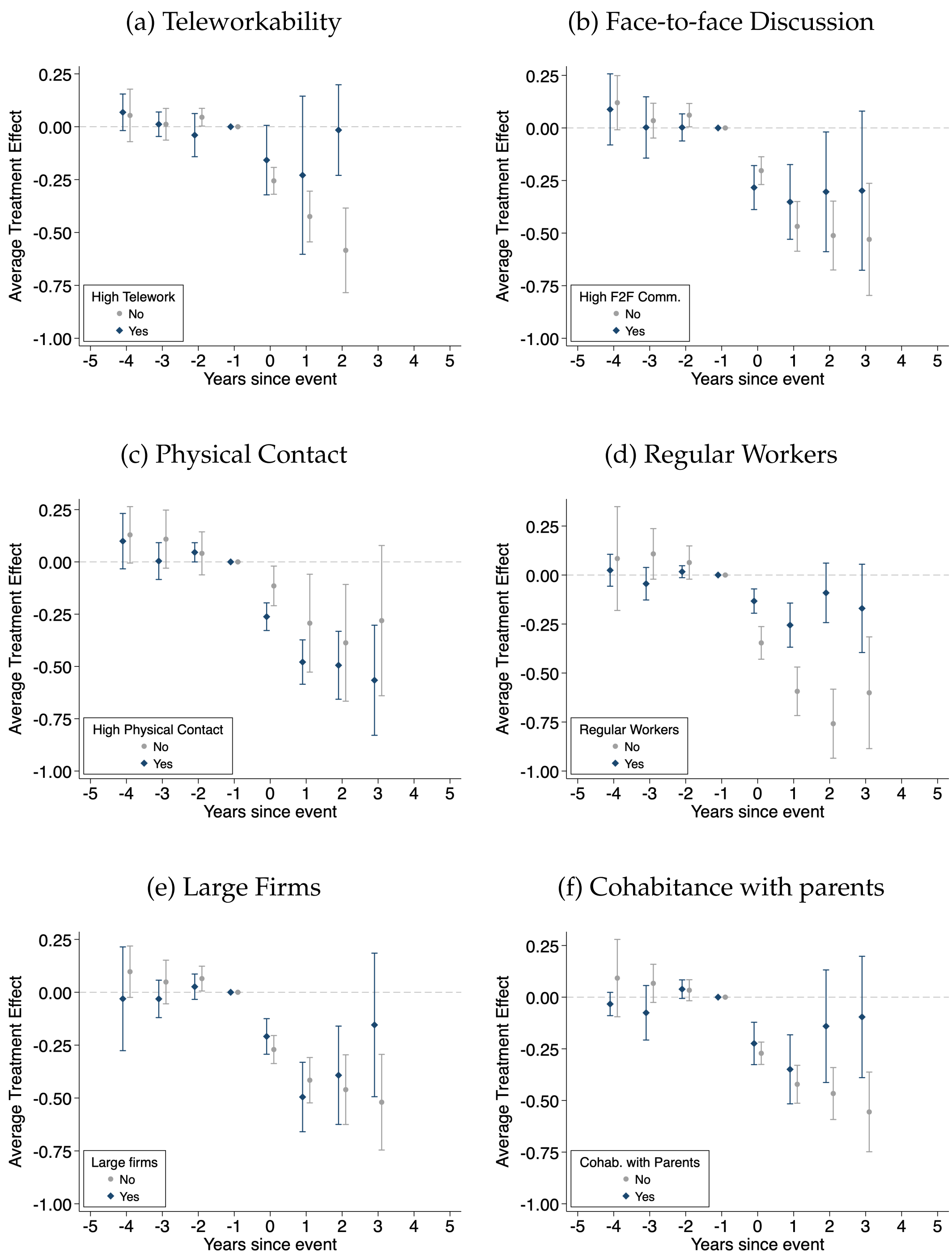
D Pre-COVID Analysis

The heterogeneity analyses in Figures 3 and 5 use occupation/task dimensions such as teleworkability and workplace interaction measures (e.g., face-to-face discussion and physical contact). A concern is that these dimensions may be correlated with COVID-era labor market shocks, so that heterogeneity patterns estimated over a sample that includes the pandemic could partly reflect pandemic-specific mechanisms rather than broader job flexibility.

Figures D.1 and D.2 replicate Figure 3 and 5 by using only the samples before 2019.[14] The main results do not change.

[14] I thank an anonymous referee for raising this issue.

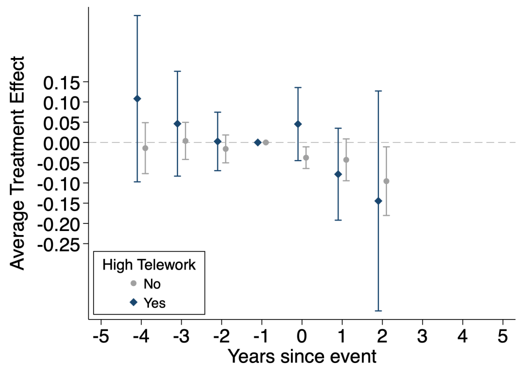
Figure D.1: Childcare Penalty, Heterogeneity (Female Employment, JPSED, Pre-COVID)



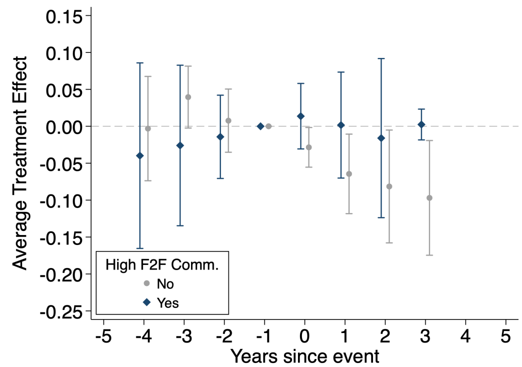
Note: This figure shows event-study coefficients for employment rates for females around the birth of their first child. The x-axis represents years since the event, with year 0 being the year of birth. Estimates are derived from JPSED data. The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals. The samples are restricted until 2019.

Figure D.2: Eldercare Penalty, Heterogeneity (Female Employment, JPSED, Pre-COVID)

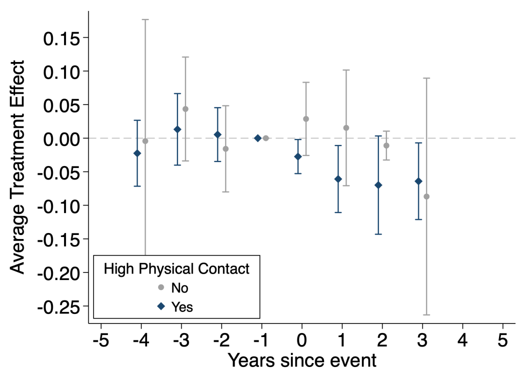
(a) Teleworkability



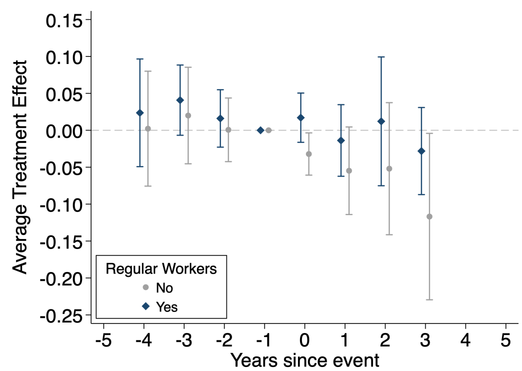
(b) Face-to-face Discussion



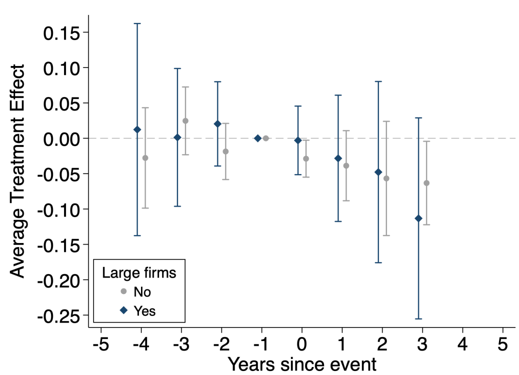
(c) Physical Contact



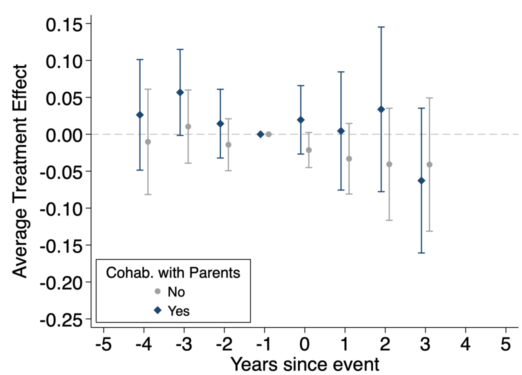
(d) Regular Workers



(e) Large Firms



(f) Cohabitance with parents



Note: This figure shows event-study coefficients for employment rates for females around the onset of caregiving needs. The x-axis represents years since the event, with year 0 being the year of the onset. Estimates are derived from JPSED data. The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. The estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals. The samples are restricted until 2019.

E Comparison to Conventional Estimator

In this section, I investigate how the estimated dynamic effects differ when using a conventional event-study estimator based on two-way fixed effects (TWFE) versus the baseline event-study estimator that accounts for staggered treatment timing. The motivation is that, under staggered adoption and heterogeneous dynamic treatment effects, conventional TWFE event studies can incorporate "forbidden comparisons" in which already-treated cohorts serve as controls for later-treated cohorts, potentially attenuating or distorting estimated post-event dynamics. By contrast, the Callaway–Sant'Anna estimator aggregates cohort-specific effects using not-yet-treated comparisons by construction.

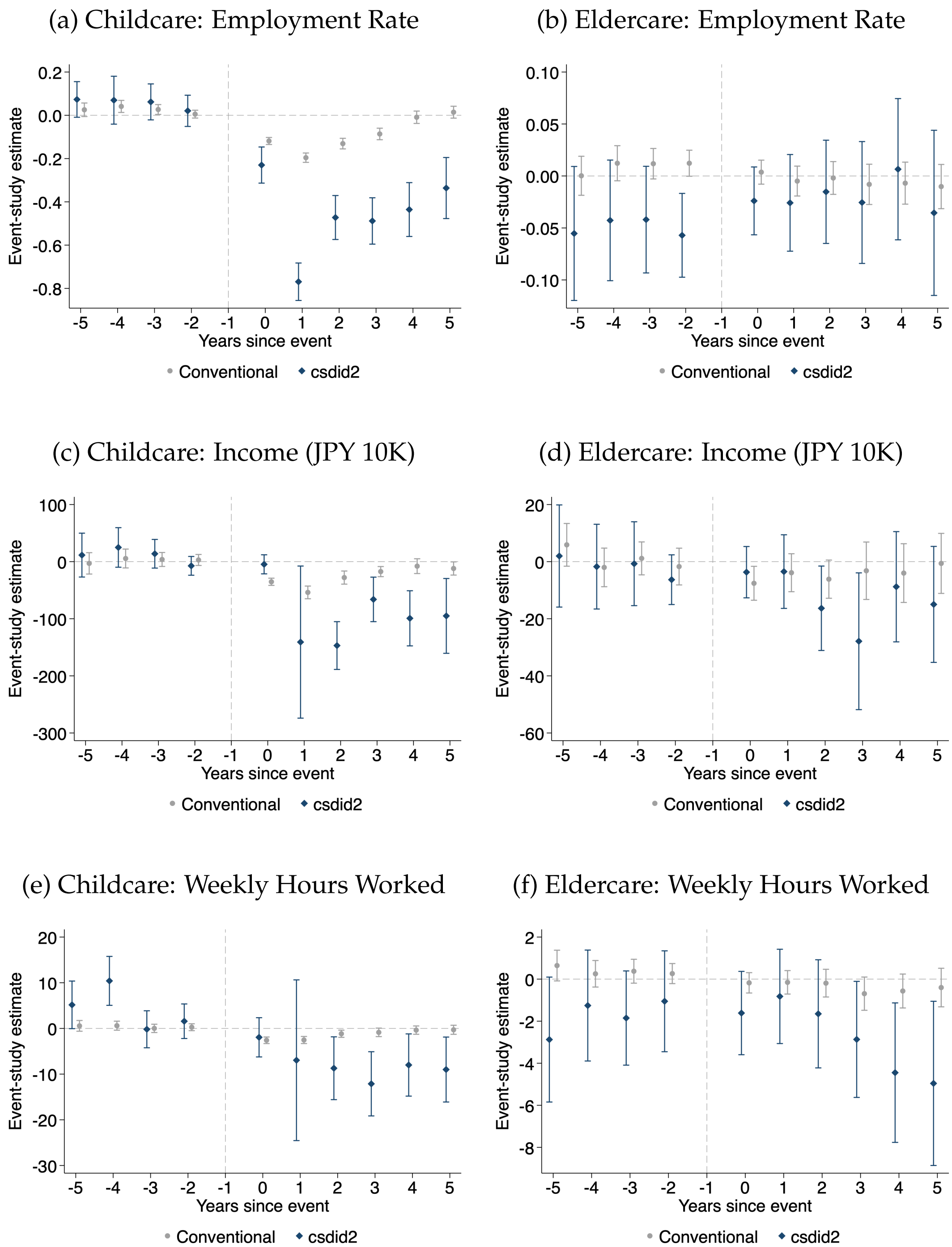
Figure E.1 compares the two event-study paths for female employment, income, and hours worked in the JHPS data for childbirth and eldercare.[15]

For employment rates, the difference is clear for the childcare penalty. The estimated penalty is very different across the two estimators: the conventional TWFE specification estimates that the declines are smaller and short-lived in female employment after the onset of childbirth. For eldercare, the choice of estimator matters less. The conventional TWFE event-study estimates are close to zero, whereas the estimator based on not-yet-treated comparisons yields more negative post-event estimates.

For annual earnings and hours worked, the conventional estimator concludes that the decline is smaller and short-lived for childcare. In the case of eldercare, the conventional estimator cannot detect the statistically significant declines in annual income and weekly hours worked.

[15] I use the JHPS data because the sample periods are longer and the issue of staggered treatment becomes more apparent.

Figure E.1: Comparison of Estimators (JHPS)



Note: This figure shows event-study coefficients for employment rates, annual income (JPY 10K), weekly hours worked for females around the onset of childcare and caregiving needs. The x-axis represents years since the event, with year 0 being the year of the onset. Estimates are derived from the JHPS data. The control groups are the not-yet-treated. The comparisons are within each group by pre-event characteristics. For the conventional estimate, I follow Kleven et al. (2019). For the csdid, the estimation follows Callaway and Sant'Anna (2021), implementing a doubly robust DiD estimator based on IPW and OLS. All regressions are weighted by sample weights. Covariates include age, age squared, and a dummy variable for having a partner. The bars represent 95% confidence intervals. Standard errors are clustered at the individual levels