WINPEC Working Paper Series No. E2601
April 2026

When Does Workplace Flexibility Help? Formal Care, Gender Norms, and Work–Care Trade-offs

Alice Chong Hiroyuki Motegi Masato Oikawa
Takumi Toyono Haruko Noguchi

Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

When Does Workplace Flexibility Help? Formal Care, Gender Norms, and Work–Care Trade-offs

Alice Chong * Hiroyuki Motegi † Masato Oikawa ‡ Takumi Toyono §
Haruko Noguchi ¶

April 3, 2026

Abstract

We provide causal evidence on whether working from home (WFH) enables workers to balance employment and eldercare, and how formal care infrastructure and gender norms shape this relationship. Exploiting Japan's COVID-19-induced remote work expansion, we find striking heterogeneity: WFH increases caregiving among part-time workers with positive health effects, but among full-time employees, only women increase caregiving — and their health deteriorates. Greater formal care availability and progressive gender norms substantially attenuate these effects. Realizing the work-care balance benefits of workplace flexibility requires complementary investments in care infrastructure and progress toward gender equality.

Keywords: working from home, informal caregiving, work-care balance, gender norms, long-term care, double burden, difference-in-differences (DID)

JEL Classification Numbers: I10, J14, J20

*Graduate School of Economics, Waseda University, and Waseda Institute of Social & Human Capital Studies (WISH), Tokyo, Japan, 1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, Japan, Email: alice.chong@fuji.waseda.jp

†National Institute of Population and Social Security Research, 2-2-3 Uchisaiwaicyo, Chiyoda-ku, Tokyo 100-0011, Japan. Email: motegihiro@gmail.com

‡Faculty of Education and Integrated Arts and Sciences, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050 Japan, and WISH, Tokyo, Japan Email: m.oikawa@waseda.jp

§WISH, Tokyo, Japan, Email: toyotaku0224@fuji.waseda.jp

¶Faculty of Political Science and Economics, Waseda University, Tokyo, Japan, and WISH, Tokyo, Japan, Email: h.noguchi@waseda.jp

1 Introduction

Population aging in developed countries poses a fundamental policy challenge: how can workers provide eldercare without sacrificing employment? This tension is intensifying globally as declining fertility reduces the ratio of working-age adults to elderly dependents, threatening both labor supply and public finances. The COVID-19 pandemic's acceleration of remote work offers a potential solution—but whether work-from-home (WFH) arrangements actually enable work-care balance, and at what cost, remains unknown.

Care can be provided through formal channels—nursing homes and professional services financed through markets—or informally by family members such as spouses and children. The balance between formal and informal care varies substantially across countries, shaped by institutions and prevailing social norms (Gruber et al., 2023).[1] Many people worldwide still rely heavily on informal care, making it crucial to understand how workers can balance employment with caregiving responsibilities.

A large health and labor economics literature has examined the impact of informal caregiving on labor supply and health using causal inference methods. Bauer and Sousa-Poza (2015) and Bom et al. (2019) provide a comprehensive survey showing that caregiving typically reduces labor supply and worsens both mental and physical health, with substantially greater burdens on women—effects driven primarily by time constraints and stress from balancing work and care.[2] Yet causal estimates vary substantially across studies—caregiving reduces employment by 5-20 percentage points and worsens mental health by 0.2-0.5 standard deviations—suggesting that modifiable institutional factors such as workplace flexibility may moderate these impacts.

Despite this large literature, direct causal evidence on whether flexible work arrangements help workers balance caregiving and employment remains surprisingly limited. Suggestive patterns point to flexibility' importance: comparing across countries, Western studies typically find caregiving reduces hours worked (intensive margin) but not employment (extensive margin), while studies from countries with rigid labor markets show the reverse—reduced employment but stable hours conditional on working.[3] This pattern suggests workplace flexibility may be critical: where workers can adjust hours or location, they remain employed while reducing intensity; where flexibility is constrained, caregiving forces binary exit decisions. Consistent with this interpretation, Niimi (2021) shows that workers granted flexible arrangements under family care leave legislation are

[1] Gruber et al. (2023) provide an international comparison of the costs of formal and informal care institutions and the labor market issues of caregivers. In particular, Fu et al. (2023) summarize the challenges facing Japan.

[2] Additional economic literature beyond Bauer and Sousa-Poza (2015) has examined these relationships. Schmitz and Stroka (2013), Oshio (2014), Niimi (2016), Wakabayashi and Kureishi (2018), Le and Ibuka (2023), and Abrahamsen and Grøtting (2023) investigate the impacts on health or well-being. Yamada and Shimizutani (2015) and Løken et al. (2017) analyzed the impacts on labor outcomes. From the perspective of substitution between formal and informal care, Fu et al. (2017), Geyer and Korfhage (2018) and Korfhage and Fischer-Weckemann (2024) examine the impacts on labor market outcomes.

[3] For example, Sugawara and Nakamura (2014), Fukahori et al. (2015), Yamada and Shimizutani (2015), Nishitateno and Shikata (2017), Oshio and Usui (2018), Ando et al. (2021), Niimi (2021) and Nishimura and Oikawa (2025).

significantly less likely to quit within one year of a parent requiring care, while Abrahamsen and Grøtting (2023) find that Norway's expansion of formal care reduced leave-taking primarily among workers without flexible job arrangement.

Yet, we lack direct causal evidence on whether and how WFH—the most widespread form of workplace flexibility post-pandemic—affects caregiving decisions and caregiver outcomes.[4] This gap has major policy implications. Economically, WFH reduces caregiving's time cost by eliminating commutes and enabling schedule flexibility (Bloom et al., 2015). However, existing evidence suggests it may reduce productivity for some workers (Morikawa (2022) and Shen (2023)) or carry career advancement penalties, particularly for women who use flexible arrangements (Goldin (2014); Kleven et al. (2019))—creating a potential equity-efficiency trade-off that varies by worker type and gender. While caregiving typically worsens caregiver mental health (Le and Ibuka, 2023), whether flexible work arrangements mitigate these health costs remains an open empirical question. The broader WFH literature has examined productivity effects (Alipour et al. (2021), Morikawa (2022) and Shen (2023)) and impacts on housework and fertility (Pabilonia and Vernon (2022) and Inoue et al. (2024)), but not eldercare.[5] [6] As remote work remains widespread post-pandemic, understanding whether—and for whom—WFH enables work-care balance is essential for designing effective workplace and long-term care policies.

This paper provides the first causal evidence on whether WFH enables workers to balance caregiving and employment, and at what cost to their health and well-being. We exploit the sudden, large-scale expansion of WFH during COVID-19 as a quasi-experiment, using a difference-in-differences (DID) design that compares workers in occupations with high versus low pre-pandemic WFH availability. Our findings inform whether promoting WFH as a caregiving accommodation—increasingly common in post-pandemic workplace policy—achieves its intended goals without unintended costs to caregivers. Our analysis examines heterogeneity by gender, employment type (regular versus non-regular), access to formal care, and prevailing social norms —dimensions that theory suggests should matter but that have received limited empirical attention in the caregiving literature.

We find striking heterogeneity in how WFH affects the care-work balance. For non-regular (part-time and contract) workers, WFH increases caregiving for both men (by approximately 24 percentage points) and women (by about 5 percentage points), with positive effects on health. For regular (full-time permanent) workers, however, WFH increases caregiving only among women (by about 13 percentage points)—and worsens their health. These patterns suggest WFH enables work-care balance when jobs are already flexible, but may intensify the "double burden" for women in

[4] Niimi (2021) examines legislated caregiver leave using cross-sectional variation; we analyze WFH using panel variation and a quasi-experimental shock, and we explicitly test for health and well-being impacts—both central to our analysis.

[5] Barrero et al. (2023) and Lee (2023) provide comprehensive reviews of the WFH literature.

[6] Lunde et al. (2022) analyzed WFH effects on health. Chong and Noguchi (2024) and Motegi and Oikawa (2024) investigate the effects of WFH on pregnancy and human capital investment respectively.

demanding careers. Importantly, we show these effects are substantially offset in regions with greater formal care availability and more progressive gender norms, pointing to complementary policies that could reduce caregiver burden. WFH has no significant effects on labor market outcomes, suggesting it may prevent the employment exits documented in prior caregiving studies.

Our findings make three contributions to the literature. First, we provide the first causal evidence on whether workplace flexibility enables work-care balance, directly addressing a key policy question raised by prior research documenting caregiving's labor costs (Bauer and Sousa-Poza (2015)).[7] Second, we demonstrate fundamental heterogeneity by gender, employment type, formal care availability, and social norms, providing among the first economic evidence on how eldercare norms interact with workplace policy (Bertrand and Pan (2015) and Kleven (2025)) for related evidence on childcare).[8] [9] Third, we contribute to the growing WFH literature (Dingel and Neiman (2020) and Barrero et al. (2023)) by showing remote work has double-edged implications for gender inequality—enabling work-care balance for some women but intensifying the double burden for others—with effects that depend critically on job characteristics and institutional context.[10]

Our analysis uses unique longitudinal data from Japan, which provides an especially informative empirical setting. Japan's historically low WFH adoption (23% of workers in 2020 compared to 58% in the United States)[11] makes the COVID-19-induced expansion a particularly large shock, likely producing more detectable effects than in countries where remote work was already common. Moreover, Japan's substantial regional variation in formal care availability and gender norms provides valuable identification for examining effect heterogeneity. While Japan's rigid labor market and employment dualism (Yokoyama et al., 2021) make it an especially clear setting for identifying flexibility effects, our findings on the interaction between workplace arrangements, formal care, and gender norms likely extend to other aging societies facing similar work-care tensions.

As remote work remains widespread post-pandemic, our results suggest WFH alone is insufficient to address caregiving challenges—particularly for women in regular employment who may face intensified work-care conflict. Complementary policies expanding formal care and challenging traditional gender norms appear necessary to prevent flexible work arrangements from reinforcing rather than reducing gender inequality in caregiving burdens.

The paper proceeds as follows. Section 2 describes our conceptual framework. Section 3 presents

[7] Niimi (2021) examines legislated caregiver leave using cross-sectional variation; we analyze WFH using panel variation and a quasi-experimental shock, and we explicitly test for health and well-being impacts—both central to our analysis.

[8] For studies on gender norms in Japan, see Rodríguez-Planas and Tanaka (2022) and Hara and Rodriguez-Planas (2023).

[9] Motegi (2025) is a recent study that examines norms concerning the elderly, such as the norm of filial responsibility.

[10] Barrero et al. (2023) and Lee (2023) provide comprehensive reviews of the WFH literature.

[11] According to the international comparative survey of four countries in the Ministry of Internal Affairs and Communications' Information and Communications White Paper 2023, the proportion of people utilising WFH and online meetings in Japan stood at 32.4% as of 2023. This is considerably lower compared to the United States (50.4%), Germany (54.6%), and China (72.7%). Furthermore, the Ministry of Land, Infrastructure, Transport and Tourism's Survey on the Actual State of the Telework Population indicates that as of 2020, the WFH population in Japan accounted for 23.0% of the workforce, whereas in the United States it was 57.9% and the EU average was 36.5%.

our dataset and variable construction. Section 4 outlines our identification strategy and validates underlying assumptions. Section 5 reports results on caregiving, health, labor market outcomes, and heterogeneity by institutional context. Section 6 concludes with policy implications.

2 Conceptual framework

This section develops a conceptual framework that (1) formalizes the mechanisms through which WFH affects time allocation and the work-care trade-off, (2) generates testable predictions about heterogeneous effects across gender, employment type, and institutional contexts, and (3) clarifies the assumptions underlying our empirical strategy. While we estimate reduced-form specifications rather than structural parameters, the framework helps interpret our findings and motivates our heterogeneity analysis.

2.1 Set up

We model an individual who derives utility from consumption C, leisure L, and the quality of care provided to an elderly parent Q:

$$U = u(C, L, Q) \tag{1}$$

The model builds on standard time allocation frameworks and the caregiving literature (Houtven et al. (2013) and Bauer and Sousa-Poza (2015)). Total available time $T(= 24)$ must be allocated across paid work hours H, informal caregiving time CG, leisure L and commuting. Let baseline commuting time be Δ, so the time constraint under office work is:

$$T = H + CG + L + \Delta \tag{2}$$

The budget constraint is:

$$C = wH - pFC \tag{3}$$

where w is the hourly wage, FC denotes formal care services, and p is their price. The prices of consumer goods are normalized to 1. Care quality is produced using informal and formal care:

$$Q = f(CG, FC) \ \ with \ \ \frac{\partial f}{\partial CG} > 0, \ \ \frac{\partial f}{\partial FC} > 0 \tag{4}$$

We assume informal and formal care are not perfect substitutes $\frac{\partial^2 f}{\partial CG \partial FC} \neq 0$ - consistent with evidence that informal care provides emotional support and monitoring that formal care cannot fully replace (Bauer and Sousa-Poza, 2015). We remain agnostic about whether they are complements or substitutes at the margin ($\frac{\partial^2 f}{\partial CG\,\partial FC} > 0$ or $\frac{\partial^2 f}{\partial CG\,\partial FC} < 0$), with our empirical results suggesting substitution $\frac{\partial^2 f}{\partial CG\,\partial FC} < 0$.

2.2 Effect of WFH

WFH affects the optimization problem through three channels. First, WFH eliminates commuting time ($\Delta = 0$), effectively relaxing the time constraint:

$$T = H + CG + L. \tag{5}$$

Second, WFH reduces the effective time cost of caregiving through two channels: it allows temporal flexibility to interleave work and care tasks during the day, and it eliminates spatial barriers to monitoring or providing care (Bloom et al., 2015). We capture this by assuming that under WFH, one unit of caregiving time is equivalent to $(1-\delta)$ units under office work, where $\delta \in (0, 1)$ represents the efficiency gain from flexibility. For example, a worker might check on an elderly parent briefly during a work break (feasible at home but not in an office), or coordinate care services via phone calls during gaps between tasks—activities that are prohibitively costly when physical office presence is required. The time constraint under WFH can thus be written as:

$$T = H + (1 - \delta)CG + L \tag{6}$$

This formulation implies that WFH effectively expands usable time for caregiving, lowering its shadow price relative to market work. Third, WFH may influence productivity and thus wages. Let wages be w_0 under office work and w_1 under WFH, where w_1 may be higher, lower, or equal to w_0 depending on job characteristics (Morikawa (2022) and Shen (2023)). The individual chooses optimal values (H^*, CG^*, FC^*, L^*) to maximize utility (1) subject to these constraints (2), (3), (4) and (6). Under standard assumptions, we expect the following comparative statics:[12]

Informal care:

$$\frac{\partial^2 CG^*}{\partial WFH} > 0 \tag{7}$$

when time constraints were previously binding (Δ large), efficiency gains are substantial (δ high), and productivity effects are modest ($w_1 \approx w_0$).

Formal care:

$$\frac{\partial^2 FC^*}{\partial WFH} < 0 \tag{8}$$

[12] An increase in WFH corresponds to a decrease in Δ.

when informal and formal care are substitutes ($\frac{\partial^2 f}{\partial CG \partial FC} < 0$).

Work hours:

$$\frac{\partial^2 H^*}{\partial WFH} \gtrless 0 \tag{9}$$

is theoretically ambiguous: time savings from eliminated commutes and increased caregiving efficiency may enable longer hours, but reduced productivity ($w_1 < w_0$) or career penalties associated with WFH may reduce labor supply. The net effect depends on the incremental effect from allocating the surplus time generated by reduced commuting time and the substitution effect from changes in productivity due to WFH.

2.3 Heterogeneity by gender and employment type

Gender Difference

We model gender differences through three channels: (1) lower market wages for women ($w_{female} < w_{male}$), (2) stronger caregiving norms (higher marginal utility from care quality for women), and (3) larger career penalties ($\pi_{female} > \pi_{male}$), where π represents implicit career costs from using workplace flexibility for caregiving, including foregone promotions, reduced workplace visibility, and statistical discrimination.[13] Prior research suggests these penalties fall disproportionately on women (Goldin (2014) and Kleven et al. (2019)). If women face lower wages or stronger social norms favoring family caregiving, their opportunity cost of informal care is lower (Carmichael and Charles, 2003). This implies the responsiveness of caregiving to WFH is larger for women:

$$\left.\frac{\partial CG^*}{\partial WFH}\right|_{\text{female}} > \left.\frac{\partial CG^*}{\partial WFH}\right|_{\text{male}} \tag{10}$$

However, if women maintain high work hours while increasing caregiving, WFH may generate a double burden, reducing health or well-being.

Employment type differences

Regular (full-time, permanent) and non-regular (part-time, contract) workers differ systematically in the Japanese labor market Yokoyama et al. (2021). Regular workers typically face higher wages, longer hours, longer commutes, and larger career penalties for flexibility. In contrast, non-regular workers typically face lower wages, shorter hours, greater baseline flexibility, and flatter career trajectories. The opportunity cost of caregiving differs systematically across these groups. For regular workers, the implicit price of informal care includes not only foregone wages (wCG) but also potential career penalties from reduced workplace presence or flexibility utilization. The full opportunity cost of caregiving for regular workers is $wCG + \pi WFH$, where πWFH captures career

[13] Long carrer penalties can be discussed within a framework nearly identical to that of significant wage reductions in this static models.

penalties that manifest when flexible work is used for care. If $\pi_{female} > \pi_{male}$ due to statistical discrimination or gender norms, then even when WFH reduces the time cost of care, it may simultaneously impose larger career penalties on women, potentially worsening their health through a double-burden effect. These differences generate two key predictions.

Non-regular workers (Prediction 1): WFH should increase caregiving for both men and women because opportunity costs are low (low w, low π) and time savings represent a net gain without major career costs:

$$\left.\frac{\partial CG^*}{\partial WFH}\right|_{\text{non-regular, male}} > 0, \quad \left.\frac{\partial CG^*}{\partial WFH}\right|_{\text{non-regular, female}} > 0 \tag{11}$$

Among non-regular workers, there is little gender difference in working arrangements; as a result, it is difficult to assess whether the estimated coefficients are larger for men or for women. Health impacts should be neutral or positive, since increased caregiving does not come at the expense of intensive work commitments.

Regular workers (Prediction 2): WFH effects depend critically on norms and career penalties. For men: If gender norms impose psychic costs κ of caregiving or if career penalties π_{male} are sufficiently large relative to time savings Δ, the time savings from WFH may be directed toward market work or leisure rather than care:

$$\left.\frac{\partial CG^*}{\partial WFH}\right|_{\text{regular, male}} \approx 0 \tag{12}$$

if $\pi_{male} + \kappa$ are sufficiently large relative to the value of time saving $\Delta + \delta$.

For women: WFH may enable "accommodating" career adjustments, raising caregiving:

$$\left.\frac{\partial CG^*}{\partial WFH}\right|_{\text{regular, female}} > 0 \tag{13}$$

but potentially at the cost of health deterioration , if π_{female} is large and women face the double burden of maintaining career hours while absorbing increased care responsibilities.

2.4 Formal care and social norms

Formal care supply

With greater availability of formal care (lower p, higher supply):

$$\frac{\partial^2 CG^*}{\partial WFH\,\partial(\text{FC})} < 0 \tag{14}$$

WFH thus has smaller caregiving effects in regions with developed long-term care infrastructure, as formal care substitutes for the informal care that WFH would otherwise enable (Fu et al. (2017)

and Løken et al. (2017)). Intuitively, when formal care is readily available and affordable, the marginal value of time savings from WFH is lower because families have already substituted toward purchased care. WFH availability and formal care supply are thus economic substitutes in producing care quality.

Social norms

Where traditional norms are stronger (filial responsibility, gender role attitude), we expect:

$$\frac{\partial^2 CG^*}{\partial WFH\, \partial(\text{Conservative norms})} > 0 \tag{15}$$

especially for women. Progressive gender norms should attenuate WFH's caregiving effects by reducing the perceived obligation or career penalty asymmetry (lowering π_{female} relative to π_{male}) (Bertrand and Pan (2015) and Kleven (2025)).

2.5 Empirical implications

This framework guides our empirical analysis in several ways. First, we estimate the causal effect of WFH on caregiving, testing whether $\frac{\partial CG^*}{\partial WFH} > 0$. Second, we examine heterogeneity by gender and employment type to test Predictions 1 and 2. Third, we investigate health impacts to test whether WFH creates a double burden among regular female workers. Fourth, we use triple-difference specifications to test whether formal care availability and social norms moderate WFH effects as predicted.

The framework clarifies our identification assumptions. Our DID strategy identifies the causal effect of WFH availability on caregiving under two key assumptions: first, occupations with different pre-pandemic WFH potential would have followed parallel trends in caregiving absent COVID-19; second, COVID-19 affected caregiving decisions primarily through its effect on WFH adoption, not through differential shocks to care needs or preferences across occupation types. The first assumption is standard for DID and testable through pre-trends analysis. The second is plausible because we condition on having family members needing care and control for occupation-invariant COVID shocks through time fixed effects. Moreover, if COVID-19 differentially affected care needs by occupation (e.g., if essential workers' parents faced higher infection risk), this would bias against finding effects, since our comparison groups would be contaminated by care-need shocks. Our results are thus likely conservative estimates of WFH's true effect. Under these assumptions, our DID estimates identify the causal effect of WFH availability through both the time savings channel (eliminating commutes, $\Delta = 0$) and the flexibility channel (reducing effective caregiving time cost through δ) specified above.

The next section describes our data and empirical setting, which provides substantial variation in both WFH adoption and the institutional factors our framework identifies as key moderators—formal care availability and social norms.

3 Data

3.1 Data source and sample construction

We use the Japanese Panel Study of Employment Dynamics (JPSED), an annual longitudinal survey of Japanese workers aged 15 and above conducted by the Recruit Works Institute. The survey is nationally representative, with sampling weights constructed to match the Labor Force Survey population composition (Recruit Works Institute, 2025). A key advantage for our analysis is that the JPSED measured WFH contemporaneously in each wave, avoiding the substantial recall bias inherent in retrospective questions about pre-COVID work arrangements (Yamagata and Miura, 2023). Furthermore, this paper applies DID as identification strategies, and it also holds an advantage from the perspective of checking the common trend assumption; pre-trend analysis.

We use waves from 2017 to 2022, as earlier waves lacked questions on WFH and well-being. COVID-19 infections in Japan were first confirmed in February 2020. Because respondents report their employment and job information as of December of the previous year, the information for 2019 and that for 2020 are completely separated with respect to the impact of COVID-19 (Kawaguchi and Motegi, 2021).

Our analysis sample includes employed individuals aged 25-65 who resided in the same prefecture at age 15 and in 2019. We exclude self-employed individuals since our focus is workplace-provided WFH arrangements. The age restriction excludes students and retirees. The residential stability restriction - requiring individuals to remain in the same prefecture during both adolescence and adulthood - increases the likelihood that individuals live near their aging parents and thus have greater potential for caregiving responsibilities.[14] Within this sample, we further restrict to individuals reporting at least one family member needing care because individuals without a care recipient have no need to WFH for caregiving purposes.

Table A.1 presents summary statistics for our analysis sample. The sample is 50% female with mean age 47. Two-thirds are married and 57% have children. Regarding employment, 66% are regular (full-time permanent) employees, while 34% are non-regular (part-time or contract) workers. Prior to COVID-19, only 11% of workers engaged in WFH, rising sharply after 2020. Among those with family members needing care, 6% provide some care, with 4% providing intensive care and 2% providing moderate care.

Like most longitudinal surveys, the JPSED experiences panel attrition over time. Section 4.2 demonstrates that pre-treatment characteristics remain balanced between treatment and control

[14]Research on Japanese family structure shows that elderly parents rarely relocate and that residential proximity between generations strongly predicts caregiving relationships (Raymo et al., 2010). Individuals who remained in the same prefecture from age 15 (when most live with parents) to 2019 are therefore more likely to live near aging parents—about 57.7% of elderly care recipients in Japan receive care from family members in the same household or nearby (Ministry of Health, Labour and Welfare, 2022). This restriction ensures our sample has high caregiving potential while avoiding endogenous migration in response to care needs, and applies uniformly across all 47 prefectures.

groups, suggesting attrition does not systematically differ by treatment status. We now describe the construction of our key variables. Details on variable construction and the questionnaire content are explained in the appendix.

3.2 Key variable construction

3.2.1 Treatment: Potential WFH availability by occupation

Our identification strategy requires a measure of WFH availability that is predetermined relative to COVID-19-era caregiving decisions. We construct occupation-level potential WFH availability using pre-COVID data to avoid endogeneity from individuals selecting into WFH-compatible occupations in response to caregiving needs. This data also includes variables indicating whether individuals actually engaged in WFH, such as the number of hours worked remotely. The rationale for using WFH availability will be discussed in Section 4.

We measure WFH availability using the question, "Had a telework system been introduced in your workplace as of last December, and were you a teleworker?" We code WFH as available if the system was introduced (regardless of personal usage) and unavailable otherwise. For each occupation, we calculate the proportion of workers with WFH availability in 2019 (pre-COVID). We then classify occupations with above-median availability as the treatment group and below-median as the control group.

Figure 1 shows this categorization yields 22 high WFH occupations (e.g., system engineers, finance professionals, media workers) and 23 low-WFH occupations (e.g., medical technicians, drivers, social welfare workers). We assign treatment status based on individuals' December 2019 occupation. This classification exploits the fact that COVID-19 expanded WFH across all occupations, but the magnitude of expansion was larger for occupations with higher pre-pandemic WFH potential.

3.2.2 Caregiving outcomes

We measure caregiving intensity using responses to: "Do you usually take care of a family member?" We construct three binary outcomes. "Provide care" equals 1 if respondents do all, most, or share caregiving equally with other family members. "Intensive care" equals if respondents do all or most caregiving. "Moderate care" equals 1 if respondents share caregiving equally. These categories capture different margins of care provision that may respond differently to WFH availability.

3.2.3 Health outcomes

We construct a health index following the Somatic Symptom Scale-8 (SSS-8) methodology (Gierk et al., 2014).[15] Respondents rate eight physical and mental symptoms on a scale from 0 (always

[15] SSS-8 is a validated self-report instrument for assessing physical symptom burden (Gierk et al., 2014). The eight symptoms assessed are: headaches/dizziness, back/waist/shoulder pain, palpitations/shortness of breath, extreme fatigue, nervousness, depression, appetite loss, and sleep difficulty. Our JPSED measures closely follow but are not

experiencing the symptom) to 4 (never experiencing it). We sum across items, yielding an index from 0 (poorest health) to 32 (best health). We create binary indicators for health exceeding the sample's first quartile (15), median (19), and third quartile (23), allowing us to examine effects across the health distribution.

3.2.4 Labor market and well-being outcomes

We examine standard labor market outcomes: employment status, annual hours worked, annual income, hourly wages, and regular employment status. Hourly wages are constructed from reported hourly pay when available or by dividing annual earnings by annual hours (assuming 50 work weeks) for salaried workers.

We measure well-being using binary indicators for happiness (feeling happy or very happy), life satisfaction (satisfied or fairly satisfied), and good work-life balance (not stressed or not really stressed about balancing work and family). These capture different dimensions of caregiver welfare beyond labor market and health outcomes.

4 Econometric Model

4.1 Econometric model

Our empirical strategy estimates the causal effect of WFH availability on caregiving decisions and caregiver outcomes using a difference-in-differences (DID) design. We exploit the COVID-19 pandemic as an exogenous shock that differentially increased WFH adoption across occupations based on their pre-pandemic WFH potential. Our baseline specification tests the primary prediction from Section 2 that WFH increases caregiving ($\frac{\partial CG}{\partial WFH} > 0$) by reducing time costs and enabling greater scheduling flexibility. The estimation equation is:

$$Care_{it} = \alpha_0 + \alpha_1 Treat_i + \alpha_2 Post_t + \beta Treat_i \times Post_t + x_{it}'\gamma + \theta_i + \phi_t + u_{it}, \tag{16}$$

where indices i and t represent the individual and time on an annual basis respectively. The dependent variable $Care_{it}$ signifies the different care intensity variables of individual i in year t, encompassing the three dummy variables taking value 1 if the individual provided any care, moderate care, or intensive care. The variable $Treat_i$ is a dummy variable that indicates whether the individual i is in the treatment group, meaning they were in an occupation with high potential WFH availability as of December 2019.[16] The variable $Post_t$ is a dummy variable equal to 1 if the survey year is after the onset of the COVID-19. The vector x_{it}' comprises control variables,

identical to the original SSS-8 items.

[16] This strategy is similar to Cowan and Spearing (2026).

including age and age squared, marriage dummy, cohabitation dummy with spouse, cohabitation dummy with child, firm size category dummies,[17] years of tenure (quadratic function), cross-terms for graduation in 2020 and year dummy variables, cross-terms for firm size in 2020 and year dummy variables, and prefecture-year fixed effects. The variables θ_i and ϕ_t represent individual and year fixed effects respectively. Individual fixed effects θ_i absorb time-invariant characteristics including baseline caregiving preferences, family structure, occupation-specific attributes, and baseline health. Year fixed effects ϕ_t absorb aggregate shocks from COVID-19 that affect all workers equally (e.g., restrictions on hospital visits, general health trends, macroeconomic conditions). u_{it} denotes the unobserved error term. The coefficient of the cross term between $Treat_i$ and post COVID-19, β, is the parameter of interest in this study. Under our identifying assumptions (detailed in Section 4.2), β identifies the average treatment effect of WFH availability on caregiving for workers in high-WFH-potential occupations relative to those in low-WFH-potential occupations. The widespread shift to remote work due to COVID-19 in 2020 was virtually impossible to predict at the individual level. The key point for identification is that the impact varies across occupations. Although its internal validity is lower than that of analyses based on RCTs such as Bloom et al. (2015), its external validity is high.

To test Predictions 1 and 2 from our conceptual framework—that WFH effects vary systematically by gender and employment type—we estimate equation (16) separately for four subgroups: non-regular male workers, non-regular female workers, regular male workers, and regular female workers. Our framework predicts that non-regular workers of both genders should increase caregiving in response to WFH (Prediction 1), while among regular workers, only women should increase caregiving, potentially at the cost of health deterioration (Prediction 2). These heterogeneous specifications allow us to test whether opportunity costs and career penalties shape the caregiving response to WFH as our model predicts.

We also estimate the effects of WFH on health outcomes, labor market outcomes, and well-being using the same specification, replacing the dependent variable with the relevant outcome measure. For health outcomes, we use both the continuous health index and binary indicators for exceeding the first quartile, median, and third quartile of the health distribution. For labor market outcomes, we examine employment status, annual hours worked, annual income, hourly wages, and regular employment status. These analyses test the model's prediction that labor supply effects are ambiguous ($\frac{\partial H}{\partial WFH} \gtrless 0$) and examine whether WFH imposes health costs, particularly for regular female workers facing a double burden. Furthermore, we restrict the sample to those already engaged in caregiving for this analysis. This allows us to examine the impact of balancing WFH and caregiving on these outcomes.

To test whether formal care availability and social norms moderate WFH effects as predicted in Section 2.4, we estimate following specifications:

[17] The categories of firm size include the following four options: small-sized (less than 100 employees), middle-sized (between 100 and 999 employees), large-sized (1,000 or more employees), and government employees.

$$
\begin{aligned}
Care_{it} = {} & \alpha_0 + \alpha_1 Treat_i + \alpha_2 Post_t + \beta Treat_i \times Post_t \\
& + \sum_{j=1}^{5} \delta_j (PrefCharacteristics_j \times Treat_i \times Post_t) + x'_{it}\gamma + \theta_i + \phi_t + u_{it}
\end{aligned} \tag{17}
$$

where $PrefCharacteristics_j$ is a vector of prefecture-level characteristics including the availability of long-term care insurance (LTCI) facilities, in-home service availability, gender norms, aging norms, and urban status. The coefficients δ_j test whether these contextual factors moderate WFH's effect on caregiving. Our framework predicts $\delta_j < 0$ for formal care availability (WFH and formal care are substitutes) and $\delta_j > 0$ for conservative norms (stronger norms amplify caregiving responses to WFH, particularly for women). We also include metropolitan area dummies and prefecture fixed effects to control for other regional differences.

To examine whether WFH's health effects operate specifically through the caregiving channel, we also estimate a DDD specification that compares caregivers versus non-caregivers:

$$
\begin{aligned}
Y_{it} = {} & \alpha_0 + \alpha_1 Treat_i + \alpha_2 Post_t + \alpha_3 CareNeed_{it} \\
& + \alpha_4 Treat_i \times Post_t + \alpha_5 Treat_i \times CareNeed_{it} + \alpha_6 Post_t \times CareNeed_{it} \\
& + \beta Treat_i \times Post_t \times CareNeed_{it} + x'_{it}\gamma + \theta_i + \phi_t + u_{it}
\end{aligned} \tag{18}
$$

where $CareNeed_{it}$ is a dummy variable equal to 1 if individual i lives with someone requiring care at time t, and 0 otherwise. The coefficient β identifies the differential effect of WFH for those with caregiving responsibilities relative to those without, netting out any direct effects of WFH on outcomes that operate independently of caregiving. This specification is particularly valuable for health outcomes, where WFH may affect caregivers and non-caregivers differently.

We use occupation-level potential WFH availability as our treatment variable rather than individual WFH adoption, an approach similar to Inoue et al. (2024) and Cowan and Spearing (2026). This yields intention-to-treat (ITT) estimates based on DID: the impact of being in a high-WFH-potential occupation, regardless of actual WFH adoption. ITT estimates avoid selection bias and are policy-relevant because they capture the effect of expanding WFH availability in an occupation—the intervention policymakers control. Moreover, using individual WFH adoption as an endogenous variable would violate the exclusion restriction, since COVID-19 affected caregiving through multiple channels beyond WFH (e.g., pandemic-induced restrictions on hospital and nursing facility visits). The ITT framework avoids this problem.

4.2 Assumption for identification

We now validate the assumptions of this identification strategy with particular focus on equation (16). First, we perform a balance check. Table 1 presents the average values of individual characteristics for the treated and control groups, along with the differences between them, to assess whether the characteristics of the two groups were statistically different in 2019. Statistically significant differences are observed for age (2.96 percentage points), living with non-spouse family (11.88 percentage points), educational attainment (31.42 percentage points), working hours (5.08 percentage points), and firm size. We control these variables either by including them directly in our specification or through fixed effects. Regarding firm size, whether the firm is in the public sector makes a significant difference, so firm size is also controlled for. The same applies to years of tenure. Since educational attainment and firm size have substantial differences, and since the impact of potential WFH availability on these variables may differ before and after COVID-19, we also control for the cross-term with year for these two variables.

Figure 2 shows the results of regression for each variable on potential WFH availability defined as a continuous variable. This figure indicates that potential WFH availability does not have a statistically significant effect for 9 out of the 16 variables. However, the remaining 7 variables are statistically significant, and since the effects of variables like educational background are large, they must be controlled for. These balance checks inform our choice of control variables in equations (16), (17) and (18), ensuring that any remaining imbalance is addressed through our empirical specification.

Next, we check the common trend assumption, which is important for identification in DID. To do this, we estimate an event study model. We focus on the three care outcome variables and actual working from home, namely, cases where WFH constitutes 40% or more of weekly working hours and those where it constitutes 60% or more. The results for WFH are shown in Figure 3 (a) and (b). For each graph, we observe that while the increase was not statistically significant prior to COVID-19, it became statistically significant after 2020. The expansion of potential WFH availability due to ITT arising from COVID-19 increased the probability of actual WFH implementation. Therefore, the common trend assumption holds. We also observe that WFH availability immediately prior to COVID-19 significantly influenced the engagement of actual WFH. Furthermore, examining care responsibilities in graphs from (c) to (e) in Figure 3, no statistically significant differences existed prior to COVID-19. However, no statistically significant differences were observed post-COVID-19 either.

We further examine heterogeneity by stratifying according to gender and employment status, shown in Figure 4. The rationale for this is explained later in the results section. Regarding (a) and (b), as with Figure 3, these concern the implementation of WFH. While female regular employment in 2017 did not satisfy the common trend assumption, the other strata appear to satisfy it without issue. Furthermore, for all employment types and both genders, the implementation of WFH has increased since 2020, although some parts are not statistically significant.

Regarding caregiving Figures 4 (c) to (e), similar to WFH, the common trend assumption holds for most caregiving types with some exceptions. Notably, it does not hold for certain categories, such as female non-regular employment in 2018. Furthermore, it is evident that male non-regular employment increases caregiving, while female regular employment increases intensive care after 2020.

Our DID strategy identifies the causal effect of WFH availability under two key assumptions, which we now state formally:

Assumption 1 (Parallel Trends): Let $Care_{it}(0)$ denote potential caregiving outcomes in the absence of COVID-19. Then:

$$\begin{aligned}&\mathbb{E}[\text{Care}_{it}(0) \mid \text{Treat}_i = 1] - \mathbb{E}[\text{Care}_{i,2019}(0) \mid \text{Treat}_i = 1] \\ &= \mathbb{E}[\text{Care}_{it}(0) \mid \text{Treat}_i = 0] - \mathbb{E}[\text{Care}_{i,2019}(0) \mid \text{Treat}_i = 0], \quad \forall\, t > 2019.\end{aligned} \tag{19}$$

This formally states that treatment and control groups would have experienced parallel trends in caregiving absent COVID-19.

This assumption requires that any pre-existing differences in caregiving trends between treatment and control groups are captured by our control variables and fixed effects. As demonstrated above, we test this assumption directly through event study specifications (Figures 3 and 4) that estimate year-specific treatment effects relative to the pre-pandemic baseline. These figures show no statistically significant differential trends in caregiving outcomes during 2017-2019, providing strong support for parallel pre-trends. The parallel trends validation is critical because it confirms that the differential changes in caregiving we observe post-2020 reflect responses to expanded WFH opportunities rather than pre-existing trajectories. This gives us confidence that our estimates capture the causal mechanisms—time savings and scheduling flexibility—specified in our conceptual framework (Section 2).

Assumption 2 (Exclusion Restriction): COVID-19 affected caregiving decisions in high-WFH-potential occupations primarily through expanding WFH opportunities, rather than through differential shocks to care needs or preferences across occupation types.

This assumption would be violated if, for example, parents of workers in high-WFH occupations faced systematically different COVID-19 health risks or access to formal care compared to parents of workers in low-WFH occupations. However, several features of our design make such violations unlikely. First, we condition on having family members needing care as of 2019 (pre-pandemic), ensuring treatment and control groups have similar baseline caregiving responsibilities. Second, prefecture-year fixed effects control for regional variation in COVID-19 severity, healthcare capacity, and long-term care restrictions. Third, our treatment is defined at the occupation level based on 2019 characteristics, predetermined relative to individual caregiving decisions during COVID-19. Any occupation-specific pandemic shocks would need to correlate with pre-pandemic WFH potential to bias our estimates. Fourth, if COVID-19 differentially affected care needs by occupation (e.g., if essential workers' parents faced higher infection risk), this would likely bias against finding positive treatment effects, since the control group would experience increased caregiving needs that are independent of WFH. Our significant positive estimates suggest this is not a primary concern.

Importantly, our identifying variation comes from differential changes in WFH adoption—not from differential exposure to COVID-19 itself. As Figure 3 demonstrates, WFH increased significantly more in high-WFH-potential occupations post-2020, validating that our treatment meaningfully captures variation in WFH expansion. The lack of pre-trends in both WFH and caregiving provides strong support that our estimates reflect causal effects of WFH availability rather than pre-existing differential trends.

Threats to internal validity

We consider several potential threats to our identification strategy. First, workers might change occupations in response to caregiving needs, selecting into high-WFH occupations to facilitate care provision (occupational sorting). However, our treatment is assigned based on December 2019 occupation (pre-pandemic), and occupation switching rates are low in Japan, particularly among our sample of workers aged 25-65 with established careers. In sensitivity analyses (Section 5.2), we exclude workers who changed occupations during our sample period and find similar results.

Second, if workers in high-WFH occupations with caregiving responsibilities are more likely to remain in the sample, this could bias our estimates (differential attrition). However, as noted in Section 3.1, attrition rates do not differ between treatment and control groups, and pre-treatment characteristics remain balanced.

Third, COVID-19 may have differentially affected employment retention across occupations (composition effects). We address this through robustness checks that exclude workers who experienced unemployment or job transitions (Section 5.2), finding substantively similar results.

Fourth, within-occupation variation in job tasks may confound our occupation-level treatment (task-based heterogeneity). In Section 5.2, we control for task-based heterogeneous trends by including interactions between individual-level task measures (from December 2019) and year fixed effects. Our results remain robust, suggesting occupation-level WFH potential captures the relevant dimension of treatment variation.

Fifth, we have checked the attrition rates by treatment status and wave. We define attrition means if an individual's final observation is before year 2022, the latest wave. For entire dataset, the total cumulative attrition rate = 57.48%. On average this was 15-17% per year, with a big jump in 2021: 26.64% of respondents in 2021 were not heard from again in 2022. There was no statistically significant difference in attrition rates between the treatment and control groups.[18] Once we restricted to caregivers and potential caregivers only, as per our analytical subset, we

[18]The attrition rates for the treatment group were 14.4%, 13.3%, and 20.2% in year 2019, 2020, and 2021, respectively, while those for the control group were 15.9%, 14.0%, and 21.3%.

found no significant difference in attrition rates between control and treatment groups.

Under these assumptions, our DID estimates identify the causal effect of WFH availability through both the time savings channel (eliminating commutes, $\Delta = 0$) and the flexibility channel (reducing effective caregiving time cost through δ) specified in our conceptual framework (Section 2). Our empirical specifications in equations (1) and (2) allow us to test the key predictions from our model: that WFH increases caregiving with heterogeneous effects by gender and employment type (Predictions 1 and 2), that formal care and WFH serve as substitutes, and that social norms moderate the caregiving response.

The next section presents our main results, heterogeneity analyses, and extensive robustness checks that further validate these identifying assumptions and test these theoretical predictions.

5 Results

5.1 Basic results

Table 2 presents DID estimates examining how WFH availability affects caregiving decisions and actual WFH adoptions. Column (1) shows results for the full sample, while Columns (2)–(5) present subgroup analyses by employment status (non-regular versus regular) and gender—the key dimensions of heterogeneity predicted by our conceptual framework (Section 2.3).

Panel A reports estimates for the WFH adoption margin, testing whether occupations with high pre-pandemic WFH potential experienced large increases in actual WFH use post-2020. The DID estimate is 0.082 for the probability of working from home 40% or more of weekly working hours, statistically significant at the 1% level, representing a 322.8% increase relative to the pre-pandemic mean among the treated group. Panel B shows similar results using the 60% threshold. These first-stage results validate our identification strategy: COVID-19 differentially expanded WFH adoption across occupations based on their pre-pandemic potential, consistent with the event study results in Figure 3. WFH adoption increased across all demographic groups, though magnitudes vary (Columns (2)–(5) in Panels A and B). Among non-regular male workers at the 40% threshold and non-regular female workers at the 60% threshold, coefficients are positive but not statistically significant, likely due to small sample sizes. For other groups, all estimates are statistically significant, and even the insignificant coefficients remain positive. While coefficient magnitudes appear modest in absolute terms, they represent substantial percentage increases relative to low pre-pandemic baselines. For example, the 0.041 coefficient for non-regular female workers at the 40% threshold represents a 204% increase from a pre-pandemic mean of 0.02. The effect for regular female workers is particularly large: a coefficient of 0.098 (Panel A) represents a 614% increase from a pre-pandemic baseline of just 0.016.

Turning to caregiving outcomes (Panels C-E in Table 2), the full sample estimates are positive but not statistically significant. This null result masks substantial heterogeneity. As our conceptual framework predicts, opportunity costs of caregiving vary systematically by employment status (Yokoyama et al., 2021) and gender (Carmichael and Charles, 2003), leading to heterogeneous treatment effects. We therefore examine subgroup results that test Predictions 1 and 2 from Section 2.3.

For non-regular workers, results strongly support Prediction 1 that WFH increases caregiving when opportunity costs are low. Among non-regular males, coefficients are 0.237 for "provide care" and 0.183 for "moderate care." For non-regular females, the coefficient for "intensive care" is 0.052. These results indicate WFH enables caregiving among workers who already have flexible schedules and face lower career penalties (Yokoyama et al., 2021), consistent with our framework's prediction that $\frac{\partial CG^*}{\partial WFH}|_{\text{non-regular}} > 0$ when opportunity costs are low. This tendency is particularly pronounced in the case of male workers.

For regular workers, results reveal striking gender differences consistent with Prediction 2. Regular male workers show no significant increases in caregiving, with most coefficients negative or near zero. In contrast, regular female workers exhibit large, statistically significant increases: 0.127 for "provide care" and 0.093 for "intensive care." This asymmetry suggests that even when time constraints are relaxed through WFH, gender norms and differential career penalties shape who responds by increasing care provision. The null effect for regular males ($\frac{\partial CG}{\partial WFH}\Big|_{(\text{regular, male})} \approx 0$) and positive effect for regular females ($\frac{\partial CG}{\partial WFH}\Big|_{(\text{regular, female})} > 0$) align precisely with our framework's predictions when career penalties π and norm costs κ differ substantially by gender.

These effect sizes are substantial relative to the existing literature. Bauer and Sousa-Poza (2015) report that caregiving typically reduces employment by 5-20 percentage points across studies, reflecting the severe trade-offs workers face when balancing work and care responsibilities. Our finding that WFH increases caregiving by approximately 13 percentage points among regular female workers—and by 24 percentage points among non-regular male workers—suggests that WFH fundamentally alters the work-care trade-off by relaxing the time constraints that previously forced binary choices between employment and caregiving. The magnitude of these effects indicates that workplace flexibility can substantially shift caregiving behavior, though as we show in subsequent sections, the welfare implications of this shift depend critically on employment type and gender.

Overall, these results demonstrate that WFH's effect on caregiving depends critically on the intersection of employment type and gender, reflecting heterogeneous opportunity costs and career constraints as formalized in Section 2.

5.2 Robustness check

Our identification strategy relies on occupation-level variation in pre-pandemic WFH potential. However, within-occupation heterogeneity in job tasks might confound our estimates (Kawaguchi and Motegi, 2021). If differences in WFH adoption underlying task variation rather than exogenous occupation-level shocks, our treatment may be endogenous. We conduct two robustness checks to address this concern and validate that our results are not driven by compositional changes or task-based heterogeneity.

First, we control for task-based heterogenous trends by adding interactions between individual-level task measures (from December 2019) and year fixed effects. This specification allows workers performing different tasks within the same occupation to follow different time trends, absorbing task-driven variation that might spuriously correlate with caregiving changes. Occupations with high versus low WFH availability may have experienced changes in job tasks before and after COVID-19, and such changes may also be reflected in the estimated treatment effect. This estimation approach addresses this concern by controlling for job tasks in 2019, thereby accounting for these potential changes.

Figure 5 presents results comparing baseline estimates to those with task-year fixed effects. Panel (a) shows estimates for non-regular workers. All coefficients that were statistically significant in Table 2 remain significant with similar magnitudes. For example, the "any care" coefficient for males is 0.237 in the baseline and 0.209 with task controls . Additionally, "moderate care" for regular females, which was not significant in Table 2, becomes statistically significant at the 10% level (0.084), suggesting WFH also promotes moderate care among this group once task heterogeneity is accounted for. Panel (b) shows regular workers follow similar patterns, with all main results robust to task controls.

Second, we address potential compositional bias from employment transitions during COVID-19. Certain industries experienced significant disruption, potentially causing workers to leave jobs, experience unemployment, or switch to WFH-compatible positions (Fukai et al. (2021), Bluedorn et al. (2023) and Kawaguchi and Motegi (2021)). Such transitions could introduce selection bias if workers who remained employed differ systematically in caregiving propensity. We re-estimate our model excluding individuals who experienced on-leave status, unemployment, or job turnover during the sample period.

Figure 6 presents results. For non-regular employees (Panel (a)), most statistically significant results from Table 2 remain significant after excluding movers. The intensive care coefficient for females loses statistical significance but maintains a similar magnitude. Only for male intensive carers does not the coefficient sign change from positive to negative, though it remains statistically insignificant. Panel (b) shows regular workers' results are similarly robust. Some coefficients, such as "any care" for females who experienced job turnover, lose statistical significance but most main findings persist.

These robustness checks confirm that our main results are not artifacts of task-based heterogeneity or compositional changes from employment transitions. The stability of estimates across specifications strengthens confidence that our DID design identifies genuine causal effects of WFH availability on caregiving decisions.

5.3 Heterogeneous impacts by social norms and formal care

Our conceptual framework (Section 2.4) predicts that WFH's effect on caregiving should vary with the availability of formal care (through substitution) and with social norms (through differential obligation or career penalties). This section tests these predictions using DID specifications (Equation (17) from Section 4), where we interact the DID term with prefecture-level measures of formal care availability and social norms.

We construct two measures of formal care availability using the Survey of Institutions and Establishments for Long-term Care. First, we calculate the number of long-term care insurance (LTCI) facility slots per population aged 65 and over, representing formal care that fully accommodates elderly individuals in residential facilities.[19] Second, we measure the number of in-home service facilities per population aged 65 and over, capturing care services that can be utilized while living at home and may either substitute for or complement family care.[20] For each measure, we create a dummy variable equal to 1 for prefectures above the 75th percentile.

For social norms, we use the National Survey on Family (Zenkoku-katei-doko-cyosa, 2018 and 2022) conducted by the Ministry of Health, Labour and Welfare. We construct two indices. The aging norm index measures agreement with three statements about filial responsibilities, such as "Older parents should live with their married children."[21] Higher values indicate more conservative attitudes. The gender norm index measures agreement with nine statements about gender roles, such as "After marriage, husbands should work outside the home and wives should focus on housework."[22] We reserve-code these so higher values indicate more progressive gender attitudes. Both indices are converted to binary variables for prefectures at or above the 75th percentile. We also control for metropolitan area status and include prefecture fixed effects.

Figure 7 plots the DID coefficients (δ_j from Equation (17)) for each subgroup. Our framework predicts ($\delta_j < 0$) for formal care variables (substitution) and ($\delta_j > 0$) for conservative aging norms but ($\delta_j < 0$) for progressive gender norms (which should reduce caregiving burdens on women).

[19] These are composed of special nursing homes for the elderly, nursing care facilities for the elderly, and medical care facilities for the elderly. These are Tokubetsu-yogo-rojin-home, Kaigo-rojin-hoken-shisetsu and Kaigo-iryo-in in Japanese respectively.

[20] This includes home care services, home bathing care, home nursing stations, day care services, day rehabilitation services, short-term residential care services, care services for residents of designated facilities, loan of welfare equipment, sale of designated welfare equipment, and home care support services.

[21] The three relevant questions are as follows. 1; "Older parents should live with their married children," 2; "Family members should be the ones to look after their elderly parents," and 3; "Financial support for the elderly should be provided by families, not by public institutions." Motegi (2025) explained the characteristics of these variables and the details of the data.

[22] The nine relevant questions are as follows. 1; "After marriage, husbands should work outside the home and wives should focus on housework." 2; Husbands and wives should prioritize their children even if it means sacrificing themselves to some extent. 3; When important decisions need to be made in the family, the father should make the final decision. 4; Until children are around three years old, mothers should not work and should focus on childcare. 5; Boys should be masculine, girls should be raised to be feminine, 6; couples are only socially recognized once they have children, 7; husbands and wives should equally share household chores and childcare, 8; husbands and wives do not need to have the same surname, 9; when work and family responsibilities conflict, husbands should prioritize work.

Non-regular Workers:

For non-regular male workers (Figure 7, Panel (a)), WFH's positive caregiving effects are substantially offset in regions with developed formal care infrastructure. The coefficients for high LTCI facilities and high in-home service facilities are -0.45 and -0.35, both statistically significant. These magnitudes are large relative to the baseline DID effect of 0.237 (Table 2), indicating that formal care availability nearly eliminates WFH's caregiving increase. Progressive gender norms also moderate effects: moderate care increases by 0.50 but intensive care decreases by -0.31 in high-gender-equality regions, suggesting that progressive norms encourage men to share care responsibilities at moderate levels while reducing expectations for intensive caregiving by any single family member.

For non-regular female workers, in-home service availability show particularly strong substitution effects. The coefficient of -0.25 (statistically significant at the 5% level) indicates that in-home care support services substantially alleviate the caregiving burden on these workers. This substitution extends to intensive care as well. Progressive gender norms also reduce intensive care by -0.14, statistically significant, the direction supports the hypothesis that gender equality norms reduce disproportionate care burdens on women. Compared to men, it becomes clear that in women's caregiving, it is not the system but the sense of norms that is important.

Regular Workers:

For male regular workers (Figure 7, Panel (b)), coefficients are generally small and mostly insignificant. Notably, in-home services availability shows a positive coefficient, indicating a complementary rather than substitution relationship for this group—suggesting that when formal services are available, regular male workers may increase their own care involvement, perhaps in coordination with purchased services. Overall, the lack of strong effects for regular male workers reinforces the interpretation from Section 5.1 that time constraints alone do not explain their limited caregiving responses; gender norms and career considerations play dominant roles.

For regular female workers, social norms show particularly striking patterns. Conservative aging norms decrease intensive care by -0.34 but increase moderate care by 0.53. While the intensive care result is surprising and merits further investigation, the moderate care result is intuitive: where traditional filial norms are strong, even regular female workers feel obligation to provide some care. One possible explanation is that when social norms are conservative, caregiving responsibilities tend to be shared among family members. Moderate care may increase, but intensive care might instead be redistributed to other family members. Although it is not statistically significant, the coefficient value of 0.19 suggests a tendency for caregiving itself to increase. Progressive gender norms reduce caregiving, with coefficients of -0.18 for "any care" and -0.10 for intensive care (the latter approaching conventional significance levels), consistent with the hypothesis that gender equality norms reduce expectations that women should be primary caregivers.

Summary:

These results strongly support our framework's predictions about institutional moderators. Formal care and informal care enabled by WFH function as substitutes ($\delta_j < 0$ for formal care), particularly for non-regular workers. Social norms play critical roles: traditional aging norms can increase caregiving obligations, while progressive gender norms reduce disproportionate burdens on women. The finding that formal care infrastructure substantially offsets WFH's caregiving effects implies that expanding public long-term care services could reduce family caregiving demands even in an era of increased workplace flexibility. The norm results suggest that cultural change toward gender equality may be as important as workplace policy in achieving sustainable work-care balance.

5.4 Effects on health outcomes

Our conceptual framework predicts ambiguous health effects of WFH-enabled caregiving: time savings and flexibility may reduce stress, but the double burden of combining intensive caregiving with demanding employment could worsen health, particularly for regular female workers (Section 2.3). We test these predictions by estimating Equation (16) restricted to caregivers, with health outcomes as dependent variables. Positive coefficients indicate health improvements.

Table 3 presents results. For non-regular male workers, coefficients for health index median (0.153) and health index Q3 (0.168) are positive and statistically significant, indicating health improvements. For non-regular female workers, health index Q1 improves significantly (0.075). These results support Prediction 1's implication that when WFH enables caregiving without intensifying work demands - as is the case for non-regular workers with already flexible schedules - caregivers experience health benefits, likely through reduced commuting stress and improved work-care coordination.

For regular workers of both genders, coefficients are generally insignificant and magnitudes are small. This null result could reflect either no effect or competing mechanisms that offset each other.

A key concern is that WFH may affect health through channels independent of caregiving—such as reduced commute stress, changed eating patterns, or altered social interactions. To isolate the caregiving-specific channel, we implement a DDD specification (Equation (18) in Section 4.1) that compares caregivers versus non-caregivers in treated versus control occupations, before versus after COVID-19. The DDD coefficient β identifies the differential health effect of WFH for those with caregiving responsibilities relative to those without, netting out any direct WFH effects on health that operate independently of caregiving.

Before turning to DDD results, we first conduct a placebo test by estimating Equation (16) on the sample of non-caregivers. Table 4 shows these results. For regular female workers, several health measures improve significantly (health index sum, Q1, and Q3), indicating that WFH directly benefits health even absent caregiving responsibilities. This finding aligns with prior evidence that WFH improves health generally (Lunde et al., 2022), and implies that the DID estimates in Table 3 can be interpreted as upper-bound estimates of caregiving-specific health effects. For other groups, most coefficients are small and insignificant, suggesting WFH's direct health effects are modest except for regular female workers.

Figure 8 presents DDD estimates. For non-regular workers, results confirm that WFH improves health beyond any general WFH health effects. For males, health index median and Q3 are 0.11 and 0.089, indicating that providing care while working from home improves health relative to non-caregivers in the same occupations. For females, health index Q1 is 0.067, also indicating health improvements. These results support the hypothesis that when WFH enables caregiving in already-flexible jobs, the net health effect is positive, consistent with reduced time stress and improved work-family coordination.

For regular female workers, however, the DDD coefficient is negative and statistically significant for the health index sum (-0.135) and approaches significance for other measures. To interpret this magnitude, the coefficient of -0.135 on the health index sum (scaled 0-32) represents approximately a 0.8% deterioration from the sample mean of approximately 17-18 points. While modest in absolute terms, this effect is notable because it represents a net health cost after accounting for the general health benefits of WFH observed among non-caregivers (Table 4). That is, regular female workers who do not provide care experience health improvements from WFH, but those who take on caregiving responsibilities while working from home experience health deterioration - a pattern consistent with the double burden hypothesis. This finding supports Prediction 2's implication that WFH may create a double burden for women in demanding careers: they increase caregiving substantially (Section 5.1) while maintaining intensive work commitments, and this combination worsens health. The negative DDD coefficient indicates that the health deterioration for regular female caregivers using WFH exceeds any general health benefits of WFH, consistent with career penalties (π_{female}) and time stress from juggling two intensive commitments.

For regular male workers, DDD coefficients are small and insignificant, consistent with the finding that this group did not significantly increase caregiving in response to WFH (Section 5.1).

We also examined multiple indicators such as the median, Q1, and Q3, and the trends were almost identical. These analyses also serve as robustness checks for the health index.

These patterns provide strong support for our framework's predictions about health effects. WFH enables healthier work-care balance for those in flexible jobs (non-regular workers), but intensifies strain for those in demanding careers who absorb caregiving responsibilities while maintaining high work effort (regular female workers). The contrasting health effects underscore that workplace flexibility alone is insufficient to achieve sustainable work-care balance; the nature of the job and the distribution of caregiving responsibilities within families critically shape whether WFH improves or worsens caregiver welfare.

5.5 Effects on labor market outcomes

A central question is whether WFH prevents the employment exits that prior research identifies as a primary cost of caregiving (Houtven et al. (2013), Bauer and Sousa-Poza (2015) and Løken et al. (2017)). If WFH enables workers to balance care and employment, we might observe stable employment outcomes despite increased caregiving. Alternatively, increased caregiving might still reduce hours or career advancement, particularly if career penalties π manifest through reduced promotion prospects rather than job loss. We test these possibilities by estimating Equation (16) with labor market outcomes as dependent variables, restricting the sample to individuals providing care in 2019 (pre-pandemic). Table 5 presents results. For non-regular male workers, the probability of transitioning to regular employment decreased by 7.4 percentage points, and while not statistically significant, working hours also decline. Combined with increased caregiving documented in Section 5.1, this suggests WFH may enable care provision but at the cost of career advancement opportunities. The transition to regular employment represents an important career step in Japan's dualistic labor market (Yokoyama et al., 2021), so reduced transitions indicate WFH may signal reduced career commitment.

For non-regular female workers, results are more nuanced. Employment probability decreases by 2.7 percentage points (not significant), but working hours increase by 1.9 percentage points and the probability of becoming a regular employee increases by 1.5 percentage points (both not significant). The pattern suggests WFH may have neutral or even slightly positive labor market effects for non-regular women, consistent with enabling rather than constraining employment. For regular male workers, employment probability increases by 3.0 percentage points, suggesting WFH helps retain workers who might otherwise exit due to caregiving demands. Hours worked show a negative but insignificant coefficient, indicating modest if any reduction in work intensity.

For regular female workers, coefficients are mostly insignificant but directionally suggest stable employment with potentially reduced hours. The employment retention is consistent with WFH preventing exits, but the lack of positive effects on hours or other margins suggests WFH facilitates continued employment without necessarily improving career prospects.

Overall, WFH appears to prevent employment exits—contradicting some prior findings that caregiving reduces labor supply. However, effects vary by worker type. For non-regular workers, increased caregiving coincides with reduced career advancement opportunities. For regular workers, WFH helps maintain employment, though whether it prevents longer-run career penalties remains unclear from these data. The absence of wage effects (Panel D) across all groups suggests WFH neither penalizes nor rewards workers in the short run, though career costs may manifest over longer horizons not observed in our data.

These results align with our framework's prediction that labor supply effects are ambiguous $\frac{\partial H}{\partial WFH} \gtrless 0:$ time savings from WFH may enable continued employment, but career penalties or reduced advancement may still emerge, particularly for groups that signal reduced career commitment by using flexibility for caregiving.

5.6 Effects on well-being

Beyond health and labor market outcomes, we examine whether WFH affects overall well-being among caregivers. We analyze three dimensions: happiness ("not reported unhappy"), life satisfaction ("not dissatisfied with life"), and work-life balance stress ("not suffered from work-life balance stress"). These capture subjective welfare beyond objective health and employment measures.

Table 6 presents DID estimates for caregivers. Across all four demographic groups, no coefficients are statistically significant. Magnitudes are generally —less than 10% of the mean for happiness and life satisfaction. Work-life balance stress shows slightly larger effects for women (non-regular; 0.043, regular; 0.033), suggesting potential improvements, but neither is statistically significant.

The null results could reflect either truly zero effects or imprecise estimation due to small samples. To assess whether WFH has general well-being effects independent of caregiving, we conduct a placebo test on non-caregivers (Table 7). Only life satisfaction for regular male workers shows a significant improvement (0.015), but the magnitude is modest. Other coefficients remain small and insignificant.

Given the null results in both caregiver and non-caregiver samples, we do not pursue DDD estimation for well-being outcomes. The evidence suggests WFH has limited effects on subjective well-being in the short run, despite its significant impacts on caregiving behavior and health. This pattern may reflect that increased caregiving enabled by WFH generates both benefits (fulfilling family obligations, reduced guilt from not providing care) and costs (time stress, foregone leisure) that offset each other in terms of overall life satisfaction. Alternatively, well-being measures may be too coarse to detect effects, or effects may manifest over longer time horizons than our data capture.

These null results contrast with the clear health effects documented in Section 5.4, suggesting that health and subjective well-being respond differently to WFH-induced changes in work-care balance. Health deterioration among regular female caregivers (Section 5.4) does not translate to reduced happiness or life satisfaction in our data, possibly because these women value fulfilling caregiving obligations even at personal health cost, or because well-being impacts take longer to manifest.

6 Conclusion

6.1 Summary of findings

This paper provides the first causal evidence on whether WFH arrangements enable workers to balance eldercare and employment, and at what cost. Using Japan's COVID-19-induced expansion of WFH as a quasi-experiment, we find striking heterogeneity in how workplace flexibility affects the work-care balance.

For non-regular workers, WFH increases caregiving for both men and women while improving health—demonstrating that workplace flexibility can successfully enable work-care balance when career structures permit. For regular workers in demanding careers, however, WFH promotes caregiving substantially but worsens health, suggesting it intensifies rather than resolve the double burden of balancing intensive work and family commitments. Regular male workers show no caregiving response despite similar time savings, indicating that gender norms and career incentives—not just time constraints—shape caregiving decisions. Critically, formal care availability and progressive gender norms substantially moderate these effects, with formal and informal care functioning as substitutes and progressive norms reducing disproportionate burdens on women.

6.2 Policy implications

Our finding carry important implications for workplace policies and public programs in aging societies. The core lesson is that workplace flexibility alone is insufficient—and may even be counterproductive—without complementary policies.

First, WFH benefits depend critically on underlying job characteristics. For non-regular workers, WFH facilitates caregiving without health costs, but for regular female workers, it intensifies work-care conflicts. Corporate human resource departments and workplace health regulators should monitor caregivers using WFH for signs of stress and burnout, rather than assuming flexibility automatically improves welfare. Policies that reduce over all work demands—not just increase schedule flexibility—may be necessary for sustainable caregiving.

Second, expanding formal care infrastructure is essential. The strong substitution effects we document suggest that investments in long-term care facilities and in-home services can substantially reduce family caregiving demands even when WFH makes informal care more feasible. Countries facing aging populations should consider formal care expansion and workplace flexibility as complements rather than substitutes.

Third, the differential caregiving responses by gender despite similar time savings indicate that cultural norms critically shape outcomes. Progressive gender norms significantly reduce caregiving burdens on women, suggesting that cultural change toward gender equality may be as important as policy reform. Public awareness campaigns, workplace training on unconscious bias, and policies that incentivize men's caregiving participation could help distribute care responsibilities more equitably.

Fourth, employers and policymakers should be attentive to potential career penalties from flexibility utilization. The finding that male non-regular workers reduce transitions to regular employment after increasing caregiving suggests WFH use may signal reduced career commitment. Policies that normalize caregiving across genders and protect workers from discrimination based on flexibility use are necessary to prevent WFH from reinforcing labor market inequalities.

6.3 Broader implications

While our analysis focuses on Japan, the work-care tensions we examine are universal features of aging societies. Japan's experience as an early-aging society with historically rigid work arrangements offers valuable lessons for other developed countries confronting similar demographic challenges. If WFH alone proves insufficient to resolve work-care conflicts in Japan, other countries should anticipate similar limitations.

A general insight emerges: workplace reforms enabling work-family balance are most effective when embedded in broader ecosystems of formal support services and egalitarian social norms. The health costs we identify for women in demanding careers who absorb caregiving while maintaining intensive work suggest that apparent labor supply benefits of WFH may be unsustainable without complementary reforms addressing both care infrastructure and gender equality.

6.4 Limitations and Future Research

Our analysis has limitations that suggest directions for future research. Our sample restriction to residentially stable workers—those who remained in the same prefecture from age 15 to 2019-while necessary for identifying caregiving responses among individuals with plausible proximity to aging parents, may limit generalizability to geographically mobile populations who face different caregiving constraints and opportunities. Mobile workers may have weaker family ties, higher education levels, or greater career orientation, and their caregiving responses to WFH availability could differ systematically from those we estimate. Future research with data on parent location and caregiving distance could relax this restriction while still identifying relevant caregiving margins.

First, our data lack detailed information on family structure and care relationships—for example, whether the care recipient is the worker's own parent or a spouse's parent. Prior research suggests this distinction is important for assessing caregiving's health impacts (Oshio, 2014). We also do not observe formal care utilization or care recipients' functional status. Linked employer-employee-family data combined with long-term care insurance claims would enable richer analysis of how family dynamics and formal-informal care substitution shape caregiving responses to WFH.

Second, we observe outcomes only through 2022, limiting evidence on long-run consequences. Career penalties and health effects may cumulate or dissipate over time. Our ITT estimates identify effects of WFH availability rather than utilization intensity, so variation in WFH dosage would enable dose-response analysis. Finally, while our robustness checks address many identification threats and event studies show parallel pre-trends, alternative identification strategies—such as firm-level WFH policy changes—would further strengthen causal inference.

These limitations notwithstanding, our analysis provides the first causal evidence that workplace flexibility enables work-care balance for some workers while intensifying gender inequality for others, with effects critically dependent on job characteristics, institutional supports, and cultural contexts. As remote work remains widespread and populations age globally, understanding when and for whom workplace flexibility supports family caregivers is essential. Our findings suggest optimism about WFH's potential must be tempered with recognition that flexibility without complementary reforms may perpetuate rather than resolve the work-care tensions facing aging societies.

Acknowledgements

The authors would like to thank Junya Kawamura, Kazuharu Yanagimoto and the participants at the European Health Economics Association Conference 2024 and research seminars at Waseda University for their invaluable comments and suggestions. The data for this secondary analysis, Japanese Panel Study of Employment Dynamics and 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. This research is partially supported by the Grants-in-Aid for Scientific Research [Research project number: 23K18822] for the Japan Society for the Promotion of Science (JSPS) Fellows by the Ministry of Education, Culture, Sports, Science and Technology in Japan. This study is also one of the research outcomes of the project "Promotion of EBPM Projects as a National Research Institute" and "The National Survey on Family" conducted by the National Institute of Population and Social Security Research.

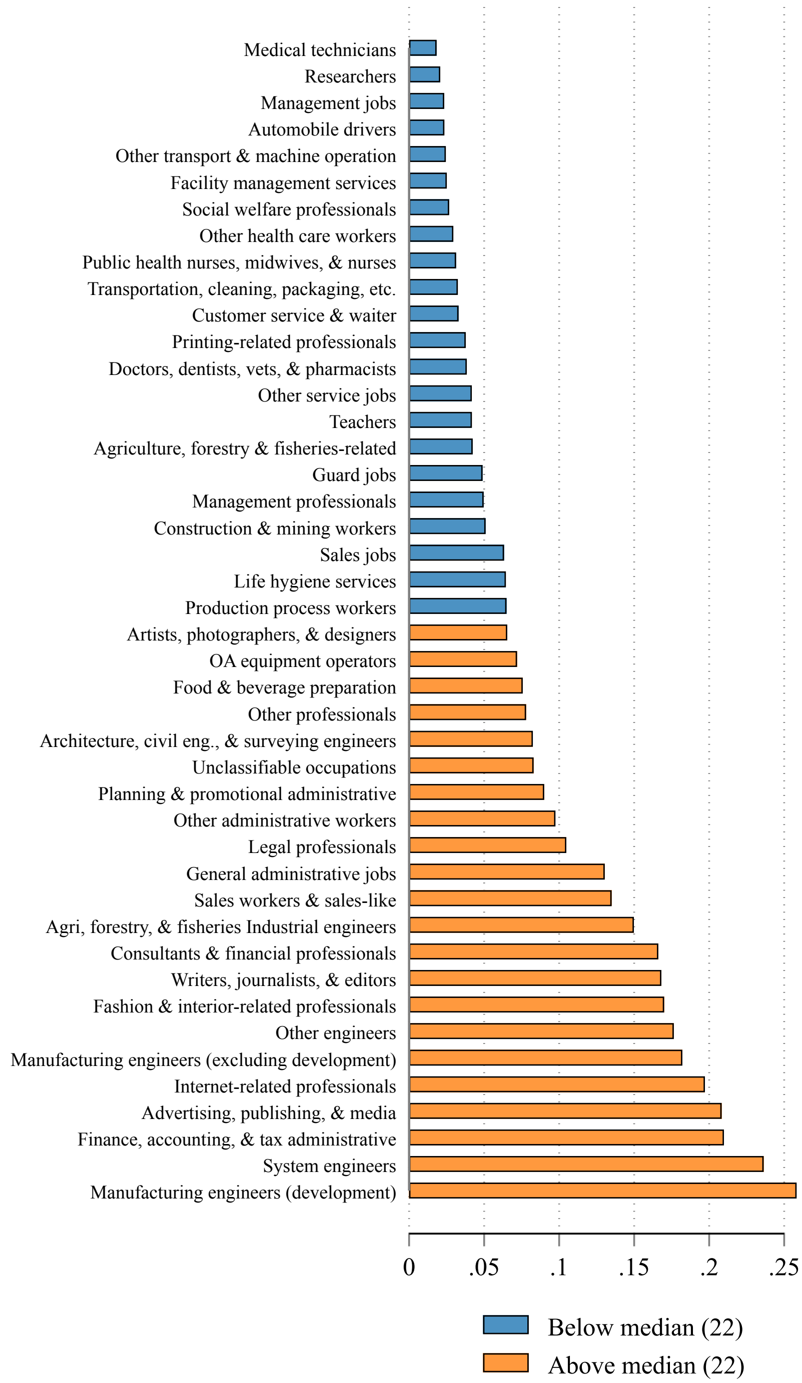


Figure 1: Proportion of workers at companies allowing them to work from home before the pandemic by occupation

Note: The sampling weight in 2016 is used for the calculations of the proportions. The occupations are divided by the median of the proportions over occupation: "below median"(blue) and "above median"(orange).

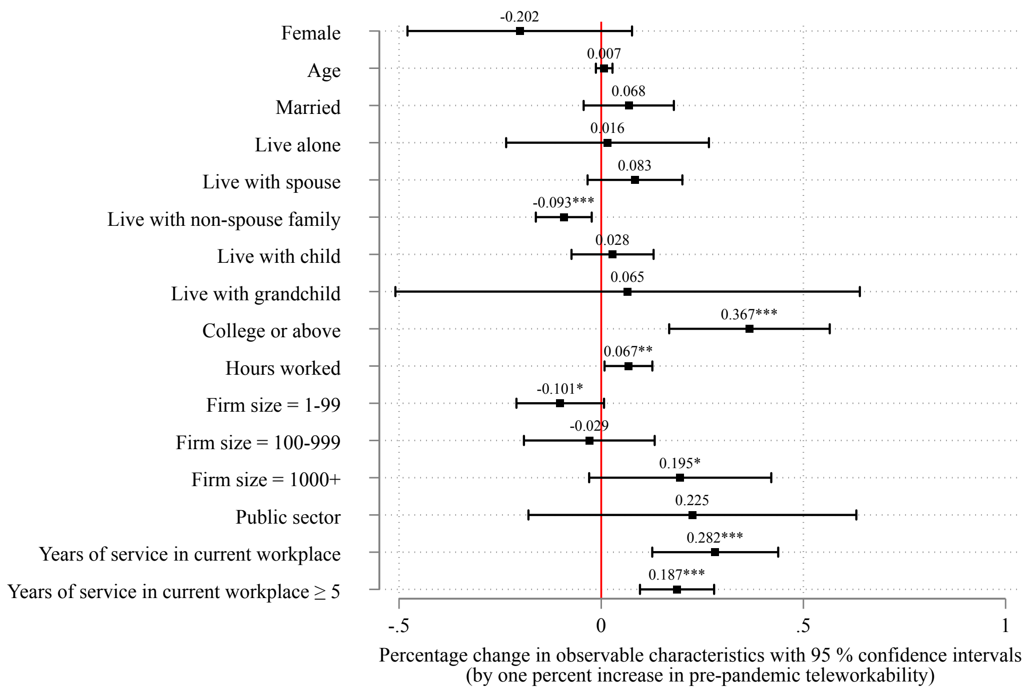
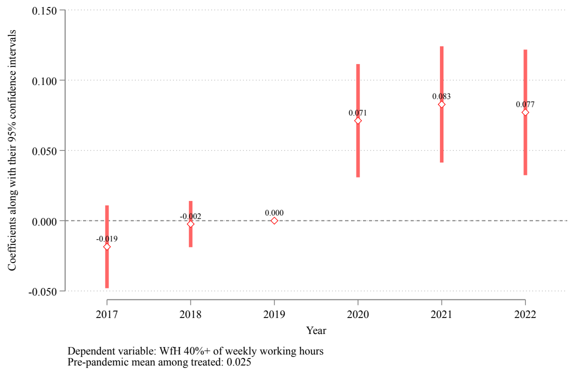
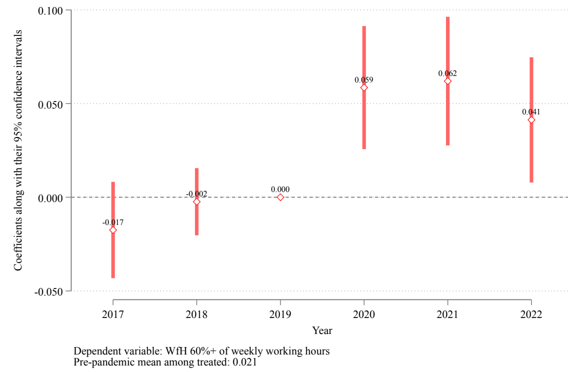


Figure 2: Observable Characteristics and Pre-Pandemic Occupation-Based WFH availability

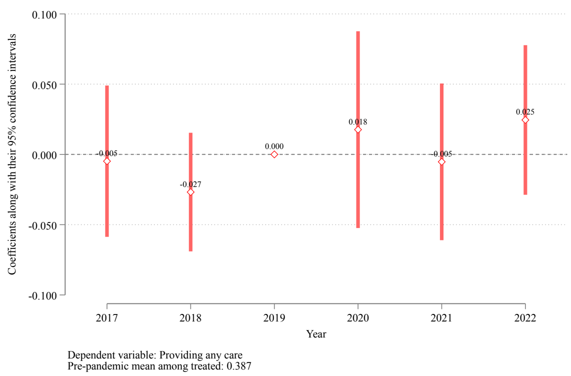
Note: The figure summarizes the relationship between individual observable characteristics and pre-pandemic occupation-based WFH availability using 2019 data. The square symbols indicate the percentage change in observable characteristics when pre-pandemic WFH availability increased by one percent, with the bars showing their 95% confidence intervals. The confidence intervals are calculated using standard errors robust against clustering at the December 2019 occupation level. The square symbols are estimated as follows: (1) regressing the observable characteristics on log-transformed pre-pandemic occupation-based WFH availability and prefecture fixed effects ($y_{ijp} = \alpha + \beta \log(teleworkability_j) + \rho_p + \epsilon_{ijp}$, where y_{ijp}, $teleworkability_j$, ρ_p, and ϵ_{ijp} are the observable characteristic, pre-pandemic occupation-based WFH availability, prefecture fixed effects, and an error term, respectively.); and (2) calculating $100 \times \frac{\hat{\beta}/100}{\bar{\mu}_y}$, where $\bar{\mu}_y$ is the average value of an observable characteristic, y_{ijp}. The sampling weight in 2017 is used for the calculations of the proportions. Inference: * $p < .1$, ** $p < .05$, *** $p < .01$.



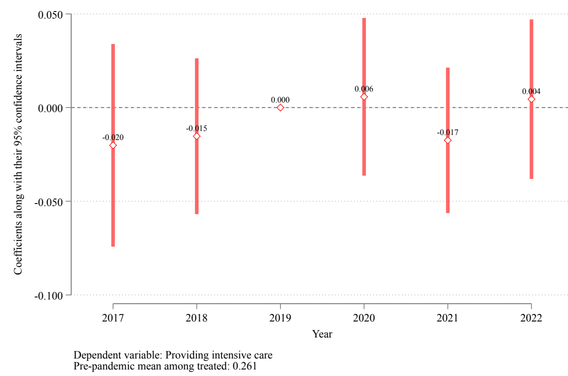
(a) Working from home for 40% or more of weekly working hours



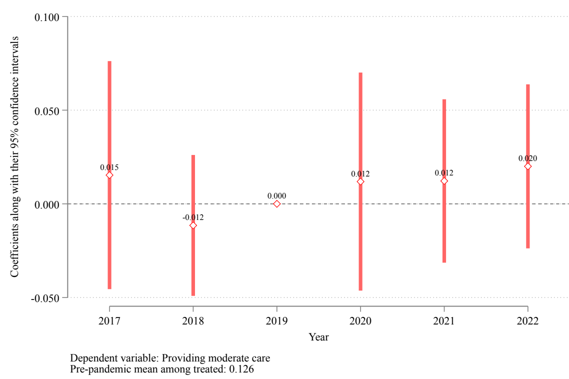
(b) Working from home for 60% or more of weekly working hours



(c) Providing any informal care



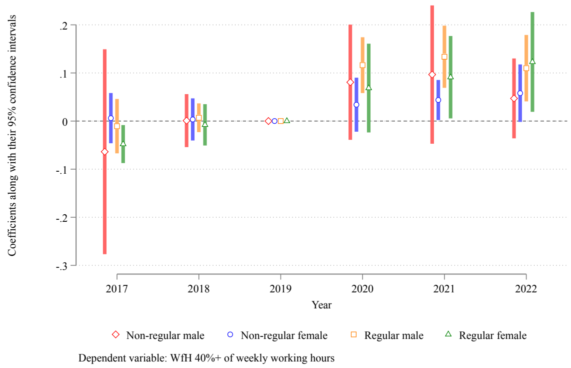
(d) Providing intensive care



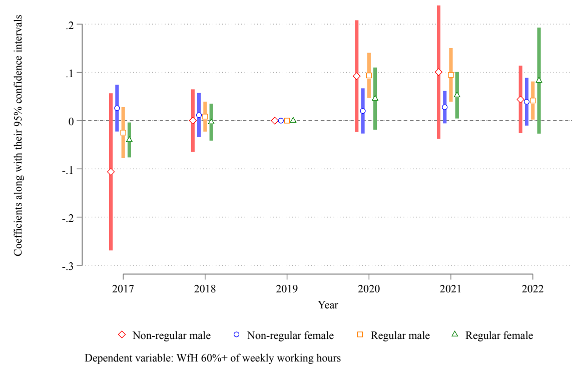
(e) Providing moderate care

Figure 3: Results of event study model

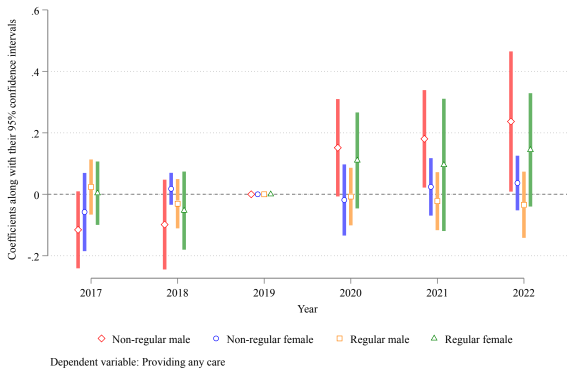
Note: The unit of observation is the individual and year. The dependent variables are dummy variables for frequency of working from home and informal care provision. The diamond symbols indicate the estimated coefficients on the cross term of the treatment dummy and the year dummy variables, and the bars represent the 95% confidence intervals for the estimates. The reference time period is 2019. Confidence intervals are calculated using standard errors robust against occupation as of December 2019 level clustering. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects.



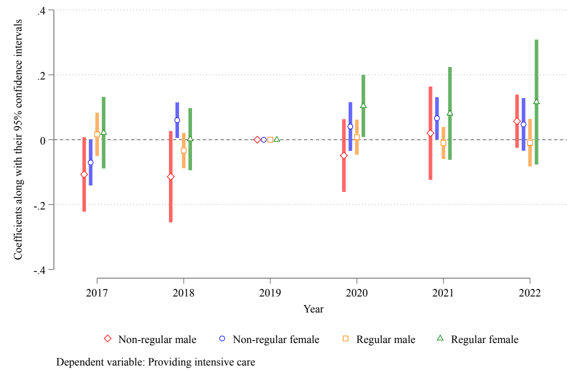
(a) Working from home for 40% or more of weekly working hours



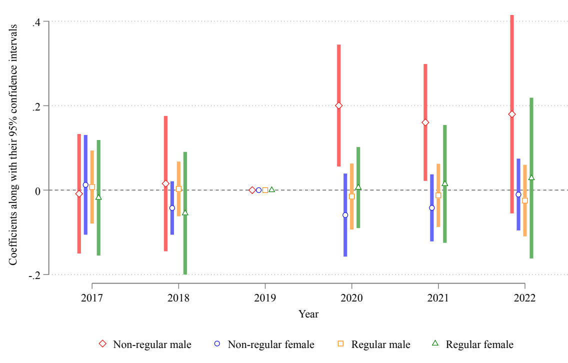
(b) Working from home for 60% or more of weekly working hours



(c) Providing any informal care



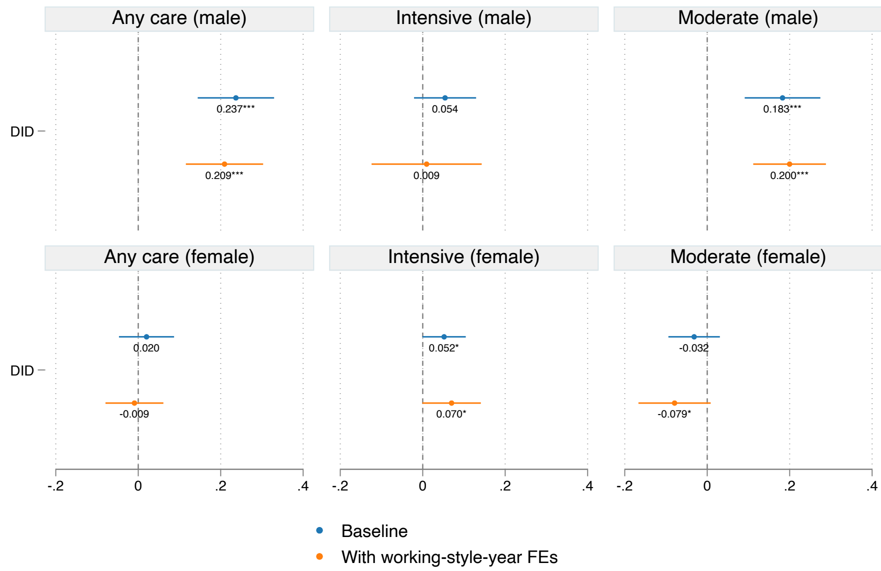
(d) Providing intensive care



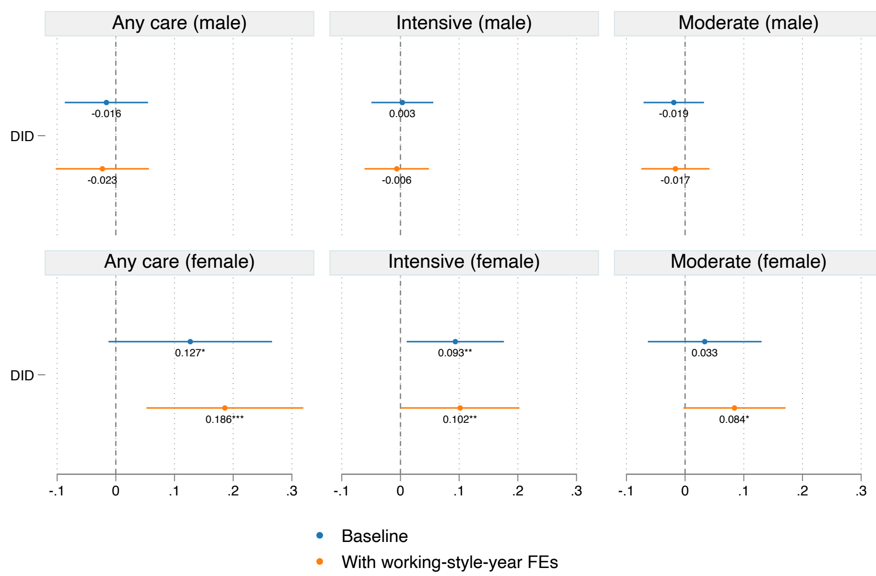
(e) Providing moderate care

Figure 4: Results of event study model by gender and employment status as of 2019

Note: The unit of observation is the individual and year. The dependent variables are dummy variables for frequency of working from home and informal care provision. The diamond symbols indicate the estimated coefficients on the cross term of the treatment dummy and the year dummy variables, and the bars represent the 95% confidence intervals for the estimates. The reference time period is 2019. Confidence intervals are calculated using standard errors robust against occupation as of December 2019 level clustering. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects.



(a) Non-regular workers

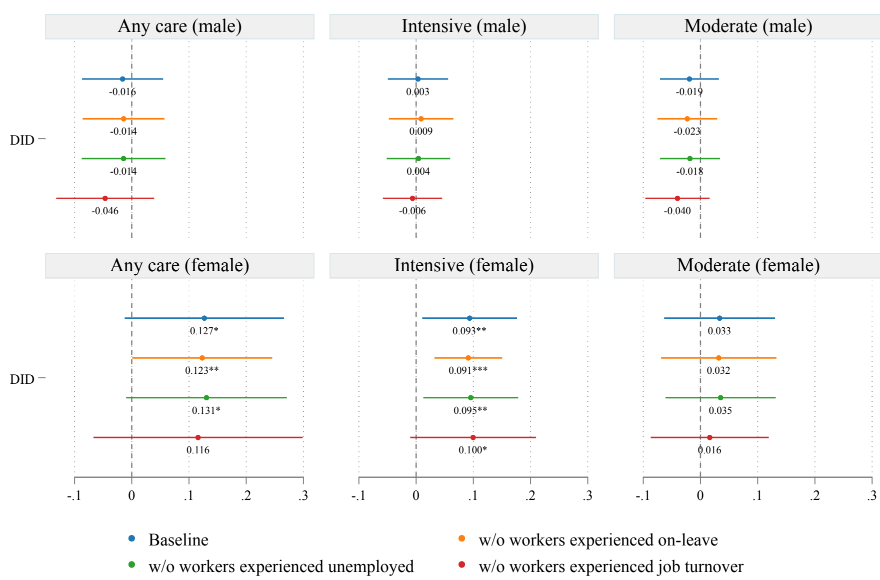


(b) Regular workers

Figure 5: Robustness check against working style heterogeneity

Note: The circle symbols indicate the estimated coefficients on the DID terms, and the bars represent the 95% confidence intervals for the estimates. Confidence intervals are calculated using standard errors robust against occupation as of December 2019 level clustering. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects.

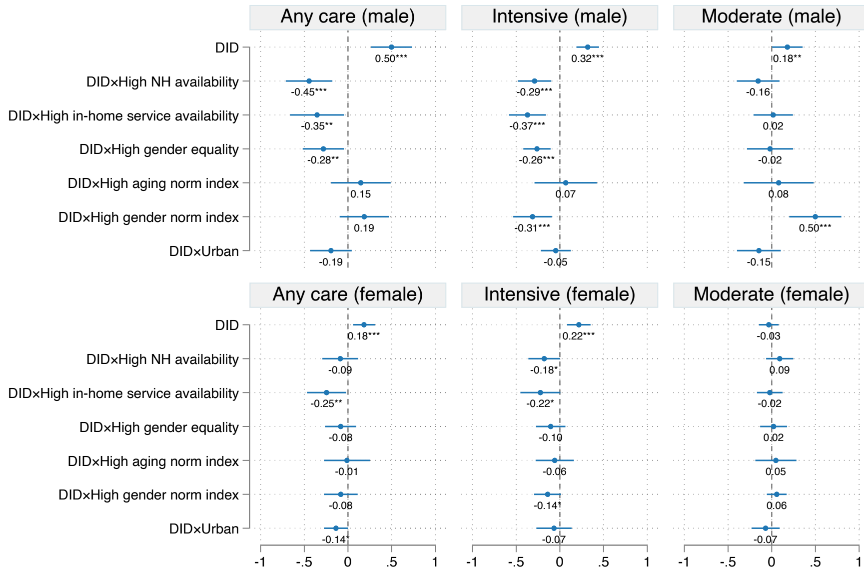
(a) Non-regular workers



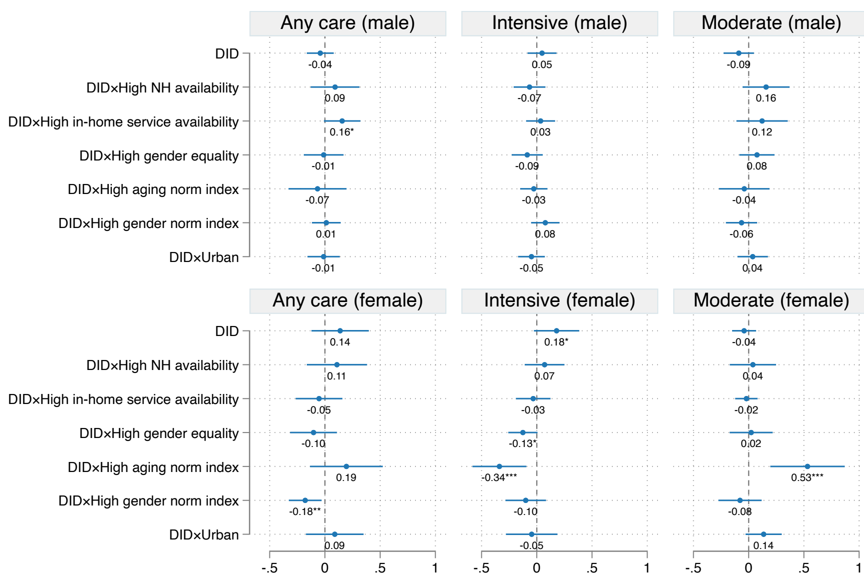
(b) Regular workers

Figure 6: Robustness check against sample restriction

Note: The circle symbols indicate the estimated coefficients on the DID terms, and the bars represent the 95% confidence intervals for the estimates. Confidence intervals are calculated using standard errors robust against occupation as of December 2019 level clustering. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects.



(a) Non-regular workers



(b) Regular workers

Figure 7: Heterogeneous impacts by gender, employment status, and regional characteristics

Note: The circle symbols indicate the estimated coefficients on the DID terms, and the bars represent the 95% confidence intervals for the estimates. Confidence intervals are calculated using standard errors robust against occupation as of December 2019 level clustering. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects.

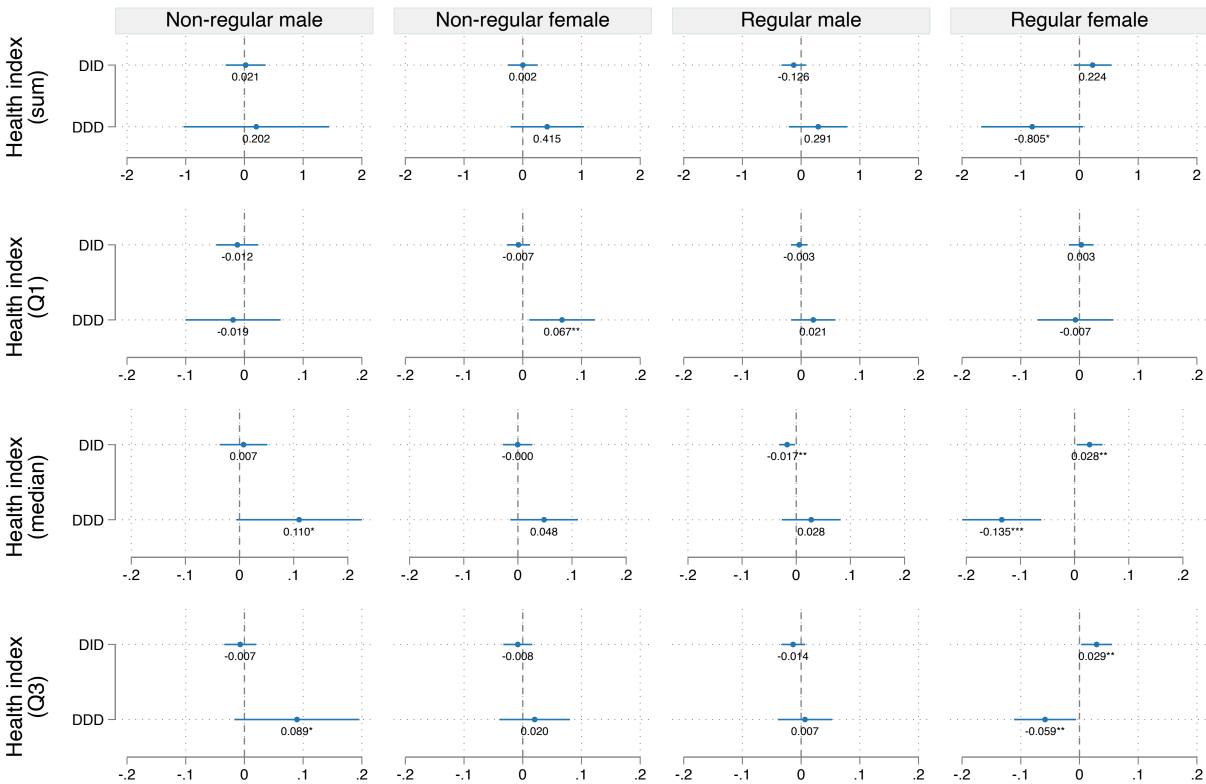


Figure 8: DDD estimation results on health outcomes

Note: The circle symbols indicate the estimated coefficients on the DID and DDD terms, and the bars represent the 95% confidence intervals for the estimates. The confidence intervals are calculated using standard errors robust against occupation as of December 2019 level clustering. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects. Inference: * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 1: Balancing between treated and control groups in December 2019

	Mean		Differences			
	Control	Treated	raw	(%)	adjusted	(%)
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.47	0.48	0.02	3.48		
			(0.02)	(5.00)		
Age	48.86	50.33	1.47***	2.96***		
			(0.49)	(0.99)		
Married	0.56	0.58	0.02	3.63		
			(0.02)	(4.13)		
Live alone	0.13	0.13	0.01	5.99		
			(0.02)	(12.29)		
Live with spouse	0.53	0.56	0.03	5.52		
			(0.02)	(4.35)		
Live with non-spouse family	0.45	0.40	-0.05**	-11.88**		
			(0.02)	(5.52)		
Live with child	0.39	0.39	0.00	0.03		
			(0.02)	(5.90)		
Live with grandchild	0.01	0.01	-0.00	-0.58		
			(0.01)	(40.11)		
College or above	0.24	0.32	0.09***	31.42***		
			(0.02)	(7.63)		
Hours worked	36.75	38.66	1.92***	5.08***		
			(0.64)	(1.70)		
Firm size						
1-99	0.48	0.44	-0.04*	-9.22*	-0.02	-4.82
			(0.02)	(5.12)	(0.02)	(5.11)
100-999	0.29	0.26	-0.03	-10.38	-0.03	-11.61
			(0.02)	(7.69)	(0.02)	(7.77)
1000+	0.18	0.20	0.01	7.25	0.01	3.47
			(0.02)	(9.88)	(0.02)	(9.95)
Public sector	0.05	0.10	0.06***	77.90***	0.05***	64.61***
			(0.01)	(17.17)	(0.01)	(17.11)
Years of service in current workplace	10.39	14.82	4.44***	35.20***	4.53***	35.98***
			(0.54)	(4.27)	(0.54)	(4.31)
Years of service in current workplace ≥ 5	0.58	0.72	0.14***	21.71***	0.15***	23.29***
			(0.02)	(3.45)	(0.02)	(3.48)

Note: the sampling weight in 2017 is used for the calculations of the proportions. Columns (1) and (2) show the mean of the variables for the control and treated groups respectively, and Columns (3) and (4) show the difference and percentage difference between the two columns respectively. Columns (5) and (6) represent the difference and percentage difference in the variables between the treated and control groups after we controlled for the level of education.

Table 2: DID Estimation Results on Working from Home (WFH) and Care Provision

		Non-regular		Regular	
	(1)	(2)	(3)	(4)	(5)
	Full sample	Male	Female	Male	Female
A. WFH 40%+ of weekly working hours					
DID estimates	0.082***	0.095	0.041**	0.122***	0.098**
	(0.019)	(0.063)	(0.017)	(0.025)	(0.040)
Number of observations	6828	580	1778	2643	1324
Pre-pandemic mean among the treated	0.025	0.037	0.020	0.024	0.016
Magnitude of the estimates (%)	322.8	258.9	204.6	506.3	613.6
B. WFH 60%+ of weekly working hours					
DID estimates	0.061***	0.112*	0.017	0.088***	0.064**
	(0.015)	(0.059)	(0.011)	(0.019)	(0.029)
Number of observations	6828	580	1778	2643	1324
Pre-pandemic mean among the treated	0.021	0.026	0.018	0.019	0.013
Magnitude of the estimates (%)	295.5	427.1	94.5	454.8	494.4
C. Providing any care					
DID estimates	0.021	0.237***	0.020	-0.016	0.127*
	(0.021)	(0.045)	(0.033)	(0.035)	(0.068)
Number of observations	7192	647	1935	2734	1373
Pre-pandemic mean among the treated	0.387	0.291	0.442	0.337	0.431
Magnitude of the estimates (%)	5.5	81.5	4.5	-4.8	29.4
D. Providing intensive care					
DID estimates	0.008	0.054	0.052*	0.003	0.093**
	(0.016)	(0.037)	(0.026)	(0.026)	(0.041)
Number of observations	7192	647	1935	2734	1373
Pre-pandemic mean among the treated	0.261	0.206	0.288	0.209	0.319
Magnitude of the estimates (%)	3.2	26.3	18.0	1.6	29.3
E. Providing moderate care					
DID estimates	0.013	0.183***	-0.032	-0.019	0.033
	(0.018)	(0.045)	(0.031)	(0.025)	(0.048)
Number of observations	7192	647	1935	2734	1373
Pre-pandemic mean among the treated	0.126	0.084	0.154	0.128	0.112
Magnitude of the estimates (%)	10.5	216.6	-20.5	-15.1	29.8

Note: The unit of observation is the individual and year. The dependent variables are dummy variables for frequency of working from home and informal care provision. Standard errors robust against occupation as of December 2019 with level clustering are shown in parentheses. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects. Inference: * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 3: DID Estimation Results on Health Outcomes

	Non-regular		Regular	
	(1) Male	(2) Female	(3) Male	(4) Female
A. Health index (sum)				
DID estimates	0.920	0.377	-0.260	0.143
	(0.688)	(0.339)	(0.417)	(0.410)
Number of observations	647	1935	2734	1373
Pre-pandemic mean among the treated	18.973	17.024	18.608	17.409
Magnitude of the estimates (%)	4.8	2.2	-1.4	0.8
B. Health index (Q1)				
DID estimates	0.025	0.075**	-0.006	0.019
	(0.062)	(0.035)	(0.033)	(0.070)
Number of observations	647	1935	2734	1373
Pre-pandemic mean among the treated	0.711	0.623	0.710	0.644
Magnitude of the estimates (%)	3.6	12.0	-0.8	2.9
C. Health index (median)				
DID estimates	0.153**	0.034	0.013	-0.070
	(0.058)	(0.031)	(0.037)	(0.047)
Number of observations	647	1935	2734	1373
Pre-pandemic mean among the treated	0.454	0.304	0.430	0.378
Magnitude of the estimates (%)	33.7	11.2	2.9	-18.5
D. Health index (Q3)				
DID estimates	0.168***	-0.004	-0.022	-0.002
	(0.058)	(0.024)	(0.031)	(0.034)
Number of observations	647	1935	2734	1373
Pre-pandemic mean among the treated	0.258	0.116	0.212	0.129
Magnitude of the estimates (%)	65.3	-3.8	-10.2	-1.3

Note: The unit of observation is the individual and year. The dependent variables are health outcomes. Standard errors robust against occupation as of December 2019 with level clustering are shown in parentheses. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects. Inference: * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 4: DID Estimation Results on Health Outcomes (Placebo)

	Non-regular		Regular	
	(1)	(2)	(3)	(4)
	Male	Female	Male	Female
A. Health index (sum)				
DID estimates	0.025	0.064	-0.106	0.258*
	(0.201)	(0.130)	(0.100)	(0.150)
Number of observations	5439	13053	25406	11899
Pre-pandemic mean among the treated	19.993	18.094	20.192	18.424
Magnitude of the estimates (%)	0.1	0.4	-0.5	1.4
B. Health index (Q1)				
DID estimates	-0.016	-0.006	-0.002	0.006
	(0.017)	(0.010)	(0.007)	(0.010)
Number of observations	5439	13053	25406	11899
Pre-pandemic mean among the treated	0.783	0.686	0.793	0.685
Magnitude of the estimates (%)	-2.1	-0.8	-0.2	0.8
C. Health index (median)				
DID estimates	0.013	0.002	-0.014*	0.030***
	(0.025)	(0.014)	(0.007)	(0.011)
Number of observations	5439	13053	25406	11899
Pre-pandemic mean among the treated	0.544	0.397	0.545	0.425
Magnitude of the estimates (%)	2.4	0.5	-2.6	7.2
D. Health index (Q3)				
DID estimates	-0.003	-0.007	-0.014	0.030**
	(0.014)	(0.013)	(0.010)	(0.013)
Number of observations	5439	13053	25406	11899
Pre-pandemic mean among the treated	0.292	0.163	0.292	0.193
Magnitude of the estimates (%)	-0.9	-4.0	-4.8	15.4

Note: The unit of observation is the individual and year. The dependent variables are health related variables. Standard errors robust against occupation as of December 2019 with level clustering are shown in parentheses. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects. Inference: * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 5: DID Estimation Results on Labor Market Outcomes

	Non-regular		Regular	
	(1) Male	(2) Female	(3) Male	(4) Female
A. Labor participation dummy				
DID estimates	-0.072	-0.027	0.030*	0.038
	(0.060)	(0.048)	(0.016)	(0.028)
Number of observations	647	1935	2734	1373
Pre-pandemic mean among the treated	0.916	0.950	0.988	0.986
Magnitude of the estimates (%)	-7.9	-2.8	3.0	3.8
B. Annual working hours (log)				
DID estimates	-0.034	0.019	-0.033	-0.078
	(0.063)	(0.028)	(0.038)	(0.058)
Number of observations	582	1779	2649	1328
Pre-pandemic mean among the treated	7.355	7.141	7.687	7.586
Magnitude of the estimates (%)	-0.5	0.3	-0.4	-1.0
C. Annual income (log)				
DID estimates	-0.053	0.078	0.019	0.063
	(0.147)	(0.063)	(0.031)	(0.099)
Number of observations	612	1824	2643	1332
Pre-pandemic mean among the treated	5.333	4.836	6.226	5.698
Magnitude of the estimates (%)	-1.0	1.6	0.3	1.1
D. Hourly wage (log)				
DID estimates	0.014	0.001	0.041	0.098
	(0.105)	(0.045)	(0.045)	(0.093)
Number of observations	482	1531	2478	1233
Pre-pandemic mean among the treated	7.279	6.978	7.759	7.330
Magnitude of the estimates (%)	0.2	0.0	0.5	1.3
E. Regular employment dummy				
DID estimates	-0.074*	0.015	-0.003	0.056
	(0.042)	(0.018)	(0.026)	(0.049)
Number of observations	580	1781	2636	1318
Pre-pandemic mean among the treated	0.111	0.025	0.994	0.961
Magnitude of the estimates (%)	-66.7	60.0	-0.3	5.8

Note: The unit of observation is the individual and year. The dependent variables are labor market outcomes. Standard errors robust against occupation as of December 2019 with level clustering are shown in parentheses. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects. Inference: * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 6: DID Estimation Results on Well-being

	Non-regular		Regular	
	(1) Male	(2) Female	(3) Male	(4) Female
A. Not reported unhappy				
DID estimates	-0.056	0.042	-0.027	0.030
	(0.050)	(0.032)	(0.043)	(0.043)
Number of observations	647	1935	2734	1373
Pre-pandemic mean among the treated	0.686	0.799	0.758	0.801
Magnitude of the estimates (%)	-8.2	5.2	-3.5	3.7
B. Not dissatisfied with life				
DID estimates	0.009	0.029	0.003	-0.001
	(0.059)	(0.047)	(0.048)	(0.060)
Number of observations	647	1935	2734	1373
Pre-pandemic mean among the treated	0.659	0.735	0.735	0.760
Magnitude of the estimates (%)	1.4	4.0	0.4	-0.1
C. Not suffered from work-life balance stress				
DID estimates	-0.061	0.043	-0.007	0.033
	(0.065)	(0.044)	(0.034)	(0.038)
Number of observations	616	1854	2687	1348
Pre-pandemic mean among the treated	0.374	0.190	0.227	0.193
Magnitude of the estimates (%)	-16.3	22.8	-2.9	17.1

Note: The unit of observation is the individual and year. The dependent variables are well-being outcomes. Standard errors robust against occupation as of December 2019 with level clustering are shown in parentheses. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects. Inference: * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 7: DID Estimation Results on Well-being (placebo)

	Non-regular		Regular	
	(1) Male	(2) Female	(3) Male	(4) Female
A. Not reported unhappy				
DID estimates	-0.019	0.015	0.009	0.008
	(0.019)	(0.011)	(0.008)	(0.008)
Number of observations	5439	13053	25406	11899
Pre-pandemic mean among the treated	0.700	0.832	0.814	0.828
Magnitude of the estimates (%)	-2.8	1.8	1.1	1.0
B. Not dissatisfied with life				
DID estimates	0.019	0.002	0.015**	0.008
	(0.023)	(0.010)	(0.007)	(0.011)
Number of observations	5439	13053	25406	11899
Pre-pandemic mean among the treated	0.655	0.780	0.789	0.795
Magnitude of the estimates (%)	2.9	0.2	1.8	1.0
C. Not suffered from work-life balance stress				
DID estimates	-0.005	0.002	0.006	-0.004
	(0.021)	(0.015)	(0.010)	(0.009)
Number of observations	5259	12577	25292	11823
Pre-pandemic mean among the treated	0.409	0.254	0.314	0.260
Magnitude of the estimates (%)	-1.1	0.9	2.0	-1.6

Note: The unit of observation is the individual and year. The dependent variables are well-being outcomes. Standard errors robust against occupation as of December 2019 with level clustering are shown in parentheses. The specifications, estimated with an individual-level fixed effects model, incorporate age (quadratic function); dummy variables for partner status, cohabitation with a partner, and cohabitation with children; cross-terms between year dummies and 2019 dummies for graduation, public sector employment, and over five years of service; and prefecture-year fixed effects. Inference: * $p < .1$, ** $p < .05$, *** $p < .01$.

Appendix

In this appendix, we explain the details of construction of variables.

A.0.1 WFH availability

The treatment variable is constructed as follows based on the JPSED question, “Had a telework system been introduced in your workplace as of last December, and were you a teleworker?” The possible answers were: 1, introduced and is a teleworker; 2, introduced but is not personally a teleworker; 3, had not been introduced; 4, don't know. First, a WFH availability variable was generated, taking the value 1 if WFH was introduced (values 1 and 2 in the original JPSED question), and 0 if it had not been introduced. Then, the average per occupation was calculated, thus creating a variable that represents the proportion of WFH availability per occupation in the sample. Finally, the treatment variable is created: it takes the value 1 for if the respondent is in an occupation that has an above-median proportion of WFH availability, and 0 if the respondent is in an occupation with a below-median proportion of WFH availability. We construct this variable using data from 2019, immediately prior to COVID-19.

A.0.2 Care variables

Questions related to caregiving are as follows: ”Do you usually take care of a family member?” There are seven responses: 1, ”I do everything”; 2, ”I do most of it”; 3, ”My family and I share the care responsibility equally”; 4, ”Another family member does most of it”; 5, ”Another family member does everything”; 6, ”Neither I nor my family do it”; 7, ”There is no family member needing care”. We define ”provide care” if the answer is 1, 2 or 3. If the respondent answers 1 or 2, it is defined as ”intensive care,” and if the respondent answers 3, it is defined as ”moderate care”. We exclude from the sample those who responded with 7 and have no family members in need of care.

A.0.3 Health outcomes

We construct a health index similar to the Somatic Symptom Scale-8 (SSS-8).[23] Specifically, the following eight items are used: ”I have a headache or feel dizzy”; ”My back, waist or shoulders hurt”; ”I feel palpitations or shortness of breath”; ”I feel extremely tired”; ”I feel nervous”; ”I feel depressed”; ”I have no appetite”; and ”I have difficulty sleeping”. For each of these, the possible responses are as follows: 0, “Always”; 1, “Often”; 2, “Sometimes”; 3, “Hardly ever”; 4, “Never”. We sum up these items' values for a maximum possible score of $8 \times 4 = 32$, creating a health index variable where 32 is the healthiest and 0 is the unhealthiest. Next, for the health index sum, we

[23] SSS-8 is a self-report questionnaire designed to assess the burden of physical symptoms in public health fields, with its characteristics summarized by Gierk et al. (2014). The JPSED questionnaire is not entirely identical to the original items, though there is considerable overlap in content. A notable difference is that the JPSED includes a question concerning mental health; ”I feel depressed.”

calculate the first quartile, median, and third quartile values for the entire sample. We obtained the values 15, 19, and 23 respectively. Then, we construct the health index Q1, health index median and health index Q3 are variables as follows. health index Q1: A dummy variable that is 1 if the health index sum is greater than 15, and 0 otherwise. health index median: A dummy variable that is 1 if the health index sum is greater than 19, and 0 otherwise. health index Q3: A dummy variable that is 1 if the health index sum is greater than 23, and 0 otherwise. These are all dummy variables that represent good health.[24]

A.0.4 Labor market outcomes

We use working status (whether working or not), working hours, annual income, hourly wages, and regular employment status as labor market outcomes. In particular, the construction of wages will be explained. For annual income, we used the sum of annual income from the main job and annual income from side jobs. For hourly wages, some respondents answered with hourly wages as the method of payment, which allowed us to use the values as is. For other salaried respondents without a reported hourly wage, it is assumed that individuals work 50 weeks per year. The weekly working hours of the main job and of the side jobs are added together and multiplied by 50. However, for those who answered "irregular, we cannot answer on a weekly basis" regarding side jobs, calculations are based solely on the main job. The sample is limited to those who worked in all months from January to December.

A.0.5 Well-being

Well-being in this study refers to happiness, life satisfaction, and work-life balance stress. First, respondents were asked, "How happy were you over the past year?", with responses ranging from 1, "very unhappy", to 5, "very happy". We create a dummy variable for happiness taking a value of 1 if the respondent answered 5 or 4, and 0 if they answered 3, 2 or 1.

Next, we create a dummy variable for life satisfaction based on the question: "How satisfied were you with your life overall during the past year?" A score of 1 is assigned if the respondent answered "Satisfied" or "Fairly satisfied"; a score of 0 is assigned if the respondent answered "Neither satisfied nor dissatisfied", "Somewhat dissatisfied", or "Dissatisfied".

Finally, for work-life balance stress, we similarly create a variable named "good WLB". This is a dummy variable that takes the value 1 if the respondent answered "Not at all" or "Not really" to

[24] The same also applies to good dummy-based health outcomes for other definitions as robustness check, where a binary variable is created with a value of 1 to cases where there was none or very little, and a value of 0 to cases where there was some, often, or always. A health outcome is then created where 8 is the healthiest and 0 is the unhealthiest. This is called the good dummy-based sum. Using the entire sample, the first quartile, median, and third quartile point (the value can be any number from 0 to 8), good dummy variables are constructed. Q1: A dummy variable that takes the value 1 if the Good dummy-based sum is greater than Q1, and 0 otherwise. Good dummy-based median: A dummy variable that takes the value 1 if the Good dummy-based sum is greater than the median, and 0 otherwise. Good dummy-based Q3: A dummy variable that takes the value 1 if the Good dummy-based sum is greater than Q3, and 0 otherwise. These are all dummy variables that represent good health.

the question "Did you feel stressed about balancing your work and family life over the past year?", and 0 if they answered "A little", "Quite a bit", or "Very much".

Table A.1: Summary Statistics

	Mean	SD
Demographics		
Female	0.50	0.50
Age	47.13	10.92
Married	0.63	0.48
Having children	0.57	0.50
Number of children	1.92	0.76
College or above	0.30	0.46
Employment		
Regular employees	0.66	0.48
Firm size		
1-99	0.43	0.50
100-999	0.26	0.44
1000+	0.23	0.42
Public sector	0.08	0.27
Years of tenure	12.36	10.88
Working from home		
working from home	0.11	0.32
working from home for 40% or more of weekly working hours	0.04	0.20
companies allowing employees to work from home	0.18	0.39
Informal care provision		
Any care	0.06	0.23
Intensive care	0.04	0.19
Moderate care	0.02	0.13

Note: The sample consists of observations from 2016 and 2022. The sampling weight in 2017 is used for the calculations.

References

Abrahamsen, Signe A. and Maja Weemes Grøtting, "Formal care of the elderly and health outcomes among adult daughters," *Health Economics*, 2 2023, *32*, 436–461.

Alipour, Jean Victor, Harald Fadinger, and Jan Schymik, "My home is my castle – The benefits of working from home during a pandemic crisis," *Journal of Public Economics*, 4 2021, *196.*

Ando, Michihito, Masato Furuichi, and Yoshihiro Kaneko, "Does universal long-term care insurance boost female labor force participation? Macro-level evidence," *IZA Journal of Labor Policy*, 5 2021, *11*, 493–505.

Barrero, José María, Nicholas Bloom, and Steven J. Davis, "The Evolution of Work from Home," *Journal of Economic Perspectives*, 2023, *37*, 23–50.

Bauer, Jan Michael and Alfonso Sousa-Poza, "Impacts of Informal Caregiving on Caregiver Employment, Health, and Family," *Journal of Population Ageing*, 9 2015, *8*, 113–145.

Bertrand, Marianne and Jessica Pan, "Gender Identity and Relative Income Within Households," *The Quarterly Journal of Economics*, 2015, *130*, 571–614.

Bloom, Nicholas, James Liang, John Roberts, and Ying Zhichun Jenny, "DOES WORKING FROM HOME WORK ? EVIDENCE FROM A CHINESE EXPERIMENT," *The Quarterly Journal of Economics*, 2015, *130*, 165–218.

Bluedorn, John, Francesca Caselli, Niels Jakob Hansen, Ippei Shibata, and Marina M. Tavares, "Gender and employment in the COVID-19 recession: Cross-Country evidence on "She-Cessions"," *Labour Economics*, 4 2023, *81.*

Bom, Judith, Pieter Bakx, Frederik Schut, and Eddy Van Doorslaer, "The Impact of Informal Caregiving for Older Adults on the Health of Various Types of Caregivers: A Systematic Review," 9 2019.

Carmichael, Fiona and Susan Charles, "The opportunity costs of informal care: Does gender matter?," *Journal of Health Economics*, 2003, *22*, 781–803.

Chong, Alice and Haruko Noguchi, "Heterogeneous Impacts of Telework on Pregnancy and Birth Rates: Evidence from Longitudinal Data on Employment Dynamics in Japan," *WINPEC Working Paper Series No.E2313*, 2024.

Cowan, Benjamin W and Joe M Spearing, "Loneliness, Mental Health and the Work-From-Home Revolution," Technical Report, National Bureau of Economic Research 2 2026.

Dingel, Jonathan I. and Brent Neiman, "How many jobs can be done at home?," *Journal of Public Economics*, 9 2020, *189.*

Fu, Rong, Haruko Noguchi, Akira Kawamura, Hideto Takahashi, and Nanako Tamiya, "Spillover effect of Japanese long-term care insurance as an employment promotion policy for family caregivers," *Journal of Health Economics*, 12 2017, *56*, 103–112.

_ , Toshiaki Iizuka, and Haruko Noguchi, "Long-term Care in Japan," Technical Report, National Bureau of Economic Research 11 2023.

Fukahori, Ryotaro, Tadashi Sakai, and Kazuma Sato, "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*, 11 2015, *62*, 518–545.

Fukai, Taiyo, Hidehiko Ichimura, and Keisuke Kawata, "Describing the impacts of COVID-19 on the labor market in Japan until June 2020," *Japanese Economic Review*, 7 2021, *72*, 439–470.

Geyer, Johannes and Thorben Korfhage, "Labor supply effects of long-term care reform in Germany," *Health Economics (United Kingdom)*, 2018, *27*, 1328–1339.

Gierk, Benjamin, Sebastian Kohlmann, Kurt Kroenke, Lena Spangenberg, Markus Zenger, Elmar Brähler, and Bernd Löwe, "The Somatic Symptom Scale-8 (SSS-8): A brief measure of somatic symptom burden," *JAMA Internal Medicine*, 2014, *174*, 399–407.

Goldin, Claudia, "A grand gender convergence: Its last chapter," *American Economic Review*, 2014, *104*, 1091–1119.

Gruber, Jonathan, Kathleen McGarry, and Charles Hanzel, "Long-term Care Around the World," Technical Report, National Bureau of Economic Research 11 2023.

Hara, Hiromi and Nuria Rodriguez-Planas, "Long-Term Consequences of Teaching Gender Roles: Evidence from Desegregating Industrial Arts and Home Economics in Japan," *Journal of Labor Economics*, 4 2023.

Houtven, Courtney Harold Van, Norma B. Coe, and Meghan M. Skira, "The effect of informal care on work and wages," *Journal of Health Economics*, 2013, *32*, 240–252.

Inoue, Chihiro, Yusuke Ishihata, and Shintaro Yamaguchi, "Working from home leads to more family-oriented men," *Review of Economics of the Household*, 6 2024, *22*, 783–829.

Kawaguchi, Daiji and Hiroyuki Motegi, "Who Can Work from Home? The Roles of Job Tasks and HRM Practices," *Journal of the Japanese and International Economies*, 2021, *62*, 101162.

Kleven, Henrik, "The Geography of Child Penalties and Gender Norms: A Pseudo-Event Study Approach," Technical Report, NBER WORKING PAPER SERIES 2025.

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

Korfhage, Thorben and Björn Fischer-Weckemann, "Long-run consequences of informal elderly care and implications of public long-term care insurance," *Journal of Health Economics*, 6 2024, *96.*

Le, Duc Dung and Yoko Ibuka, "Understanding the effects of informal caregiving on health and well-being: Heterogeneity and mechanisms," *Social Science and Medicine*, 1 2023, *317.*

Lee, Kangoh, "Working from home as an economic and social change: A review," *Labour Economics*, 12 2023, *85.*

Lunde, Lars Kristian, Lise Fløvik, Jan Olav Christensen, Håkon A. Johannessen, Live Bakke Finne, Ingrid Løken Jørgensen, Benedicte Mohr, and Jolien Vleeshouwers, "The relationship between telework from home and employee health: a systematic review," *BMC Public Health*, 12 2022, *22.*

Løken, Katrine V., Shelly Lundberg, and Julie Riise, "Lifting the burden: Formal care of the elderly and labor supply of adult children," *Journal of Human Resources*, 2017, *52*, 247–271.

Morikawa, Masayuki, "Work-from-home productivity during the COVID-19 pandemic: Evidence from Japan," *Economic Inquiry*, 4 2022, *60*, 508–527.

Motegi, Hiroyuki, "Understanding social norms for women toward the elderly: Evidence from Japan," *Journal of the Economics of Ageing*, 12 2025, *32.*

_ **and Masato Oikawa**, "Work from Home and Human Capital Investment," *IPSS working paper series*, 2024.

Niimi, Yoko, "The "Costs" of informal care: an analysis of the impact of elderly care on caregivers' subjective well-being in Japan," *Review of Economics of the Household*, 2016, *14*, 779–810.

_ , "Juggling paid work and elderly care provision in Japan: Does a flexible work environment help family caregivers cope?," *Journal of the Japanese and International Economies*, 12 2021, *62.*

Nishimura, Yoshinori and Masato Oikawa, "Who has benefited from nursing home expansion in Japan?: The effects of government supply-side intervention," *Journal of the Economics of Ageing*, 12 2025, *32.*

Nishitateno, Shuhei and Masato Shikata, "Has improved daycare accessibility increased Japan's maternal employment rate? Municipal evidence from 2000–2010," *Journal of the Japanese and International Economies*, 6 2017, *44*, 67–77.

Oshio, Takashi, "The association between involvement in family caregiving and mental health among middle-aged adults in Japan," *Social Science and Medicine*, 2014, *115*, 121–129.

_ **and Emiko Usui**, "How does informal caregiving affect daughters' employment and mental health in Japan?," *Journal of the Japanese and International Economies*, 9 2018, *49*, 1–7.

Pabilonia, Sabrina Wulff and Victoria Vernon, "Telework, Wages, and Time Use in the United States," *Review of Economics of the Household*, 9 2022, *20*, 687–734.

Raymo, James M., Letizia Mencarini, Miho Iwasawa, and Rie Moriizumi, "Intergenerational proximity and the fertility intentions of married women: A Japan-Italy comparison," *Asian Population Studies*, 2010, *6*, 193–214.

Rodríguez-Planas, Núria and Ryuichi Tanaka, "Gender norms and women's decision to work: evidence from Japan," *Review of Economics of the Household*, 3 2022, *20*, 15–36.

Schmitz, Hendrik and Magdalena A. Stroka, "Health and the double burden of full-time work and informal care provision - Evidence from administrative data," *Labour Economics*, 10 2013, *24*, 305–322.

Shen, Lucas, "Does working from home work? A natural experiment from lockdowns," *European Economic Review*, 1 2023, *151.*

Sugawara, Shinya and Jiro Nakamura, "Can formal elderly care stimulate female labor supply? The Japanese experience," *Journal of the Japanese and International Economies*, 2014, *34*, 98–115.

Wakabayashi, Midori and Wataru Kureishi, "Differences in the effects of informal family caregiving on health and life satisfaction between wives and husbands as caregivers," *Review of Development Economics*, 8 2018, *22*, 1063–1080.

Yamada, Hiroyuki and Satoshi Shimizutani, "Labor market outcomes of informal care provision in Japan," *Journal of the Economics of Ageing*, 2015, *6*, 79–88.

Yamagata, Mei and Asako Miura, "Retrospective bias during the COVID-19 pandemic," *Japanese Journal of Experimental Social Psychology*, 2023, *62*, 234–239.

Yokoyama, Izumi, Kazuhito Higa, and Daiji Kawaguchi, "Employment Adjustments of Regular and Non-Regular Workers to Exogenous Shocks: Evidence from Exchange-Rate Fluctuation," *ILR Review*, 3 2021, *74*, 470–510.