Discussion Paper Series – CRC TR 224

Discussion Paper No. 403
Project C 01

Spillover Effects of Old-Age Pension Across Generations: Family Labor Supply and Child Outcomes

Katja M. Kaufmann [1]
Yasemin Özdemir [2]
Han Ye [3]

October 2025
(First version: March 2023)

[1] Bayreuth University, briq, CESifo, HCEO, IZA, Email: katja.kaufmann@uni-bayreuth.de
[2] Bayreuth University, Email: yasemin.oezdemir@uni-bayreuth.de
[3] University of Mannheim, IZA, ZEW, Email: han.ye@uni-mannheim.de

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Collaborative Research Center Transregio 224 - www.crctr224.de
Rheinische Friedrich-Wilhelms-Universität Bonn - Universität Mannheim

Spillover Effects of Old-Age Pension across Generations: Family Labor Supply and Child Outcomes*

Katja M. Kaufmann† Yasemin Özdemir ‡ Han Ye §

First version: June 2022

Current Version: September 2025

Abstract

We study the impact of grandparental retirement decisions on family members' labor supply and child outcomes by exploiting a Dutch pension reform and a fuzzy Regression Discontinuity design. We find that a one-hour increase in grandmothers' hours worked causes adult daughters with young children to work 40 minutes less. Daughters without children, with older children and sons/daughters-in-law are not affected. We find positive reform effects on test scores of grandchildren aged 4-7, who experienced a substitution from grandparental to maternal care, and no effects for grandchildren aged 8-12, for whom grandparental childcare was substituted for by formal care.

JEL Codes: J13, J22, J26, I38, D64
Keywords: spillover effects, retirement, pension reforms, grandparental childcare, maternal labor supply, child development

*Katja Kaufmann (PI), Yasemin Özdemir and Han Ye gratefully acknowledge funding from the German Research Foundation (DFG) through CRC TR 224 (Project C01). We also thank Andreas Backhaus, Sandra Black, Kasey Buckles, Emilia Del Bono, Thomas Cornelissen, David Dorn, Libertad González, Daniel Hamermesh, Hilary Hoynes, Eckhard Janeba, Giovanni Mastrobuoni, Todd Morris, Cheti Nicoletti, Matthew Notwidigdo, Daniele Paserman, Gerard Pfann, Barbara Petrongolo, Paolo Pinotti, Uta Schönberg, Michéle Tertilt, Marian Vidal-Fernández, Riley Wilson and seminar and conference participants at the CRC annual meeting, the IZA gender workshop 2022, IZA/SOLE Transatlantic Meeting 2022, SEHO 2022, EALE 2023, CESifo Area Conference on Labor Economics 2024 and at the University of Bamberg, University of Bonn, Duesseldorf University, Technical University Dortmund, Free University of Berlin, Johannes Gutenberg University Mainz, University of Mannheim, Friedrich-Alexander-University Erlangen-Nürnberg, Pompeu Fabra University, University of Strathclyde Glasgow, University of Würzburg, University of St. Gallen, ETH Zurich, CReAM at University College London, Ca' Foscari University of Venice, the University of Essex and University of Basel for their helpful insights and suggestions.

†University of Bayreuth, CESifo, HCEO, IZA, *Email:* katja.kaufmann@uni-bayreuth.de

‡University of Bayreuth, *Email:* yasemin.oezdemir@uni-bayreuth.de

§University of Mannheim, IZA, ZEW *Email:* han.ye@uni-mannheim.de

1 Introduction

Evaluating the overall impact of public policies is complex. In addition to direct effects, there is a myriad of potential spillover effects, both within and across generations. Among the most important public policy changes in developed countries in recent years have been changes in retirement policies, aimed at increasing labor supply in old age. Recent evidence shows that such policies also have indirect effects on spouses/partners, which tend to exacerbate the direct labor supply response of the older generation.[1] However, relatively little is known about the cross-generational spillover effects of older people working longer on their children and grandchildren.

Grandparents make up a large share of the older population and play an essential role in childcare. In Europe from 2004 to 2015, 74% of women and 70% of men aged 64 were grandparents (Backhaus and Barslund, 2021). In 2023, around 1.5 billion grandparents make up 20% of the world population (The Economist, 2023). In most OECD countries, more than 45% of grandparents took care of grandchildren in 2006 (OECD, 2012).[2] A delay in grandparents' retirement may thus lead to spillover effects across multiple generations, because of a reduction in grandparental childcare.

One important consequence of pension reforms increasing grandparents' age at retirement may be a decrease in the labor supply of adult children (in particular daughters), which may counteract the direct effects on the overall labor supply. There may be further important ramifications for long-run labor market outcomes of mothers and for the implied child penalty and gender wage gap of a society. Lastly, there may be impacts on long-run outcomes of (grand)children due to changes in childcare modes. Despite these critical implications at the individual and societal levels, the relevance of such spillover effects across generations is relatively understudied.

This paper aims to fill this gap by investigating the importance of multigenerational spillover effects of old-age pension. In particular, we study a Dutch pension reform in 2006, which made early retirement less attractive for people born in 1950 or later. Using administrative data covering the universe of the Dutch population and a fuzzy Regression Discontinuity (RD) design, we provide first evidence of the spillover effects of a pension reform across three generations of household members in one common setting. The main analysis focuses on the grandmother-mother-grandchild linkage, and on families with mothers whose youngest child is of primary school age during the sample period (i.e., aged 4-12 when grandmothers are aged 60-64). We focus on ages 60 to 64 because the pension reform effects are concentrated between these ages (Rabaté et al., 2024).

First, we employ the cohort-based pension reform in an RD design and show that it led to a

[1] See, for example, Coile and Gruber (2007), Mastrobuoni (2009), Manoli and Weber (2016), Blundell et al. (2016), for recent evidence on the direct effects of recent pension reforms and see, for example, Hurd (1990), Coile (2004), Stancanelli and Van Soest (2012), Lalive and Parrotta (2017) for indirect effects on spouses/partners.

[2] In the Netherlands, the country under study in this paper, this fraction was 60% in 2006. For children aged 4 to 12, who attend primary school, 20% of parents relied solely on grandparental childcare in 2008 (see Section 2.1).

considerable increase in grandmothers' labor supply. To simultaneously capture the extensive and intensive margin of labor supply, we primarily focus on the total number of hours worked (including zeros). We find that grandmothers increase their work hours between ages 60 and 64 by 6.4 hours per month (19%), while their likelihood of being employed increases by six p.p. (15%).

In the main analysis, we examine the spillover impacts on adult children's labor supply, employing the same RD approach. The reform decreased the labor supply of daughters with young children by 4.9 hours per month (6%) when their mothers are between the ages of 60 and 64. To scale the effects and investigate intergenerational labor supply spillovers, we also employ a fuzzy RD design. We find that a one-hour increase in grandmothers' work hours decreases the work hours of adult daughters with young children by forty minutes.

To probe the mechanisms, we show that the effects are linked to general caregiving patterns and childcare needs, suggesting the reduction in grandmothers' time availability for childcare as the main channel. First, we find only small spillover effects of grandfathers' labor supply, despite large direct effects, consistent with grandmothers providing more childcare.[3] Second, we find that an increase in grandmothers' labor supply decreases the labor supply of daughters, but not of sons or daughters-in-law, consistent with maternal grandparents providing more childcare (see Section 2.1). The effect on daughters is mirrored in the effect of opposite sign on their husbands, who increase their labor supply, while household income remains unchanged. The increased specialization within the household is likely to affect intra-household decision-making. Third, we find no effect on adult daughters without childcare needs, that is daughters without children or with children older than 12. Instead, results are driven by mothers with a youngest child below 12, with the largest effect for children aged 4 to 7 when in pre-/primary school. This is consistent with grandparental childcare being particularly important for families with children in this age group (see Section 2.1). Lastly, we investigate heterogeneous effects by grandmothers' time availability. We find that grandmothers living in the same neighborhood as their daughters have a larger impact, and that only grandmothers with healthy partners and with only one young maternal grandchild lead to a decrease in mothers' labor supply, underlining the importance of time transfers as mechanism.[4]

Furthermore, we study the reform's impact on the educational performance of (grand)children, in particular their performance in a high-stakes test (*Cito*) taken at the end of primary school. We find positive effects on children aged 4 to 7 when their grandmothers were affected by the reform. These children, who experienced a substitution from grandparental to maternal care, scored 31% of a std. dev. higher and are 6.4 p.p. more likely to receive an academic track recommendation (prerequisite

[3]This is consistent with the literature according to which grandmothers' but not grandfathers' labor supply responds to the arrival of a grandchild (Rupert and Zanella, 2018; Backhaus and Barslund, 2021). Similarly, there is only a weak linkage between paternal labor supply and childcare (Kleven et al., 2019b; Huebener et al., 2020).

[4]We provide evidence against other channels, such as monetary transfers from grandparents to children, "reminder effects" about pension policies, and "role model effects" of grandmothers working more (see Section 4.3).

for university enrollment).[5] For children aged 8 to 12, we find no effects overall, but some negative effects for boys, who are less likely to receive the academic track recommendation. For this age group, grandmothers' time availability decreases, while mothers do not change their labor supply and after-school care increases, pointing towards formal care substituting for grandmothers' care.

Lastly, based on the framework of Kleven et al. (2019a), we show that changes in grandmothers' labor supply also have dynamic long-run effects on child penalties. The labor supply of women whose mothers are affected by the pension reform recovers more slowly after the birth of the first child than that of women with unaffected mothers. While the child penalty remains at 30% for the latter, the child penalty continues to widen and reaches 36% eight years after the birth of the first child for the former. Thus, the pension reform magnifies the already existing child penalty and gender gaps, underlining the importance of unintended distributional consequences.

Our paper contributes to the following four strands of literature. First, it speaks directly to the scarce literature on spillover effects of pension policies, which focuses on the effects on spouses (see Hurd (1990), Coile (2004), Stancanelli and Van Soest (2012), Lalive and Parrotta (2017)), rather than on spillover effects of pension policies across generations.[6] There is a small body of recent work using the retirement eligibility age as an instrument for grandmothers' labor market participation to study the impact on maternal labor supply. Aparicio Fenoll and Vidal-Fernandez (2015) and Bratti et al. (2018) examine the changes in retirement eligibility ages of pension reforms in Italy and find that grandmothers participating longer in the labor market reduce their daughters' labor force participation. Aparicio Fenoll (2020) and Zamarro (2020) use the retirement eligibility ages across countries in Europe as instruments for grandmothers' labor supply and find positive effects on their daughters' labor force participation. We aim to contribute to this literature by going beyond the effects on adult daughters and providing a more complete picture, causally estimating impacts of pension reforms on three generations in an integrated perspective. Moreover, the high-quality Dutch data allow us to measure a variety of labor supply outcomes, including total hours worked per month, to identify effects on (grand)children's academic performance, and to investigate dynamic effects on the child penalty and gender gaps.

Second, we contribute to the general literature that studies the responses of maternal labor supply to various care provisions.[7] Our paper provides causal evidence of strong responses to the

[5] Our results are in line with recent findings on the importance of maternal care for children's cognitive and non-cognitive skills (see, e.g., Fort et al. (2020) and Baker et al. (2008)). Their counterfactual is, however, formal care. We provide suggestive evidence that our results appear to be driven by higher-quality maternal care substituting for lower-quality grandmother care. Also, in our context household income remains unchanged.

[6] Papers examining intergenerational spillovers of other types of policies are, e.g., Dahl et al. (2014), Aizer et al. (2016), and Hoynes et al. (2016a) on welfare programs and Black et al. (2005) on education policies. Moroni et al. (2023) examine the spillover effects of grandfathers working more, induced by a change in job search requirements for the unemployed aged 57.5 and older, on the educational performance of their grandchildren. Our paper is also related to the literature on peer effects in maternal labor supply decisions (see Nicoletti et al. (2018) and Olivetti et al. (2018)).

[7] For the effect of formal childcare, see, e.g., Baker et al. (2008); Cascio (2009); Fitzpatrick (2010); Bauernschuster

availability of grandparental care, contributing to the limited evidence on the effects of grandparents on maternal labor supply, as discussed above (also see Posadas and Vidal-Fernandez (2013)).

Third, our paper relates to the literature on parental investments, childcare choices, and skill development in childhood, which finds –as we do– that maternal time is an important determinant for children's cognitive development (Bernal and Keane, 2011; Carneiro et al., 2013; Del Bono et al., 2016; Francesconi and Heckman, 2016; Bastian and Lochner, 2022).[8] Our paper is most closely related to the following three studies on the impact of grandparental care. Del Boca et al. (2018) show that UK children cared for by their grandparents (instead of formal childcare) have higher verbal skills, but are worse on other cognitive tests. Zhang et al. (2021) and Wang and Bansak (2024) find that grandparents as primary care takers (instead of parents) delay the achievement of children aged 1 to 5 and have adverse impacts on test scores of children aged 10 and above in China, respectively.[9] We contribute to this scarce literature by showing that a substitution from grandparental to maternal care has important positive effects on young children's educational performance in the Netherlands, while a substitution from grandparental to formal care or no adult supervision can have negative effects.

Lastly, our paper is connected to research on gender inequality in the labor market (see reviews by, e.g., Altonji and Blank (1999); Olivetti and Petrongolo (2016) and Blau and Kahn (2017)). We contribute to this literature by showing that changes in grandmothers' labor supply have dynamic effects on child penalties, a question on which evidence is mostly lacking (a notable exception is Marcos (2023) on Mexico). In particular, lack of care support from grandmothers slows down the recovery of women's earnings and working hours to pre-birth levels.

Our results show that public policies, such as pension reforms, can trigger multigenerational spillover effects with important distributional consequences. While the studied reform has reached the intended goal of increasing the labor supply in old age, maternal labor supply has decreased. This has critical implications for women's long-run labor market outcomes, for the child penalty, and gender gaps within households and society overall. On the other hand, younger children appear to have benefited from the increase in maternal time. Our paper thereby adds to a recent strand of the literature that estimates the long-term costs and benefits of public policies (see, e.g., Hoynes et al. (2016b), Aizer et al. (2022) and Bailey et al. (2023) on safety net programs).

and Schlotter (2015); Baker et al. (2019) and for the effect of parental leave policies, see, e.g., Gruber (1994); Schönberg and Ludsteck (2014); Kleven et al. (2024); Bailey et al. (2025).

[8]Studies on extensions in paid maternity leave find no overall effects on child well-being (Dustmann and Schönberg, 2012; Dahl et al., 2016; Danzer and Lavy, 2018), but positive effects for children from privileged families (Danzer and Lavy, 2018; Ginja et al., 2020). Studies on formal childcare tend to find small overall effects on skills (e.g., Cornelissen et al., 2018; Felfe and Lalive, 2018) or sometimes even negative effects for girls/children from privileged families (e.g., Baker et al., 2008; Fort et al., 2020; Baker et al., 2019), but important positive effects for children from less privileged families (e.g., Bailey et al., 2021; Andrew et al., 2024). Also see the survey article by Duncan et al. (2023).

[9]Zhang et al. (2021) instrument with local demand conditions that influence childcare choices, while Wang and Bansak (2024) use as instruments parents' number of brothers and sisters or employ fixed-effects models.

2 Institutional Setting and Empirical Methods

2.1 Grandparents and childcare modes

Grandparents play an essential role in childcare in the Netherlands and many other countries. In the Netherlands, 60% of grandparents take care of at least one grandchild. In most OECD countries, this fraction is between 45 and 55%, while Ireland has an even higher fraction of grandparents providing care (65%) (OECD, 2012). In the US, according to the Survey of Income and Program Participation, 23.4% of all children under 5 years old benefited from regular grandparent-provided childcare in 2011. In fact, for 93% of these, grandparents were the primary childcare arrangement (Laughlin (2010) and Rupert and Zanella (2018)). Furthermore, a growing literature has shown that the arrival of a grandchild reduces grandmothers' employment (Rupert and Zanella, 2018; Gørtz et al., 2025; Frimmel et al., 2020; Karademir et al., 2024) and the provision of formal childcare increases grandmothers' employment (Karademir et al., 2024). This evidence points to the importance of grandmothers in childcare provision.

The relevance of grandparental childcare is linked to the education and childcare system in the Netherlands. Children under the age of four can attend childcare centers, which charge an hourly rate of around 7 euros on average. These daycare centers typically have opening hours that mirror working hours, making it easier for mothers of very young children to balance family and work without the help of grandparents. From the age of 4, most children start primary school (mandatory at age 5), and at age 12, they start attending secondary school. Although primary schools are free of charge, they provide fewer hours and a less flexible schedule. Most primary schools close around 2 p.m. and on Wednesday afternoons. Parents thus have to reduce work hours, buy additional hours in the private market, or ask for additional hours from other helpers (such as grandmothers) or send their children to formal out-of-school/ after-school care (buitenschoolse opvang, OSC). See Appendix A.4 for a detailed description of the childcare system. The survey evidence below indeed shows that grandparents matter most when children are aged 4 to 8, exactly reflecting the complementarity between free half-day school care and grandparental care for children of this age.

To further illustrate the importance of different childcare modes in the Netherlands, we explore the 2008 wave of the Longitudinal Internet Studies for the Social Sciences (LISS) data.[10] Overall, there are four types of childcare modes: parental care, grandparental care, formal (institutionalized) childcare, and informal childcare (other than grandparental care). First, according to the LISS data – and similarly to most countries – mothers spend more time in terms of childcare than fathers.

[10]The LISS (Longitudinal Internet Studies for the Social Sciences) panel is a representative sample of Dutch individuals who participate in monthly internet surveys which are administered by CentERdata (Tilburg University, The Netherlands). We use the 2008 wave because it is the wave shortly before our sample period. For more details, see Section B.1.

Even conditional on both parents working, 48% of mothers with young children state that they currently work less to care for their children, compared to only 8% of fathers. Conditional on working less, mothers state working 14 hours less per week to care for their children, while those fathers who state that they reduce their work hours to care for their children reduce their working time by 8 hours per week.

Next, Figure A1 shows the distribution of the different types of childcare (other than parental childcare), in particular, formal care, grandparental care, and other types of informal care. Panel (a) of Figure A1 displays what fraction of parents use a particular mode of childcare (potentially in combination with other modes), while Panel (b) displays the fraction of parents using a particular combination of childcare modes (presenting the most common combinations). According to Panel (a) of Figure A1, around 35 to 40% of parents report using grandparental care in the past week, while 60-80% of them use some formal care. The two most common care arrangements for children younger than 4 years old are only formal care and a combination of formal daycare and grandparental care. According to Panel (b), children aged between 4 and 12 need less childcare overall because primary school (which includes pre-school) provides a considerable amount of free care. From the perspective of the paper, it is important to note that 20% of parents with children aged 4 -12 rely *solely* on grandparental care, which is only true for less than 10% for younger children below age 4. Lastly, the data show that maternal grandparents are more important in terms of caregiving than paternal grandparents, as more than 60% of the caregiving grandparents are maternal grandparents.

2.2 Early retirement schemes and the 2006 Dutch pension reform

The Dutch pension system consists of three pillars: Pay-as-You-Go state pensions (AOW, *Algemene Ouderdomswet*), occupational pensions, and individual savings. The first pillar, the state pensions, provides all Dutch residents with a flat-rate pension once they reach their AOW claiming age. The second pillar, occupational pensions, which we focus on in our analysis, is a collective pension scheme connected to a specific industry or company, capital-funded and managed by pension funds. Contribution to the second pillar is mandatory. Retiring early (i.e., before the statutory AOW claiming age) was and is still only possible through the early pension scheme, which is part of occupational pensions. The third pillar consists of non-mandatory savings. See Appendix A.1 for more details on the Dutch pension system.

The reform we explore in this paper is the 2006 pension reform, which made early retirement less attractive. Before 2006, the earliest possible age to claim occupational pensions was between ages 55 and 60, depending on the sectoral schemes. Early retirement was attractive, offering a replacement rate of around 80%. The years of early retirement were counted as accumulated years of work. At age 65, the early retirement benefit was replaced by the regular AOW and an old-age occupational pension, the benefit of which depends on years worked and average lifetime earnings. Consequently, around 80% of all workers retired at the age of 62 or younger before 2006 (Statistics Netherlands, 2009).

In 2006, there was a major reform of the early retirement schemes. The goal was to encourage labor market participation of the elderly by speeding up the transition towards an actuarially fair early retirement system. The reform package was announced in 2005 and came into effect on January 1, 2006. However, people who were 55 years or older before January 1, 2005, are not affected by the new bill. Thus, people born before 1950 are exempted from the changes, while those born since 1950 are facing substantial financial incentives to postpone early retirement as of January 1, 2006. Even though the general topic of eliminating early retirement tax benefits has been discussed since 2000, the sharp differential treatment by birth date was unexpected by the public and spurred heated public debate.[11]

This cohort-based reform creates a sharp discontinuous drop in early retirement incentives for people born since January 1950. Figure A2 shows the distribution of age at exiting employment for women born in 1949 and 1950. There is a clear shift towards later retirement, with most of the change being concentrated between ages 60 and 64. This is consistent with the findings of Rabaté et al. (2024), which show that the reform largely increased employment and decreased retirement after age 60 and before age 65. Therefore, the reform led to a quasi-exogenous change in the early retirement incentives of the older generations, which allows us to causally estimate the impact of the reform on grandmothers' and mothers' labor supply and on children's outcomes. Specifically, we employ a Regression Discontinuity Design based on grandmothers born shortly before and after the cutoff of January 1950, and compare outcomes (of grandmothers, mothers, and children) when grandmothers are aged 60 to 64.

2.3 Empirical Methods

We investigate the impact of the pension reform on three generations. First, we show that the reform creates a sharp discontinuous increase in labor supply for grandmothers born since January 1950. The direct effect of the reform on grandmothers' outcome y^{GM} is modeled in the following Regression Discontinuity (RD) framework:

$$y_i^{GM} = \alpha_0^{GM} + \alpha_1^{GM} D_i^{GM} + \alpha_2^{GM} r_i^{GM} + \alpha_3^{GM} D_i^{GM} \times r_i^{GM} + \delta^{GM} X_i + \epsilon_i^{GM} \tag{1}$$

[11] See Appendix A.2 for details on the evolution of early retirement schemes and the reform. See also Euwals et al. (2010) for a summary of sectoral rules for the period 1989–2000. For earlier analyses of this reform, see e.g., Lindeboom and Montizaan (2020); Rabaté et al. (2024). In 2006, a "Life course savings" scheme was also introduced, which allows workers to save part of their gross salary to finance a period of unpaid leave. All individuals were eligible to participate in this savings scheme, independent of their birth cohorts. See Appendix A.3 for further discussions.

where r_i^{GM} is the running variable defined as grandmothers' birth month, c_i^{GM}, centered around the cutoff c, $r_i^{GM} = (c_i^{GM} - c)$. c is set to January 1950. The treatment indicator D_i^{GM} is defined as $D_i^{GM} = \mathbb{1}(r_i^{GM} \geq 0)$. α_2^{GM} and α_3^{GM} allow for cohort trends in the outcome variables to differ by treatment status. The coefficient $\widehat{\alpha_1^{GM}}$ is the estimated impact of the reform on grandmothers' labor supply outcomes. X_i contains demographic characteristics of the grandmother and the mother, including the mother's age and migration background, the number of her siblings and sisters, the age at first birth, and a list of predetermined characteristics of mothers and grandmothers, including marital status, number of (grand)children, disability status of the grandmother's partner, employment probability of the grandmother, and whether mother and grandmother lived in the same district before the analysis period. We also include sector fixed effects to control for sector-specific pension rules.[12] All predetermined variables are averaged over the period, when grandmothers were 50 to 53 years old. All outcome variables and contemporary controls are averaged over the period when grandmothers were 60 and 64 years old. To estimate Equation (1) and the equations below, we use a local linear specification with a triangular kernel (Hahn et al., 2001; Gelman and Imbens, 2019), and the mean squared error optimal bandwidths (Calonico et al., 2014, 2018) of each outcome.

Second, we investigate the middle generation: the mothers. The reform allows us to causally estimate the reduced-form impact on mothers' labor supply, as well as the spillover effect of grandmothers' labor supply on the labor supply of the mothers. The corresponding reduced-form model for mothers' outcome y^M is:

$$y_i^M = \alpha_0^M + \alpha_1^M D_i^{GM} + \alpha_2^M r_i^{GM} + \alpha_3^M D_i^{GM} \times r_i^{GM} + \delta^M X_i + \epsilon_i^M \tag{2}$$

where y_i^M is a list of mothers' labor supply outcomes, which are the average values over the period when the grandmother is between the ages of 60 and 64. The coefficient $\widehat{\alpha_1^M}$ is the reduced-form effect of the reform on the outcomes of mothers.

To quantify the effect of grandmothers' labor supply on mothers' labor supply, we report the two-stage least squares (2SLS) fuzzy RD estimate. The fuzzy RD estimates help us to understand to what extent and through which channels grandmothers' labor supply affects the maternal labor supply of their adult daughters. In general, it is difficult to causally estimate this effect for two reasons. First, unobserved variables may affect the employment decisions of both grandmothers and mothers. For example, grandmothers' gender identity can be transmitted to their adult daughters (Fernández et al. (2004); Kleven et al. (2019a)). Second, there may be reversed causality as grandmothers' retirement decisions can be affected by childcare decisions. The timing of grandparenthood can

[12] In the case where a grandmother has changed her sector at some point during her employment history, we consider the one in which she was employed the longest.

cause a reduction in the labor supply of grandmothers (Rupert and Zanella (2018); Frimmel et al. (2020); Backhaus and Barslund (2021)). The cohort-based pension reform employed in this paper allows us to address these endogeneity issues as follows.

$$y_i^M = \beta_0 + \beta_1 \widehat{Y_i^{GM}} + \beta_2 r_i^{GM} + \beta_3 D_i^{GM} \times r_i^{GM} + \theta X_i + \eta_i \tag{3}$$

$\widehat{Y_i^{GM}}$ is the predicted value of grandmothers' labor supply in Equation 1. The coefficient $\widehat{\beta_1}$ measures the local average treatment effects (LATEs) of grandmothers' labor supply on mothers' labor supply. The fuzzy RD estimate is analogous to a two-stage least squares (2SLS) estimate with imperfect compliance.

Lastly, we focus on the youngest generation, the (grand)children, and examine the reform's impact on children's educational performance. The reduced-form model for children's outcome y^C is

$$y_i^C = \alpha_0^C + \alpha_1^C D_i^{GM} + \alpha_2^C r_i^{GM} + \alpha_3^C D_i^{GM} \times r_i^{GM} + \delta^C X_i + \epsilon_i^C \tag{4}$$

where y_i^C is a list of children's outcomes, which are measured at age 12 when they took the Cito test. The treatment indicator D_i^{GM} is defined as a grandmother born since January 1950. X_i includes (in addition to the controls used in the analysis of mothers) the child's birth cohort and month, and treatment duration (i.e., number of years the child is exposed to the grandmothers' labor response when aged 60-64). The coefficient $\widehat{\alpha_1^C}$ is the estimated reform impact on children.[13]

As mentioned above, in the baseline analysis, we estimate a local linear regression with a triangular kernel (Hahn et al., 2001; Gelman and Imbens, 2019), and use the mean squared error optimal bandwidths generated by the Calonico et al. (2014) and Calonico et al. (2018) procedures for the different outcomes. Various robustness checks (including varying the bandwidth/polynomial order, placebo checks and using alternative/augmented methods) are conducted in Section 4.3.

For an RD design to be valid, individuals must not manipulate the assignment variable, which, in our case, is the grandmother's birthdate. In Section 4.3, we verify that the density and the predetermined covariances are smooth around the cutoff. We also demonstrate that the three conditions necessary to interpret a valid 2SLS estimate are satisfied in Section 4.3.

3 Data and Sample

We use Dutch administrative data maintained by Statistics Netherlands (Centraal Bureau voor de Statistiek, CBS), which covers the entire Dutch population and contains information that allows

[13] We present results from regressions with clustered standard errors at the primary school level. The clustering allows for correlations of test performance within schools. Results are also robust to clustering at the mother level and to two-way clustering at the mother and primary school levels.

us to follow families across generations and over time.[14] Birth and marriage records enable us to link three generations and create extended family networks (for more details on the data, specific variables, and data sets, see Appendix B). We link individuals born around January 1950 to their two descendant generations and refer to this "first generation" as the *grandmothers* or *grandfathers*. Their adult children, i.e., the "middle generation", are referred to as *mothers/adult daughters* or *fathers/adult sons*, and the "third generation" is referred to as *(grand)children*. We also analyze the effects on the partners of the middle generation and refer to them as *sons-in-law* and *daughters-in-law*. Since the pension reform affects the first generation's labor supply mainly between ages 60 and 64 (see Figure A2), we examine the average labor market outcomes of grandparents and their adult children (i.e., mothers and fathers), while the "first generation" is between ages 60 and 64.

Baseline sample: Because the main goal of our analysis is to estimate spillovers of grandmothers on their adult daughters, we focus on the baseline sample of grandmothers born between 1948 and 1951 who have at least one daughter. We then exclude grandmothers who are migrants due to missing birth records. We further drop grandmothers who are unlikely to be affected by the reform. These include grandmothers who are self-employed, family workers, and inactive grandmothers, i.e., those who exited the labor force before age 50, those who have never been active in the labor market, and those who have claimed disability before age 55. We also drop those who died before age 65. We are thus left with 62% of the 1948-1951 generation of women.

Moreover, since we aim to investigate spillover effects due to changes in grandparental childcare provision, we further restrict our sample to adult daughters who have at least one child, i.e., who are mothers.[15] To focus on maternal labor supply, we exclude mothers who are studying (less than 1%) or who have incomplete employment histories due to work/ study abroad (around 3%).

We further restrict our baseline sample to mothers whose youngest child is of primary school age during the sample period (i.e., aged 4-12 when grandmothers are aged 60-64) because grandparents are particularly important as the sole source of childcare for this age group. This is because the need for childcare and the compatibility of grandparental childcare and formal care vary according to the child's age, as discussed in Section 2.1. Appendix B.2 provides more details on the sample construction. Table A1 shows that being affected by the pension reform has no significant impact on each restriction (i.e., on the likelihood of being dropped from the sample) that we impose in terms of sample selection.

In addition to our baseline sample, we also analyze samples of adult daughters without children, mothers with a toddler (below age 4), mothers with an adolescent (ages 13 to 18), mothers with

[14] Under certain conditions, these non-public microdata are accessible for statistical and scientific research. For further information: microdata@cbs.nl

[15] 29% of adult daughters do not have any children, while grandmothers are aged 60 to 64. In Section 4.3.7 we show that the reform had no effects on adult daughters' fertility. For the group of adult daughters without children, we provide labor supply results in Section 4.4.

inactive grandmothers, and other family members.

Summary statistics: Table A2 shows the characteristics of families in the *RD sample*, which contains families with grandmothers who are within the optimal bandwidth of 7 months around January 1950. In the RD sample, grandmothers have on average 2.5 adult children and 1.7 adult daughters. On average, mothers are 38 years old, entered the labor market at age 25, had their first child at age 28, have two children, and 66% are married. Our main outcome variables capture the labor supply of grandmothers and mothers, which are measured when the grandmothers are between the ages of 60 and 64. Grandmothers work on average 37 hours per month and earn 638 euros per month (all income measures are CPI-adjusted for the year 2015). These grandmothers are on average employed 45% of the time (27 months between 60 and 64), and are full-time employed for 5% of the time (3 months between 60 and 64). On average, grandmothers exit the labor force at age 61 and start claiming pensions at age 63. Mothers work 78 hours and earn 1533 euros per month. They are employed for 78% of the time (47 months between ages 60 and 64 of the grandmother), and for 6% of the time, employment is full-time (3.5 months between ages 60 and 64 of the grandmother). The *baseline sample* and *RD sample* are comparable to the *all mothers* sample (which consists of all adult daughters of grandmothers born between 1948 and 1951 who have a child, but not necessarily in the age range of 4 to 12), except for mothers in our (baseline and RD) sample working slightly less. This is consistent with the fact that mothers in our sample have at least one young (primary school age) child (for details, see Appendix B).

Children's sample: We also examine the reform effect on children's educational outcomes. In line with our main analysis, we study educational outcomes for the youngest child who was of primary school age when their grandmother was aged 60. To measure educational performance, we use data from a high-stakes standardized test (called Cito test) administered at the end of primary school (around age 11/12) to sort students into different secondary school tracks.

We merge the youngest children in our baseline sample with the test score data to obtain information on the number of correct answers on the Cito test, and the number of correct answers in math and verbal skills. Children in our sample took the Cito test between 2009 and 2019. We exclude children under the age of 4, because they are either too young to have taken the test or because their Cito test did not take place due to the COVID-19 pandemic. While the Cito test is used in the majority of schools to determine students' secondary school track, schools can opt for alternative tests, for which we do not have data. It is important to note that the schools, not the parents or children, select the type of test. Among all children aged 4 to 12 (of our baseline mothers), 50% attend schools that administer the Cito test (as opposed to alternative tests). Table A3 compares the characteristics of children (and their families) who can be matched with the Cito test data with those of all children in our baseline sample (see Appendix B.3 for a detailed description of the Cito linkage). The characteristics are very similar, suggesting no differential selection into taking the

Cito test. Indeed, Table A4 shows that the probability of a match with the data on the Cito test is not affected by the reform. We supplement our analysis of child outcomes with annual data on the childcare allowance that families receive for childcare usage, which contains information on the probability of childcare take-up, the type of childcare, and the hours requested.

4 Spillover Effects on Maternal Labor Supply

In this section, we study the direct and spillover effects of the old-age pension reform on maternal labor supply. First, we study the direct reform impact on grandmothers' labor supply. Then we analyze the indirect effect on the labor supply of adult daughters with the youngest child of primary school age, 4 to 12, when grandmothers are aged 60-64.[16] After conducting a number of validity and robustness checks, including whether the reform affected mothers' fertility, we shed light on the underlying mechanisms of the intergenerational spillover effects. We show that they are linked to likely caregivers (i.e., to grandmothers instead of grandfathers and to maternal instead of paternal grandparents) and to childcare need (such as to the age and existence of (grand)children).

4.1 Reform Impacts on Grandmothers' Labor Supply: First-Stage

First, we provide graphical evidence of the first-stage reform impact on grandmothers' total hours worked. Figure 1 (a) shows the bin scatter plot of total monthly hours worked as a function of the distance of grandmothers' birth month to the cutoff, which is January 1950. While we employ local linear regressions in the estimation (see Section 2.3), in the figures we add a solid line for illustrational purposes, which is a linear polynomial fit of the outcome on the running variable, given the optimal bandwidth generated by Calonico et al. (2014) and fit separately left and right of the cutoff. According to Figure 1 (a), grandmothers born before January 1950 (more specifically, born between June 1949 and December 1949) work on average 33 hours per month between the ages of 60 and 64, while grandmothers born between January 1950 and July 1950 work 42 hours between the same ages. Moreover, there is a clear jump at the cutoff from 35 to 42.5 hours per

[16] We hypothesize that mothers with children of primary school age (4-12) are most strongly affected by changes in the informal care provided by grandmothers, a hypothesis we test in Section 4.4 by comparing the effects for different age groups. From age 4 on, children can attend primary school, which offers around 30 to 35 hours of free care per week. In contrast, children below age 4 require care full-time, which grandparents are rarely able to provide in its entirety (according to the LISS data, grandparents provide about 9 hours of childcare per week). In addition, daycare schedules tend to be more flexible than school schedules, making it easier for mothers of very young children to balance family and work without the help of grandparents. In fact, for children aged 4 to 12, 20% of families rely solely on grandparental childcare, while for children below age 4, this is only the case for less than 10% of families (see Section 2.1). As a result, grandparents' time availability is less critical for the latter age group and hence less likely to alter mothers' labor supply decisions.

month, an increase of around 23%.[17]

Table 1 (Panel B) presents first-stage estimates of the pension reform. Columns (1), (2), and (3) show the results for a local linear regression without controls, with controls, and with controls and sector fixed effects, respectively. Standard errors are clustered at the grandmother level, since grandmothers may be in the sample multiple times (if they have several daughters with children in the relevant age range). The regression results lend support to the graphical analysis. We find that the reform increases grandmothers' monthly hours worked by around 6.4 hours (18%). This effect is robust across specifications and is highly significant at the one percent level. Moreover, the estimates are robust to varying bandwidths (see Panel (a) of Figure A3).

4.2 Effects on Mothers' Labor Supply

What are the implications of the increase in grandmothers' work hours for the labor supply of their adult daughters with young children? How important are spillover effects, particularly given that grandmothers tend to play an important role in childcare? Graphical evidence and our regression analysis show that there are indeed important spillover effects on mothers' labor supply.

Working hours Figure 1 (b) shows that mothers' working hours drop sharply at the cutoff, i.e., mothers with (grand)mothers affected by the reform work less. The regression results of the reform impact on mothers' labor supply are presented in Panel A of Table 1. The pension reform leads to a reduction of 4.9 hours (6%) in monthly working hours of mothers with treated grandmothers. The effect is robust across specifications and highly significant at the one percent level. Panel B also shows the corresponding 2SLS/fuzzy RD estimates. They show that a one-hour increase in grandmothers' monthly working hours induces a decline in mothers' monthly hours worked of around 38 minutes (0.63 hours) at the five percent significance level.

Other labor supply measures The patterns are similar for other measures of labor supply, such as the probability of being employed and the probability of working full-time (see Table A5 and Figure A4). The reform causes mothers with treated (grand)mothers to reduce the probability of being employed by 2.5 p.p. (significant at the five percent level), but has no significant impact on the probability of working full-time. This is not surprising, given that an important feature of female employment in the Netherlands is flexible working hours and the dominance of part-time work, with 70% of women working part-time in 2021 (Statistics Netherlands, 2022). The IV estimates indicate that when grandmothers work one hour more per month, mothers are 0.3 p.p. less likely to engage in formal employment (significant at the 10 percent level).

[17] Average employment rates of women between ages 60 and 64 are 38.7%, and full-time employment rates are only 5%, reflecting the weaker attachment of older women to the labor market.

Since a change in grandmothers' labor supply affects mothers' employment probability (and thus changes differentially who is working on the two sides of the cutoff), we cannot estimate the effect of grandmothers' labor supply on mothers' hours worked *conditional* on working. However, back-of-the-envelope calculations suggest that there is both an extensive and an intensive margin response.[18]

Characteristics of compliers To interpret the IV results, it is important to understand who the compliers are. As compliers cannot be individually identified, we follow Abadie (2003) and calculate the fraction of compliers in different subsamples to recover their characteristics (see Appendix C.1 for details). In Table A6, we characterize the compliers in our baseline sample based on mothers' and grandmothers' characteristics. While complier families are relatively similar to the overall sample, they are characterized by grandmothers who are slightly more attached to the labor market and by mothers who are somewhat more educated and who have slightly fewer children before the reform announcement.

4.3 Validity Checks

4.3.1 RD Assumption I: Smoothness in Density

For an RD design to be valid, individuals must not manipulate the assignment variable, which, in our case, is the grandmother's birthdate. Since the timing of grandmothers' birth cannot be affected by a pension reform more than 50 years later, and since we are using administrative birth records from the Netherlands, there is little to no room for manipulation. Figures A5a and b show the density plot of grandmothers' birth month, 24 months and 8 months around the cutoff. Figures A5c and d show the density plot of mothers' birth month, 24 months and 8 months around the cutoff. The fluctuating pattern of the density plots is similar when we compare grandmothers and their adult daughters of our sample as well as comparing them to the pattern for women (elderly and the middle generation) without (grand)children (see Figures A5e and f). This suggests seasonal patterns which commonly occur in terms of birth rates (and which are not driven, for example, by

[18] Table A5 shows that 78.5% of women are employed, while the remaining 22.5% work zero hours. Since the average number of total hours worked per month is 78.8 (see Table 1), we can infer that the employed women work on average 100 hours per month. How might our result of a decrease in the probability of employment of 0.3 p.p. translate into a change in total hours worked, if the entire response in hours was driven by the extensive margin response? A back-of-the-envelope calculation suggests that the reduction in the probability of employment of 0.003 translates into a decrease in total hours worked of 100 hours*(0.003)=0.3 hours. According to Table 1, we find a reduction in mothers' monthly hours worked of 0.63 hours, twice as large as 0.3 hours. This suggests that the estimated decrease in hours worked results from both an extensive and an intensive margin response. In principle, those working women who reduce their participation could have worked an above-average number of hours per month. This is, however, not very plausible. On the contrary, if those women who reduce their participation work a below-average number of hours, the back-of-the-envelope calculation would indicate an even larger role of the intensive margin response.

the sample restriction of having a (grand)child). Moreover, Haandrikman and van Wissen (2008) and Calot and Blayo (1982) show that in the Netherlands, birth rates peak in spring and are the lowest around November, which is consistent with the density plots described above.

4.3.2 RD Assumption II: Smoothness in Covariates

Next, we test for continuity of covariates in the RD Design by presenting graphs of the covariates against the running variable, and by performing statistical tests to check for discontinuities at the cutoff point. Based on Equations 1 and 2 with covariates instead of outcomes as dependent variables, we estimate the impact of grandmothers being born since January 1950 (as opposed to before) on a list of predetermined characteristics of grandmothers (including employment and marital status, partner being disabled, birth cohort of their partners and more) and of mothers (including the employment and marital status, distance to grandmothers, age at first employment, age at first birth and more). All variables are predetermined and refer to the time period when the grandmothers were aged 50 to 53 (between years 2000 and 2003, thus before the announcement of the reform), with the exception of grandmothers' income, which refers to ages 53 to 56, because income data is only available from 2003 onwards. All specifications use the optimal bandwidth selection algorithm and a local linear specification. Table A7 and Figure A6 show that covariates are smooth across the cutoff. In particular, there are no significant differences in the covariates above versus below the cutoff, with one exception out of 20 variables (significant at the 10 percent level), lending support to the RD assumption of smoothness in covariates.

4.3.3 LATE Assumptions: Instrument Validity

There are three conditions necessary to interpret the 2SLS estimate as a spillover effect of grandmothers' labor supply on their daughters' labor supply. First, grandmothers' birthdates are strongly associated with grandmothers' labor supply. We have shown the validity and magnitude of the first-stage relationship in Section 4.1. Second, a grandmother's birthdate only impacts her adult daughter's labor supply through changes in her own labor supply. The exclusion restriction could be violated if a grandmother who was born before or after 1950 affects her adult daughters' labor supply through channels other than her own labor supply. This assumption is fundamentally untestable. We argue that the exclusion restriction assumption is reasonable because there are no other reforms with the same grandmother birthdate cutoff. Moreover, to provide empirical support for the validity of the exclusion restriction, we rule out other channels, including monetary transfers, a "reminder effect" in terms of changes in pension policies, and "role model effects" of grandmothers working more/longer, by examining –among other points– the effects on other outcomes (such as grandparents' gross (household) income) or on groups without childcare needs.[19]

Third, the monotonicity condition requires that the changes in early retirement incentives (in our case, a change towards a less generous early retirement policy) always induce grandmothers to increase their labor supply or at least maintain the same level of labor supply as under the old regime. Given the nature of the 2006 pension reform, this condition is likely to be satisfied. Moreover, Figure A7 shows the cumulative distribution function of hours worked for women born before versus since January 1950. We can see that the distribution of hours worked for women born since January 1950 is of first-order dominance over the distribution for women born before.

4.3.4 Robustness: Bandwidth Choices and Polynomial Order

Several exercises further establish the robustness of the estimates. The RD estimates are robust with respect to the bandwidth choice, as depicted by Figure A3, which plots the point estimates and their 95% confidence intervals with bandwidths from 4 to 13 at one-month intervals using the local linear specification and a full list of controls (see Section 2.3). Table A9 shows that the RD estimates are also robust to different polynomial orders: local linear specification, local quadratic specification, and quadratic polynomials.[20] The Ataike information criterion (AIC), Bayesian information criterion (BIC), and AICc (AIC with a correction for small sample sizes) are also reported. Panel A of Table A9 presents the reduced-form reform impact on mothers' total monthly hours worked and shows that the estimates are robust to the choice of polynomial orders. According to the AIC, BIC, and AICc criteria, the local linear specification is the best fit in terms of the reduced-form impact on mothers' labor supply. Comparing the estimate based on this (preferred) specification with -for example- the estimate using quadratic polynomials and the same bandwidth, the coefficients are extremely similar (-4.912 versus -5.094, respectively). Panel B shows the IV estimates, which are also robust to using different polynomial orders.

[19]Table A8 shows that grandmothers' monthly gross income and household gross income are not affected by the reform (i.e., grandparents do not have more money available), which suggests that the monetary transfer channel is unlikely. In case of the reform impacting adult children's labor supply due to reminder effects (anticipating more costly early retirement in the future), we would expect adult daughters to work *more* to save up for retirement, not less. Moreover, we would expect to find effects on adult children more generally, i.e., also for adult sons and for adult daughters without childcare need, which we do not (see Sections 4.4.2 and 4.4.3, respectively). A similar argument applies to a "role model effect" of grandmothers working more, which should induce adult daughters to work *more*, not less, and apply to adult daughters more generally. These analyses suggest that it is the change in grandmothers' labor supply/time availability that causes mothers to work less.

[20]Gelman and Imbens (2019) suggest that a high-order polynomial regression is a poor choice in regression discontinuity analyses. They recommend local linear or quadratic polynomials for RD designs for causal inference.

4.3.5 Placebo Tests

We conduct two placebo exercises to further support the credibility of our estimates and to show that results are not driven, for example, by the seasonality of grandmothers' birthdates. First, we use a sample of families with grandmothers who have little attachment to the labor market during their lifetime. Specifically, we take families with grandmothers who exit the labor force before age 50.[21] We would not expect the pension reform to affect the labor supply of these grandmothers, and it should therefore not affect their daughters' labor supply. Table A10 shows the estimated reform impacts on the labor supply of these grandmothers and their adult daughters (mothers). As expected, we find no impact on grandmothers' total monthly hours worked. Moreover, all of the estimates on mothers are small and insignificant (also compared to Table 1), lending further support to the identifying assumptions. The results suggest that the estimated changes in mothers' labor supply in our baseline analysis are not caused by any other policy changes at the cutoff or by differences in unobserved characteristics of mothers above and below the cutoff.

Second, we further test the validity of our results by using placebo cutoffs up to 24 months prior and 24 months post the actual cutoff, at a four-month frequency. Figure A8 plots the RD-estimates for the placebo test for grandmothers' and mothers' labor supply in Panels (a) and (b), respectively. We do not find discontinuities at the placebo cutoffs for grandmothers' total monthly hours worked (Panel (a) of Figure A8) with one exception. There is a small increase in grandmothers' total hours worked at the January 1949 placebo cutoff, which is likely due to a reform of the state pension age (AOW age), which increased by one month for people born since November 1949. We conduct an additional robustness exercise with respect to this aspect below. Panel (b) of Figure A8 displays the RD estimates on mothers' monthly hours worked. We find no significant discontinuities at the placebo cutoffs for maternal labor supply. Figure A9 further illustrates graphically the reform impact on grandmothers' and mothers' total monthly hours worked (in Panels (a) and (b), respectively) around the January cutoffs in 1948, 1951, and 1952.

4.3.6 Robustness: Alternative Methods

In this section, we employ alternative/augmented methods to further test the robustness of our results, in particular with respect to seasonality effects, as investigated above, and with respect to the fact that the running variable is discrete, since we have data on the birth month rather than the exact birth date.

To address the two concerns of seasonality and discreteness of the running variable jointly, we estimate a Difference-in-Differences model making use of time variation (birth of grandmother after versus before the reform cutoff of January 1950) and variation between the treatment group

[21] We define exiting the labor force when labor earnings are no longer the main income source.

(mothers with young children between 4 and 12) and a control group. We use families with the youngest child between the ages of 13 and 18 as a control group, as the older children need little or no childcare. Using families with children not in the target age group helps to reveal and control for any potential mechanical correlation between birth month and labor supply decisions. Moreover, in this approach, we control for grandmothers' month-by-year of birth fixed effects and allow for differential linear cohort trends between the treated and control groups. Table A11 shows that being treated, i.e., having a (grand)mother born after 1950 and having a youngest child aged 4-12, leads to a significant reduction in mothers' labor supply. For a sample window of 8 months around the cutoff (i.e., the optimal bandwidth in our RD approach), we find a similar effect of a reduction of 4.8 hours per month (compared to 4.9 hours with the RD approach), while the effect size is even larger for smaller windows (and significant at the 1 percent level). To lend support to the parallel trend assumption, we run a placebo DID. Panel B of Table A11 shows that pre-trends are similar and not statistically different between the two groups.

Furthermore, to address the concern of seasonality within our RD framework, taking advantage of the discrete change in eligibility at the January cutoff, we also augment our RD design with a DID model using two different control/placebo groups. First, and similar in spirit to the DID approach above, we use families with the youngest child between the ages of 13 and 18 as a comparison ("placebo") group to control for any potential mechanical correlation between birth month and labor supply decisions. The RD-DID estimates in Columns (1)-(3) of Table A12 show that effects on grandmothers' labor supply in the main sample (youngest child aged 4-12) and in the placebo group (youngest child aged 13-18) are very similar (and not statistically different from each other). Columns (4)-(6) of Table A12 display the fuzzy RD-DID estimates on mothers' labor supply, according to which a one-hour increase in grandmothers' monthly work hours reduces mothers' monthly work hours by 0.67 hours, a value which is very similar to our baseline RD estimates in Table 1 (0.63 hours).

Second, we explore the RD-DID specification by using as a control/placebo group families with grandmothers from other cohorts (also born around the January cutoff, but in the years 1948, 1949, 1951, and 1952). We thereby estimate the difference between the discontinuity of labor supply responses in families with grandmothers born around the reform cutoff of January 1950 and any potential discontinuity in families with grandmothers born around January of the other years. This exercise should address any potential seasonality bias and is in the same spirit as the placebo analysis in Figure A9. In Table A13, we report reduced-form effects and not 2SLS estimates, because there are no first-stage effects for the control group (i.e., at the placebo cutoffs). The RD-DID coefficient suggests the reform leads to an increase of grandmothers' labor supply by 5.1 hours (significant at the one-percent level) and a reduction of mothers' total working hours by 4.6 hours (significant at the five-percent level). These RD-DID estimates are again very similar to our baseline RD estimates (see Table 1).

Lastly, to address concerns with respect to the discrete running variable within the RD framework, we adopt the local randomization approach to RD analysis (Cattaneo et al., 2024). This approach leads to treatment and control groups very close to the cutoff that can be treated as if they were randomly assigned. More specifically, the specified window is $[-1, 0]$.[22] Table A14 compares the estimates based on the local randomization estimation with our baseline RD estimates. In terms of the direct effect on grandmothers, the local randomization approach leads to a similar effect as our RD analysis, namely an increase of 8.2 hours worked by grandmothers (compared to 6.8 hours), both significant at the one-percent level and the difference is not significant. Also, the reduced-form effect on mothers and the 2SLS/LATE estimates on mothers' labor supply are very similar (and not significantly different) when comparing the two methods. According to the local randomization approach, mothers with young children decrease their labor supply by 4.9 hours in response to the reform (identical to our RDD estimate), and the 2SLS/LATE suggest that a one-hour increase of grandmothers' labor supply leads to a reduction of 0.6 hours (compared to the RDD results of -0.63 hours).

4.3.7 Fertility Responses

In our main analysis, we investigate the importance of spillover effects of a pension reform on the labor supply of mothers (adult daughters) with young children. However, given the importance of grandparental childcare, the reform may also affect the fertility decisions of adult daughters. The limited evidence in the literature on this effect is mixed in terms of sign and size (e.g. Battistin et al., 2014; Eibich and Siedler, 2020).

Analyzing whether there are reform effects on adult daughters' fertility decisions is not only interesting, but also important, because fertility responses may indicate problems of sample selection (for example, mothers with (grand)mothers born since 1950 may be less likely to have children or may time the births differently). We test this by estimating the reform impact on the fertility outcomes of adult daughters. We start with the same sample of (native Dutch) grandmothers born around January 1950 as in our main analysis, but now use all adult daughters of these grandmothers, including those without children. Table A16 shows the effect on a number of different outcomes. In terms of total fertility, we look at the probability of ever having a child, the total number of children, and the probability of having at least two children. In terms of fertility timing, we examine age at first birth, age at last birth, the average age gap between children, the average age gap between children born after grandmothers turned age 55, and the probability of having their first child after

[22] In Table A15 we test for balancedness of characteristics based on this approach (Cattaneo et al., 2024) and show that pre-determined characteristics of grandmothers and mothers are indeed balanced between the treatment and control groups very close to the cutoff, within the specified window of $[-1, 0]$.

grandmothers turned age 55. Moreover, to investigate possible short-, medium-, and longer-run effects on fertility, we also examine effects on the number of births up to 3, 6, and 9 years post reform and the probability of having a birth up to 3, 6, and 9 years post reform. None of these fertility measures is affected by the reform.[23] Lastly, the results in Tables A1 and A4 indicate that the probability of having the youngest child(ren) in different age groups is not affected by the reform, which also implies no fertility responses at these margins.

4.4 Mechanisms and Heterogeneity

We hypothesize that the pension reform affects mothers' labor supply through changes in the time grandmothers can devote to childcare — a "time transfer channel". To test this hypothesis, we examine the reform's impact on other family members and explore heterogeneity by childcare needs (indicated by age and existence of the (youngest) child) and grandmothers' availability for childcare (proxied by grandfather's health, geographic proximity, and family composition).[24]

4.4.1 Spillover Effects of Grandfathers

While both grandparents can provide childcare, previous studies show that grandmothers are substantially more likely to be engaged in childcare activities (Jappens and Van Bavel, 2012; Janta, 2014). Therefore, we expect grandfathers to have smaller impacts on daughters' labor supply if the mechanism behind our results is indeed the "time transfer" channel. If instead "monetary transfers" (or "reminder/role model" effects) are the main underlying factor, we expect to see a similar impact by grandfathers, as they could provide money equally well as grandmothers.

Table 2 displays the reform effects on grandmothers and grandfathers and the respective spillover effects on their adult daughters with young children. In Column (1), we present baseline estimates on treated grandmothers (in Panel A) and on treated grandfathers (in Panel B), which are irrespective of whether their partner/spouse is also treated. Treatment status is based on the birthdate, i.e., Panel A focuses on families with grandmothers born around the January 1950 cutoff (compare our baseline analysis), while Panel B studies families with grandfathers born around the same cutoff. In Column (2), we show estimates for when only the grandmother (grandfather) is treated (in Panels

[23] Ilciukas (2023) studies the impact of the same pension reform on fertility outcomes and finds a reduction in fertility among women with reform-affected (grand)mothers. However, his sample is different from ours. Ilciukas (2023) uses a restricted sample — adult daughters of grandmothers born around the cutoff who were married or cohabiting and had no children before the reform. We do not impose restrictions based on marital status or fertility prior to the reform. When including adult daughters with children before the reform (the same general sample as our analysis), Ilciukas (2023) also finds no fertility responses.

[24] We rule out the 'monetary transfer channel' by showing that grandmothers' monthly gross income and household gross income are not affected by the reform. Additionally, we argue that alternative channels, such as being reminded of pension planning or grandmothers serving as role models by working longer, are unlikely, since effects on mothers' labor supply would go in the opposite direction of what we find. See Section 4.3.3 for a detailed discussion.

A and B, respectively). For this analysis, we exclude families in which the grandfathers (Panel A) or grandmothers (Panel B) were born since January 1950. In Column (3), we present estimates on grandmothers (Panel A) and grandfathers (Panel B) when both are treated. We thus compare families in which both grandparents were born since January 1950, with those where both were born before January 1950.

Table 2 Column (1) shows that grandfathers' labor supply response to the pension reform is substantially larger than grandmothers' due to their stronger attachment to the labor force (comparing Panel A and B). In terms of spillover effect on adult daughters, however, we find that the per-hour effect of grandmothers' labor supply (as opposed to grandfathers') on their daughters is an order of magnitude larger. While a grandmother working one hour more per month induces her daughter with young children to work close to 40 minutes less (-0.6), this figure is less than 10 minutes (-0.1) for the grandfather-daughter pair (and the difference is significant at 10%).

Column (2) of Table 2 shows the impact on families where only the grandmother (Panel A) or only the grandfather (Panel B) is treated. In this case, the conclusion in terms of spillover effects on adult daughters is even more extreme. For these families, there is only a significant spillover effect for grandmothers (with a coefficient of -0.7), but no effect for grandfathers, for whom the coefficient is -0.026 and statistically insignificant (the p-value of the difference is 0.106).

Column (3) suggests why we find any spillover effect of grandfathers' labor supply in the baseline estimates. These effects appear to be driven by families in which both grandparents are treated, since we only find spillover effects of grandfathers' labor supply on adult daughters for those families, albeit still substantially smaller for grandfathers compared to grandmothers in these families. In particular, a one-hour increase in work hours by grandmothers leads to a decrease in adult daughters' labor supply of 0.7 hours, while the same increase in grandfathers' work hours leads to an effect of only 0.18 hours (the difference is significant at 5%).

These findings lend further support to the "time transfer" channel. We show that grandfathers play a substantially smaller role than grandmothers in terms of providing childcare. Moreover, the grandfather-daughter spillover effects appear to be driven by families in which both grandparents are treated.

4.4.2 Effects on Other Family Members and Within-Household Inequality

Descriptive evidence suggests that the majority of grandparents providing child care are maternal grandparents (more than 60%) (see Section 2.1 and, for example, Danielsbacka et al. (2019)). Thus, reform effects due to changes in child care availability should imply larger effects for maternal grandparents and thus smaller effects on daughters-in-law.

Table 3 compares the impact of grandmothers' labor supply on their daughters, sons-in-law, sons, and daughters-in-law, whose youngest child is aged between 4 and 12. Table A17 reports the pair-wise p-values testing the hypothesis that the coefficients are equal across subgroups. While the direct reform impact on grandmothers is very similar, we only find labor supply spillovers on daughters and their husbands, but not on daughters-in-law/sons. Moreover, the impact on daughters-in-law is less than a third of the effect on daughters, with a coefficient of -0.19 as opposed to -0.63 (the difference is, however, not significant at conventional levels). The evidence is thus largely consistent with maternal grandmothers playing the central role in providing childcare and thus with the time/childcare channel.

As for the core families with the youngest child aged 4 to 12, we have shown that in response to the increases in grandmothers' labor supply, mothers work less. Table 3 shows that their husbands (the sons-in-law) work more, most likely to compensate for the loss in their wives' labor earnings.[25] Indeed, we find that there is no impact on overall household income. The finding that mothers of young children reduce their labor supply, while their husbands work more, has important implications for gender inequality within the household as well as in society overall.

4.4.3 Effects by Age of the Youngest Child

To provide further direct evidence on the relevance of the "time transfer channel", we compare reform effects on adult daughters with different childcare needs. More specifically, we split the sample of adult daughters into those *with* childcare needs (with children younger than 13) and those *without* childcare needs (children older than 13 or no children). In case of the "time transfer channel" being the relevant mechanism, we would expect to find reform impacts only for mothers with childcare needs, while in case of a "monetary transfer channel" (or reminder/role model effects), one would expect effects also for mothers with older/no children.

In Table 4, we classify mothers with their youngest child below age 13 into the group *with* childcare needs (Columns (1) to (3)) and the ones with their youngest child above age 13 or without children into the group with *little or no* childcare needs (Columns (4) and (5)). Moreover, we further divide mothers with the youngest child between ages 1 to 12 into three different categories: 1 to 3, 4 to 7, and 8 to 12. Children aged 1 to 3 require the most care. Starting at age 4, children attend (pre-)school for around 6 to 7 hours per day (see Appendix A.4). A mother working part-time thus requires only a few additional hours of help, grandmothers may fill this gap. The hours attended in school increase with age, and at some point, children are able to spend some time unsupervised. Thus, we also show results separately for children aged 8 to 12. Table A4 shows that the probability of having the youngest child(ren) in different age groups is not affected by the reform (also see

[25] As De Nardi et al. (2021) show for the Netherlands and the U.S., the presence of spousal earnings reduces the variability of household income and provides an important insurance mechanism. See Section 2.1 for supporting evidence by the LISS panel that, in the majority of cases, it is the mother who reduces her work hours to care for young children instead of the father.

Section 4.3.7). Table A18 reports the pair-wise p-values testing the hypothesis that the coefficients are equal across subgroups.

Table 4 presents the estimates of the reform effect on grandmothers' and mothers' labor supply and the spillover effects on mothers with children of different ages (or without children). Figure 2 displays the corresponding scatter plots of mothers' total working hours by age of the youngest child. In terms of first-stage/direct effects, we find that the reform led to a significant and similar increase in grandmothers' work hours for all subgroups (except for the group without grandchildren, where the first-stage effect is somewhat smaller). As hypothesized under the premise of a time transfer channel, only mothers with childcare needs reduce their labor supply due to the reform (see Columns (1)-(3)). Instead, there is no effect on adult daughters without childcare needs (see Columns (4) and (5)). Table A18 shows that the effects on mothers' labor supply are significantly different between mothers with the strongest childcare need (youngest child aged 4-7) and the groups without childcare needs. These results further support our interpretation that the changes in mothers' labor supply are indeed related to the time availability of grandmothers and their childcare responsibilities.

Among mothers with childcare needs, we find the strongest effects on the labor supply of mothers with a primary school-aged child between 4 and 7. For this group, the reform leads to a reduction of nearly 7 work hours (significant at the one percent level). The corresponding IV estimate indicates that a one-hour increase in grandmothers' monthly work hours induces a decline in mothers' monthly hours worked of 0.8 hours, i.e., close to 50 minutes (significant at the 5-percent level). The reform also leads to a significant decrease in work hours of mothers with a youngest child aged 1 to 3, but the effect is significantly smaller, consistent with families with children below 4 being less likely to rely solely on grandparental childcare (as discussed in Section 2.1).[26] These findings are robust to including mothers with a youngest child under the age of one. Furthermore, the effect on mothers with a youngest child aged 8 to 12 is also smaller and not statistically significant.

In principle, the different impacts by age groups may also be related to the mothers having different characteristics depending on the age of their youngest child when grandmothers are 60 to 64 (see Table A19). In particular, mothers with a youngest child aged 1 to 3 are, on average, younger, more educated, marry later, and have their first child later than mothers whose youngest child is aged 4 to 7, consistent with a stronger attachment to the labor force. However, Table A19 shows that these characteristics develop monotonically across the three age groups 1-3, 4-7, 8-12 (e.g., the age at first birth and the education level are highest for mothers of the 1-3 group and lowest for mothers of the 8-12 group). Instead, the reform effects we find are non-monotonic

[26] In addition to the reasons discussed above for different effects depending on age of the youngest child, daycare schedules tend to be more flexible than school schedules, making it easier for mothers of very young children to balance family and work without the help of grandparents.

(largest effects for the middle group), suggesting that the heterogeneity in reform effects is unlikely to be only due to differences in mothers' characteristics.[27]

To shed light on the dynamic labor supply adjustments of mothers in response to the reform and how they are linked to changes in grandmothers' labor supply, we display the reform impacts for each age of the grandmother (between ages 58 and 64). The effects on mothers' and grandmothers' labor supply are depicted as dots and squares, respectively, with 95 percent confidence intervals, and we present reduced-form estimates for mothers (and grandmothers) with the youngest child aged 4-7 on their working hours and the probability of employment in Figure A10. We find that the labor supply of mothers and grandmothers respond at similar times and in opposite directions. As expected, given Table 4 (which shows the strongest effects for mothers with youngest children aged 4-7 and weaker or no effects for other age groups), the dynamic patterns for mothers (and grandmothers) with children in other age groups are similar, but less precisely estimated (results available upon request).

4.4.4 Heterogeneity by Grandmothers' Time Availability

In this section, we test for differences in effects based on grandmothers' availability for childcare, as proxied by the following three indicators: the care needs of their partner, proximity to their daughter, and the number of young grandchildren. Since grandmothers with healthier partners, who live closer to their daughters, and have fewer young grandchildren are more likely to provide regular childcare in the absence of the reform, the "time transfer channel" predicts that delaying their retirement will have a larger impact on their daughters' labor supply.

Health status of partner Table 5, Columns (1) and (2), show results for grandmothers whose partner (mostly the grandfather) is in poor health or healthy, respectively. We define the grandfather to be healthy if he has not claimed any disability insurance before age 50. As predicted, we only find significant effects on maternal labor supply for grandmothers who have a healthy partner. Differences in effects between the two groups are, however, not significant. Also, it should be noted that the first-stage effect on grandmothers with sick partners is very noisy and insignificant, i.e., these grandmothers do not change their labor supply in response to the reform in the first place.

Proximity of grandmothers Columns (3) and (4) of Table 5 show results for grandmothers living further away from (closer to) their adult daughter, respectively, defined by whether prior to the reform they lived in a different (the same) neighborhood (see CBS definition of *buurt*). The impact on grandmothers' labor supply is similar in magnitude and significant for both groups. In

[27] We find suggestive evidence that only mothers with some college decrease their labor supply in Table A20. See Appendix Section C.2 for more details.

terms of reform effects on mothers' work hours, the coefficient for mothers living nearby is twice as large (albeit not statistically different). Similarly, the reduction in the probability of employment is more than three times larger for this group, and the difference in effects is significant at ten percent. These findings lend support to the time transfer channel.

Family composition Columns (5) and (6) of Table 5 explore the dimension of competition for grandmothers' time. For this purpose, we compare (among grandmothers with at least one daughter who has a child) grandmothers with exactly one maternal grandchild aged between 4 and 7 with the remainder of grandmothers.[28] Indeed, we find important effects on mothers for the group of grandmothers with only one maternal grandchild in the relevant age range, while the impact on the other group is small and insignificant. More specifically, the reform causes the former group of mothers to work about 5.7 hours less per month and reduces their probability of being employed by 3 p.p.. However, differences between the two groups are not statistically significant.

5 Reform Effects on Grandchildren's Educational Achievement

5.1 Main Results and Heterogeneity

We have shown that the pension reform and the resulting increase in grandmothers' labor supply led to a decrease in the labor supply of mothers with young children. This suggests a substitution effect away from grandparental care to maternal care, raising the question of whether this change affects children's educational performance.

For this purpose, we make use of data on children's performance on the Cito test, which is a high-stakes test taken at the end of primary school to place children into different tracks in secondary school (vocational, technical, academic). The performance on the test and the resulting track assignment have important long-run implications in terms of the likelihood of enrolling in college (which requires completing the academic track) and in terms of earnings and family formation outcomes (see, e.g., Kaufmann et al. (2021) on the marriage market effects of university education). We explore the reform impact on educational achievement in terms of the following outcomes: the number of correct answers on the test overall and in the math and verbal components, and the likelihood of receiving the recommendation for the highest (academic) track in secondary school. We examine the implications for children's educational performance a few years after having been exposed to the change in childcare mode. Since we find effects on mothers' labor supply that depend on the age of the youngest child, we investigate reform spillovers on the performance of the

[28] We focus on maternal grandmothers since they are most relevant in terms of taking care of grandchildren (see Sections 2.1 and 4.4.2). Moreover, we focus on mothers with the youngest child aged 4 to 7, since we find the strongest effects for this subsample (see Section 4.4.3).

youngest child based on its age when its grandmother was treated and aged 60 (while the Cito is administered when the child is age 11-12). Table A21 supports the smoothness condition in that covariates for the Cito sample are balanced across the cutoff.

Pooled Effects Table 6 and Figure A11 present the estimated reform impacts on children's educational performance. Panel A of Table 6 and Figure A11 (a) show the results for children who were the youngest child aged 4 to 12 when their grandmother was 60. We find that the reform has positive effects on children's Cito performance. In particular, for children with grandmothers affected by the reform, the number of correct answers on the test is 14.6% of a standard deviation higher overall and by 10% of a standard deviation higher on the verbal part of the test (significant at the 5 and 10 percent level, respectively).

Effects By Grandmothers' Time Availability In Table A22, we show the heterogeneous effects on children's educational performance when splitting the sample by whether the grandmother's partner is healthy, whether the grandmother lives in the same municipality, and by family composition (as in Section 4.4.4). We find statistically significant improvements in test scores for families with only one maternal grandchild in the relevant age range. The difference in reform effects in terms of the number of correct answers in math and overall is statistically significant at the 1 and 10 percent level, respectively. Although we do not find statistically significant differences between other subgroups, partly due to sizable differences in sample size between them, the estimated coefficients tend to be larger for families with healthy grandfathers and for families with grandmothers who live nearby. The findings are largely consistent with mothers' labor supply responses in Section 4.4.4.

Effects By Age Groups We further split the sample by age groups, since the effects of grandmothers' labor supply on mothers' labor supply depended in important ways on the age of the youngest child. In particular, the increase in grandmothers' hours worked -and the resulting decrease in their availability for childcare- led to a strong decrease in mothers' hours worked, but only for children aged 4 to 7 (see Table 4). We therefore split the sample into children aged 4 to 7 years and those aged 8 to 12 years.[29]

Panel B of Table 6 shows that the results in Panel A hide a substantial amount of heterogeneity. In particular, we find strong positive effects on children who were between 4 and 7 years old, i.e., on those children whose mothers' labor supply decreased in response to the reform. Their overall number of correct answers in the Cito test increases by 31% of a standard deviation, and the number

[29] Panel B of Table A4 shows that there is no selection into taking the Cito test for the different age groups. We exclude children under the age of 4, because they are either too young to have taken the test or because their Cito test did not take place due to the COVID-19 pandemic.

of correct answers on the verbal and the mathematical components increases by 25 and 27% of a standard deviation, respectively (all significant at the one percent level). Moreover, we find that the youngest children aged 4 to 7 in families affected by the reform are 6.4 p.p. more likely to get a recommendation for the academic track in secondary school. For a detailed discussion and interpretation of the effect sizes, see Section 5.3. The bottom part of Panel B in Table 6 presents reform impacts on children aged 8 to 12. For this age group, the estimates are small and insignificant (and significantly smaller than for the age group 4-7, except for math). The patterns by age groups are also confirmed by the scatter plots in Figures A11 (b) and (c).

It is worth noting that mothers with children in different age groups have different characteristics (such as education level) and different counterfactual care modes (discussed in Section 5.2). Both factors can contribute to the different impacts on children's educational outcomes.

Effects By Gender Our findings by gender paint a more nuanced picture for the latter age group. In Table A23, we split the sample by gender. Figures A12 and A13 display the scatter plots by gender for the two age groups (age 4-7 and 8-12, respectively). While the reform effects are strongly positive for both girls and boys between ages 4 and 7, girls improve more in math, while boys improve more in verbal test scores.

For children aged 8-12, we find important negative spillover effects in terms of educational performance, but only for boys. The number of correct answers on the verbal component decreases by 23% of a standard deviation for boys in this age group, and their likelihood of receiving a recommendation for the highest (academic) track decreases by 9.1 p.p. (significant at the 5 percent level and in this case the difference by gender is significant). This has critical implications for their future life outcomes because of the decreased chance of enrolling in college. One potential explanation might be that grandmaternal supervision time (which decreases in response to the reform) is substituted for (at least in part) by unsupervised time for these older children (compare this to Aizer (2004), who finds that a lack of adult supervision after school has important negative consequences for human capital development). See a more detailed discussion in Section 5.3.

5.2 Counterfactual Care

Formal Care and Substitution Patterns While we have provided strong evidence that there has been a substitution away from grandmaternal care to maternal care for children aged 4 to 7, it is less clear what has happened in the case of children aged between 8 and 12. For these children, on average, mothers' labor supply did not decrease in response to an increase in grandmothers' labor supply. However, we find a strong negative reform effect on these children, especially boys, suggesting that there was a change in the mode of supervision in response to the decrease in grandmaternal time. This raises the question as to whether substitution has taken place towards formal after-school care. We, therefore, supplement our analysis with data on whether parents applied for childcare subsidies, for which type (daycare or after-school care), and for how many hours.

Table A24 presents the estimated reform impact on the probability of taking up and the hours of daycare and of after-school care. For families with a youngest child aged 4 to 7, we do not find a significant change in the daycare or after-school care arrangement. Thus, grandmaternal care is substituted for by maternal care in that mothers reduce their labor supply in response to the decrease in grandmothers' availability.

For children aged 8-12, we find that the pension reform leads to a 3.2 p.p. increase in the probability of using subsidies for after-school care. In addition, these families also increase the number of hours of after-school care by 15 hours per month, which is significant at the ten percent level. Thus, for children of this age group, there appears to have been some substitution away from grandmother supervision towards after-school care. It is not clear whether this increase fully makes up for the reduction in time availability of grandmothers, and it is possible that for the remaining time, children aged 8 and older are at home unsupervised for a few hours in the afternoon, as hypothesized above.

Grandmother-Mother Care Qualities Our results underscore the critical importance of high-quality childcare. For children between 4 and 7, we show that substituting grandparental care with maternal care improves test scores. Importantly, for this age group, we do not find a significant change in formal care. Maternal care in our setting appears to provide better quality, possibly linked to higher education of mothers, grandparents overindulging grandchildren, and/or placing less weight on fostering cognitive skills. To further study the importance of care quality, we show the heterogeneous impacts on children between 4 and 7 by different grandmother-mother care quality pairs in Table A25.

We show the impact on the Cito test separately for families with four different grandmother-mother care quality pairs: high-high, high-low, low-high, and low-low.[30] We find that the positive effects of switching from grandmaternal to maternal care appear to be primarily driven by families switching from low-quality grandmaternal care to high-quality maternal care, since we only find significant positive effects for this group of families. This finding is consistent with the previous literature (Bernal and Keane, 2010, 2011) showing that, for children of single mothers, switching

[30] We proxy for the mother's care quality with her educational attainment. Care quality is classified as high when the mother attended at least some college, including academic colleges (WO) and colleges of applied sciences (HBO). Since the education variable is often missing for grandmothers, we proxy for the grandmother's care quality with her socio-economic status, which is classified as high if her average earnings between ages 53-56 (predetermined) are above the median.

from maternal care to informal care (most often grandparents) has adverse effects on children's test scores. This result is also consistent with the recent literature on the importance of high-quality elementary education. For example, Cascio (2023) reports that pre-kindergarten attendance raises test scores of low-income 4-year-olds in U.S. states with universal pre-K programs by 1.16 standard deviations.

Grandmother-Formal Care Qualities For children aged 8-12, we show that there is a substitution from grandmaternal to formal care; however, we cannot exclude the possibility that the children previously cared for by their grandmothers spend some time unsupervised at home after the reform. The negative effects we find are supported by the findings of Aizer (2004), who shows that a lack of adult supervision after school has important negative consequences for children's human capital development. Moreover, the fact that this has a particularly negative effect on boys is consistent with the findings of Bertrand and Pan (2013), who show that boys tend to underperform relative to girls in several non-cognitive areas. Consequently, boys' educational outcomes are more sensitive to negative environments or shocks, including poor school quality or living in single-parent households, etc. (e.g., Bertrand and Pan, 2013; Autor et al., 2016). Additionally, the quality of childcare in the Netherlands during our study period may explain these results. The expenditures on and the teacher-child ratio in after-school care in the Netherlands were below the OECD average and comparable to levels in the U.S. (OECD, 2017).

5.3 Discussion and Comparison to the Literature

To sum up, interpreting the results in terms of the educational performance of children of different ages points to the following conclusions: children who are aged 4 to 7 when their grandmothers are affected by the reform benefit from the fact that the mother spends more time with them, as a substitute for grandmothers' care. For children aged 8 to 12, we find no effects overall, but a negative effect on the high-track recommendation of boys, suggesting that the substitution away from grandmothers' care towards either after-school care or no adult supervision (for a few hours after school) during the years prior to the high-stakes test may have negative effects on the performance of boys. How do our results compare to the literature (in terms of sign and magnitude)?

As discussed in the introduction, the literature analyzing the effect of grandparental care as opposed to parental/maternal care (as in the case of our 4 to 7-year-olds) is extremely limited. Two papers we are aware of are Zhang et al. (2021) and Wang and Bansak (2024), which compare the effect of grandparents versus parents as primary caregivers in China. Both papers find that compared with parental care, grandparental care delays the achievement of children between ages 1 and 5 and has adverse impacts on the test scores of children aged 10 and above, respectively, consistent with our findings. In terms of magnitudes, Wang and Bansak (2024) find that -using 2SLS estimators and FE estimators- grandparental instead of parental care has negative effects on children's test scores of 33% and 27% of a std. dev. Their estimates are thus very similar to our estimates.[31] When interpreting the positive effect of a reduction in maternal labor supply on child outcomes, it is important to note that in our context household income remains unchanged due to an accompanying increase in fathers' labor supply. In contrast, in the literature maternal labor supply changes often lead to accompanying changes in household income (e.g., Nicoletti et al. (2023)).

To provide further comparisons in terms of the magnitude of effects of maternal care, we briefly discuss two other papers, where the counterfactual childcare mode is, however, formal childcare. For example, Fort et al. (2020) find that one month less of maternal care decreases IQ scores by 4.7 percent of a std. dev. Extrapolating, this would imply that six additional months of maternal care increase IQ scores by nearly 30% of a std. dev., aligning with our findings of a similar increase in test scores (albeit our counterfactual is grandmaternal care). Compared to Fort et al. (2020), in our case, children are exposed to additional hours of maternal care for a much longer period of time. Similarly, Baker et al. (2008) analyze a Canadian universal childcare reform that increased the use of formal childcare and reduced maternal care time. They show that eligibility for the subsidized childcare plan leads to a decline in motor and social development scores by more than 10% of a std. dev., equating to a treatment-on-the-treated effect of 75 to 133% of a std. dev.

The literature comparing grandparental to formal care (as is the case for our 8 to 12-year-olds) is extremely limited. A notable exception is Del Boca et al. (2018), who analyze the effect of early childcare provision by grandparents and by formal childcare centers on child cognitive outcomes in the UK. They find that, on average, children looked after by their grandparents at the age of 18 months have better vocabulary skills at later ages. This effect is driven by higher socioeconomic status families. The authors argue that it may be related to the more stable relationships and one-on-one interactions with an adult provided by the grandparents. However, they find a negative association of grandparental care with other cognitive outcomes driven by less privileged families. This is also consistent with the findings of a recent paper by Moroni et al. (2023), which shows that replacing low-quality care provided by unemployed grandfathers with formal childcare between the ages of 0 and 4 improves the test scores of grandchildren at later ages. While we investigate the effects of substituting grandparental care for formal care (or no adult supervision) for older children, the effects we find are mostly driven by more privileged families with working grandmothers and mothers. For this group, our results are consistent with the results of Del Boca et al. (2018). Interestingly, also in our case, the main positive effects of grandparents (or negative effects of a

[31] Zhang et al. (2021) investigate the effect of grandparents instead of parents as primary caregivers in China for children aged 1 to 5. However, it is difficult to compare their estimates to ours, since they employ Hazard models to explain by what age very young children are able to walk or are toilet-trained.

decrease of grandmaternal care) are on verbal skills.

From a policy perspective, our findings strongly underline the importance of high-quality childcare (compare, for example, Cascio, 2014; Andrew et al., 2024). Such high-quality childcare can be made possible and shared by both parents through more generous parental leave policies or by improving the quality of formal care provision.

6 Reform Effects on Child Penalty

Reducing gender inequality in the labor market is high on the policy agenda. The existing literature has shown that children have a large and persistent impact on the gender gap in labor market outcomes (Kleven et al. (2019a,b, 2024)). The Netherlands faces a similar situation to that of the U.S. and other developed countries. In particular, the monthly gender wage gap in 2014 was 41.8% (women earn EUR 580 for every EUR 1,000 earned by a man), and thereby the Netherlands is among the OECD countries with a large gender gap (OECD, 2019a).

In this section, we connect our empirical findings to this debate and address the question of whether grandmothers' retirement decisions affect the gender gap and child penalties. Building on Kleven et al. (2019a), we first estimate the causal long-run reform impact by comparing the effect of having a child on the labor market trajectories of mothers with treated grandmothers to mothers with untreated grandmothers. Second, we compare the differences in child penalties (i.e., relative loss women experience compared to men at a given year due to children) between mothers with treated and untreated grandmothers (for details on the estimation, see Appendix C.3). While the literature on gender gaps and child penalty shows whether and to what extent women's labor market outcomes converge to men's outcomes, we are interested in whether the pension reform causally leads to a *slower* convergence due to its spillovers on maternal labor supply.

The child penalty results are based on a sample of parents for whom we observe labor market outcomes in all years from four years prior to eight years after the birth of their child. This leads to a balanced sample of women and men whose first child was born between the years 2010 and 2014.[32] We compare women and men, with treated (grand)mothers born between January 1950 and December 1951, with those with control (grand)mothers born between January 1948 and December 1949. This yields a sample of around 12,000 men and 12,000 women for each treatment group, whom we can follow over 13 years.[33] Since we only found negative effects on daughters' labor

[32] We exclude teenage births by dropping observations with first birth before age 20, and exclude late entry into parenthood after the age of 40. Note that using a different event window changes the composition of birth cohorts, but leads qualitatively to the same findings.

[33] For the child penalty analysis, we use an event-study design rather than the RD design. Here, we take a two-year window around 1950 for sample size considerations. Our findings are robust to using only families with grandmothers who were born closer to January 1950, leading to a smaller sample size. Results are available upon request.

supply (as opposed to daughters-in-law), we focus on the middle generation that has a direct relation to the treatment generation, i.e., (adult) daughters and sons.

Panel (a) of Figure 3 compares the development of mothers' total hours worked around the birth of their first child (marked as event time zero) for women with treated (blue dots) and untreated grandmothers (black triangles). The 95% confidence intervals are shown by the shaded area. We find that women experience a sharp drop in their monthly hours worked within the first year after birth, which amounts to a 20% drop relative to their pre-birth work hours.

However, women with treated grandmothers recover more slowly than women with untreated grandmothers. This difference becomes significant at the five percent level one year after birth, and the gap increases substantially in the following years. The pension reform not only led to a short-run decline in maternal labor supply but also had dynamic effects, as it led to a slower recovery of the working hours of mothers with young children.

Panel (b) of Figure 3 compares the estimated child penalties for women (blue dots) relative to men (black triangles) separately for treated (left panel) and control groups (right panel). The long-run relative child penalty faced by women eight years after birth is reported at the bottom of each panel. We observe that the gender gap in total hours worked starts to flatten and remains at 30% eight years after birth for the control group (i.e., with grandmothers not treated by the reform). In contrast, for the treatment group, the gap continues to widen, and the long-run gap in total hours worked reaches 36% eight years after the birth of the first child. Similarly, we also find that the reform leads to a slower recovery of monthly labor earnings. Figure A14 shows that the dynamic treatment effects and the relative child penalty in terms of earnings are indeed similar to the ones for hours.[34]

In general, countries with more generous family policies and child care access have lower child penalties for mothers (see, e.g., Olivetti and Petrongolo, 2017; Kleven et al., 2019a). However, evaluations of family policies provide mixed evidence. Kleven et al. (2024) find no evidence of effects of child care expansions in Austria. Based on a more recent reform, Krapf et al. (2020) show that child care availability increases women's earnings in the first year after childbirth, implying a decrease in the child penalty of 4.5 p.p. or 6.3%. Andresen and Nix (2022) find that a Norwegian reform expanding child care availability reduced the child penalty by 23% during the years of early child care use. In comparison, we find that the reform led to a 6.1 p.p. (16%) decrease in the hours penalty and a 7.1 p.p. (19%) decrease in the earnings penalty 8 years after the birth of the first child. It is, however, important to note that while the literature has focused (mostly) on formal care provision, we analyze the effect of changes in informal (grandparental) childcare availability.

Overall, we find that the changes in grandmothers' labor supply decisions not only affect maternal

[34] In this analysis, we use grandmothers' sons in the relevant age group to construct men's labor market outcomes. However, if we were to use the women's husbands (i.e. the grandmothers' sons-in-law) to construct the gender gap, the difference in the gap would be even larger because sons-in-law increase their hours worked (see Column 2 of Table 3), suggesting that the reform had an even stronger impact on the gender gap within the household.

labor supply in the short run, but also have dynamic spillover effects in the long run. The decrease in time availability of grandmothers to provide childcare leads to a significant reduction in mothers' long-run labor supply and to a substantial increase in the child penalty and in the gender gap within households and in society overall.

7 Conclusion

This paper provides the first evidence of spillover effects of old age pension across three generations. Specifically, we study the impact on the labor supply of family members of the middle generation and show the resulting effects on grandchildren's academic performance by exploiting a cohort-based pension reform in the Netherlands. We show that a one-hour increase in grandmothers' hours worked causes their adult daughters with young school-age children to work 40 minutes less. In contrast, the change in grandmothers' labor supply does not affect adult daughters without children or older children, nor sons or daughters-in-law. Moreover, a change in grandfathers' labor supply has only small spillover effects. Combined with a heterogeneity analysis, our evidence indicates the importance of time transfers provided by grandparents, and in particular by maternal grandmothers.

In addition to the impacts on the middle generation, we also investigate reform effects on grandchildren. We find a sizeable positive impact on the educational performance of children who were aged 4 to 7 when their grandmothers were most affected by the pension reform. These children experienced an increase in grandmothers' labor supply and a resulting reduction in their mothers' labor supply, suggesting a substitution away from grandparental to maternal care. For older children aged 8 to 12 who experienced a substitution away from grandparental to formal (after-school) care, we find no overall effects, but some adverse effects on boys.

Our findings have important policy implications. First, our results show that pension reforms aimed at increasing labor market attachment for the elderly generation can have unintended and critical consequences for younger generations, including adult daughters and grandchildren. To further illustrate this point, we follow the framework proposed by Hendren and Sprung-Keyser (2020) and calculate –under different sets of assumptions– the Marginal Value of Public Funds of the Dutch reform studied in this paper. Under the strong assumption (made for illustrative purposes only) that the government only cares about income tax revenue, and the impact on maternal labor supply lasts for up to eight years, we show that the loss in tax revenue from the drop in maternal labor supply of the adult daughters would outweigh the gain in tax revenue from delaying the retirement of grandmothers (see Appendix D for detailed steps of the calculation). Second, the positive educational outcomes of children aged 4 to 7 point to the importance of the quality of care. High-quality childcare provided by mothers can be made possible and shared by both parents through more generous parental leave policies or by improving the quality of formal care provision.

Third, the negative impact on boys aged 8-12 suggests that unsupervised time at home due to a lack of grandparental care and/or low-quality after-school care can negatively affect children's performance in high-stakes tests, with decisive long-run implications. When reforming the pension system, governments should take such spillover effects on childcare arrangements into account and improve childcare provision, especially after-school care. Finally, our results are meaningful for recent policy discussions on the gender gap and the child penalty. We show that women whose mothers delay retirement due to pension reforms face a much larger child penalty eight years after the first childbirth, relative to women whose (grand)mothers could retire earlier. We thereby provide the first evidence that pension reforms can have unintended implications for the child penalty and gender gap within households and in society overall.

Two features of the Netherlands are important in interpreting the external validity of our findings. First, while Dutch women have a high labor market participation rate, most women work part-time. Especially, mothers work part-time, while fathers usually work full-time. This may explain why we find a sizable adjustment in mothers' working hours, as adjusting hours in part-time work is common in the Dutch labor market. Second, more than half of Dutch primary schools have so-called traditional school hours, without school on Wednesday afternoons and sometimes Friday afternoons, and with long lunch breaks at home. This situation is not unique; many European countries, such as Germany, France, and Italy, have a similar organization of the school day (Kamette, 2011; Felfe et al., 2016; Dehos and Paul, 2023). Grandparental care can be complementary to these unfilled childcare hours. Thus, grandmothers working more leads to mothers needing to work less to care for their children. Had the school hours mirrored the regular working hours, the impact on the labor supply of mothers with young school-age children might have been smaller.

Although at first glance the estimates are only applicable to old-age pension reforms, the actual relevance extends further. Our paper points to an essential policy implication: public policies can trigger multigenerational spillover effects with important distributional consequences. While outside the scope of our paper, we believe examining such spillover effects across generations for other public policies is a fruitful avenue for future research.

References

Abadie, Alberto, "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, apr 2003, *113* (2), 231–263.

Aizer, Anna, "Home alone: Supervision after school and child behavior," *Journal of Public Economics*, 2004, *88* (9-10), 1835–1848.

_, Hilary Hoynes, and Adriana Lleras-Muney, "Children and the US Social Safety Net: Balancing Disincentives for Adults and Benefits for Children," *Journal of Economic Perspectives*, 2022, *36* (2), 149–74.

_, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney, "The Long-Run Impact of Cash Transfers to Poor Families," *American Economic Review*, apr 2016, *106* (4), 935–971.

Altonji, Joseph G and Rebecca M Blank, "Race and gender in the labor market," *Handbook of Labor Economics*, 1999, *3*, 3143–3259.

Andresen, Martin Eckhoff and Emily Nix, "Can the child penalty be reduced? Evaluating multiple policy interventions," Technical Report, Discussion Papers 2022.

Andrew, Alison, Orazio P Attanasio, Raquel Bernal, Lina Cardona Sosa, Sonya Krutikova, and Marta Rubio-Codina, "Preschool quality and child development," *Journal of Political Economy*, 2024, *132* (7), 000–000.

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly harmless econometrics: An empiricist's companion*, Princeton university press, 2009.

Aparicio Fenoll, Ainoa, "The uneven impact of women's retirement on their daughters' employment," *Review of Economics of the Household*, 2020, *18* (3), 795–821.

— **and Marian Vidal-Fernandez**, "Working women and fertility: the role of grandmothers' labor force participation," *CESifo Economic Studies*, 2015, *61* (1), 123–147.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman, "School Quality and the Gender Gap in Educational Achievement," *American Economic Review: Papers & Proceedings*, 2016, *106* (5), 289–295.

Backhaus, Andreas and Mikkel Barslund, "The effect of grandchildren on grandparental labor supply: evidence from Europe," *European Economic Review*, 2021, *137*, 103817.

Bailey, Martha J, Hilary Hoynes, Maya Rossin-Slater, and Reed Walker, "Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence From the Food Stamps Program," *The Review of Economic Studies*, 06 2023, *91* (3), 1291–1330.

— **, Shuqiao Sun, and Brenden Timpe**, "Prep School for poor kids: The long-run impacts of Head Start on Human capital and economic self-sufficiency," *American Economic Review*, 2021, *111* (12), 3963–4001.

Bailey, Martha, Tanya Byker, Elena Patel, and Shanthi Ramnath, "The long-run effects of california's paid family leave act on women's careers and childbearing: New evidence from a regression discontinuity design and us tax data," *American Economic Journal: Economic Policy*, 2025, *17* (1), 401–431.

Baker, Michael, Jonathan Gruber, and Kevin Milligan, "Universal child care, maternal labor supply, and family well-being," *Journal of Political Economy*, 2008, *116* (4), 709–745.

— **,** — **, and** — , "The long-run impacts of a universal child care program," *American Economic Journal: Economic Policy*, 2019, *11* (3), 1–26.

Bastian, Jacob and Lance Lochner, "The Earned Income Tax Credit and Maternal Time Use: More Time Working and Less Time with Kids?," *Journal of Labor Economics*, 2022, *40* (3), 573–611.

Battistin, Erich, Michele De Nadai, and Mario Padula, "Roadblocks on the road to grandma's house: fertility consequences of delayed retirement," *IZA Discussion Paper*, 2014, *No. 8071.*

Bauernschuster, Stefan and Martin Schlotter, "Public child care and mothers' labor supply—Evidence from two quasi-experiments," *Journal of Public Economics*, 2015, *123*, 1–16.

Bernal, Raquel and Michael P Keane, "Quasi-structural estimation of a model of childcare choices and child cognitive ability production," *Journal of Econometrics*, 2010, *156* (1), 164–189.

— **and** — , "Child care choices and children's cognitive achievement: The case of single mothers," *Journal of Labor Economics*, 2011, *29* (3), 459–512.

Bertrand, Marianne and Jessica Pan, "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior," *American Economic Journal: Applied Economics*, 2013, *5* (1), 32–64.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes, "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital," *American Economic Review*, feb 2005, *95* (1), 437–449.

Blau, Francine D and Lawrence M Kahn, "The gender wage gap: Extent, trends, and explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

Blundell, Richard, Eric French, and Gemma Tetlow, "Retirement incentives and labor supply," in "Handbook of the economics of population aging," Vol. 1, Elsevier, 2016, pp. 457–566.

Bovenberg, Lans and Raymond Gradus, "Reforming occupational pension schemes: the case of the Netherlands," *Journal of Economic Policy Reform*, 2015, *18* (3), 244–257.

Bratti, Massimiliano, Tommaso Frattini, and Francesco Scervini, "Grandparental availability for child care and maternal labor force participation: pension reform evidence from Italy," *Journal of Population Economics*, 2018, *31* (4), 1239–1277.

Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell, "On the effect of bias estimation on coverage accuracy in nonparametric inference," *Journal of the American Statistical Association*, 2018, *113* (522), 767–779.

_ , _ , and Rocio Titiunik, "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Calot, G. and Chantal Blayo, "Recent course of fertility in Western Europe," *Population Studies*, November 1982, *Vol. 36, No. 3*, 349–372.

Carneiro, Pedro, Costas Meghir, and Matthias Parey, "Maternal education, home environments, and the development of children and adolescents," *Journal of the European Economic Association*, 2013, *11* (suppl_1), 123–160.

Cascio, Elizabeth U, "Maternal labor supply and the introduction of kindergartens into American public schools," *Journal of Human resources*, 2009, *44* (1), 140–170.

_ , "The promises and pitfalls of universal early education," *IZA World of Labor*, 2014.

_ , "Does Universal Preschool Hit the Target?: Program Access and Preschool Impacts," *Journal of Human Resources*, 2023, *58* (1), 1–42.

Cattaneo, Matias D, Nicolás Idrobo, and Rocío Titiunik, *A practical introduction to regression discontinuity designs: Extensions*, Cambridge University Press, 2024.

CentERdata, "Longitudinal Internet studies for the SocialSciences," *LISS*, 2008.

Coile, Courtney, "Retirement incentives and couples' retirement decisions," *The BE Journal of Economic Analysis & Policy*, 2004, *4* (1), 1–30.

_ **and Jonathan Gruber**, "Future social security entitlements and the retirement decision," *The Review of Economics and Statistics*, 2007, *89* (2), 234–246.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg, "Who benefits from universal child care? Estimating marginal returns to early child care attendance," *Journal of Political Economy*, 2018, *126* (6), 2356–2409.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad, "Family Welfare Cultures," *The Quarterly Journal of Economics*, aug 2014, *129* (4), 1711–1752.

Dahl, Gordon B, Katrine V Løken, Magne Mogstad, and Kari Vea Salvanes, "What is the case for paid maternity leave?," *Review of Economics and Statistics*, 2016, *98* (4), 655–670.

Danielsbacka, Mirkka, Antti O Tanskanen, David A Coall, and Markus Jokela, "Grandparental childcare, health and well-being in Europe: A within-individual investigation of longitudinal data," *Social Science & Medicine*, 2019, *230*, 194–203.

Danzer, Natalia and Victor Lavy, "Paid parental leave and children's schooling outcomes," *The Economic Journal*, 2018, *128* (608), 81–117.

De Nardi, Mariacristina, Giulio Fella, Marike Knoef, Gonzalo Paz-Pardo, and Raun Van Ooijen, "Family and government insurance: Wage, earnings, and income risks in the Netherlands and the US," *Journal of Public Economics*, 2021, *193*, 104–327.

Dehos, Fabian T and Marie Paul, "The effects of after-school programs on maternal employment," *Journal of Human Resources*, 2023, *58* (5), 1644–1678.

Del Boca, Daniela, Daniela Piazzalunga, and Chiara Pronzato, "The role of grandparenting in early childcare and child outcomes," *Review of Economics of the Household*, 2018, *16* (2), 477–512.

Del Bono, Emilia, Marco Francesconi, Yvonne Kelly, and Amanda Sacker, "Early maternal time investment and early child outcomes," *The Economic Journal*, 2016, *126* (596), F96–F135.

Duncan, Greg, Ariel Kalil, Magne Mogstad, and Mari Rege, "Investing in early childhood development in preschool and at home," *Handbook of the Economics of Education*, 2023, *6*, 1–91.

Dustmann, Christian and Uta Schönberg, "Expansions in maternity leave coverage and children's long-term outcomes," *American Economic Journal: Applied Economics*, 2012, *4* (3), 190–224.

Eibich, Peter and Thomas Siedler, "Retirement, intergenerational time transfers, and fertility," *European Economic Review*, 2020, *124*, 103392.

Euwals, Rob, Daniel Van Vuuren, and Ronald Wolthoff, "Early retirement behaviour in the

Netherlands: Evidence from a policy reform," *De Economist*, 2010, *158* (3), 209–236.
Felfe, Christina and Rafael Lalive, "Does early child care affect children's development?," *Journal of Public Economics*, 2018, *159*, 33–53.
_ , Michael Lechner, and Petra Thiemann, "After-school care and parents' labor supply," *Labour Economics*, 2016, *42*, 64–75.
Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti, "Mothers and sons: preference formation and female labor force dynamics," *The Quarterly Journal of Economics*, 2004, *119* (4), 1249–1299.
Fitzpatrick, Maria D, "Preschoolers enrolled and mothers at work? The effects of universal prekindergarten," *Journal of Labor Economics*, 2010, *28* (1), 51–85.
Fort, Margherita, Andrea Ichino, and Giulio Zanella, "Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families," *Journal of Political Economy*, 2020, *128* (1), 158–205.
Francesconi, Marco and James J Heckman, "Child development and parental investment: Introduction," *The Economic Journal*, 2016, *126* (596), F1–F27.
Frimmel, Wolfgang, Martin Halla, Bernhard Schmidpeter, and Rudolf Winter-Ebmer, "Grandmothers' labor supply," *Journal of Human Resources*, 2020, pp. 0419–10144R1.
Gelman, Andrew and Guido Imbens, "Why high-order polynomials should not be used in regression discontinuity designs," *Journal of Business & Economic Statistics*, 2019, *37* (3), 447–456.
Ginja, Rita, Jenny Jans, and Arizo Karimi, "Parental leave benefits, household labor supply, and children's long-run outcomes," *Journal of Labor Economics*, 2020, *38* (1), 261–320.
Gørtz, Mette, Sarah Sander, and Almudena Sevilla, "Does the child penalty strike twice?," *European Economic Review*, 2025, *172*, 104942.
Gruber, Jonathan, "The incidence of mandated maternity benefits," *The American Economic Review*, 1994, pp. 622–641.
Haandrikman, Karen and Leo J. G. van Wissen, "Effects of the fertility transition on birth seasonality in the Netherlands," *Journal of Biosocial Science*, jan 2008, *40* (05).
Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 2001, *69* (1), 201–209.
Hendren, Nathaniel and Ben Sprung-Keyser, "A unified welfare analysis of government policies," *The Quarterly Journal of Economics*, 2020, *135* (3), 1209–1318.
Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond, "Long-run impacts of childhood access to the safety net," *American Economic Review*, 2016, *106* (4), 903–34.
_ , _ , and _ , "Long-Run Impacts of Childhood Access to the Safety Net," *American Economic Review*, apr 2016, *106* (4), 903–934.
Huebener, Mathias, Astrid Pape, and C Katharina Spiess, "Parental labour supply responses to the abolition of day care fees," *Journal of Economic Behavior & Organization*, 2020, *180*, 510–543.
Hurd, Michael D, "The joint retirement decision of husbands and wives," in "Issues in the Economics of Aging," University of Chicago Press, 1990, 1990, pp. 231–258.
Ilciukas, Julius, "Fertility and parental retirement," *Journal of Public Economics*, 2023, *226*, 104928.
Janta, B, "Caring for children in Europe: How childcare, parental leave and flexible working arrangements interact in Europe. Report prepared for the European Commission, Directorate-General for Employment, Social Affairs and Inclusion," *RAND Corporation*, 2014, *2*.
Jappens, Maaike and Jan Van Bavel, "Regional family norms and child care by grandparents in Europe," *Demographic Research*, 2012, *27*, 85–120.
Kamette, Florence, "Organisation of school time in the European Union," *European issues*, 2011, *212*, 1–6.
Karademir, Sencer, Jean-William P Laliberté, and Stefan Staubli, "The multigenerational impact of children and childcare policies," Technical Report, National Bureau of Economic Research 2024.
Kaufmann, Katja M, Matthias Messner, and Alex Solis, "Elite higher education, the marriage market and the intergenerational transmission of human capital," *CRCTR 224 Discussion Paper*, 2021, *No. 269.*

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

— , — , **Johanna Posch, Andreas Steinhauer, and Josef Zweimüller**, "Child penalties across countries: Evidence and explanations," in "AEA Papers and Proceedings," Vol. 109 2019, pp. 122–26.

— , — , — , — , **and** — , "Do family policies reduce gender inequality? Evidence from 60 years of policy experimentation," *American Economic Journal: Economic Policy*, 2024, *16* (2), 110–149.

Krapf, Matthias, Anja Roth, and Michaela Slotwinski, "The effect of childcare on parental earnings trajectories," 2020.

Lalive, Rafael and Pierpaolo Parrotta, "How does pension eligibility affect labor supply in couples?," *Labour Economics*, 2017, *46*, 177–188.

Laughlin, Lynda L, "Who's Minding the Kids? Child Care Arrangements: Spring 2005/Summer 2006. Current Population Reports. P70-121.," *US Census Bureau*, 2010.

Lindeboom, Maarten and Raymond Montizaan, "Disentangling retirement and savings responses," *Journal of Public Economics*, 2020, *192*, 104297.

Manoli, Day and Andrea Weber, "Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions," *American Economic Journal: Economic Policy*, November 2016, *8* (4), 160–82.

Marcos, Miguel Angel Talamas, "Grandmothers and the gender gap in the Mexican labor market," *Journal of Development Economics*, 2023, *162*, 103013.

Mastrobuoni, Giovanni, "Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities," *Journal of Public Economics*, dec 2009, *93* (11-12), 1224–1233.

Moroni, Gloria, Anne C Gielen, Marike Knoef, and Jim Been, "Prolonged Worklife Among Grandfathers: Spillover Effects on Grandchildren'S Educational Outcomes," *Available at SSRN 4828698*, 2023.

Nicoletti, Cheti, Kjell G Salvanes, and Emma Tominey, "The Family Peer Effect on Mothers' Labor Supply," *American Economic Journal: Applied Economics*, 2018, *10* (3), 206–34.

— , — , **and** — , "Mothers working during preschool years and child skills: does income compensate?," *Journal of Labor Economics*, 2023, *41* (2), 389–429.

OECD, *OECD Family database PF1.7: Intergenerational solidarity* 2012.

— , *Starting strong 2017: Key OECD indicators on early childhood education and care*, OECD Publishing, 2017.

— , *Part-time and partly equal: gender and work in the Netherlands* 2019.

— , *Pensions at a glance* 2019.

Olivetti, Claudia and Barbara Petrongolo, "The evolution of gender gaps in industrialized countries," *Annual review of Economics*, 2016, *8*, 405–434.

— **and** — , "The economic consequences of family policies: lessons from a century of legislation in high-income countries," *Journal of Economic Perspectives*, 2017, *31* (1), 205–230.

— **, Eleonora Patacchini, and Yves Zenou**, "Mothers, Peers, and Gender-Role Identity," *Journal of the European Economic Association*, dec 2018, *18* (1), 266–301.

Posadas, Josefina and Marian Vidal-Fernandez, "Grandparents' childcare and female labor force participation," *IZA Journal of Labor Policy*, 2013, *2* (1), 1–20.

Rabaté, Simon, Egbert Jongen, and Tilbe Atav, "Increasing the retirement age: policy effects and underlying mechanisms," *American Economic Journal: Economic Policy*, 2024, *16* (1), 259–291.

Rupert, Peter and Giulio Zanella, "Grandchildren and their grandparents' labor supply," *Journal of Public Economics*, 2018, *159*, 89–103.

Schönberg, Uta and Johannes Ludsteck, "Expansions in maternity leave coverage and mothers' labor market outcomes after childbirth," *Journal of Labor Economics*, 2014, *32* (3), 469–505.

Stancanelli, Elena and Arthur Van Soest, "Retirement and home production: A regression discontinuity approach," *American Economic Review*, 2012, *102* (3), 600–605.

Statistics Netherlands, "Population Register, Labor Supply Data, Education Data," *CBS*, 1995-2018.

— , "Labour force survey," *Statline*, 2009.

—, *The Netherlands in numbers: Who are most likely to work part-time?* 2022.
The Economist, "The age of the grandparent has arrived," *The Economist*, 2023, *Jan 12th.*
Wang, Sophie Xuefei and Cynthia Bansak, "Are grandparents a good substitute for parents as the primary caregiver? The impact of grandparents on Children's academic performance," *Labour Economics*, 2024, *88*, 102545.
Zamarro, Gema, "Family labor participation and child care decisions: the role of grannies," *SERIEs*, 2020, *11* (3), 287–312.
Zhang, Jing, Simon Appleton, Lina Song, and Bing Liu, "Who looks after the kids? the effects of childcare choice on early childhood development in China," *Oxford Bulletin of Economics and Statistics*, 2021, *83* (3), 619–640.

8 Tables and Figures

Table 1: Impact on mothers' labor supply

	Fuzzy RD estimates			
	(1)	(2)	(3)	Means at cutoff
Panel A: Reduced-form estimates - Mothers' labor supply outcomes				
Total monthly hours worked	-4.443**	-4.774***	-4.912***	78.876
	[1.767]	[1.777]	[1.772]	[47.744]
Optimal bandwidth	5.882	5.367	5.347	
Obs. Mothers	15156	15156	15156	4018
Panel B: Fuzzy RD estimates -				
First-stage: Impact on Grandmother's (GM) total labor supply				
Total monthly hours worked	7.080***	7.264***	6.367***	34.418
	[1.720]	[1.690]	[1.593]	[47.608]
LATE: Impact on mothers' labor supply				
Total monthly hours worked	-0.553**	-0.552**	-0.630**	78.876
	[0.249]	[0.247]	[0.282]	[47.744]
Optimal bandwidth	7.952	7.126	7.324	
F-Stat	16.949	18.474	15.979	
Obs. Mothers	20711	20711	20711	4018
Controls	NO	YES	YES	
Sector FE	NO	NO	YES	

Source: Authors' calculations from the CBS data.

Note: Panel A reports reduced-form estimates of the reform impact on mothers' total monthly hours worked, and Panel B reports 2SLS fuzzy RD estimates (first-stage and LATE) of grandmothers' labor supply on mothers' total monthly hours worked. The running variable is the grandmother's birthdate (month-year), centered around January 1950. Columns 1, 2, and 3 show the results without controls, with controls, and with both controls and sector fixed effects, respectively. For a description of the full list of controls, see Section 2.3. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure. Sample means at the cutoff (measured three months before the cutoff) are reported in Column 4. All outcomes are measured when the grandmothers are between the ages of 60 and 64. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Mechanisms: Grandfathers' vs. grandmothers' effect

	baseline	only one treated	both treated
	(1)	(2)	(3)
Panel A: Grandmothers' adult daughters			
First-stage: Impact on Grandmother's labor supply			
Total monthly hours worked	6.367***	5.272***	14.553***
	[1.593]	[1.786]	[2.804]
LATE: Impact on Daughters' labor supply			
Total monthly hours worked	-0.630**	-0.701*	-0.692***
	[0.282]	[0.403]	[0.219]
Optimal bandwidth	7.324	7.770	7.873
F-Stat	15.979	8.715	26.940
Obs. Daughters	20711	13990	8842
Panel B: Grandfathers' adult daughters			
First-stage: Impact on Grandfather's labor supply			
Total monthly hours worked	26.075***	25.122***	25.967***
	[2.186]	[4.186]	[3.837]
LATE: Impact on Daughters' labor supply			
Total monthly hours worked	-0.110**	-0.026	-0.175*
	[0.055]	[0.104]	[0.072]
Optimal bandwidth	8.062	10.372	8.672
F-Stat	142.283	36.016	45.788
Obs. Daughters	23609	7245	19469
Differences Panel A vs. Panel B effects			
P-value: difference in first-stages	0.000	0.000	0.012
P-value: difference in LATE	0.066	0.106	0.033

Source: Authors' calculations from the CBS data.

Note: This table shows the effect of grandparents' total monthly hours worked on their adult daughters' (mothers') labor supply (Fuzzy RD estimates). Panel A reports first- and second-stage coefficients of grandmothers born around the cutoff, and panel B considers grandfathers born around the cutoff. Column 1 considers the treatment of grandmothers (in panel A) and grandfathers (in panel B), irrespective of whether their partner was also in the relevant age range to be treated. Column 2 compares families where no grandparent is treated to those where only the grandmother (panel A) or only the grandfather (panel B) is treated. In Column 3, families with no treated grandparent are compared to families where both grandparents are treated. An indicator for the grandparent being born since January 1950 serves as the instrument for the grandparent's total monthly hours worked. All outcomes are measured when the grandparent affected by the reform is between the ages of 60 and 64. All columns consider mothers with a youngest (grand)child aged 4-12 when the grandparent is aged 60-64. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure, including controls and sector fixed effects. Robust standard errors clustered at the grandparent level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Mechanisms: Impact on other family members

Family member	Grandmothers' Daughters (1)	Sons-in-law (2)	Sons (3)	Daughters-in-law (4)
Panel A: Hours worked				
First-Stage: Impact on GM's total labor supply				
	6.367***	6.374***	5.233***	5.449***
	[1.593]	[1.460]	[1.684]	[1.691]
LATE: Impact on family members' labor supply				
Total monthly hours worked	-0.630**	0.729**	0.619	-0.191
	[0.282]	[0.340]	[0.435]	[0.313]
Optimal bandwidth	7.324	9.094	8.651	8.659
F-Stat	15.979	19.048	9.661	10.388
Obs. Family members	20711	24297	16773	16596
Panel B: Other labor supply measures				
LATE: Prob(employed)	-0.003*	0.004**	0.002	-0.002
	[0.002]	[0.002]	[0.002]	[0.002]
Optimal bandwidth	7.915	8.906	8.762	8.906
F-Stat	16.36	19.604	9.749	10.215
Obs. Family members	20711	21549	16773	16550
LATE: Prob(full-time employed)	-0.002	0.003	0.002	-0.001
	[0.001]	[0.002]	[0.003]	[0.002]
Optimal bandwidth	8.764	8.678	9.366	9.107
F-Stat	18.40	18.259	11.092	11.092
Obs. Family members	23497	21767	18767	18570
LATE: HH labor income	21.721		18.456	
	[13.258]		[15.136]	
Optimal bandwidth	8.323		9.132	
F-Stat	17.519		11.182	
Obs. Family members	21767		18570	

Source: Authors' calculations from the CBS data.

Note: This table shows the coefficient estimates of grandmothers' total monthly hours worked on different family members' labor supply, namely adult daughters (Col. 1), sons-in-law (Col. 2), adult sons (Col. 3) and daughters-in-law (Col. 4) of the grandmothers (Fuzzy RD estimates). Panel A reports first- and second-stage estimates for total hours worked, and Panel B reports LATE coefficients for other labor supply outcomes (first-stage results are available upon request). An indicator for the grandmother being born since January 1950 serves as the instrument for the grandmother's total monthly hours worked. All outcomes are measured when the grandmother affected by the reform is between the ages of 60 and 64. Household income is only considered for daughters/ sons with a partner, i.e., for this outcome the number of observations and F-statistics of Column 2/4 apply. All columns consider family members with a youngest (grand)child aged 4-12 when the grandmother is aged 60-64. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure, including controls and sector fixed effects. For a description of the full list of controls, see Section 2.3. All income measures are CPI-adjusted for the year 2015. Robust standard errors clustered at grandmother level are in parentheses. Table A17 reports p-values for the differences in coefficients across columns. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Mechanisms: Effects by age of the youngest child

	Childcare need				
	more			less/no	
	Age of the youngest child				
	1-3	4 - 7	8-12	13 - 18	No child
	(1)	(2)	(3)	(4)	(5)
Panel A: Reduced-form estimates -					
Mothers' labor supply outcomes					
Total monthly hours worked	-3.202*	-6.855***	-2.051	1.257	-1.197
	[1.747]	[2.051]	[2.028]	[3.551]	[2.023]
Optimal Bandwidth	5.082	4.740	8.657	7.301	6.847
Obs. Mothers	15799	10887	11378	4392	9368
Panel B: Fuzzy RD estimates -					
First-stage: Impact on GM's total labor supply					
Total monthly hours worked	6.589***	6.841***	7.256***	5.782**	3.781*
	[1.513]	[1.835]	[2.068]	[2.927]	[2.031]
LATE: Impact on mothers' labor supply					
Total monthly hours worked	-0.300	-0.808**	-0.276	0.221	-0.133
	[0.212]	[0.328]	[0.292]	[0.585]	[0.484]
Optimal bandwidth	8.219	6.701	8.503	8.337	8.450
F-Stat	18.981	13.898	12.310	3.902	3.464
Obs. Mothers	24429	15668	11378	4984	12289

Source: Authors' calculations from the CBS data.

Note: Panel A reports reduced-form estimates on the mother's monthly hours worked by the age of the youngest child. Panel B reports 2SLS fuzzy RD estimates (first-stage and LATE) of grandmothers' total monthly hours worked on the mother's monthly hours worked by the age of the youngest child. The running variable is the grandmother's birthdate (month-year), centered around January 1950. Columns 1 - 3 show the results for families with childcare needs based on the age of the youngest child (1-3, 4-7, 8-12, respectively). Columns 4 and 5 show results for families with little or no childcare need (youngest child aged 13 - 18 and without children, respectively). All outcomes are measured when the grandmothers are between the ages of 60 and 64. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure, including controls and sector fixed effects. For a description of the full list of controls, see Section 2.3. Robust standard errors clustered at the grandmother level are in parentheses. Table A18 reports p-values for the differences in coefficients across columns. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Mechanisms: Heterogeneous effects (reduced-form)

	Grandmother's partner is		Grandmother's residence neighborhood		Number of maternal grandchildren aged 4-7	
Subgroups	unhealthy	healthy	different	same	more or other age	only one
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Impact on GM' labor supply						
Total monthly hours worked	5.473	5.956***	6.770***	5.098*	8.496***	5.050**
	[5.972]	[1.774]	[1.920]	[3.084]	[2.886]	[2.066]
p-value	0.939		0.631		0.331	
Optimal bandwidth	6.865	6.865	6.865	6.865	6.865	6.865
Obs. Grandmothers	1333	16597	13409	4521	7420	10510
Panel B: Impact on mothers' labor supply						
Total monthly hours worked	-9.104	-4.642**	-3.891*	-8.512**	-4.310	-5.668**
	[6.165]	[1.837]	[2.048]	[3.464]	[2.870]	[2.236]
p-value	0.500		0.253		0.711	
Optimal bandwidth	5.403	5.403	5.403	5.403	5.403	5.403
Obs. Mothers	1129	14027	11308	3848	6273	8883
Other labor supply measures						
Prob(employed)	-0.008	-0.026**	-0.013	-0.060**	-0.016	-0.031**
	[0.045]	[0.129]	[0.014]	[0.024]	[0.020]	[0.016]
p-value	0.717		0.096		0.561	
Optimal bandwidth	6.980	6.980	6.980	6.980	6.980	6.980
Obs. Mothers	1333	16597	13409	4521	7420	10510
Prob(full-time employed)	-0.046*	-0.007	-0.009	-0.018	-0.010	-0.011
	[0.024]	[0.007]	[0.008]	[0.013]	[0.011]	[0.008]
p-value	0.123		0.558		0.915	
Optimal bandwidth	7.817	7.817	7.817	7.817	7.817	7.817
Obs. Mothers	1553	19158	15505	5206	8547	12164

Source: Authors' calculations from the CBS data.

Note: This table shows heterogeneous effects of the reform on grandmothers' and mothers' labor supply (reduced-form). Columns 1 and 2 show the effects by health status of the grandmother's partner. Partners are defined as healthy if they have not claimed any disability insurance before age 54. Columns 3 and 4 show effects by the proximity of adult daughters (mothers) to grandmothers. We define the grandmother to be nearby when the mother and grandmother live in the same neighborhood. Columns 5 and 6 show effects by number of maternal grandchildren aged 4-7. All outcomes are measured when grandmothers are between the ages of 60 and 64. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure, including controls (mother's age, the age at first birth, and grandmothers' predetermined employment status) and sector fixed effects. Robust standard errors clustered at the grandmother level are in parentheses. The p-values are from a test of the hypothesis that the coefficients are equal. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Effects on children's educational performance (reduced-form)

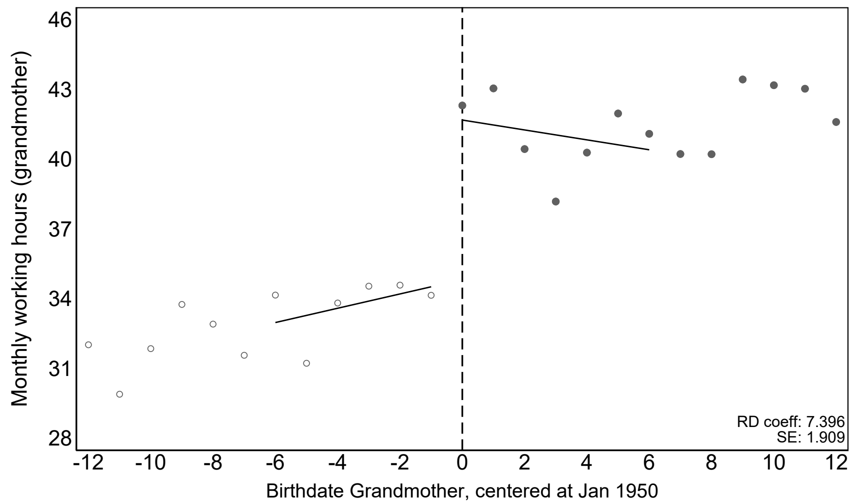
RD estimates	Number of correct answers (Cito)			High
	Verbal	Math	Overall	track
	(1)	(2)	(3)	(4)
Panel A: All youngest children				
Aged between 4 - 12	0.101*	0.084	0.146**	0.021
	[0.057]	[0.056]	[0.068]	[0.022]
Optimal bandwidth	5.998	6.291	5.320	6.471
Obs. Children	4232	5025	4232	5025
Panel B: By age groups				
Aged between 4 - 7	0.249***	0.271***	0.305***	0.064*
	[0.089]	[0.078]	[0.096]	[0.034]
Optimal bandwidth	5.495	6.652	5.511	5.589
Obs. Children	2358	2795	2358	2358
Aged between 8 - 12	-0.013	0.017	0.013	-0.032
	[0.093]	[0.100]	[0.114]	[0.028]
Optimal bandwidth	7.136	5.933	5.599	8.364
Obs. Children	1744	1268	1268	1984
P-value: difference between age groups	0.052	0.136	0.057	0.091

Source: Authors' calculations from the CBS data.

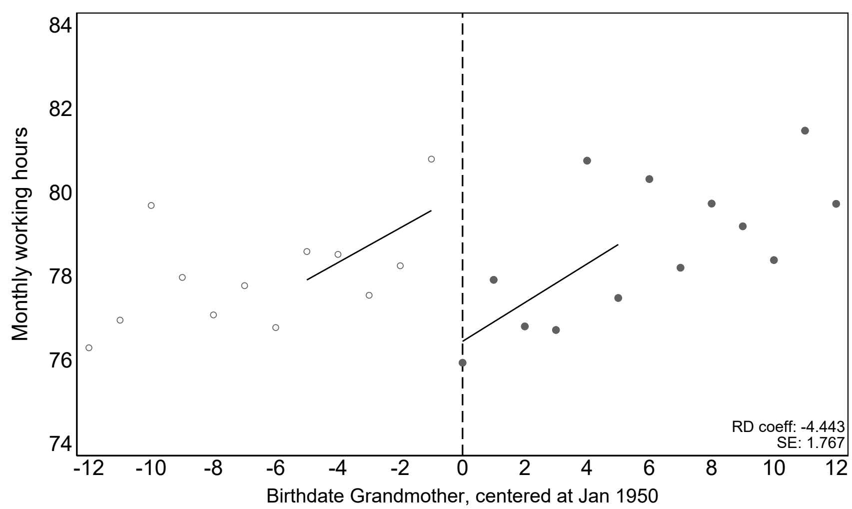
Note: This table shows reduced-form impacts on the education outcomes of children. Panel A shows results for the youngest child aged 4 - 12. Panel B presents results separately for the youngest child aged 4 - 7 and 8 - 12. Columns 1 - 3 report effects on the number of correct answers in the verbal part, the mathematical part, and the overall Cito test, respectively. These columns are based on standardized outcomes and thus measure effects in percent of the standard deviation. Column 4 shows the impact on the probability of obtaining a secondary school recommendation for the highest (academic) track (VWO). All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure and include controls, including the child's birth cohort and month, and treatment duration. For a description of the full list of controls, see Section 2.3. Robust standard errors (clustered by the child's primary school) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: RD plots: Grandmothers' and mothers' total hours worked

(a) Total hours worked by grandmothers (first-stage)



(b) Total hours worked by mothers (reduced-form)

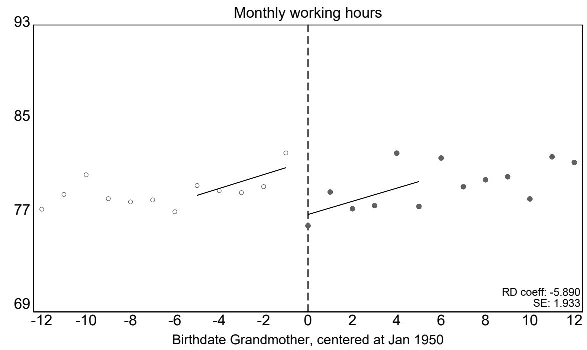


Note: Panel (a) of Figure 1 shows the scatter bin plot of grandmothers' total monthly hours worked as a function of distance to the cutoff, which is grandmothers' birth month being January 1950. Panel (b) of Figure 1 shows the scatter bin plots of the mother's labor supply as a function of distance to the cutoff, which is the grandmother's birth month being January 1950. While we employ local linear regressions in the estimation, in this figure we add a solid line for illustrational purposes, which is a linear polynomial fit of the outcome on the running variable, given the optimal bandwidth generated by Calonico et al. (2014) and fit separately left and right of the cutoff. Reported coefficients are RD estimates without controls. For estimations including controls, see Tables 1 and A8.

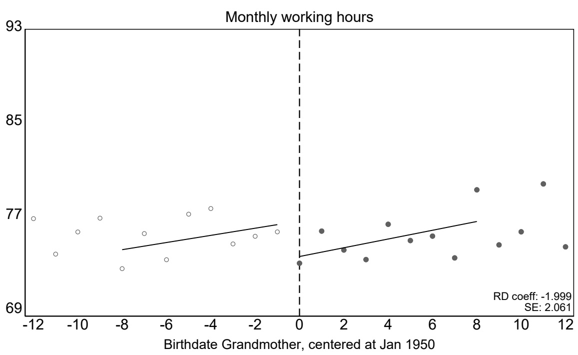
Figure 2: RD plots: Mothers' total hours worked by age of the youngest child

(a) Mothers with youngest child aged 1-3

(b) Mothers with youngest child aged 4-7



(c) Mothers with youngest child aged 8-12

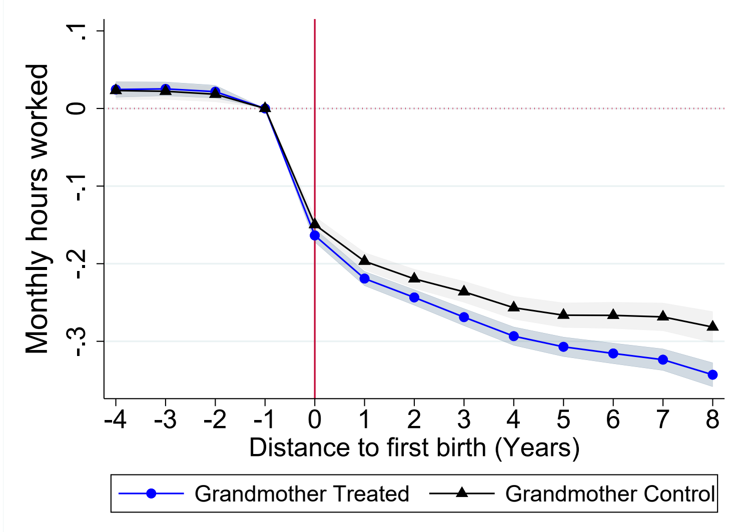


(d) Mothers with youngest child aged 13-18

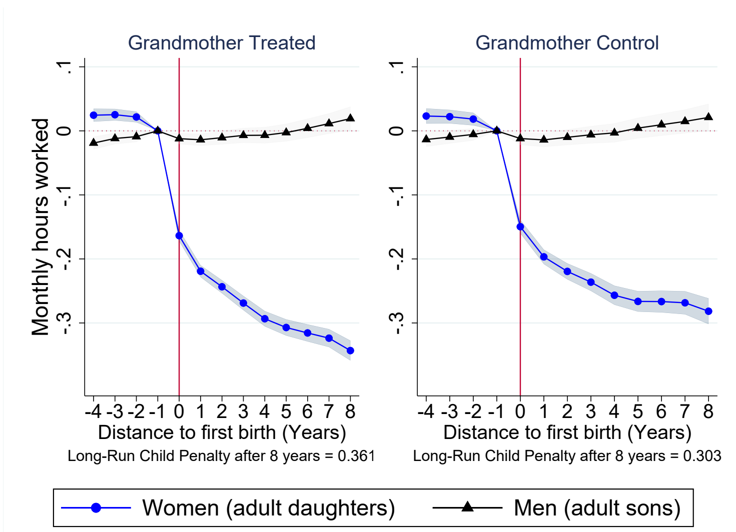
Note: Figure 2 shows the scatter bin plot of mothers' total monthly hours worked as a function of distance to the cutoff, considering mothers with a youngest child aged 1-3, 4-7, 8-12, and 13-18, respectively. Each plot considers the cutoff of the grandmother's birth month as January 1950. The solid line is a linear polynomial fit of each outcome on the running variable based on the optimal bandwidth generated by Calonico et al. (2014) and fit separately left and right of the cutoff. Reported coefficients are RD estimates without controls. For estimations including controls, see Table 4.

Figure 3: Dynamic treatment effects and child penalty

(a) Dynamic treatment effects on monthly working hours



(b) Relative child penalty by treatment status



Note: Panel (a) of Figure 3 shows the evolution of mothers' total monthly hours worked from 4 years before and to 8 years after they gave birth to their first child. It compares the monthly working hours of treated mothers (blue dots), whose (grand)mothers were born between January 1950 and December 1951 and thus treated by the pension reform, to those of control mothers (black triangles), with untreated (grand)mothers born between January 1948 and December 1949. Event time 0 marks the birth of the first child. Panel (b) of Figure 3 depicts the child penalty in total monthly working hours (including zeros) by treatment status. The left figure presents the child penalty for men and women with treated grandmothers and the right figure for men and women with control grandmothers. Blue dots document women's and black triangles indicate men's monthly working hours, the difference in between represents the child penalty. The long-run relative child penalty after 8 years (i.e., the relative loss women experience compared to men) is reported below each sub-graph. The value at t = -1 is normalized to zero so that coefficients measure the impact of the first child relative to the year before birth. The shaded areas indicate the 95 percent confidence interval.

For Online Publication

Spillover Effects of Old-Age Pension across Generations: Family Labor Supply and Child Outcomes

Online Appendix

Katja M. Kaufmann
University of Bayreuth,
CESifo, HCEO and IZA

Yasemin Özdemir
University of Bayreuth

Han Ye
University of Mannheim,
IZA, ZEW

List of Tables

List of Figures

A Additional Details on the Institutional Background

A.1 The Dutch pension system

The Dutch pension system consists of three pillars: the Pay-as-You-Go state pensions (AOW), occupational pensions, and individual savings. The first pillar, the state pensions, provide all Dutch residents a flat-rate pension at their birth cohort specific AOW claiming age.[1] AOW benefits depend on years of residence and partnership status and are not related to earnings and contributions paid

[1] The AOW claiming age was set at age 65. Since 2012, the state pension claiming age was set to gradually increase, reaching 66 in 2018 and 67 in 2021. For our baseline sample of grandmothers, their state pension claiming ages are between 65 and 2 months and 65 and 3 months. The AOW age increased from 65 and 2 months to 65 and 3 months for people born since November 1949. See Rabaté et al. (2024) for more details of the AOW reform.

before retirement. They are financed by income taxes and are linked to the minimum wage (OECD, 2019b). Individuals can not claim AOW pension benefits before when they retire earlier.

The second pillar, the occupational pensions, which we focus on in our analysis, are collective pension schemes connected to a specific industry or company, capital-funded, and managed by pension funds. The majority of these schemes are of the defined benefit type. Contribution to the second pillar is mandatory, and more than 90 percent of the workers in the Netherlands contribute to a collective pension fund via their employer. The contribution rate is 14% of gross wages, of which 70% is contributed by the employers and 30% by the employees. These schemes typically aim at a replacement rate of about 80% (including the AOW benefits) of average pay after 40 years of service (Bovenberg and Gradus, 2015). Retirement before the statutory AOW claiming age is only possible through the occupational pensions, which have sectoral early pension schemes as part of the collective agreements.

The third pillar consists of non-mandatory savings. It is relatively small in the Netherlands and provides around 5% of pension income.

A.2 The evolution of Dutch early retirement schemes

The early retirement (ER) schemes are part of the collective labor agreements, which constitute the basis of the second pillar occupational pension schemes. In the Netherlands, the early retirement schemes were first introduced in the 1970s at a flat rate and were financed on a pay-as-you-go basis. The replacement rates vary by sectors and even by firms within sectors but are generally considered financially attractive. The average replacement rate is 80 percent of previous gross earnings. The flat-rate ER schemes were attractive and not actuarially fair.[2] The earliest possible age to claim occupational pensions is between ages 55 and 60, depending on the sectoral schemes.

In the early 1990s, the Dutch social partners started to replace the flat-rate ER schemes with actuarially adjusted schemes due to concerns about the long-run financial sustainability. The ER scheme started initial transitioning from the generous and actuarially unfair VUT schemes towards capital-funded, actuarially fair, and less generous schemes. Under the new ER schemes, workers receive lower pension benefits if they retire earlier than the statutory retirement age. However, the adjustment is slow and early retirement is still very attractive. The years spend on early retirement counted for the accrual of pensions via accumulated years of work experience. At age 65, the early retirement benefit was replaced by the regular AOW and an old-age occupational pension. The benefit level depends on years worked (including the years in early retirement) and average lifetime earnings. Moreover, contributions to the ER schemes were tax-deductible. The tax advantage amounted to about 25% of the net early retirement allowance (Euwals et al. (2010)). Therefore, retiring early was common. Approximately 80% of all workers retired at the age of 62 or younger before 2006 (Statistics Netherlands (2009)).

In 2006, there was a major reform of the early retirement schemes. The goal was to encourage labor market participation of the elderly by speeding up the transition towards an actuarially fair early retirement system. The two-tiered system was unified by applying the benefit formula to all pensioners and adjusting early retirement in an actuarially fair manner. Moreover, the early retirement tax advantages was eliminated. The general plan of the reform was announced in

[2]The flat-rate ER schemes were also called "VUT schemes". In Dutch, VUT stands for "Vervroegde Uittreding" in Dutch, which means "early retirement".

2000 by the first Balkenende cabinet. The goal was to encourage labor market participation of the elderly by speeding up the transition towards an actuarially fair early retirement system. The second Balkenende cabinet made several proposals to speed up the cancellation of the favorable tax treatment of the ER schemes in 2004, which has entailed one of the largest union demonstrations in Dutch history in October 2004. In November 2004, the proposal of bill No. 29760 was passed by the House of Representatives and adopted by the Senate in February 2005. The bill was published in the Official Gazette 115 of March 10, 2005.

From that date onwards, all sectors and industries introduced new pension schemes that are more actuarially fair and flexible. For example, the Dutch government announced to increase early retirement age from 55 to 60 and abolished the use of annuity to bridge the gap between early retirement and age 65 for cohorts born after 1949. The reform bill no. 29760 includes a clause to adjust fiscal policy VUT and prepension (Wet voor aanpassing fiscal behandeling VUT and pre-pension) to be more actuarially fair and is sometimes referred to as the 56-plus scheme (de 56-plusregeling). People, who were 55 years or older before January 1, 2005 (i.e. born before 1950) are not affected by the reform, while for people who turned 55 since January 1, 2005 (i.e. who were born since January 1, 1950) tax benefits for early retirement schemes were eliminated.

A.3 The life course savings program

In 2006, the Dutch government introduced the "Life course savings" (Levensloopregeling, LCS) program. This tax-facilitated savings program allows workers to save for periods of unpaid leave or early retirement. Employees can save up to 210 percent of their last wages, which equates to around two years of full income or two years with 70% of previous income. Each year employees can save up to 12% of annual earnings. This life-course savings program was abolished in 2012. However, people who started participating in the program prior to 2012 were still able to save tax-free in life course savings programs until 2021.

All individuals in our sample were eligible to participate in the life course savings program, which means both the grandmothers born before 1950 and since 1950 can use this new tax-facilitated saving scheme. However, individuals who were at least 50 years old but not yet 55 on 1 January 2005 (born since 1950) could save more than 12% per year. The policy intention was to provide a slight advantage for people aged 50 to 55 in 2005 to save quicker. This favorable treatment might wane the reform-induced rise in grandmother's labor supply because it was perceived as a way out of the labor market for the ones affected by the 2006 reform. However, we are not worried about the LCS plan as a confounding factor. First, both treated and control can use this new tax-facilitated saving scheme. If anything, the availability of the LCS plan makes our first stage estimates smaller. Moreover, in practice, only some high-wage workers manage to retire early using the LCS plan. Lindeboom and Montizaan (2020) shows around 15% of the 1950 cohort participated in the LCS plan, among which only 16% managed to counter the reform effect and maintain their previously planned retirement dates.

A.4 Childcare and primary education in the Netherlands

In the Netherlands, children aged 1 to 3 can go to center-based childcare and informal care. Childcare centers charge an hourly rate of between 6 and 8 euros on average.[3] From age 4 onwards, most children start primary school (mandatory at age 5) and at age 12 they go to secondary school. Primary schools are free of charge and provide around 30 to 35 hours of free care per week. The number of hours in school increases as children grow older.[4] School starts at around 8 am and ends at around 2 or 3 pm and at some schools finish early on Wednesday afternoons after the lunch break. In case families take the option of after-school (also called out-of-school) care (buitenschoolse opvang, OSC), which is generally provided by center-based out-of-school care providers, they need to pay for it. Parents who do not send their children to OSC, need to arrange other types of care. A portion of the daycare and after-school care costs is reimbursable for working parents. More specifically, the Dutch Childcare Allowance reimburses part of the childcare costs for dual-earner couples and single working parents who sent their children aged 0-12 years to registered daycare and after-school care facilities and certified childminders. Depending on gross household income, around 30 to 96 percent of the costs will be reimbursed.

At age 12, pupils at the vast majority of primary schools participate in an aptitude test called the Cito primary education final test (Cito Eindtoets Basisonderwijs, Cito test). Performance on the Cito test is one of the key determinants of the track the child attends in secondary education (such as vocational, technical and academic track).

B More Details on Data

The administrative records allow us to follow the entire Dutch population (more specifically, those individuals still alive in October 1994, when official records start being available). Basic demographics, labor market participation and the main source of income is available since 1994, detailed labor market histories including working hours, employment sector, and employment contract details are available since 2006. For the analysis of the third generation, we exploit official records of the "Cito" test results (nationwide standardized test) and data on childcare usage (both the type and hours) related to childcare subsidies, which are available from 2007 onward. Data availability does not represent a constraint for us, since our main sample period is between 2009 and 2015, when the grandparents born around January 1950 were between ages 60 and 64.

Summary statistics: Table A2 presents summary statistics. Columns 1 and 2 ("All") consist of all (extended) families (who are not necessarily living together) with grandmothers born between 1948 and 1951 who are Dutch, have worked at least one month in their lives, have not claimed disability insurance before age 55, have not exited the labor market before age 50, and are still alive at age 65. Columns 3 and 4 ("Full sample") restrict the sample to (extended) families with a

[3] In the Netherlands, mothers are entitled to fully-funded maternity leave 6 weeks before and 10 weeks after childbirth. Partners are entitled to two days of fully paid paternity leave at the time of childbirth, and they can extend this up to 5 weeks of unpaid leave (there were some changes to this in 2019). After childbirth, each parent can take up to 26 weeks of unpaid parental leave per child. The parental leave period can be taken at any time up to the 8th birthday of the child with flexibility in terms of the exact arrangement, either in blocks or several hours per week.

[4] According to the overview of teaching hours on the official Dutch government website, which provides information on Dutch central government policy, pupils must be taught at least 3,520 hours in the first four school years (lower secondary) compared to 3,760 hours in the last four school years (senior years).

youngest child aged 4 to 12 when their grandmothers are aged 60 to 64. Columns 5 and 6 ("RD sample") are the baseline analysis sample which is the full sample (of Columns 3 and 4) restricted to families with grandmothers who were born within the bandwidth of 8 months around January 1950. In the baseline RD sample, grandmothers have on average 2.5 adult children and 1.7 adult daughters, similar to the "Full sample" and the sample "All", since we condition on grandmothers having at least one daughter. The mothers in this sample are on average 38 years old, entered the labor market on average at age 25, had their first child at age 28, 66% are married, and they have on average two children. Since, in the Full sample and the RD sample, we condition on mothers having a child, mothers in these samples are slightly older, are more likely married, were younger when they had their first child, and have two children on average instead of one compared to the sample "All".

B.1 Data sources

Below we list the different data sources used in the analysis. All datasets used are provided by Statistics Netherlands (CBS). Documentation for each of the files below can be found at the embedded link. Please note that these are only available in Dutch.

CBS data: Personal background information, death and birth dates are combined using gpapersoontab, ogbaoverlijdentab, and kindoudertab. Linkages within households and information on the residence location come from gbahuishoudenbus, gbaadresobjectbus, and vslgwbtab. Labor market histories and income data are extracted from Official documentation of secmbus, integraal persoonlijk inkomen, kinderopvang, and spolisbus. Education degrees of mothers and children's academic performance is documented in CITOtab and hoogsteopltab.

LISS panel: The LISS panel is an online household panel. The panel consists of around 5,000 households in the Netherlands, comprising approximately 7500 individuals over the age of 16. The panel is based on a true probability sample of households drawn from the population register by Statistics Netherlands. Every year, a longitudinal survey is fielded in the panel, covering many domains, including health, work, education, income, housing, time use, political views, values, and personality. More information about the LISS panel can be found at: www.lissdata.nl

We use the first wave collected in 2008 and restrict our focus on parents (i.e., individuals with children) whose own mother (i.e., the grandmother) is still alive. Parents are asked about childcare arrangements separately for their children below age four and children aged 4-12 who do not attend secondary school yet. Parents of children were asked several questions in the Family and Household component of the survey. We make use of questions on the regular choice of childcare (questions cf238-cf248), and questions on labor supply adjustment in response to childcare responsibilities (questions cw446-cw449) from the Work and Schooling component. Following the embedded links, detailed information on the questions as well as LISS survey can be retrieved.

B.2 Sample selection

In Table A1 we illustrate our sample construction step-by-step and show that our sample restrictions are smooth around the RD cutoff. Starting with all native Dutch grandmothers born seven months around January 1950 with at least one adult daughter, we show that exiting the labor market before age 50 is smooth around the cutoff and not very common with a likelihood of 38% (step 1). We exclude inactive grandmothers and test in step 2 whether the probability of living up to age 65 differs by treatment. Among our sample, the death rate before age 65 is 2% and does not differ between treated and control grandmothers. We exclude the small fraction of deceased grandmothers. In step 3, we test whether there is evidence of self-selection based on restrictions in terms of health status. Among both treated and non-treated grandmothers, 8% claim disability insurance before age 55. After excluding grandmothers claiming disability before age 55, sample restrictions based on grandmothers' characteristics are complete.

To ensure the focus on the relevant sample, we make additional restrictions based on mothers' characteristics. Step 4 shows that almost 60% of mothers have a youngest child aged 4-12 when the grandmother is aged 60-64. Keeping only mothers with a youngest child aged 4-12 gives us a baseline sample of 23,497 mothers (and 19,548 grandmothers).

B.3 Linkages to Cito data

At the end of primary school, schools administer a standardized test, the *Cito test*, to determine the secondary school track children will be admitted to. Since the academic year 2014/15, schools can choose between three different tests, the most important of which is the central final test administered by Cito. It is important to note that the schools, not the parents or children, select the type of test. The administrative data includes information on the Cito test for schools that permitted Cito to pass on data to Statistics Netherlands. Overall 50% of our sample of children aged 4 to 12 can be matched. The Cito-sample uses the youngest children aged 4-12 to mothers of our baseline sample that can be matched to their Cito test results. As the Cito test result is a time-invariant outcome, we take the youngest child in the family who was between 4 and 12 years old when their grandmother was 60 years old. This results in a sample of children born between 1998 and 2006 for treated families and children born between 1999 and 2007 for control families.

Table A3 compares the characteristics of all youngest children aged 4-12 (see Columns (1) and (2)), to characteristics of the youngest children aged 4-12 in the Cito-sample (Columns (3) and (4)), and to the characteristics of the children in the Cito-sample aged 4-7 (Columns (5) and (6)) and aged 8-12 (Columns (7) and (8)). The characteristics of "all youngest children aged 4-12" and of those matched to their Cito outcomes are extremely similar, both in terms of child and family characteristics. Comparing youngest children aged 4-7 and youngest children aged 8-12 in the Cito-sample, we find that children in the 8-12 sample have a slightly higher likelihood of being firstborn, are more likely to have married parents, and have mothers that are slightly less educated.

Panel B of Table A4 tests whether the Cito-sample restrictions and matching rates differ by treatment status. We show that among all children aged 4-12 when the grandmother is aged 60, the likelihood of being the youngest child in the given age range and matched to Cito results are not affected by grandmothers' treatment status. These results provide evidence that the restricted data availability of test scores does not lead to selection problems. Further support is provided by the results in Table A21, which reports the estimated impact of a grandmother being treated (i.e. born since January 1950) on a list of predetermined characteristics of the Cito-sample. Treated and non-treated grandchildren are comparable in terms of children's and family characteristics (such as, for example, age, gender, birth order, mother's education, marital status etc.).

C Details of Additional Analysis

C.1 Complier characteristics

Even though compilers cannot be identified individually, Abadie (2003) provides a general method for recovering the distribution of covariates of compliers using the kappa-weighting scheme. Suppose D_{1i} and D_{0i} denote individual i's treatment status when $Z_i = 1$ and $Z_i = 0$, respectively. Then, the mean of a characteristic X_i for compliers, i.e., $D_{1i} > D_{0i}$, is given by $E[X_i|D_{1i} > D_{0i}] = \frac{\kappa_i X_i}{\kappa_i}$. The kappa weight is obtained via $\kappa_i = 1 - \frac{D_i(1-Z_i}{1-P(Z_i=1|X_i)} - \frac{(1-D_i)Z_i}{P(Z_i=0|X_i)}$. In the case of binary characteristics, Angrist and Pischke (2009) show that instead of kappa-weighting, complier characteristics can be described by the ratio of the first-stage among those with a certain characteristic over the whole sample. Which means, $\frac{P(X_i=1|D_{1i}>D_{0i})}{P(X_i=1)} = \frac{E[D_i|Z_i=1,X_i=1]-E[D_i|Z_i=0,X_i=1]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$. Since our main outcome of interest is the total labor supply of grandmothers, we define compliance as working on average more than the median working hours of 34.41 hours per month when aged 60 to 64. Moreover, defining compliance using the employment probability leads to a similar complier distribution.

In Table A6 we report in column (1) the distribution of characteristics among the baseline sample, in column (2) the distribution of characteristics among the compliers is reported, and column (3) shows the relative distribution of compliers among the baseline sample. Comparing compliers to the entire baseline sample, we find that the two samples are very comparable. The only noticeable difference is that compiler grandmothers are more attached to the labor market, as measured by their employment probability prior to the reform in 2006. Mothers among compiler families are on average more likely to have higher education as measured by some college experience. Also, complier families are less likely to have grandmothers who are cohabiting and have on average fewer grandchildren prior to the reform compared with the overall sample.

C.2 Heterogeneity by Socioeconomic Status

In Table A20, we explore whether the direct and indirect effects of the pension reform vary by socioeconomic status (SES) of the (extended) family. We define socioeconomic status by the education level of the mothers and by whether grandmothers' income prior to the reform (when she is aged 53 to 56) is above or below the median. Because data on education are limited for the grandmothers' cohorts, we use mother's educational level as a proxy for the socio-economic status of the extended family.

Columns (1) and (2) of Table A20 explore heterogeneous impacts by whether the mother has higher education. Higher education is defined as having attended some form of higher education, including universities and universities of applied sciences (higher professional education). We find that only grandmothers with highly educated adult daughters increase their labor supply, by 9 hours per month. This may be due to the fact that grandmothers with more educated daughters are themselves more educated as well. Highly educated women are likely to be more strongly attached to the labor force, working in an environment where extending employment is easier. Education is also an important indicator for knowledge of the pension reform. Consistently, we find that only mothers with higher education are affected in their labor supply (although the difference between the education groups is not significant). In particular, higher educated mothers work on average 8.5 hours less per month in response to the reform and are on average 4.4 p.p. less likely to be employed and 2.6 p.p. less likely to be full-time employed.

In Columns (3) and (4), we investigate the reform effects depending on the socioeconomic status of the grandmother, proxied by her income between the ages of 53 and 56. Similarly, we find that direct and spillover effects of the reform are larger for higher income grandmothers. High SES grandmothers work on average 12 hours more per month in response to the reform, while their adult daughters work on average 7 hours less and are on average 2 p.p. less likely to be full-time employed. Again, this is likely to be due to the fact that higher SES grandmothers are more aware of the reform and more attached to the labor market, and therefore react more strongly to the pension reform.

C.3 Reform effects on child penalty

We build on the framework developed by Kleven et al. (2019a) and estimate the following regression separately by gender (g) and treatment status (d):

$$Y_{ist}^{gd} = \sum_{j \neq -1} \alpha_j^{gd} I[t = j] + \sum_k \beta_k^{gd} I[age = k] + \sum_s \gamma_s^{gd} I[year = s] + \nu_{ist}^{gd} \tag{A1}$$

Hereby Y_{ist}^{gd} denotes the labor market outcome of individual i, in calendar year s, at event time t. The first term captures a full set of event time dummies, where event time $t = 0$ marks the birth of the first child. We exclude $t = -1$ so that the coefficients measure the impact of the first child relative to the year before birth. To control for life-cycle and time trends, the second and third terms include sets of dummies for the age of individual i and calendar year, respectively. Conditional on age and year, there is variation in the age at first childbirth, which identifies the effects of all three sets of dummies (see Kleven et al. (2019a) for details of the method).

Since our main interest lies in measuring changes in total labor supply (total monthly hours worked), we keep zeros (i.e., non-participation), and specify Equation A1 in levels. First, we estimate the effect of children on men and women separately by converting estimated level effects into percentages: $P_t^{gd} = \frac{\hat{\alpha}_t^{gd}}{E[\tilde{Y}_{ist}^{gd}|t]}$ with $\tilde{Y}_{ist}^{gd}$ capturing the predicted labor market outcome without the contribution of the event time dummies (i.e., excluding the first term from Equation A1). This transformation allows us to interpret P_t^{gd} as the percentage loss of average labor market outcomes due to having a child that individual i of gender g with treatment status d experiences.

Second, to compare penalties between women and men, we calculate the relative child penalty, P_t^d, measuring the relative loss women experience at event time t due to children: $P_t^d = \frac{\hat{\alpha}_t^{md} - \hat{\alpha}_t^{wd}}{E[\tilde{Y}_{ist}^{wd}|t]}$.

D Calculation of Marginal Value of Public Fund

To provide a comprehensive assessment of the benefits of pension reforms incentivizing later retirement relative to the costs, we follow the framework proposed by Hendren and Sprung-Keyser (2020) to calculate the Marginal Value of Public Funds (MVPF). The MVPF is the ratio of society's willingness to pay for incentivizing later retirement to the net cost to the government of implementing this policy.

At first glance, it seems unnecessary to calculate the MVPF for the policy of incentivizing later retirement, because the government's budget constraint is expanded mechanically by a less generous pension and behaviorally by the resulting prolonged working life of the elderly. However, we find that adult daughters reduce their labor supply due to the pension reform, which could potentially offset the gain in the government's budget. Therefore, we calculate the MVPF for the policy of incentivizing later retirement, taking into account grandmothers' and mothers' labor supply responses (as well as of other family members).

The net cost of a policy incentivizing later retirement is the sum of the net costs of the government or -in this case- benefits saved by the government plus any fiscal externalities from both the grandmother and the indirect impacts on her family members. The mechanical net cost of incentivizing later retirement is the change in pension generosity. By reducing pension replacement rates, the government saves €7013 per person. More specifically, the 2006 reform reduced pension replacement rates between ages 60 and 64 from 70% to 64% (Lindeboom and Montizaan (2020) Table A.1). For a typical woman with average monthly labor earnings of 573 euros and an average pension claim duration of 17 years,[5] we calculate that the government saves about €7013 per person (6%*573*12*17). The government saves €7013 per person by paying less generous pension.

The behavioral costs consist of the direct impact on grandmothers and the indirect spillover effects on mothers. First, the government saves additional €1833 per person on pension payments due to delayed claiming by around 5 months. The government collects 489 euros more income taxes per person per year, because grandmothers earn 117 euros more per month between age 60 and 64 (Table A8), using the Dutch Income Tax Calculator. The total increase in tax income is €2445 over these five years. Moreover, the government saves additional €1833 per person on pension payments due to delayed claiming by around 5 months.

Second, the government forgoes additional income taxes of 1585 euros per person due to negative spillover effects on the earnings of their adult-daughters. We find that mothers whose youngest child is between 4 and 12 years old earn 58 euros less between ages 60 and 64. Using the average monthly labor earnings around the cutoff of 2064 euros, we show that the government losses 317 euros per person per year, which adding up 1585 euros per person during these five years.

Therefore, the net fiscal cost is negative. The government saves €9706 (€7013 + €1833 + €2445 - €1585). The total net cost is negative. The government gains 9706 per family of an affected grandmother with adult daughter whose youngest child is between age 4 and 12.

Although not the main focus of our paper, we can also include the spillovers on the sons-in-law, who experience an increase in earnings, thus might offset the loss in income taxes due to adult-daughters working longer.

Note that if we consider the long-run reform effects on mothers' lifetime income and assume that government only cares about income tax revenue, the loss in tax revenue due to the drop in maternal labor supply would outweigh the gain in tax revenue from delaying the retirement of the grandmothers if the impact on maternal labor supply lasts for up to eight years. However, when taking into account the changes in pension payments, the net costs remain negative.

Furthermore, we calculate the MVPF from a perspective of the affected grandmother's family. If we consider the family of the affected grandfathers, the net costs will be even larger, as grandfathers delay retirement more and the spillover effects on his daughter are smaller.

The willingness to pay of the grandmother for a policy incentivizing later retirement is the sum

[5] The duration of pension claim is the length of the period between pension claim age (65 years old) and death (life expectancy of 82 years old).

of what would be their forgone public pension (- €7013 - €1833) and their increased earnings from working more (€7020=117*12*5), for a total of -€1826.

To account for the spillover effect on mothers, we also include the willingness to pay of the adult daughters whose youngest child is between 4 and 12 years. Their WTP for a policy incentivizing later retirement is the reduced earnings of - €3480 (58*12*5).

The total willingness to pay per person is then - €1826- €3480 = -€5306. In other words, the family would be willing to pay 5306 euro to avoid the policy incentivizing later retirement.

If we were to also add the willingness to pay of the sons-in-law, who experienced an increase in earnings. The total willingness to pay would be smaller but still remain negative.

To sum up, the estimated MVPF of a policy incentivizing later retirement is then (- €5306/ -€9706)) = 0.55. The paper highlights the importance of incorporating the spillover effects by noting the MVPF would be smaller (-€1826/-€11291)=0.16 if the spillover effects on adult daughters are excluded from the calculation. In our case, although the fiscal gain is smaller when considering the spillover effects, the WTP to avoid this policy also gets larger.

Note that we could also include the spillover impacts on grandchildren in the MVPF calculation. However, we find that children treated when they are young (ages 4-7) perform better, and boys treated when older (ages 8-12) perform worse. Therefore, we do not include the spillovers on grandchildren on tax revenue. Based on the early childhood development literature, we know that the return to investment on younger children is higher and can have a long-term impact on their lifetime earnings. Moreover, young girls and boys are positively affected, while the negative effect on older children is only on boys. Therefore, if we were to include the spillover effects on grandchildren, we would expect the impact on government tax revenue, in the long run, to be positive or at least non-negative due to the reform. The grandchildren's willingness to pay for policy incentivizing later retirement is positive.

To conclude, our MVPF exercise highlights the importance of consider the spillover effects across generations in cost-benefit analyses in order to optimally design public policies. Moreover, the different types of spillover effects point to the possibility of complementing the original policy with additional policies counteracting the unintended "side effects", such as –say- complementing an early retirement reform with better access to high-quality childcare.

E Appendix Tables and Figures

Table A1: Impacts on sample selection

	RD estimates (1)	Mean at cutoff
Restrictions by Grandmothers' characteristics		
Step 1: reform relevance		
Exit labor force before age 50	0.004	0.385
	[0.012]	[0.487]
Obs. Mothers	55525	12307
Step 2: alive during treatment period		
Dead before age 65	0.002	0.023
	[0.005]	[0.148]
Obs. Mothers	34270	7564
Step 3: health status/ relevance for care responsibility		
Claim disability before age 55	0.000	0.081
	[0.008]	[0.272]
Obs. Mothers	33414	7394
Restrictions by Mothers' characteristics		
Step 4: Keep by relevance of childcare		
Youngest child aged 4-12	-0.005	0.595
	[0.013]	[0.491]
Obs. Mothers	30649	6799

Source: Authors' calculations from the CBS data.
Note: This table tests the impact of grandmothers being born since January 1950 on a list of sample selection variables. Step 1 is based on all women (grandmothers) born 7 months around the January 1950 cutoff who have at least one daughter. In step 2, we show that for all grandmothers with at least one adult daughter and still in the labor force by age 50, the probability of death before age 65 is smooth around the RD cutoff. Each further step builds on the previous one. Steps 1-3 test groups to drop from the sample and step 4 tests for groups to keep in the baseline sample. Regressions are estimated via a local linear regression with an optimal bandwidth of 7 months. For the full list of controls, see section 2.3. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Summary statistics

	All		All Mothers		Baseline sample		RD sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Grandmothers' characteristics								
Birth cohort	1949.51	[1.121]	1949.50	[1.119]	1949.48	[1.117]	1949.53	[0.499]
Age	62.516	[0.292]	62.73	[0.583]	62.878	[1.110]	62.958	[1.122]
Number of adult children	2.510	[1.016]	2.518	[1.008]	2.475	[0.982]	2.469	[0.982]
Number of adult daughters	1.729	[0.800]	1.734	[0.798]]	1.712	[0.785]	1.701	[0.782]
Mothers' characteristics								
Age	35.254	[4.521]	36.484	[3.723]	37.884	[2.938]	37.923	[2.956]
Age at first child birth	29.691	[4.200]	29.175	[3.930]	28.337	[3.477]	28.368	[3.531]
Age at first employment	23.589	[3.916]	24.087	[3.996]	24.855	[3.888]	24.734	[3.733]
Married	0.466	[0.470]	0.621	[0.463]	0.661	[0.460]	0.657	[0.461]
Age gap to partner	2.663	[4.395]	2.635	[4.026]	2.809	[3.995]	2.806	[4.009]
Number of children	1.643	[1.100]	2.141	[0.788]	2.141	[0.786]	2.146	[0.790]
Education not missing	0.729	[0.444]	0.712	[0.453]	0.684	[0.465]	0.689	[0.463]
Higher education (some college)	0.607	[0.448]	0.585	[0.493]	0.576	[0.492]	0.573	[0.495]
Outcomes: grandmothers' labor supply								
Monthly hours worked	44.023	[48.245]	39.853	[47.299]	37.230	[48.712]	37.427	[48.651]
Prob (Employed)	0.474	[0.419]	0.438	[0.423]	0.415	[0.446]	0.418	[0.447]
Prob (Full-time employed)	0.067	[0.212]	0.058	[0.200]	0.055	[0.204]	0.053	[0.201]
Monthly labor income	803.52	[1019.71]	710.923	[968.142]	638.99	[947.15]	637.59	[934.17]
Monthly gross income	1635.15	[1552.12]	1534.39	[1472.44]	1419.83	[1344.81]	1394.20	[1313.26]
Monthly HH labor earnings	1725.80	[1966.94]	1520.79	[1846.43]	1307.63	[1739.54]	1280.61	[1709.23]
Age at exiting employment	61.103	[4.304]	60.960	[4.314]	60.812	[4.388]	60.777	[4.453]
Age at claiming pension	63.039	[3.171]	62.981	[3.186]	62.954	[3.230]	62.916	[3.182]
Outcomes: mothers' labor supply								
Monthly hours worked	97.255	[51.482]	85.706	[47.630]	78.498	[47.458]	78.149	[47.352]
Prob (Employed)	0.816	[0.334]	0.801	[0.350]	0.784	[0.377]	0.783	[0.378]
Prob (Full-time employed)	0.232	[0.355]	0.105	[0.247]	0.063	[0.209]	0.061	[0.205]
Monthly labor income	1844.71	[1219.59]	1697.11	[1213.47]	1531.13	[1188.53]	1533.07	[1193.45]
Monthly HH labor income	4280.80	[2477.81]	4695.58	[2585.33]	4525.05	[2677.43]	4556.66	[2697.82]
Obs. Mothers	147858		105798		66252		20711	
Obs. Grandmothers	106036		81725		55055		17256	

Source: Authors' calculations from the CBS data.

Note: This table reports means and standard deviations. Columns 1 and 2 consist of all (extended) families – not necessarily living in the same household – with grandmothers born between 1948 and 1951 with at least one adult daughter, who are Dutch, have worked at least one month in their lives, have not claimed disability insurance before age 55, and who are still alive by age 65. Columns 3 and 4 restrict the sample in addition to the adult daughters having at least one child, i.e., being mothers, when the grandmother is aged 60 to 64. Columns 5 and 6 additionally restrict the sample to (extended) families with grandmothers with the youngest grandchild aged 4-12 when the grandmother is between 60 and 64. Columns 7 and 8 are the RD sample, which is the sample of Columns 5 and 6 restricted to families with grandmothers born within a bandwidth of 7 months before and after January 1950. Grandmothers' and mothers' labor supply is measured when the grandmother is between the ages of 60 and 64. All income measures are CPI-adjusted for the year 2015.

Table A3: Summary statistics of children

	Youngest child aged 4-12		Cito-sample: Youngest child aged 4-12		4-7		8-12	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Children's characteristics								
Birth cohort	2002.70	[2.146]	2002.55	[2.070]	2004.04	[0.979]	1999.98	[1.212]
Age (when GM is aged 60)	6.837	[2.090]	6.974	[2.019]	5.475	[0.874]	9.538	[1.111]
Girl	0.494	[0.500]	0.508	[0.500]	0.507	[0.500]	0.509	[0.500]
Size of total sibship	2.004	[0.751]	1.993	[0.739]	2.027	[0.734]	1.909	[0.724]
Birthorder	1.942	[0.724]	1.932	[0.716]	1.957	[0.712]	1.878	[0.703]
Prob(First-born child)	0.262	[0.440]	0.264	[0.441]	0.247	[0.431]	0.290	[0.454]
Age at Cito	-	-	11.912	[0.449]	11.838	[0.429]	12.028	[0.467]
Cito year	-	-	2014.84	[2.006]	2016.26	[1.001]	2012.39	[1.229]
Obs Children	9147		5835		3241		1751	
Mothers' characteristics								
Prob(married)	0.579	[0.443]	0.589	[0.441]	0.516	[0.435]	0.718	[0.419]
Live in same municipality as GM	0.559	[0.470]	0.558	[0.472]	0.542	[0.469]	0.591	[0.473]
Age gap to partner	3.007	[4.059]	3.001	[3.932]	2.946	[3.775]	3.058	[4.074]
Age (when GM 60-64)	38.634	[2.779]	38.761	[2.729]	37.233	[2.677]	39.222	[2.591]
Number siblings	2.429	[0.949]	2.410	[0.958]	2.424	[0.943]	2.391	[0.978]
Number sisters	1.680	[0.764]	1.676	[0.772]	1.684	[0.771]	1.659	[0.777]
Age at first birth	27.065	[3.202]	27.143	[3.106]	27.819	[3.047]	26.041	[2.935]
Education not missing	0.637	[0.481]	0.626	[0.484]	0.638	[0.481]	0.603	[0.490]
Higher education (some college)	0.436	[0.496]	0.451	[0.498]	0.528	[0.499]	0.318	[0.466]
Obs. Mothers	9147		5835		3241		1751	

Source: Authors' calculations from the CBS data.

Note: This table reports means and standard deviations for the samples used in the analysis for child outcomes. Columns 1 and 2 consist of the youngest children aged 4-12 of mothers in the baseline sample (i.e., with grandmothers born 7 months before and after January 1950). Columns 3 and 4 restrict the sample to the youngest children aged 4-12 which can be linked to their Cito test scores. Columns 5 to 8 summarize the characteristics of the youngest children which can be linked to their Cito test scores and are aged 4-7 and 8-12, respectively. The probability of parents being married, and living in the same municipality as the grandmother are predetermined (i.e., measured when grandmothers are aged 50-53).

Table A4: Impacts on sub-sample selection

	RD estimates (1)	Mean at cutoff
Panel A: Sub-sample selection criteria		
Restrictions by age of youngest child		
Youngest aged 1-3	-0.007	0.294
	[0.007]	[0.285]
Youngest aged 4-7	-0.001	0.525
	[0.013]	[0.499]
Youngest aged 8-12	0.007	0.293
	[0.012]	[0.455]
Youngest aged 13-18	-0.003	0.130
	[0.009]	[0.336]
Obs. Mother	34623	6663
Panel B: Cito test score availability		
Youngest child 4-7 and Cito available	-0.014	0.218
	[0.012]	[0.413]
Obs. Children	13597	3077
Obs. Mother	8579	1866
Youngest child 8-12 and Cito available	-0.014	0.236
	[0.022]	[0.425]
Obs. Children	6417	1454
Obs. Mother	5048	840
Youngest child 4-12 and Cito available	-0.016	0.255
	[0.012]	[0.436]
Obs. Children	20345	4605
Obs. Mother	12575	2845

Source: Authors' calculations from the CBS data.
Note: Panel A tests the impact of grandmothers being born since January 1950 on a list of sub-sample selection variables. All regressions are based on the sample selected after completing Steps 1 to 4 displayed in Table A1. Panel B tests the impact of grandmothers being born since January 1950 on a list of Cito-sample selection variables. Regressions are based on the youngest child in a given family aged 4-7, 8-12, and 4-12 when the grandmother is aged 60. Column 1 is estimated via a local linear regression with a bandwidth of 7 months. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Impact on mothers' labor supply (other outcomes)

	Fuzzy RD estimates			
	(1)	(2)	(3)	Means at cutoff
Panel A: Reduced-form estimates				
Prob(Employed)	-0.021*	-0.023*	-0.025**	0.785
	[0.012]	[0.013]	[0.013]	[0.378]
Optimal bandwidth	8.143	6.829	6.811	
Obs. Mothers	23497	17930	17930	4018
Prob(Full-time employed)	-0.010	-0.010	-0.010	0.066
	[0.007]	[0.007]	[0.007]	[0.214]
Optimal bandwidth	8.112	7.519	7.518	
Obs. Mothers	23497	20711	20711	4018
Panel B: Fuzzy RD estimates				
First-stage:				
Total monthly hours worked (GM)	7.043***	7.059***	6.222***	34.418
	[1.660]	[1.606]	[1.538]	[47.608]
LATE:				
Prob(Employed)	-0.003*	-0.003*	-0.003*	0.785
	[0.002]	[0.002]	[0.002]	[0.378]
Optimal bandwidth	8.354	7.807	7.915	
F-Stat	17.997	19.317	16.359	
Obs. Mothers	23497	20711	20711	4018
Prob(Full-time employed)	-0.002	-0.001	-0.002	0.066
	[0.001]	[0.001]	[0.001]	[0.214]
Optimal bandwidth	8.640	8.572	8.764	
F-Stat	18.54	21.18	18.40	
Obs. Mothers	23497	23497	23497	4018
Controls	NO	YES	YES	
Sector FE	NO	NO	YES	

Source: Authors' calculations from the CBS data.
Note: Panel A reports reduced-form reform impact on mothers' probability of being employed and of being full-time employed. Panel B reports 2SLS fuzzy RD estimates (incl. first- and second-stage) of the impact of grandmothers' labor supply on mothers' labor supply. The running variable is the grandmother's birthdate (month-year), centered around January 1950. Columns 1, 2, and 3 show the results without controls, with controls, and with both controls and sector fixed effects, respectively. For a description of the full list of controls, see section 2.3. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure. Sample means at the cutoff (measured in the three months before the cutoff) are reported in Column 4. All outcomes are measured when the grandmothers are between the ages of 60 and 64. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Characteristics of compliers

	$P[X = x]$	$P[X = x \mid complier]$	$\frac{P[X=x \mid complier]}{P[X=x]}$
Grandmothers' characteristics			
Age	62.958	63.109	1.002
Number of adult children	2.469	2.304	0.933
Number of adult daughters	1.701	1.600	0.941
Prob(Married)	0.829	0.838	1.011
Prob(Cohabit)	0.039	0.035	0.875
Prob(disabled partner)	0.068	0.066	0.966
Birth cohort of partner	1947.63	1945.16	0.999
Pre-determined Prob(Employed)	0.802	1.000	1.247
Pre-determined total gross income (53-56)	16409.62	21802.29	1.329
Mothers' characteristics			
Age	37.923	38.001	1.002
Birth cohort	1974.58	1974.38	1.000
Prob(Married)	0.361	0.330	0.914
Live in same neighborhood as GM	0.237	0.233	0.987
Age at first birth	28.368	28.693	1.011
Age at first employment	24.734	24.630	0.996
Number of children (when GM 55)	0.826	0.729	0.884
Education not missing	0.689	0.658	0.981
Higher education (some college)	0.573	0.710	1.139
Pre-determined Prob(Employed)	0.774	0.778	1.006

Source: Authors' calculations from the CBS data.

Note: This table shows the marginal distribution in column 1, the complier distribution in column 2, and the relative likelihood of different subgroups in column 3. The sample is restricted to mothers with a youngest child aged 4-12 (when the grandmother is aged 60-64) focusing on a bandwidth of 7 months around the cutoff. Compliance is defined as having a grandmother who is working on average more than the median working hours (34.41) per month when she is aged 60-64. Pre-determined employment (income) refer to the employment probability (income) when the grandmother is aged 50 to 53.

Table A7: Balance test: reform impacts on covariates

	RD estimates (1)	Mean at cutoff
Grandmothers' characteristics		
Age	0.044	62.934
	[0.041]	[1.144]
Number of adult children	0.027	2.451
	[0.042]	[0.892]
Number of adult daughters	0.002	1.712
	[0.036]	[0.761]
Prob (Employed)	0.014	0.793
	[0.017]	[0.338]
Prob (Married)	0.017	0.835
	[0.017]	[0.355]
Prob (Partner disabled)	0.016*	0.062
	[0.009]	[0.235]
Birthcohort of partner	-0.069	1947.50
	[0.184]	[3.849]
Total gross income (when aged 53-56)	236.20	16189.69
	[587.42]	[14291.44]
Mothers' characteristics		
Age	0.116	37.867
	[0.099]	[2.899]
Birth cohort	-0.072	1974.45
	[0.112]	[3.211]
Prob (Married)	0.003	0.369
	[0.014]	[0.434]
Prob (Employed)	-0.007	0.772
	[0.011]	[0.333]
Live in the same neighborhood as GM	-0.019	0.244
	[0.013]	[0.377]
Age at first child birth	0.147	28.268
	[0.137]	[3.609]
Age of youngest child	0.026	2.059
	[0.086]	[2.022]
Age of oldest child	-0.081	3.790
	[0.153]	[3.164]
Number of children	-0.012	0.842
	[0.033]	[0.969]
Age of first employment	0.047	24.891
	[0.142]	[3.809]
Education not missing	-0.001	0.683
	[0.016]	[0.465]
Higher education (some college)	0.022	0.564
	[0.021]	[0.496]
Average optimal bandwidth	6.516	
Average Obs. Mothers	16388	4018

Source: Authors' calculations from the CBS data.
Note: In this table, we test the impact of grandmothers being born since January 1950 on a list of the grandmothers' and mothers' characteristics. All variables are predetermined and refer to the time when the grandmothers were 50 to 53 years old before the announcement of the reform. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Reform impacts on grandmothers' labor supply

	RD estimates			Means at
	(1)	(2)	(3)	cutoff
Total monthly hours worked	7.396***	7.771***	6.750***	34.418
	[1.910]	[1.927]	[1.795]	[47.608]
Optimal bandwidth	6.482	5.652	6.061	
Obs. Grandmothers	17930	15156	17930	4005
Prob(Employed)	0.070***	0.069***	0.057***	0.387
	[0.018]	[0.018]	[0.015]	[0.438]
Optimal bandwidth	6.435	6.000	6.832	
Obs. Grandmothers	17930	17930	17930	4005
Prob(Full-time employed)	0.007	0.008	0.008	0.054
	[0.007]	[0.007]	[0.007]	[0.202]
Optimal bandwidth	8.883	7.242	7.746	
Obs. Grandmothers	23497	20711	20711	
Monthly labor income	133.281***	138.981***	116.864***	573.065
	[34.150]	[34.544]	[32.179]	[887.669]
Optimal bandwidth	6.842	6.023	6.196	
Obs. Grandmothers	17930	17930	17930	4005
Monthly HH labor income	119.237*	133.746**	111.339*	1211.88
	[64.055]	[64.010]	[62.639]	[1647.46]
Optimal bandwidth	6.438	6.256	6.304	
Obs. Grandmothers	17930	17930	17930	4005
Monthly gross income	80.511	65.252	41.409	1361.66
	[53.941]	[40.415]	[37.412]	[1304.62]
Optimal bandwidth	6.002	8.186	8.348	
Obs. Grandmothers	17860	23406	23406	4005
Monthly gross HH income	19.910	28.587	7.595	4082.91
	[88.124]	[86.657]	[86.579]	[2144.97]
Optimal bandwidth	6.150	6.038	5.711	
Obs. Grandmothers	17860	17860	15093	4005
Controls	NO	YES	YES	
Sector FE	NO	NO	YES	

Source: Authors' calculations from the CBS data.

Note: This table shows the reduced-form impacts on grandmothers' labor supply and income measures. Columns 1, 2, and 3 show the results without controls, with controls, and with both controls and sector fixed effects, respectively. For a description of the full list of controls, see section 2.3. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure. All outcomes are measured when the grandmothers are between the ages of 60 and 64. All income measures are CPI-adjusted for the year 2015. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A9: Fuzzy RD estimates for different polynomial orders

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Reduced-form estimates -						
Mothers' labor supply outcomes						
Total monthly hours worked	-4.912***	-5.348***	-3.234**	-5.132***	-5.094**	-5.963**
	[1.779]	[2.001]	[1.608]	[1.781]	[2.059]	[2.527]
AIC	159332	275546	348906	291697	232486	174621
BIC	159660	275914	349049	291837	232622	174752
AICs	159332	275546	348906	291697	232486	174621
Panel B: Fuzzy RD estimates -						
First-stage: Impact on Grandmother's total labor supply						
Total monthly hours worked	6.367***	6.644***	6.520***	6.475***	6.227***	7.592***
	[1.589]	[2.131]	[1.712]	[1.900]	[2.184]	[2.683]
LATE: Impact on mothers' labor supply						
Total monthly hours worked	-0.630**	-0.805**	-0.496*	-0.793**	-0.818*	-0.785*
	[0.283]	[0.399]	[0.281]	[0.363]	[0.441]	[0.436]
AIC	224314	288137	356145	304778	243464	182289
BIC	224656	288505	356279	304910	243592	182321
AICs	224314	288137	356145	304778	243464	182189
Polynomial	LLR	LQR	2	2	2	2
Bandwidth	7.324	9.697	12	10	8	6
F-Stat	16.058	9.725	14.509	11.607	8.134	8.005
Obs. Mothers	20711	66525	33213	27767	22132	16623
Controls	YES	YES	YES	YES	YES	YES
Sector FE	YES	YES	YES	YES	YES	YES

Source: Authors' calculations from the CBS data.

Note: Panel A reports reduced-form estimates on mothers' labor supply outcomes. Panel B shows the Fuzzy RD (2SLS) estimates of grandmothers' total monthly hours worked on mothers' labor supply outcomes using different bandwidths and polynomial orders. The running variable is the grandmother's birthdate (month-year), centered around January 1950. All outcomes are measured when the grandmother is between 60 and 64. Column 1 is estimated via a local linear regression (LLR), column 2 is estimated based on a local quadratic regression (LQR), and columns 3 to 6 are estimated using a quadratic spline with different bandwidths. Columns 1 and 2 are based on the optimal bandwidth generated by the Calonico et al. (2014) procedure, columns 3 through 6 use the indicated bandwidths. All columns consider mothers with a youngest child aged 4-12 when the grandmother is 60-64. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A10: Placebo test: reform impacts on labor supply of mothers with inactive grandmothers

	RD estimates (1)	(2)	(3)	Means at cutoff
Impact on Grandmother's labor supply				
Total monthly hours worked	0.058	0.071	0.020	0.373
	[0.155]	[0.149]	[0.135]	[3.225]
Optimal bandwidth	7.437	7.795	5.967	
Obs. Grandmothers	15174	15174	11035	
Impact on mothers' labor supply				
Total monthly hours worked	-1.801	-1.204	-1.307	67.530
	[2.037]	[1.971]	[1.959]	[48.821]
Optimal bandwidth	6.708	6.657	6.742	
Obs. Mothers	13079	13079	13079	
Prob(Employed)	-0.014	-0.010	-0.010	0.712
	[0.019]	[0.019]	[0.019]	[0.422]
Optimal bandwidth	5.907	5.963	5.966	
Obs. Mothers	11035	11035	11035	
Prob(Full-time employed)	0.003	0.003	0.003	0.046
	[0.007]	[0.007]	[0.007]	[0.177]
Optimal bandwidth	6.605	6.510	6.647	
Obs. Mothers	13079	13079	13079	
Controls	NO	YES	YES	
Sector fixed effects	NO	NO	YES	

Source: Authors' calculations from the CBS data.
Note: This table shows the reduced-form impacts on adult daughters (mothers) whose mothers (grandmothers) exited the labor force by age 50. Column 1 shows results without any controls, Column 2 includes controls, and Column 3 includes controls and sector fixed effects. For a description of the full list of controls, see section 2.3. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A11: DID estimation (reduced-form)

	(1)	(2)	(3)	(4)	(5)
Panel A: Main DID					
Impact on Mother's Total monthly hours worked					
GM born since 1950 * Youngest child 4-12	-2.448**	-3.863***	-4.835***	-5.390***	-6.926***
	[1.061]	[1.089]	[1.164]	[1.406]	[1.381]
Youngest child 4-12	2.262***	2.977***	3.805***	3.688***	5.513***
	[0.659]	[0.708]	[0.831]	[0.933]	[0.987]
Sample window (left/right of cutoff)	12	10	8	6	4
Obs.	40319	33683	26822	20178	13335
Panel B: Placebo DID					
Impact on Mother's Total monthly hours worked					
GM born since 1949 * Youngest child 4-12	-0.497	-0.186	0.904	-0.652	2.472
	[1.061]	[1.214]	[1.314]	[1.503]	[2.025]
Youngest child 4-12	0.149	-0.019	-1.021	-0.029	-1.746
	[0.654]	[0.765]	[0.838]	[0.819]	[1.472]
Sample window (left/right of cutoff)	12	10	8	6	4
Obs.	41002	34274	27599	20677	13693

Source: Authors' calculations from the CBS data.

Note: Panel A reports coefficients from a DID estimation, where the treatment indicator is the interaction between "GM born since 1950" and "Youngest child 4-12". Panel B reports results from a placebo test based on the interaction between "GM born since 1949" and "Youngest child 4-12". "GM born since 1950 (1949)" takes the value 1 when the grandmother is born since January 1950 (1949) and 0 otherwise. "Youngest child 4-12" takes the value 1 when the grandmother's adult daughter has a youngest child aged 4-12 and 0 when the grandmother's adult daughter has a youngest child aged 13-18. In all regressions, we control for grandmother month-by-year of birth fixed effects and allow for a different linear trend for the treated group (i.e., mothers with the youngest child aged 4-12). We display results for different sample windows around the cutoff. A sample window of 12 implies that the sample contains grandmothers born 12 months before to 12 months since January 1950 (Panel A) or since January 1949 (Panel B). Standard errors (two-way clustered at treatment status and grandmother's birthdate (month-by-year)) are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A12: RD - DID estimates using families with older children

	First-stage: Impact on GM's labor supply			LATE: Impact on mothers' labor supply		
	(1)	(2)	(3)	(4)	(5)	(6)
Main sample (youngest child 4-12)	7.080***	5.976***	6.358***	-0.553**	-0.671**	-0.627**
	[1.720]	[1.598]	[1.590]	[0.249]	[0.300]	[0.281]
Youngest child 13 - 18	8.858**	6.602*	6.182*	0.183	0.045	0.046
	[3.453]	[3.170]	[3.178]	[0.397]	[0.439]	[0.456]
RD-DID	-1.778	-0.627	0.177	-0.735*	-0.716*	-0.673
	[3.214]	[3.020]	[3.037]	[0.412]	[0.420]	[0.425]
Optimal bandwidth	7.952	7.999	7.324	7.952	7.999	7.324
Observations	25103	25103	25103	25103	25103	25103
Controls	NO	NO	YES	NO	NO	YES
Sector FE	NO	YES	YES	NO	YES	YES

Source: Authors' calculations from the CBS data.

Note: This table reports RD-DID estimates testing the difference in effects between the main sample of mothers with the youngest child aged 4 to 12 and the placebo group of mothers with older children. Columns (1) to (3) report the first-stage results on grandmothers' total monthly hours worked, and Columns (4) to (6) report the corresponding LATE estimates of mothers' total monthly hours worked. The last row reports the difference in the effects on the main sample (mothers with youngest child aged 4-12) as opposed to the placebo group results (mothers with a youngest child older than 12). All outcomes are measured when the grandmothers are between 60 and 64. The running variable is the grandmother's birthdate (month-year), centered around January 1950. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A13: RD - DID estimates using placebo cohorts of grandmothers

	Total monthly hours worked of Grandmothers			Total monthly hours worked of Mothers		
	(1)	(2)	(3)	(4)	(5)	(6)
Reform cutoff (Jan. 1950)	7.396***	6.137***	6.634***	-4.443**	-4.812***	-5.001***
	[1.909]	[1.740]	[1.800]	[1.767]	[1.766]	[1.782]
Placebo Jan. cutoffs	0.658	1.058	1.572*	-0.089	0.009	-0.386
	[0.957]	[0.870]	[0.912]	[0.893]	[0.893]	[0.900]
RD-DID	6.738***	5.079***	5.062***	-4.354**	-4.821**	-4.615**
	[2.139]	[1.949]	[2.018]	[1.979]	[1.979]	[1.996]
Optimal bandwidth	6.482	6.880	6.061	5.882	5.818	5.347
Observations	90640	90640	90640	76414	76414	76414
Controls	NO	NO	YES	NO	NO	YES
Sector FE	NO	YES	YES	NO	YES	YES

Source: Authors' calculations from the CBS data.

Note: This table reports RD-DID estimates testing the difference in effects between the main sample of families with a grandmother born around the January 1950 cutoff and families with grandmothers born around January 1948, 1949, 1951, or 1952. Both the main sample and placebo group include mothers with the youngest child aged 4 to 12, when the grandmother is between 60 and 64. Columns (1) to (3) test the effects on grandmothers' total monthly hours worked, and Columns (4) to (6) report results on mothers' total monthly hours worked. The last row reports the difference in the effect on the main sample as opposed to the placebo group. All outcomes are measured when the grandmothers are between 60 and 64. The running variable is the grandmother's birthdate (month-year), centered around January 1950. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A14: RDD local randomization

	Randomized RDD	Baseline RD (Tables 1 and A8)	P-value of difference
Panel A: Reduced-form estimates-			
Total monthly hours worked of grandmothers	8.160***	6.750***	0.494
		[1.795]	
Specified Window/ Optimal Bandwidth	[-1,0]	6.061	
Panel B: Reduced-form estimates-			
Total monthly hours worked of mothers	-4.863***	-4.912***	0.938
		[1.772]	
Specified Window/ Optimal Bandwidth	[-1,0]	5.347	
Panel C: Fuzzy RD estimates-			
Total monthly hours worked of mothers	-0.596**	-0.630**	0.896
		[0.282]	
Specified Window/Optimal Bandwidth	[-1,0]	7.324	

Source: Authors' calculations from the CBS data.

Note: Randomized RDD coefficients are based on the local randomization approach to RD analysis (Cattaneo et al., 2024), which specifies a window of $[-1, 0]$ and allows for masspoints. Main RDD results (reported in the second column, also see Tables 1 and A8) are based on local linear regressions with the optimal bandwidth generated by the Calonico et al. (2014) procedure. We test for the balancedness of the characteristics of grandmothers and mothers for the local randomization sample (i.e., for the window $[-1, 0]$ specified by the approach) in Table A15. *** p<0.01, ** p<0.05, * p<0.1.

Table A15: RDD local randomization: sample balancedness

	RD estimates (1)	Mean at cutoff
Grandmothers' characteristics		
Age	0.038	62.938
	[0.045]	[1.165]
Number of adult children	0.069	2.461
	[0.044]	[0.885]
Number of adult daughters	0.027	1.702
	[0.038]	[0.756]
Prob (Employed)	0.000	0.788
	[0.016]	[0.344]
Prob (Married)	0.023	0.824
	[0.016]	0.366
Prob (Partner disabled)	0.020*	0.060
	[0.011]	[0.230]
Birthcohort of partner	-0.016	1947.73
	[0.185]	[4.094]
Total gross income (when aged 53-56)	11.728	16350.5
	[620.01]	[14355.6]
Mothers' characteristics		
Age	0.181	37.836
	[0.117]	[2.910]
Birth cohort	-0.063	1974.57
	[0.131]	[3.261]
Prob (Married)	0.019	0.351
	[0.017]	[0.430]
Prob (Employed)	-0.004	0.777
	[0.013]	[0.330]
Live in the same neighborhood as GM	-0.014	0.239
	[0.014]	[0.374]
Age at first child birth	0.125	28.191
	[0.142]	[3.507]
Age of youngest child	0.042	2.040
	[0.094]	[1.694]
Age of oldest child	-0.061	3.834
	[0.164]	[3.129]
Number of children	0.003	0.843
	[0.038]	[0.982]
Age of first employment	0.012	24.804
	[0.153]	[3.840]
Education not missing	-0.009	0.694
	[0.018]	[0.461]
Higher education (some college)	0.026	0.562
	[0.024]	[0.496]
Specified Window	[-1,0]	
Obs. Mothers	2730	

Source: Authors' calculations from the CBS data.

Note: Based on the local randomization approach (Cattaneo et al., 2024), we test the impact of grandmothers being born since January 1950 on a list of the grandmothers' and mothers' characteristics for the local randomization sample (i.e., for the window $[-1, 0]$ specified by the approach). All variables are predetermined and refer to the time when the grandmothers were 50 to 53 years old. Robust standard errors clustered at the grandmother level are in parentheses.*** p<0.01, ** p<0.05, * p<0.1.

Table A16: Reform impacts on mother's fertility outcomes

	RD estimates (1)	Mean at cutoff
Mothers' fertility outcomes		
Prob(Ever child)	0.008	0.771
	[0.008]	[0.420]
Prob(At least 2 children)	-0.005	0.626
	[0.007]	[0.488]
Total number of children	-0.002	1.640
	[0.018]	[1.158]
Age at first birth	-0.023	29.309
	[0.086]	[4.430]
Age at last birth	-0.068	32.827
	[0.069]	[4.128]
Average age gap of children	0.004	3.124
	[0.038]	[1.796]
Average age gap after GM age 55	-0.045	3.302
	[0.043]	[2.078]
Prob(First child after GM age 55)	-0.001	0.425
	[0.008]	[0.494]
Births up to 3 years post-reform	-0.002	0.378
	[0.009]	[0.622]
Prob(Births up to 3 years post-reform)	-0.005	0.306
	[0.007]	[0.461]
Births up to 6 years post-reform	-0.012	0.629
	[0.015]	[0.835]
Prob(Births up to 6 years post-reform)	-0.006	0.424
	[0.007]	[0.494]
Births up to 9 years post-reform	-0.015	0.801
	[0.018]	[0.948]
Prob(Births up to 9 years post-reform)	-0.007	0.490
	[0.008]	[0.500]
Average optimal bandwidth	7.206	
Average obs. Mothers	77556	

Source: Authors' calculations from the CBS data.

Note: This table tests the impact of grandmothers being born since January 1950 on the adult daughters' (mothers') fertility outcomes. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A17: P-value matrix for Table 3

Family Member		Grandmothers' Sons-in-law	Grandmothers' Sons	Grandmothers' Daughters-in-law
Panel A: Total monthly hours worked				
First-Stage: Impact on GM's total hours				
	Daughters	0.475	0.951	0.965
	Sons-in-law		0.805	0.830
	Sons			0.591
LATE: Impact on Family members' total hours				
	Daughters	0.004	0.014	0.224
	Sons-in-law		0.693	0.048
	Sons			0.149
Panel B: Other labor supply measures				
LATE: Prob(employed)				
	Daughters	0.008	0.050	0.470
	Sons-in-law		0.455	0.063
	Sons			0.229
LATE: Prob(fulltime)				
	Daughters	0.085	0.223	0.527
	Sons-in-law		0.730	0.191
	Sons			0.455
LATE: Household income				
	Daughters		0.980	

Source: Authors' calculations from the CBS data.
Note: This table shows the p-values of the differences in the effects on the respective family members. The underlying estimates measure the reform effect on grandmothers' total monthly hours worked and the effect of grandmothers' hours worked on their daughters', sons', and sons- and daughters-in-laws' labor supply (Fuzzy RD estimates). Panel A reports p-values for the first-stage and the LATE estimates of the main outcome, total monthly hours worked. Panel B reports p-values for the LATE estimates of other labor supply measures. For details, see Table 3.

Table A18: P-values for Table 4

	Childcare need			
	more		less/no	
	Age of the youngest child			
	4 - 7	8-12	13 - 18	No child
Panel A: Reduced-form estimates -				
Mothers' total monthly working hours				
Youngest child aged:				
1-3	0.057	0.803	0.373	0.184
4-7		0.076	0.089	0.015
8-12			0.271	0.315
13-18				0.805
Panel B: Fuzzy RD estimates -				
First-stage: GMs' total monthly working hours				
Youngest child aged:				
1-3	0.864	0.629	0.812	0.017
4-7		0.881	0.959	0.113
8-12			0.780	0.263
13-18				0.317
LATE: Mothers' total monthly working hours				
Youngest child aged:				
1-3	0.100	0.772	0.379	0.315
4-7		0.198	0.080	0.120
8-12			0.223	0.310
13-18				0.951

Source: Authors' calculations from the CBS data.
Note: This table shows the p-values of the differences in the effects on mothers with their youngest child of different ages (or without a child). Panel A reports p-values corresponding to the reduced-form effects on mothers' total monthly hours worked. Panel B reports p-values for the first-stage and LATE coefficients (Fuzzy RD estimates) measuring the effects of grandmother's total monthly hours worked on mother's total monthly hours worked. For further details, see Table 4.

Table A19: Summary statistics of mothers by age of the youngest child

	Age of youngest child		
	1-3	4-7	8-12
	(1)	(2)	(3)
Grandmothers' characteristics			
Birth cohort	1949.54	1949.54	1949.53
	[0.498]	[0.499]	[0.499]
Age	62.389	62.766	63.119
	[1.283]	[1.428]	[1.357]
Number of adult children	2.550	2.475	2.432
	[1.044]	[0.980]	[0.949]
Number of adult daughters	1.7467	1.7055	1.6814
	[0.820]	[0.785]	[0.763]
Mothers' characteristics			
Age	34.913	37.314	39.596
	[3.011]	[2.877]	[2.663]
Age at first child birth	30.114	28.733	27.258
	[3.588]	[3.454]	[3.297]
Age at first employment	22.849	24.206	26.258
	[2.897]	[3.253]	[3.968]
Married	0.624	0.663	0.656
	[0.472]	[0.463]	[0.465]
Age gap to partner	2.464	2.763	2.979
	[3.893]	[3.935]	[4.118]
Number of children	1.829	2.034	2.004
	[0.790]	[0.776]	[0.737]
Education not missing	0.783	0.698	0.639
	[0.412]	[0.459]	[0.481]
Higher education (some college)	0.731	0.607	0.450
	[0.443]	[0.488]	[0.498]
Obs. Mothers	21508	18101	10045

Source: Authors' calculations from the CBS data.
Note: This table reports the means and standard deviations of variables related to education, marital status, and fertility. Column 1 is based on mothers with the youngest child aged 1-3, Column 2 is based on mothers with the youngest child aged 4-7, and Column 3 is based on mothers with the youngest child aged 8-12 when the grandmothers are between the ages of 60 and 64. All samples consider a bandwidth of 7 months. *** p<0.01, ** p<0.05, * p<0.1..

Table A20: Heterogeneous effects by socioeconomic status (reduced-form)

	Mother has higher education		Grandmother has above median income	
	NO	YES	NO	YES
	(1)	(2)	(3)	(4)
Panel A: Impact on GM' labor supply				
Total monthly hours worked	-0.920	9.018**	2.557*	11.736***
	[2.944]	[2.495]	[1.363]	[2.869]
p-value	0.008		0.004	
Optimal Bandwidth	6.882	6.882	6.861	6.861
Obs. Mothers	5251	7077	8984	8908
Panel B: Impact on Mothers' labor supply				
Total monthly hours worked	-3.312	-8.471***	-2.966	-7.048***
	[3.066]	[2.604]	[2.472]	[2.508]
p-value	0.203		0.249	
Optimal Bandwidth	5.571	5.571	5.420	5.420
Obs. Mothers	4427	5985	7635	7488
Prob(Employed)	-0.043	-0.044**	-0.022	-0.028
	[0.026]	[0.019]	[0.018]	[0.017]
p-value	0.980		0.801	
Optimal Bandwidth	5.793	5.793	7.008	7.008
Obs. Mothers	4427	5985	10454	10214
Prob(Full-time employed)	0.004	-0.026**	0.002	-0.022**
	[0.009]	[0.012]	[0.008]	[0.010]
p-value	0.045		0.075	
Optimal Bandwidth	8.265	8.265	7.838	7.838
Obs. Mothers	6872	9323	10454	10214

Source: Authors' calculations from the CBS data.
Note: This table shows heterogeneous effects of the reform on grandmothers' and mothers' labor supply (reduced-form). Columns 1 and 2 show the results by mothers' education. Mothers with higher education are mothers with some college education, incl. academic colleges (WO) and colleges of applied sciences (HBO). Columns 3 and 4 show the results by whether grandmothers' average income between aged 53-56 (predetermined) is above the median or not. All outcomes are measured when the grandmothers are between the ages of 60 and 64. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure, including controls (mother's age, the age at first birth, and grandmothers' predetermined employment status) and sector fixed effects. Robust standard errors clustered at the grandmother level are in parentheses. The p-values are from a test of the hypothesis that the coefficients are equal. *** p<0.01, ** p<0.05, * p<0.1.

Table A21: Smoothness of children's covariates (reduced-form)

	RD estimates Youngest child aged		
	4-12	4-7	8-12
	(1)	(2)	(3)
Birth Cohort	0.053	0.081	-0.020
	[0.144]	[0.077]	[0.118]
Age	-0.059	-0.104	0.078
	[0.144]	[0.093]	[0.116]
Girl	0.047*	-0.001	0.122**
	[0.027]	[0.039]	[0.056]
Children in Household	-0.022	-0.014	-0.039
	[0.049]	[0.062]	[0.074]
Birthorder	-0.041	-0.046	-0.035
	[0.047]	[0.062]	[0.073]
Prob(First-born child)	0.023	0.044	-0.015
	[0.027]	[0.036]	[0.047]
Prob(Parents married)	-0.004	-0.008	-0.026
	[0.028]	[0.032]	[0.050]
Live in same neighborhood as GM	-0.041*	-0.052	0.008
	[0.021]	[0.032]	[0.042]
Parents' age difference	-0.430	-0.287	0.069
	[0.278]	[0.332]	[0.555]
Age mother (GM 60-64)	0.044	0.039	0.068
	[0.171]	[0.208]	[0.316]
Mother age first child	0.252	0.319	0.015
	[0.224]	[0.300]	[0.331]
Mother's education not missing	0.000	0.019	-0.002
	[0.033]	[0.041]	[0.063]
Mother has higher education (some college)	0.055	0.036	0.075
	[0.043]	[0.049]	[0.077]
Grandmother's total gross income (53-56)	-156.9	384.3	-388.8
	[816.5]	[1265.6]	[1381.5]
Average optimal Bandwidth	6.78	7.325	7.219
Average Obs. Children	5111	3070	1727
Average Obs. Mothers	5093	3006	1769

Source: Authors' calculations from the CBS data.
Note: This table tests the impact of grandmothers being born since January 1950 on a list of children's characteristics. Results in Columns 1-3 are based on the youngest children aged 4-12, 4-7, and 8-12, respectively, when the grandmother is aged 60 who can be matched to their Cito results. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure. Robust standard errors clustered at the mother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A22: Effects on children's educational performance (4-12): Heterogeneous effects (reduced-form)

	Grandmother's partner is		Grandmother's residence neighborhood		Number of maternal grandchildren aged 4-7	
Subgroups	unhealthy	healthy	different	same	more or other age	only one
	(1)	(2)	(3)	(4)	(5)	(6)
Number of correct answers:						
overall	0.028	0.152**	0.146*	0.174	-0.003	0.258**
	[0.201]	[0.072]	[0.076]	[0.143]	[0.087]	[0.102]
p-value	0.643		0.870		0.074	
Optimal bandwidth	5.412		5.412		5.412	
Obs. Children	354	3885	3394	845	2074	2165
Number of correct answers:						
verbal	0.192	0.092	0.087	0.135	0.010	0.173*
	[0.162]	[0.060]	[0.063]	[0.120]	[0.070]	[0.088]
p-value	0.642		0.746		0.184	
Optimal bandwidth	6.080		6.080		6.080	
Obs. Children	423	4610	4035	998	2468	2565
Number of correct answers:						
math	0.037	0.092	0.081	0.121	-0.116*	0.259***
	[0.169]	[0.058]	[0.061]	[0.118]	[0.070]	[0.080]
p-value	0.800		0.778		0.001	
Optimal bandwidth	6.605		6.605		6.605	
Obs. Children	423	4610	4035	998	2468	2565
Recommendation:						
high track	0.082	0.024	0.024	0.001	0.018	0.012
	[0.069]	[0.022]	[0.024]	[0.044]	[0.029]	[0.031]
p-value	0.500		0.680		0.906	
Optimal bandwidth	6.771		6.771		6.771	
Obs. Children	423	4610	4035	998	2468	2565

Source: Authors' calculations from the CBS data.

Note: This table shows heterogeneous effects of the reform on children's educational performance (reduced-form). Columns 1 and 2 show the results by the health status of the grandmother's partner. Partners are defined as healthy if they haven't claimed any disability insurance before age 54. Columns 3 and 4 show the results by the proximity of adult daughters (mothers) to grandmothers. We define the grandmother to be nearby when the mother and grandmother live in the same neighborhood. Columns 5 and 6 show the results by the number of maternal grandchildren aged 4-7. All outcomes are measured when children take the Cito test at the end of primary school. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure and include controls, including the child's birth cohort and month, and treatment duration. For a description of the full list of controls, see Section 2.3. Robust standard errors (clustered by the child's primary school) are in parentheses. The p-values are from a test of the hypothesis that the coefficients are equal. *** p<0.01, ** p<0.05, * p<0.1.

Table A23: Effects on children's educational performance by gender (reduced-form)

RD estimates	Number of correct answers (Cito)			High
	Verbal	Math	Overall	track
	(1)	(2)	(3)	(4)
By age groups and gender				
Girls aged between 4 - 7	0.163	0.300***	0.289**	0.070
	[0.116]	[0.109]	[0.133]	[0.044]
Boys aged between 4 - 7	0.305**	0.181	0.263*	0.040
	[0.134]	[0.111]	[0.143]	[0.051]
p-value (difference by gender)	0.444	0.458	0.910	0.667
Obs. Girls	1213	1421	1213	1213
Obs. Boys	1145	1374	1145	1145
Girls aged between 8 - 12	0.061	0.086	0.154	0.012
	[0.119]	[0.136]	[0.152]	[0.039]
Boys aged between 8 - 12	-0.234*	-0.073	-0.232	-0.091**
	[0.131]	[0.138]	[0.161]	[0.036]
p-value (difference by gender)	0.118	0.456	0.107	0.075
Obs. Girls	892	782	648	1022
Obs. Boys	859	739	626	969

Source: Authors' calculations from the CBS data.
Note: This table shows reduced-form impacts on the education outcomes of children by age groups and gender of the children. Columns 1 - 3 report effects on the number of correct answers in the verbal part, mathematical part, and the overall Cito test, respectively. These columns are based on standardized outcomes and thus measure effects in percent of a standard deviation. Column 4 shows the impact on the probability of obtaining a secondary school recommendation for the highest (academic) track (VWO). All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure and include controls, including the child's birth cohort and month, and treatment duration. For a description of the full list of controls, see Section 2.3. Robust standard errors (clustered by the child's primary school) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A24: Impacts on subsidy take-up (reduced-form)

RD estimates	Daycare		After-school care	
	Prob.	Hours	Prob.	Hours
	(1)	(2)	(3)	(4)
Age between 4-7	0.004	-2.632	0.014	6.489
	[0.003]	[3.563]	[0.0285]	[13.380]
Optimal bandwidth	4.600	4.603	6.559	9.047
Obs. Mothers	3708	3708	5348	7850
Age between 8-12			0.032*	15.377*
			[0.019]	[7.963]
Optimal bandwidth			7.141	7.867
Obs. Mothers			3467	3467

Source: Authors' calculations from the CBS data.

Note: This table shows reduced-form reform impacts on childcare subsidy take-up in families with the youngest child aged 4 - 7, and 8 - 12 when the grandmothers are aged 60. Subsidy take-up is shown for any child within the indicated age range. Columns (1) and (2) show effects on the probability of daycare take-up and the average hours of daycare usage, respectively. Columns (3) and (4) show effects on the probability of after-school care take-up and the average hours of after-school care usage, respectively. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure and include controls. For a description of the full list of controls, see section 2.3. Robust standard errors clustered at the grandmother level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A25: Effects on children's educational performance by care quality (reduced-form)

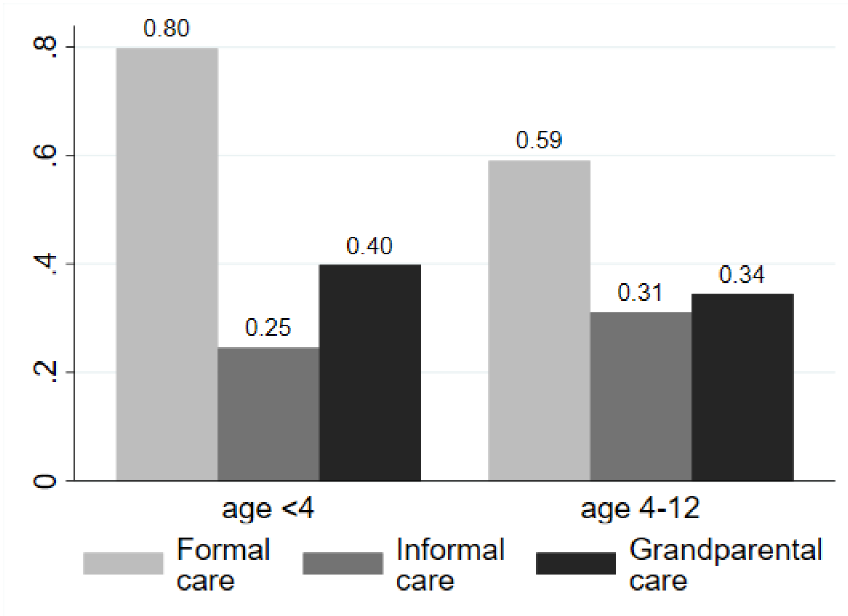
Grandmother's care quality	High		Low	
Mother's care quality	High	Low	High	Low
	(1)	(2)	(3)	(4)
Number of correct answers (Cito)				
Overall	0.279	0.316	0.644***	-0.017
	[0.219]	[0.252]	[0.228]	[0.217]
P-value difference	0.940		0.272	
Optimal bandwidth	5.413	5.413	5.413	5.413
Obs. Children	453	289	332	421
Number of correct answers (Cito)				
Verbal	0.123	0.224	0.501**	-0.081
	[0.187]	[0.230]	[0.219]	[0.199]
P-value difference	0.571		0.143	
Optimal bandwidth	5.731	5.731	5.731	5.731
Obs. Children	453	289	332	421
Number of correct answers (Cito)				
Math	0.277	0.256	0.573***	0.119
	[0.199]	[0.210]	[0.182]	[0.189]
P-value difference	0.919		0.191	
Optimal bandwidth	6.097	6.097	6.097	6.097
Obs. Children	531	360	394	493
Recommendation				
High track	0.095	0.080	0.140	0.033
	[0.083]	[0.060]	[0.097]	[0.060]
P-value difference	0.684		0.483	
Optimal bandwidth	6.563	6.563	6.563	6.563
Obs. Children	531	360	394	493

Source: Authors' calculations from the CBS data.

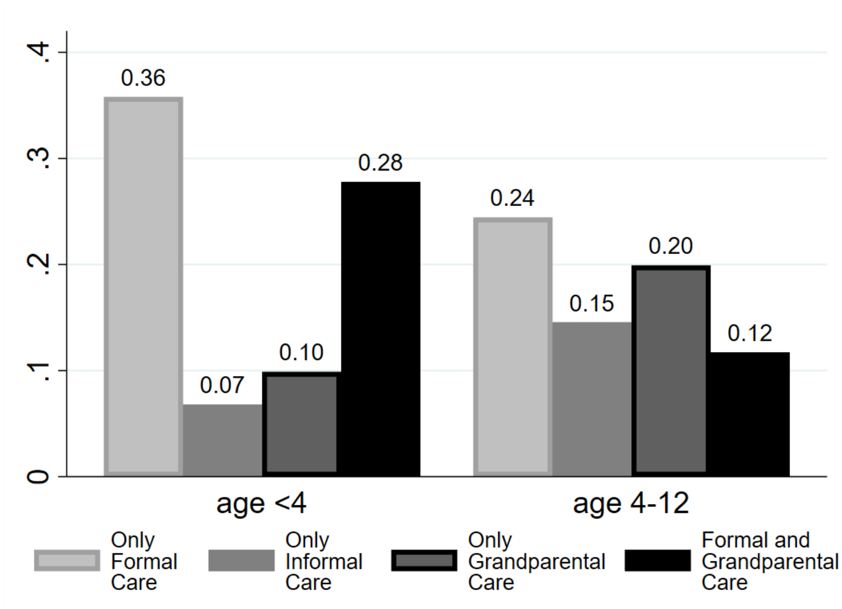
Note: This table shows the reduced-form impact on the education outcomes of the youngest child aged 4 - 7. Column (1) reports results of children with a high SES grandmother and a highly educated mother, Column (2) is based on children with high SES grandmothers and a low educated mother, Column (3) is based on children with a low SES grandmother and a highly educated mother, and Column (4) reports results on children with a low SES grandmother and low educated mother. Highly educated mothers are mothers with some college education, including academic colleges (WO) and colleges of applied sciences (HBO). Grandmothers' SES is determined by whether their average income between ages 53-56 (predetermined) is above (or below) the median. Effects on the number of correct answers in the verbal part, mathematical part, and the overall Cito test are based on standardized outcomes and thus measure effects in percent of the standard deviation. All specifications are estimated via a local linear regression with the optimal bandwidth generated by the Calonico et al. (2014) procedure and include controls, including the child's birth cohort and month, and treatment duration. For a description of the full list of controls, see Section 2.3. Robust standard errors (clustered by the child's primary school) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure A1: Survey evidence on childcare modes

(a) Distribution of childcare modes

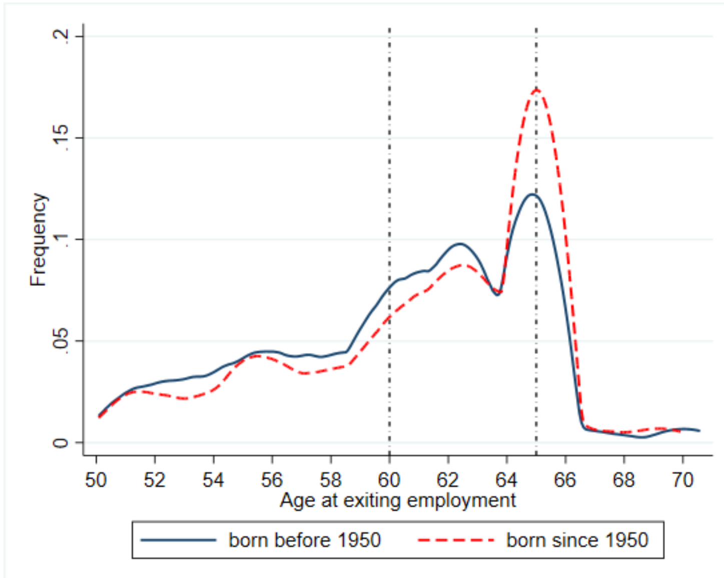


(b) Distribution of care mix



Source: Authors' own calculations from LISS panel administered by CentERdata (Tilburg University, The Netherlands).
Note: Figure A1 shows childcare modes employed by parents in the LISS panel. Parents are asked separately for their children below 4 and between ages 4 and 12 whether they make regular use (at least once a week) of various types of childcare. Panel (a) shows childcare take-up allowing for multiple answers so that the three categories are not mutually exclusive. Panel (b) shows the four most common combinations of childcare modes with mutually exclusive categories.

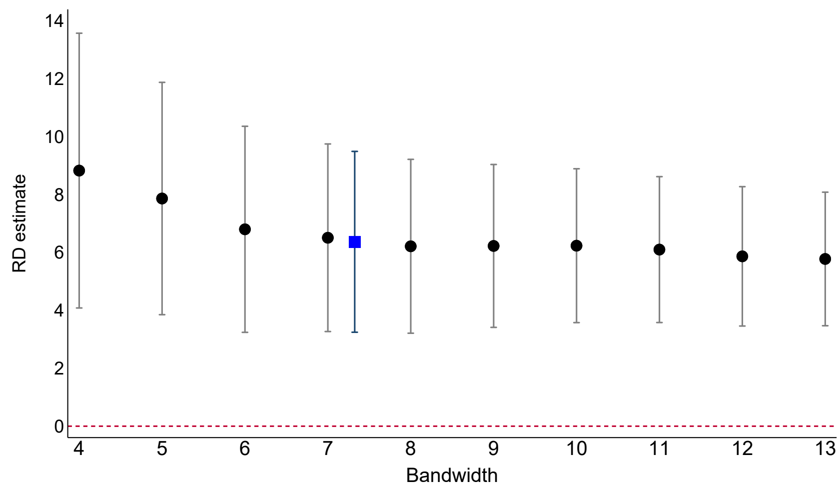
Figure A2: Distribution of age at exiting employment by treatment status



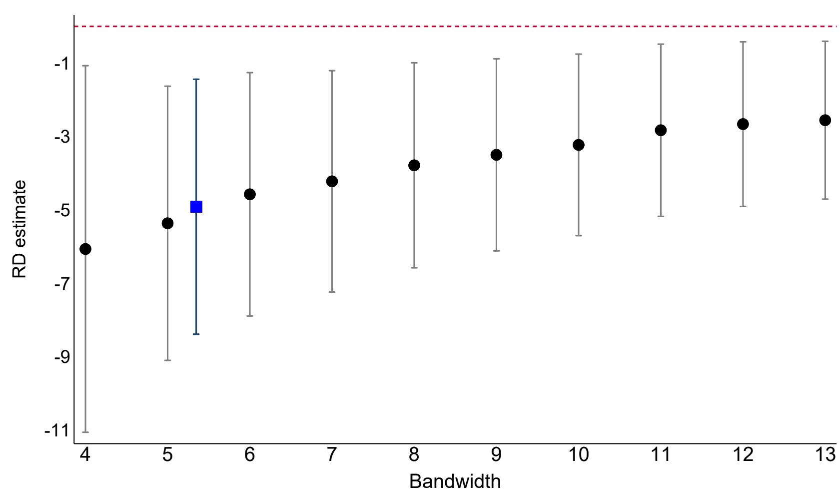
Source: Authors' calculations from the CBS data.
Note: Figure A2 shows the distribution of age at exiting employment for the cohorts born before (blue solid line) and since 1950 (red dashed line) in the baseline sample. We can clearly see a shift towards later retirement for the treated cohorts.

Figure A3: RD coefficients by varying bandwidths

(a) Total monthly hours worked by grandmothers



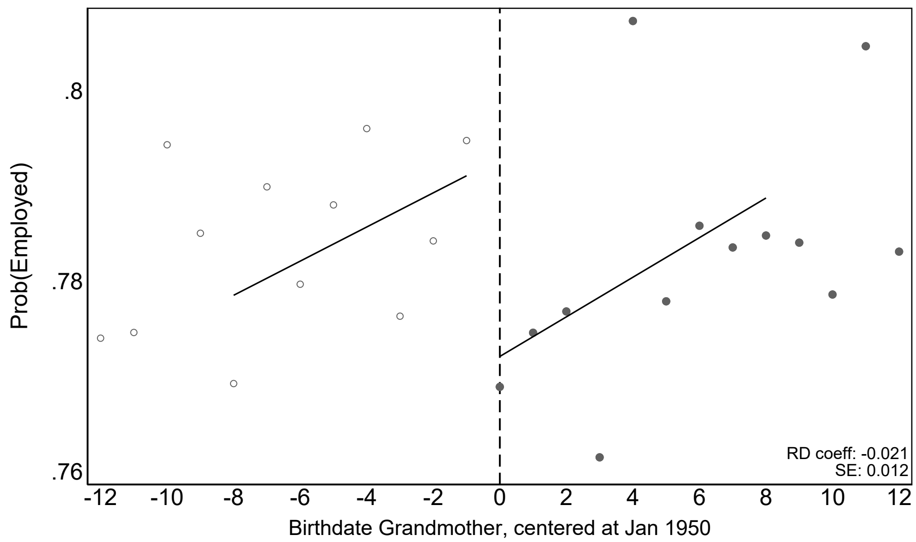
(b) Total monthly hours worked by mothers



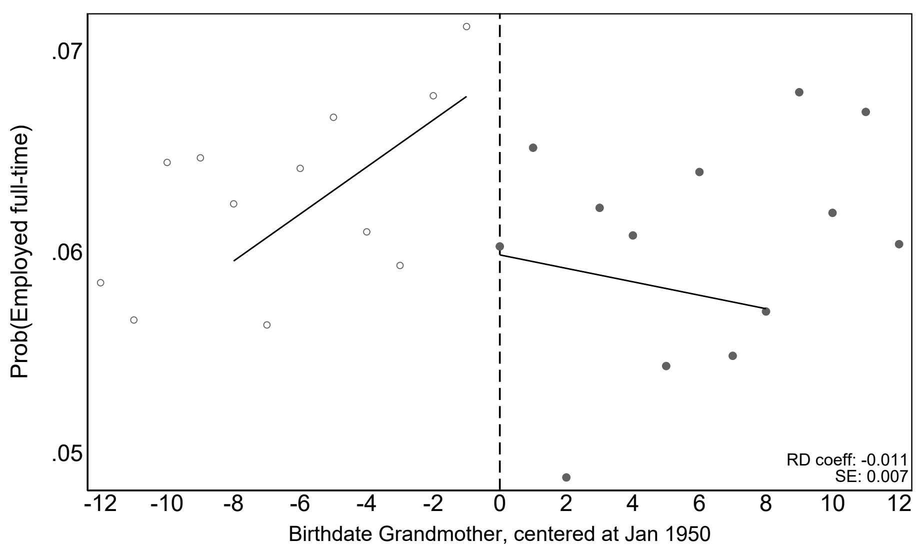
Note: Figure A3 plots the RD coefficients for varying bandwidths for grandmothers' total monthly hours worked in panel (a) and for mothers' total monthly hours worked in panel (b). Each regression includes controls and fixed effects as in the main specification. 95% confidence intervals using clustered standard errors at the grandmother level are depicted as well. The coefficients highlighted as a blue square depict the estimates based on the optimal bandwidth generated by the Calonico et al. (2014) procedure.

Figure A4: RD plots: Mothers' additional labor supply outcomes

(a) Mothers' employment probability (reduced-form)



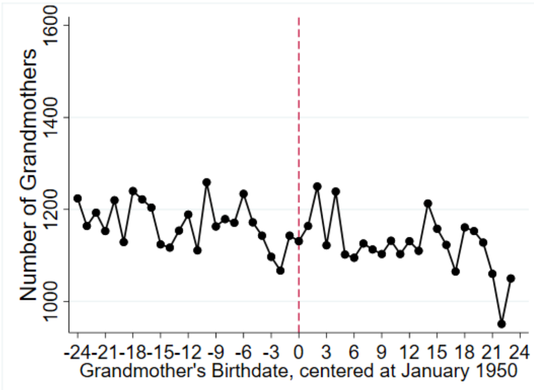
(b) Mothers' fulltime employment probability (reduced-form)



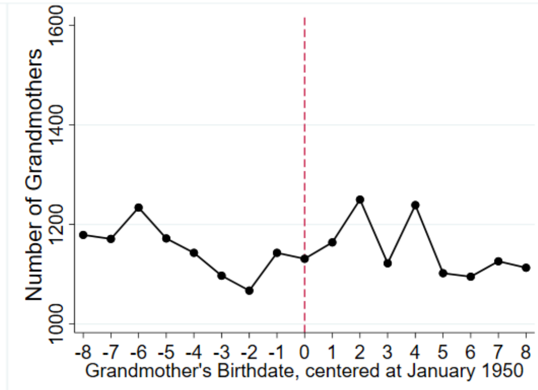
Note: Panel (a) of Figure 1 shows the scatter bin plot of the mother's employment probability and panel (b) mother's full-time employment probability as a function of distance to the cutoff, which is the grandmother's birth month being January 1950. The solid line is a linear polynomial fit of each outcome on the running variable based on the optimal bandwidth generated by the Calonico et al. (2014) procedure and fit separately left and right of the cutoff. Reported coefficients are RD estimates without controls. For estimations including controls, see Table A5.

Figure A5: Smoothness of the density around the cutoff

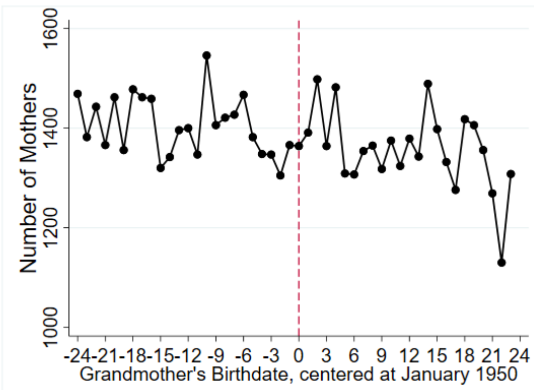
(a) Density of grandmothers (24 months)



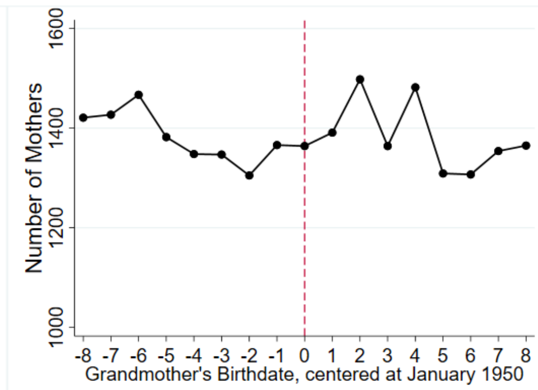
(b) Density of grandmothers (8 months)



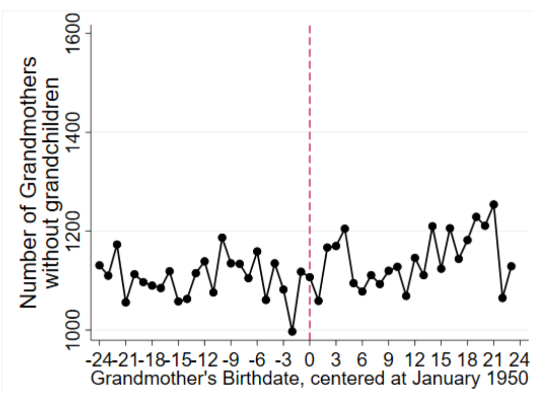
(c) Density of mothers (24 months)



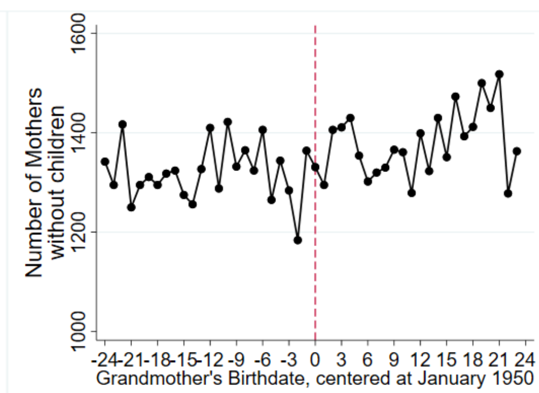
(d) Density of mothers (8 months)



(e) Density of grandmothers without grandchildren



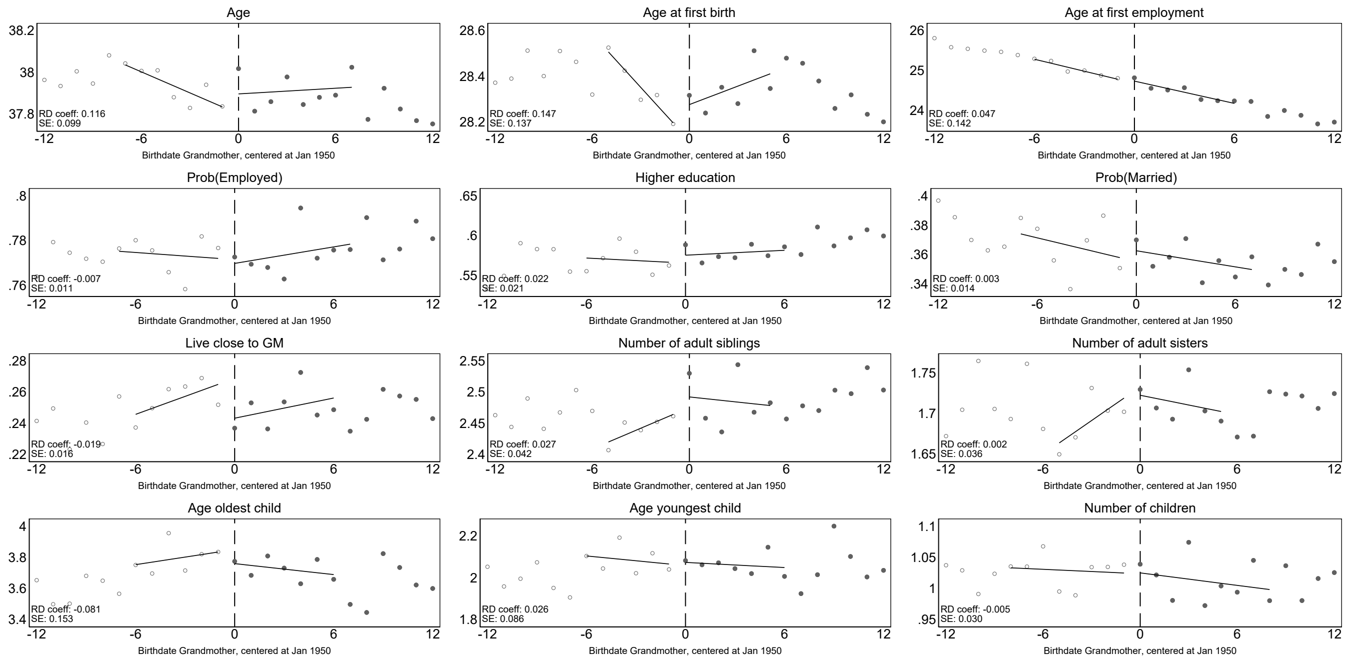
(f) Density of mothers without children



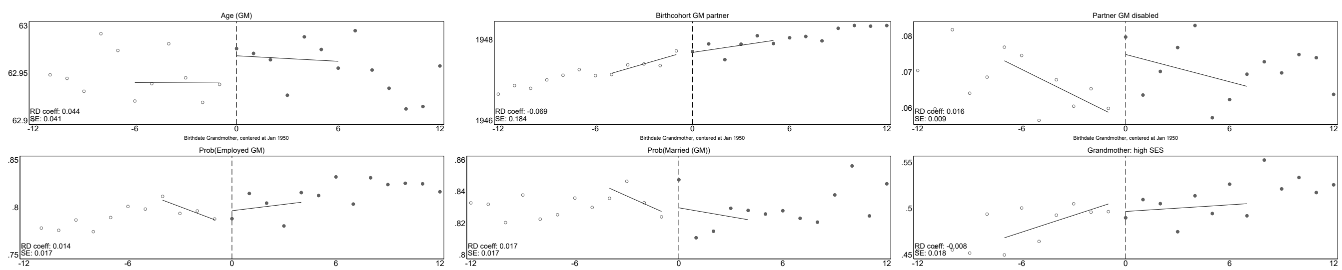
Note: The (bin size/running variable) in Figure A5 is grandmothers' birth date/months. Figure A5a and A5b show the density plot of grandmothers 24 months and 8 months around the cutoff. Figure A5c and A5d show the density plot of mothers whose mothers' ('grandmother') birth month is 24 months and 8 months around the cutoff. Figure A5e and A5f show that the fluctuating patterns of the density plots for grandmothers and mothers of our baseline sample are not unique, but a pattern that also shows up for 'grandmothers' and 'mothers' without (grand)children.

Figure A6: Balance test: covariate scatter plots

(a) Mothers' characteristics

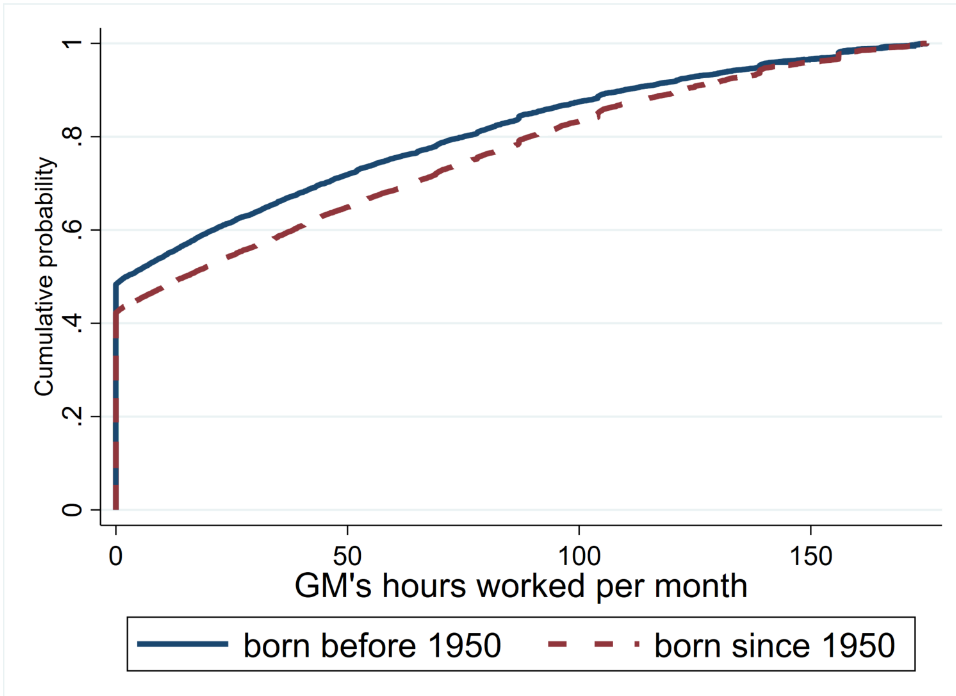


(b) Grandmothers' chracteristics



Note: Panel (a) of Figure A6 shows the scatter bin plot of mothers' predetermined characteristics as a function of distance to the cutoff. Panel (b) of Figure A6 shows the scatter bin plots of grandmothers' predetermined characteristics as a function of distance to the cutoff. Within each panel, each plot considers the cutoff of the grandmother's birth month as January 1950. All variables are predetermined and refer to the time when the grandmothers were 50 to 53 years old. The solid line is a linear polynomial fit of each outcome on the running variable based on the optimal bandwidth generated by the Calonico et al. (2014) procedure and fit separately left and right of the cutoff. Reported coefficients are RD estimates without controls. For estimations including controls, see Table A7.

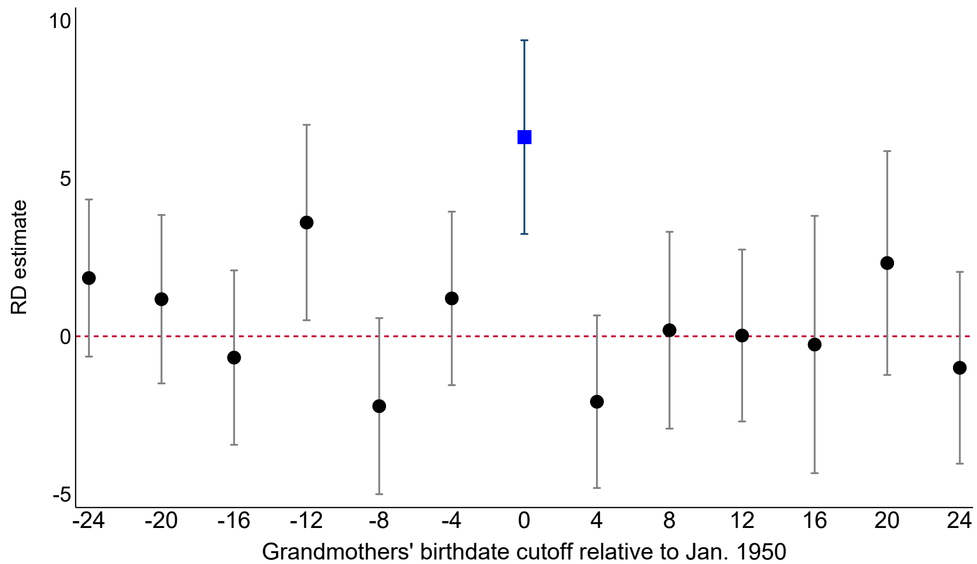
Figure A7: Cumulative Distribution of grandmothers' hours worked by treatment status



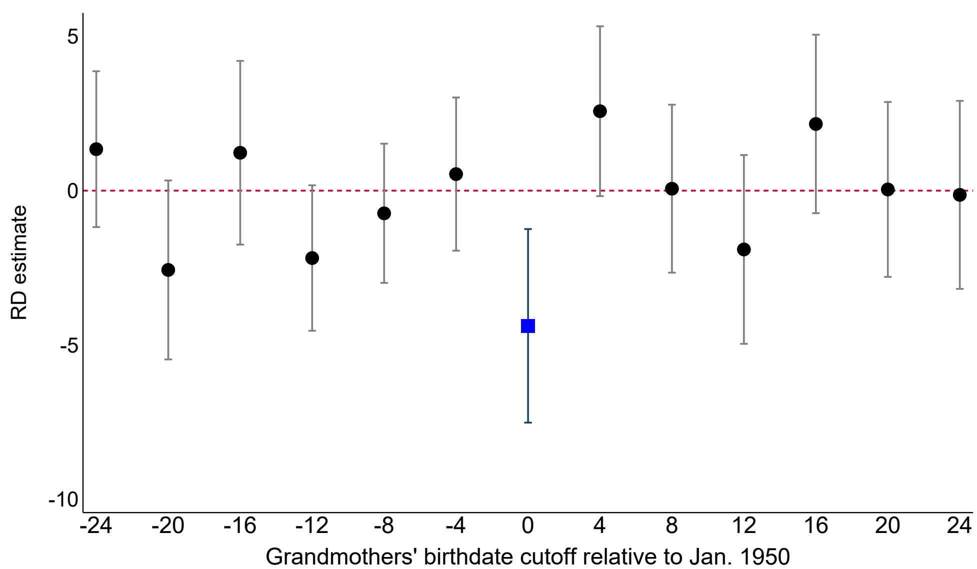
Source: Authors' calculations from the CBS data.
Note: Figure A7 shows the cumulative distribution of total hours worked per month for the cohorts born before (the blue line) and since 1950 (the red line) in the baseline sample (7 months around the cutoff).

Figure A8: RD estimate plots: Placebo cutoffs

(a) Monthly working hours of grandmothers



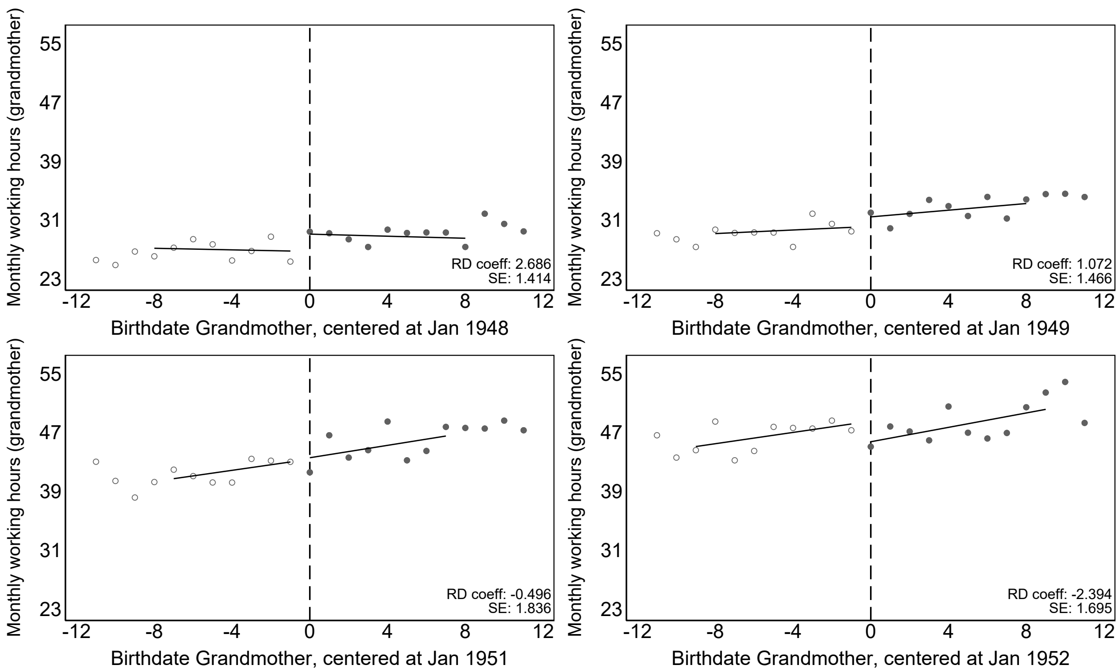
(b) Monthly working hours of mothers



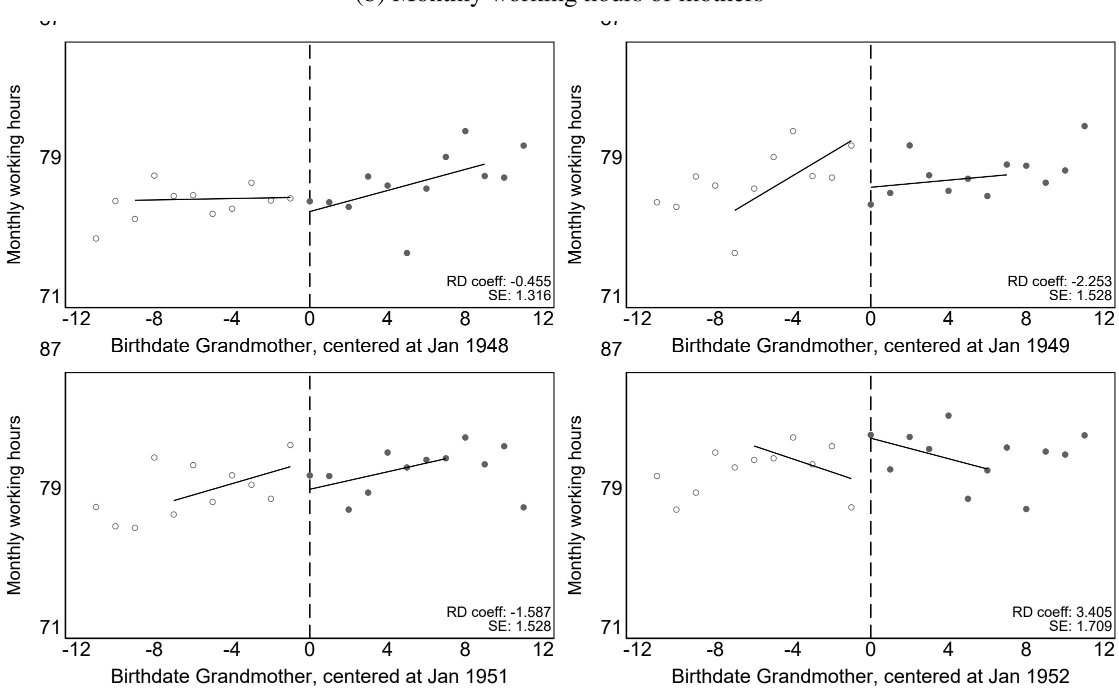
Note: Panels (a) and (b) of Figure A8 show RD estimates of total monthly hours worked of grandmothers and mothers, using varying cutoffs as the grandmothers' birth month. 95% confidence intervals are plotted around the point estimates. The blue square indicates the main result (with the grandmother‘s birth month centered at January 1950). Each estimation includes controls and uses the optimal bandwidth generated by the Calonico et al. (2014) procedure.

Figure A9: RD plots: Placebo January cutoffs

(a) Monthly working hours of grandmothers



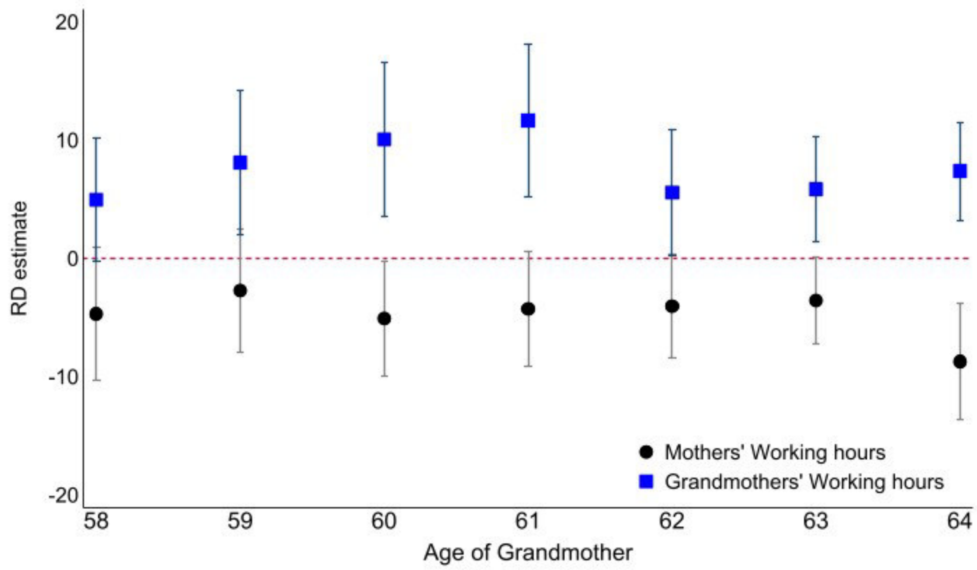
(b) Monthly working hours of mothers



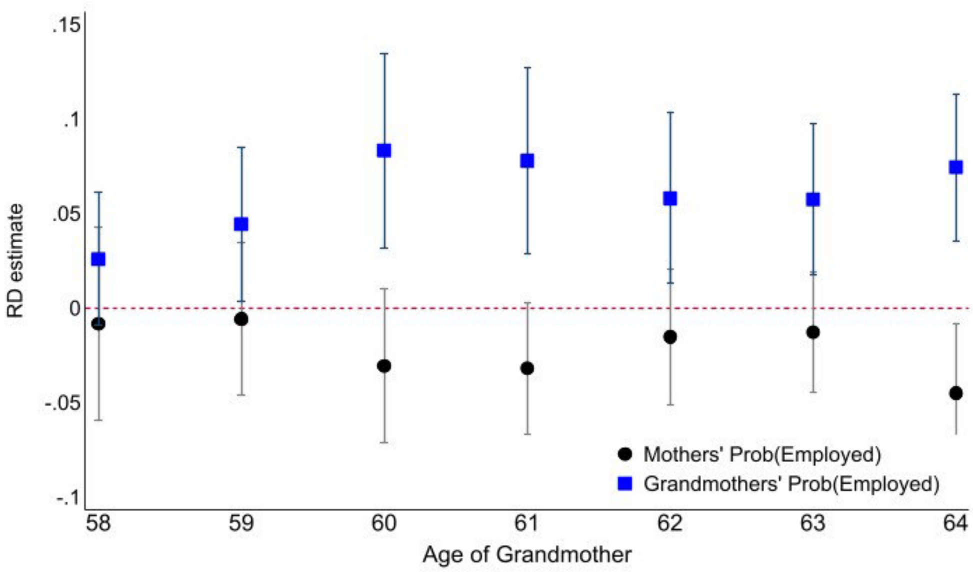
Note: Panels (a) and (b) of Figure A9 show the scatter bin plot of total monthly hours worked by grandmothers and mothers as a function of distance to the cutoff. Within each panel, the different plots consider placebo cutoffs of the grandmother's birth months, in particular January 1948, 1949, 1951, and 1952. The solid line is a linear polynomial fit of each outcome on the running variable based on the optimal bandwidth generated by the Calonico et al. (2014) procedure and fit separately left and right of the cutoff. Reported coefficients are RD estimates without controls.

Figure A10: Effects by Age of Grandmother (Youngest Child Aged 4-7)

(a) Working hours



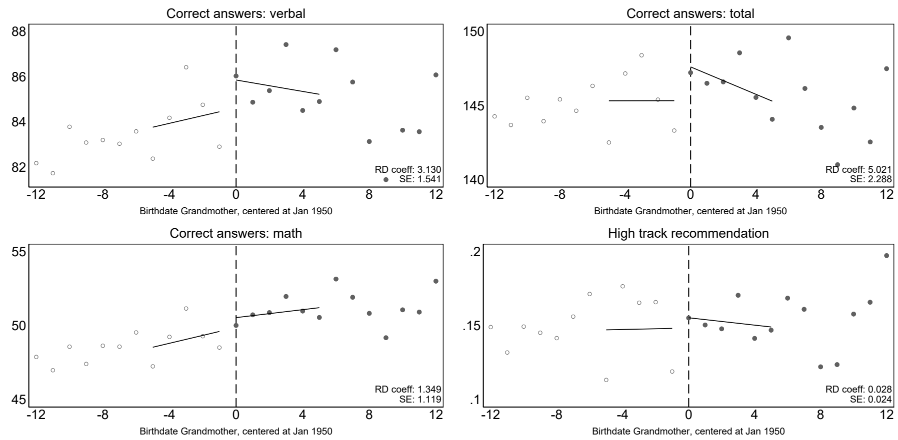
(b) Probability of employment



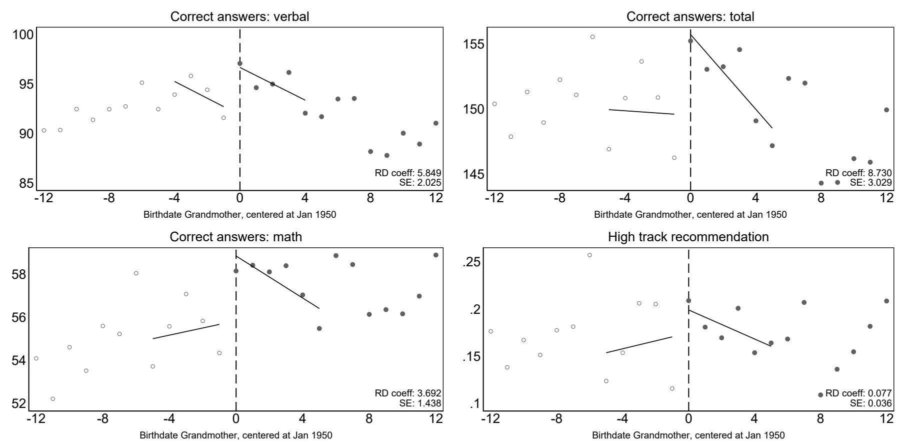
Note: Panel (a) shows the RD estimates of the mothers' total monthly hours worked (black dots) and the grandmothers' total monthly hours worked (blue squares) by the age of the grandmother. Panel (b) shows the RD estimates of the mothers' employment probability (black dots) and the grandmothers' employment probability (blue squares) by the age of the grandmother. 95% confidence intervals are plotted around the point estimates. Each estimation includes controls and uses the optimal bandwidth generated by the Calonico et al. (2014) procedure and is based on families with the youngest (grand)child aged 4-7 when the grandmother is between 60 and 64.

Figure A11: RD plots: Children outcomes

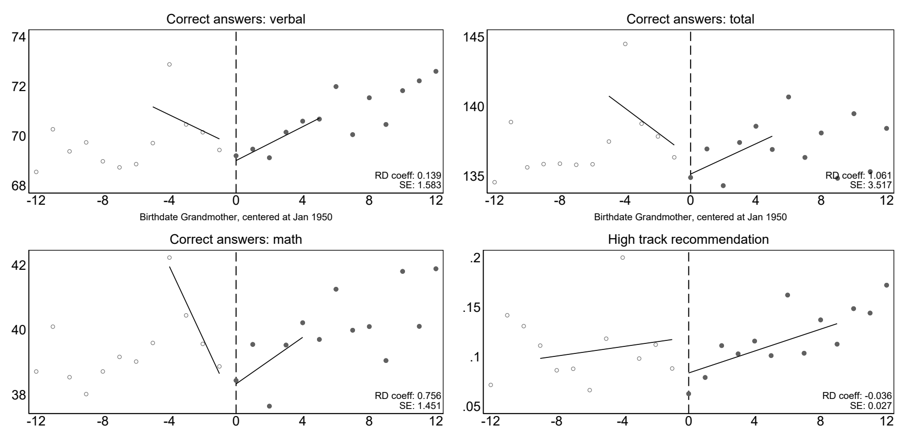
(a) Youngest children aged 4-12



(b) Youngest children aged 4-7



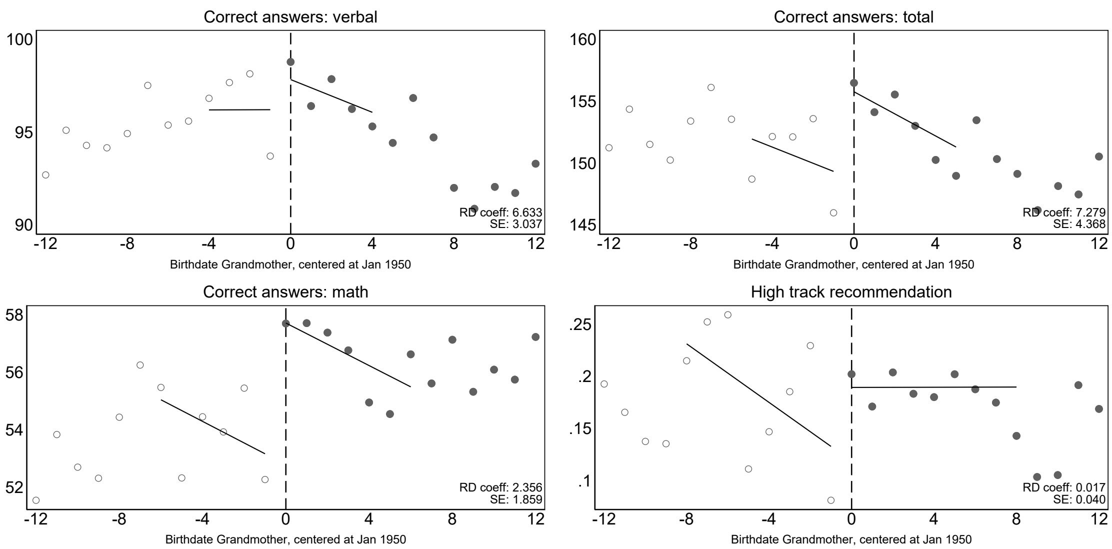
(c) Youngest children aged 8-12



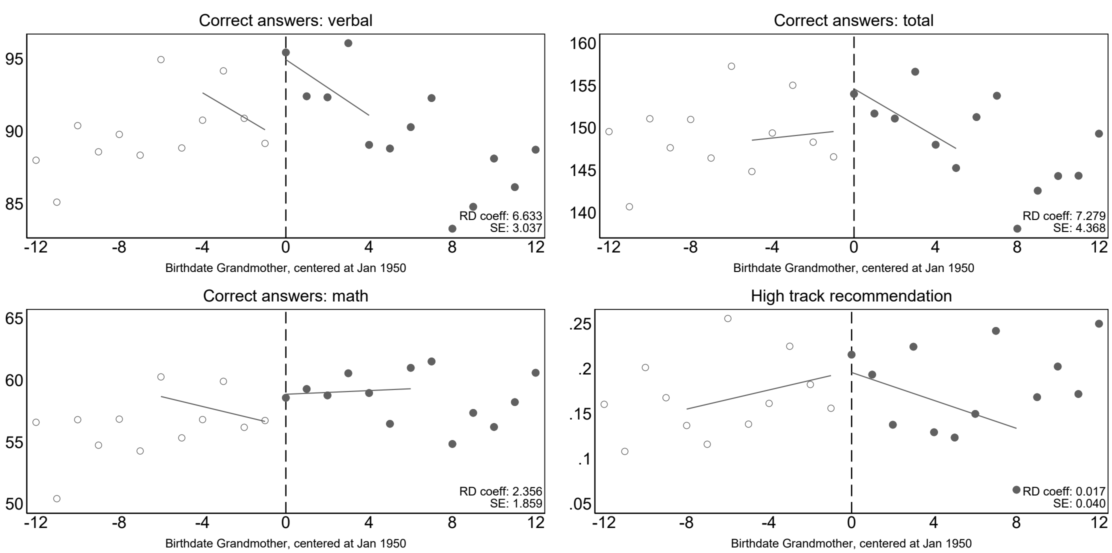
Note: Panels (a), (b), and (c) show the scatter bin plots of performance in the standardized test as a function of distance to the cutoff for the youngest children aged 4-12, 4-7, and 8-12, respectively. Each plot considers the cutoff of the grandmother's birth month as January 1950. The solid line is a linear polynomial fit of each outcome on the running variable based on the optimal bandwidth generated by the Calonico et al. (2014) procedure and fit separately left and right of the cutoff. Reported coefficients are RD estimates without controls. For estimations including controls, see Panels A and B of Table 6.

Figure A12: RD plots: Children's outcomes by gender (aged 4 - 7)

(a) Girls



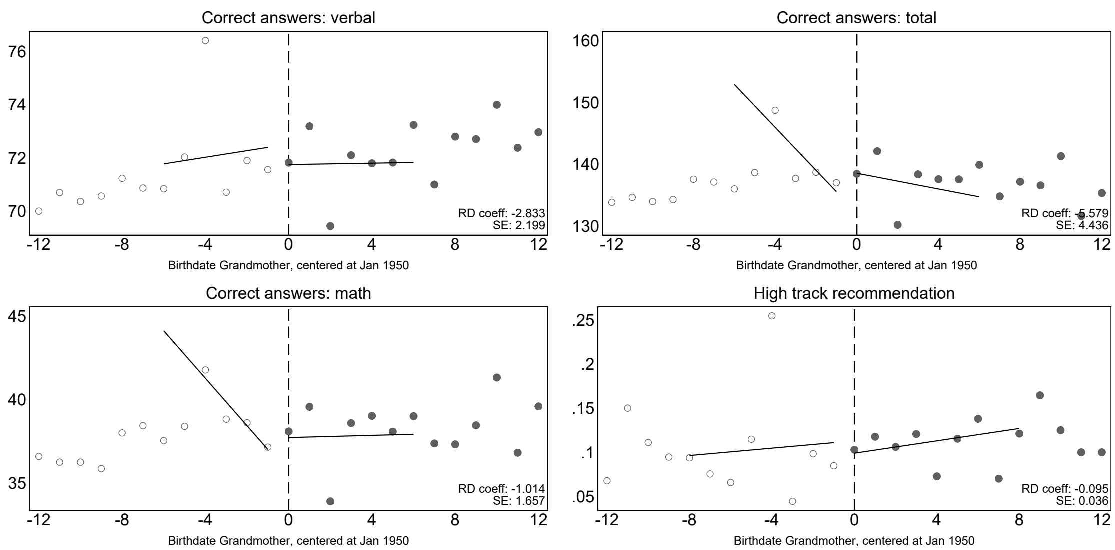
(b) Boys



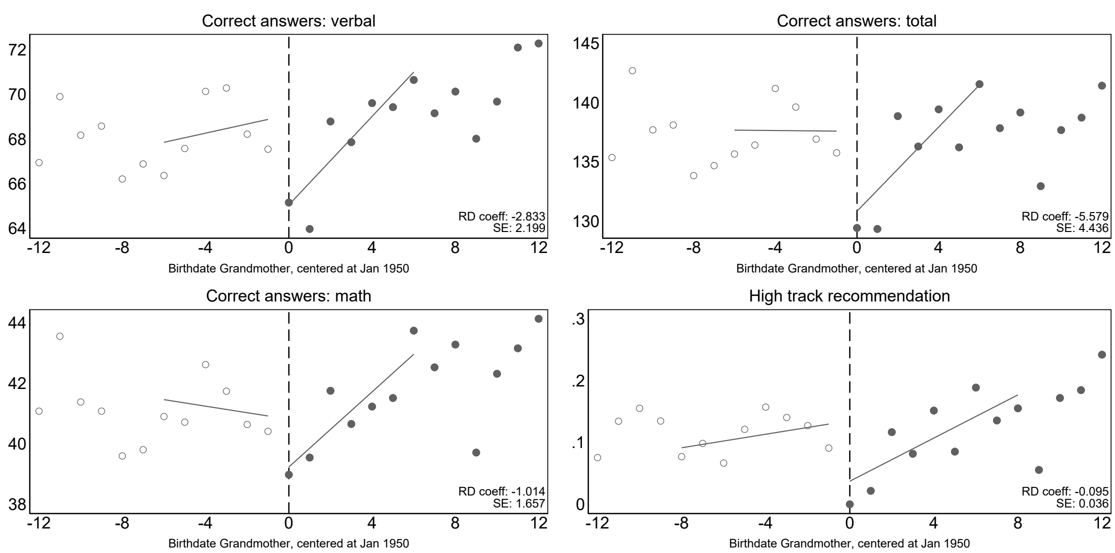
Note: Panel (a) and (b) of Figure A12 show the scatter bin plots of performance in the standardized test as a function of distance to the cutoff for youngest girls and youngest boys aged 4-7, respectively. Each plot considers the cutoff of the grandmother's birth month as January 1950. The solid line is a linear polynomial fit of each outcome on the running variable based on the optimal bandwidth generated by the Calonico et al. (2014) procedure and fit separately left and right of the cutoff. For estimations including controls, see Table A23 .

Figure A13: RD plots: Children's outcomes by gender (aged 8 - 12)

(a) Girls



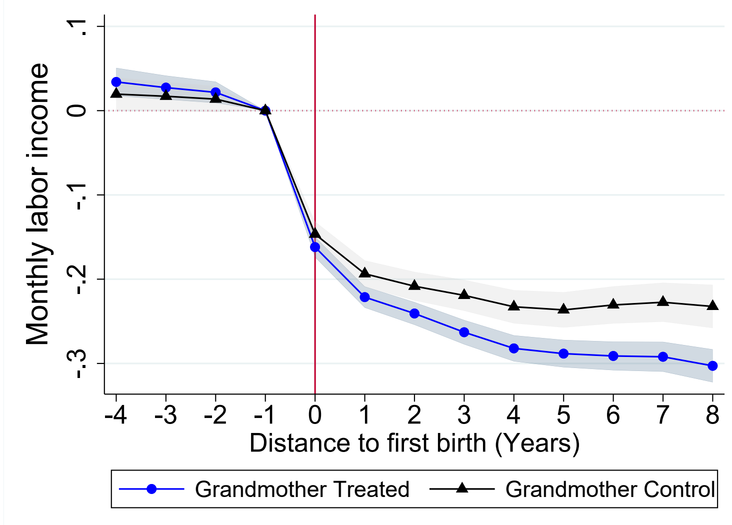
(b) Boys



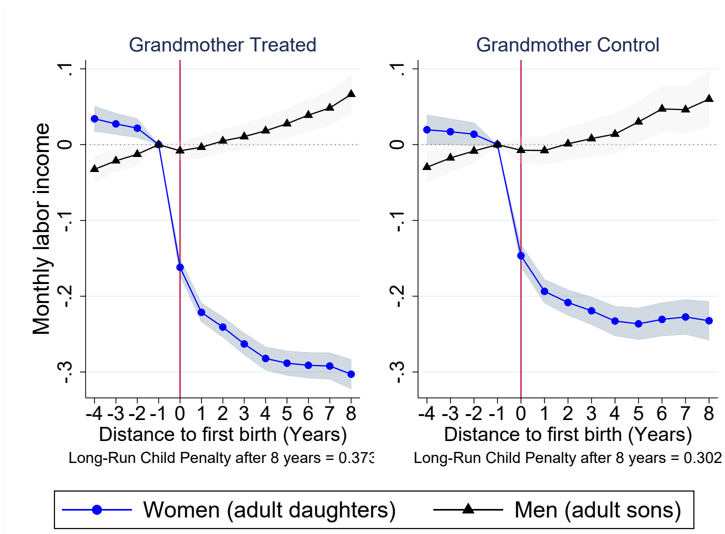
Note: Panels (a) and (b) of Figure A13 show the scatter bin plots of performance in the standardized test as a function of distance to the cutoff for youngest girls and youngest boys aged 8-12, respectively. Each plot considers the cutoff of the grandmother's birth month as January 1950. The solid line is a linear polynomial fit of each outcome on the running variable based on the optimal bandwidth generated by the Calonico et al. (2014) procedure and fit separately left and right of the cutoff. Reported coefficients are RD estimates without controls. For estimations including controls, see Table A23 .

Figure A14: Dynamic treatment effects and child penalty in labor earnings

(a) Dynamic treatment effects on monthly labor earnings



(b) Relative child penalty by treatment status



Note: Panel (a) of Figure A14 shows the evolution of mothers' total monthly labor earnings from four years before and to eight years after they gave birth to their first child. It compares the monthly labor earnings of treated mothers (blue dots), whose (grand)mothers were born between January 1950 and December 1951 and thus treated by the pension reform, to those of control mothers (black triangles), with untreated (grand)mothers born between January 1948 and December 1949. Event time 0 marks the birth of the first child. Panel (b) depicts the child penalty in total monthly labor income by treatment status. The left figure presents the child penalty for women and men with treated grandmothers and the right figure for women and men with control grandmothers. Blue dots document women's and black triangles indicate men's monthly labor earnings (including zeros), the difference between which represents the child penalty. The long-run relative child penalty after 8 years (i.e., the relative loss women experience compared to men) is reported below each sub-graph. The value at t = -1 is normalized to zero so that coefficients measure the impact of the first child relative to the year before birth. The shaded areas indicate the 95 percent confidence interval.