CESifo WORKING PAPERS

9813
2022

June 2022

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

Katja M. Kaufmann, Yasemin Özdemir, Han Ye

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9813*

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

Abstract

We study the impact of grandparental retirement decisions on family members' labor supply and child outcomes by exploiting a Dutch pension reform in a fuzzy Regression Discontinuity design. A one-hour increase in grandmothers' hours worked causes adult daughters with young children to work half an hour less. Daughters without children, with older children and sons/daughters-in-law are not affected. We show important long-run impacts on maternal labor supply and on the child penalty. Test score effects are positive for children aged 4-7 (substitution from grandparental to maternal care), and negative for children aged 11-12 (substitution from grandparental to formal childcare).

JEL-Codes: J130, J220, J260, I380, D640.

Keywords: spillover effects, retirement, grandparental childcare, maternal labor supply, child development.

Katja M. Kaufmann
Johannes Gutenberg University Mainz / Germany
kkaufman@uni-mainz.de

Yasemin Özdemir
Johannes Gutenberg University
Mainz / Germany
yasemin.oezdemir@uni-mainz.de

Han Ye
University of Mannheim / Germany
han.ye@uni-mannheim.de

June 2022
We gratefully acknowledge funding from the German Research Foundation (DFG) through CRC TR 224 (Project C01). We also thank Kasey Buckles, Emilia Del Bono, David Dorn, Libertad González, Daniel Hamermesh, Hilary Hoynes, Eckhard Janeba, Giovanni Mastrobuoni, Todd Morris, Cheti Nicoletti, Daniele Paserman, Gerard Pfann, Barbara Petrongolo, Michéle Tertilt and seminar and conference participants at the CRC annual meeting, the IZA gender workshop 2022, IZA/SOLE Transatlantic Meeting 2022, SEHO 2022 and at Bonn University, Duesseldorf University, Technical University Dortmund, Johannes Gutenberg University Mainz, University of Mannheim, Pompeu Fabra University, University of Würzburg, University of St. Gallen for their helpful insights and suggestions.

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

Katja M. Kaufmann† Yasemin Özdemir ‡ Han Ye §

June 2022

Evaluating the overall impact of public policies is complex, given a myriad of potential spillover effects, both within and across generations. Among the most important public policy changes in recent years have been changes in early retirement policies and policies related to (female) labor supply. Many OECD countries have reformed their public pension systems to reduce the incentive for early retirement and increase labor supply in old age. Recent evidence shows that such policies can have indirect effects on spouses/partners, which tends to exacerbate the direct labor supply response of the elderly generation.[1]

However, little is known about the cross-generational spillover effects of prolonged work life in old age. A delay in the timing of grandparents' retirement might lead to spillover effects across generations, because of a resulting reduction in grandparental childcare. In fact, grandparents

[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.

play an essential role in childcare in many countries.[2] Therefore, pension reforms that increase grandparents' age at retirement might lead to a decrease in the labor supply of their adult children (in particular of daughters) or an increase in childcare costs. Such intergenerational spillovers might thereby counteract or even reverse the direct effects on overall labor supply. They might even have important consequences for the long-run labor market outcomes of mothers, and for the implied child penalty and gender wage gap within a society. Lastly, there might 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 importance of such spillover effects across generations is largely understudied.

We aim to fill this gap by investigating the importance of multigenerational spillover effects of retirement and labor supply decisions, exploiting pension reform-induced variation in the retirement age in the Netherlands. In particular, in 2006, the Dutch government abolished the tax-favored voluntary early retirement schemes for employees born in 1950 or later, which had strong early retirement incentives for the earlier cohorts. Lindeboom and Montizaan (2020) examine the direct effect of the same pension reform on retirement expectations. They show that the affected workers born since 1950 face a substantial loss in pension wealth and expect to work longer, validating our design. Using administrative data covering the universe of the Dutch population[3] and a fuzzy Regression Discontinuity (RD) design, we provide the first evidence of the spillover effects of such a reform across household members and generations. We thereby shed light on the relevance of intergenerational spillover effects in terms of labor supply decisions and of the multigenerational effects of grandparents' labor supply decisions on children's educational outcomes.

First, we show that the pension reform has led to an important increase in grandparents' labor supply. We find that grandmothers increase their total hours worked (including zeros) between age 60 and 64 by 6.1 hours per month (equivalent to an 18-percent increase), while their likelihood of still being employed increases by six percentage points (14 percent). While grandfathers are not the main focus of our analysis, we find that they work 26 hours more per month (45 percent), and that their likelihood of being employed increases by 14 percentage points (32 percent). To simultaneously capture the extensive and intensive margins of labor supply, we primarily focus in our analysis on the total number of hours worked (including zeros) as a measure of grandparents' time availability.

In our main analysis we employ a fuzzy RD design to investigate the importance of intergener-

[2] In most OECD countries, more than 45% of grandparents take care of at least one grandchild (see OECD (2012)). In the Netherlands (studied in this paper) this fraction is 60%. For children aged 4 to 12 who attend primary school, 20% of parents rely solely on grandparental childcare (see Section 1.2).

[3] Results are based on calculations by the authors using non-public microdata from Statistics Netherlands. Under certain conditions, these microdata are accessible for statistical and scientific research. For further information: microdata@cbs.nl.

ational labor supply spillovers. In particular, we use the first-stage reform estimate to instrument for the labor supply decision of the elderly generation affected by the pension reform. In terms of short-run labor supply spillovers, we find that a one-hour increase in grandmothers' hours worked causes their adult daughters with young children to work around half an hour less. This result is robust to the inclusion of controls and sector fixed effects as well as to variation of the bandwidth. The usual validity tests provide supporting evidence for the necessary RD assumptions. While the reform has important direct effects on grandfathers' labor supply, we do not find spillover effects on their adult children.

To investigate the underlying mechanisms and for a more complete picture of who is affected by labor supply spillovers, we first show that the effects we find are linked to childcare needs and a reduction in grandmothers' childcare availability. In particular, we find no effect on adult daughters without children or with children older than 12. Instead, we find that results are driven by mothers with their youngest child aged 4 to 12 (with the largest effect for children aged 4 to 7) when their children are in primary school (including pre-school). This is consistent with grandparental childcare being particularly important for families with children in this age group, as 20% of these families use grandparents as the sole caretaker (instead of daycare or after-school care), compared to less than 10% of families with a child younger than age 4.[4]

In support of our findings in terms of channels, we explore heterogeneity in terms of distance between where grandmothers and their adult daughters live, the health of grandmothers' partners, and the number of young maternal grandchildren. Our results strongly point to the importance of intergenerational time transfers (i.e., child care provided by grandmothers), as the main channel. We rule out other channels, including money transfers and a "reminder effect", by investigating the effects on other outcomes (such as gross (household) income of grandparents) or on groups without childcare needs.

For a more complete picture of labor supply spillovers, we also examine the impact of grandmothers' retirement on other extended family members, such as adult sons, daughters-in-law, and sons-in-law, and the impacts of grandfathers' retirement. We find that while an increase in grandmothers' labor supply decreases the labor supply of (adult) daughters with young children between ages 4 and 12, it does not have an impact on (adult) sons or daughters-in-law with children in the same age range. This is consistent with maternal grandparents being those who provide childcare in about two-thirds of cases (own calculation based on LISS data). Interestingly, the effects on adult daughters are mirrored in the effect of the opposite sign on their husbands (i.e., the sons-in-law), who increase their labor supply, while the overall effect on household income is zero. This has important implications for the gender gap within households. We do not find spillover effects of

[4]These figures are based on own calculations from the Dutch LISS (Longitudinal Internet studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands). Also see Section 1.2.

grandfathers' labor supply on their adult children, despite large direct reform effects, consistent with grandmothers being more important in providing childcare.

Next, we analyze the importance of long-run spillover effects. We first show the reform impact on (grand)children's educational performance and then investigate dynamic spillover effects on mothers' long-run labor market outcomes and on the child penalty and gender gap. As a measure of children's educational outcomes, we use their performance on a high-stake test (*Cito* test), taken at the end of primary school to determine the school track children follow in secondary school. The school track is critical for individuals' long-run educational and labor market outcomes, since enrolling in university is only possible with a degree from the academic track. Interestingly, we find positive reform effects on the educational performance of children who were aged 4 to 7 when their grandmothers were between ages 60 and 64. These children, who experienced a substitution away from grandparental care towards maternal care, scored on average 17 percent of a standard deviation higher on the Cito test. Effects are particularly strong for girls.[5] While we do not find effects for children aged 8 to 10, children aged 11 to 12 are strongly negatively affected. This effect is entirely driven by boys. Their performance on the Cito test decreases by more than 20 percent of a standard deviation and they are 7 percentage points less likely to receive a recommendation for the highest track in secondary school. For this age group, grandmothers' time availability decreases while mothers do not change their labor supply. Using supplementary data on childcare take-up, we show that after-school care substitutes for the decrease in grandparental care (the probability of taking up after-school care increases by 2 percentage points and the number of hours by 9). In addition, some children in this age group may stay at home alone without supervision for a few hours, with negative consequences in particular for boys (consistent with findings by Aizer (2004)). While there is no change in formal childcare for children aged 8 to 10, the use of formal childcare is reduced for children aged 4 to 7. This suggests that the substitution of grandmothers' care and formal care through maternal care positively affects children and particularly girls in this age group.

Lastly, we investigate the importance of dynamic spillover effects on mothers' labor supply and earnings and the reform impacts on the child penalty and gender gap. Building on the framework by Kleven et al. (2019a), we show that in addition to the immediate impacts, changes in grandmothers' labor supply have dynamic long-run effects on child penalties. In particular, a significant gap in mothers' hours worked and earnings opens up between mothers with (grand)mothers who are affected versus those unaffected by the reform. While mothers who have (grand)mothers unaffected by the reform start to recover about 5 or 6 years after the birth of their child, hours worked and earnings remain at the same (low) level for mothers whose (grand)mothers are affected by the

[5]These results are in line with recent findings in the literature showing that children receiving formal childcare instead of maternal care have worse cognitive skills, in particular in the case of girls (see, e.g., Fort et al. (2020)), and worse non-cognitive skills (see, e.g., Baker et al. (2008) and Baker et al. (2019)).

reform. These dynamic effects on the gender gap in labor supply are quite substantial: while mothers with treated (grand)mothers face a child penalty of 30% in terms of working hours seven years after the first birth, mothers whose (grand)mothers are not treated only face a penalty of 11%. The reform thus magnifies the already existing child penalty and gender gaps (both within households and overall) and has important distributional consequences.

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 on their adult daughters and grandchildren. Spillover effects of such policy reforms might counteract (or even reverse) direct reform effects. 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 assumption 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 outweighs the gain in tax revenue from delaying the retirement of grandmothers.[6] Second, our findings in terms of children's educational outcomes suggest the importance of quality of care. We show that unsupervised time at home due to a lack of grandparental care and/or low-quality formal care can negatively affect the performance of children in high-stake tests, with decisive long-run implications. The positive effects we find for children aged 4 to 7 underline the importance of high-quality care for children's cognitive development. Such high-quality care options can be made possible and shared by both parents through generous parental leave policies or by improving the quality of formal care choices.

Our results show that public policies, such as pension reforms, can trigger multigenerational spillover effects with important distributional consequences. In particular, while the reform has reached the intended goal of increasing labor supply in old age, maternal labor supply has in fact decreased. This has critical implications for the long-run labor market outcomes of women and for the child penalty and gender gap within households and the society overall. On the other hand, children –in particular younger ones– appear to have benefited from the increase in maternal care time. Our paper thereby adds to a recent strand of the literature that estimates the long-term costs and benefits of public policies and examines how they can have opposing effects on different generations (see, e.g., Hoynes et al. (2016a), Bailey et al. (2020) and Aizer et al. (2022) on safety net programs).

Our paper adds to the following four strands of literature. First, it speaks directly to the scarce literature on the spillover effects of pension policies. Apart from effects on spouses (see Hurd (1990), Coile (2004), Stancanelli and Van Soest (2012), Lalive and Parrotta (2017)) and work place peers (see Duflo and Saez (2002), Duflo and Saez (2003) and Brown and Laschever (2012)),

[6] See Appendix E for detailed steps of the calculation.

the literature on spillovers of such policies across generations is extremely scarce.[7] One notable exception is Bratti et al. (2018), who study the impact of grandparental availability on maternal labor force participation by exploring an Italian pension reform. Using an instrumental-variable approach, they find that mothers of children under age 15 whose own mothers are retirement eligible have a 11% higher probability of being in the labor force than those whose mothers are ineligible. Our paper goes beyond the effects on adult daughters and provides a more complete picture of intergenerational spillover effects. To the best of our knowledge, our paper is the first to causally estimate impacts on all family members' labor supply and to show evidence of changes in childcare choices and the resulting effects on children's academic performance. Moreover, the high-quality Dutch data allow us to show labor supply effects beyond the extensive margin responses and to investigate longer-run 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, such as formal childcare (e.g., Baker et al., 2008; Fitzpatrick, 2010; Bauernschuster and Schlotter, 2015; Baker et al., 2019) and parental leave policies (e.g. Gruber, 1994; Schönberg and Ludsteck, 2014; Kleven et al., 2020). Our paper provides causal evidence of strong responses to the availability of grandparental care and thereby contributes to the limited evidence available on the effects of grandparents on maternal labor supply (Posadas and Vidal-Fernandez, 2013; Bratti et al., 2018; Fenoll, 2020). Moreover, we provide detailed evidence on the underlying mechanisms, including an analysis of labor supply spillovers for all extended family members. Our paper also relates to the literature on peer effects in maternal labor supply decisions (see Nicoletti et al. (2018) on sibling spillovers and Olivetti et al. (2018) on peer effects of classmates' mothers). These papers do not look at spillovers of a public policy but use the strategy of overlapping peer groups to identify peer effects. Our contribution to this literature is to provide comprehensive insights into spillover/peer effects across generations.

Third, our paper relates to the literature on parental investments, childcare choices and skill development in childhood and adolescence, which shows that maternal time is an important determinant for children's cognitive development (Carneiro et al., 2013; Del Bono et al., 2016; Francesconi and Heckman, 2016; Bastian and Lochner, forthcoming). Studies examining extensions in paid maternity leave (usually when the child is 0-2 years old) overall find no effects on child well-being (Dustmann and Schönberg, 2012; Dahl et al., 2016; Danzer and Lavy, 2018), but a positive effect of a longer period of maternity leave for children from privileged families (Danzer and Lavy, 2018; Ginja et al., 2020). Studies on the effect of formal childcare tend to find small or no overall effects on children's cognitive and non-cognitive skills (e.g., Cornelissen et al., 2018; Felfe and Lalive,

[7]Papers examining intergenerational spillovers of other types of policies are, for example, Dahl et al. (2014), Aizer et al. (2016) and Hoynes et al. (2016b) on intergenerational effects of welfare programs and Black et al. (2005) on intergenerational effects of education policies. Papers that investigate labor supply spillovers across spouses are, for example, Goux et al. (2014) and Fadlon and Nielsen (2019).

2018) or even negative effects on children's skills and well-being (e.g., Baker et al., 2008; Fort et al., 2020; Baker et al., 2019), in particular for girls and/or children in more financially advantaged families. Consistent with many of these studies, we find that an increase in maternal time spent with the children has positive effects on their cognitive skills and educational performance, in particular for girls. Our paper also directly contributes to the very few studies on the impact of grandparental care, which have inconclusive findings.[8] In particular, we show that a substitution from formal and grandparental care to maternal care tends to have positive effects on children's educational performance, while a substitution from grandparental care to formal care and/or no adult supervision has strong negative effects particularly for boys (consistent with the findings of Aizer (2004)).

Lastly, our paper is connected to research on gender inequality in the labor market (see reviews by, e.g., Altonji and Blank (1999); Blau and Kahn (2017) and Olivetti and Petrongolo (2016)). A large body of studies show the career costs of children (Bertrand et al., 2010; Adda et al., 2017; Lundborg et al., 2017) and the large and persistent impact of children on the gender earnings gap (Daniel et al., 2013; Angelov et al., 2016; Kleven et al., 2019a,b; Cubas et al., 2021). We contribute to this literature by evaluating the impact of grandmothers' labor supply on child penalties. Building on the framework developed by Kleven et al. (2019a), we show that in addition to the immediate impact, changes in grandmothers' labor supply have dynamic long-run effects on child penalties. The pension reform, which aims to prolong the working life of the older generation, has unintended consequences on the labor supply of adult daughters with young children, in that the lack of care support from grandmothers slows down the recovery of earnings and working hours to women's pre-birth levels.

The paper proceeds as follows: Section 2 describes the core features of the Dutch pension system, the pension reform, and details of childcare arrangements in the Netherlands, while Section 3 presents the data and the empirical design. Section 4 presents the short-run spillover effects, and Section 5 discusses the mechanisms and threats to validity. Section 6 presents the long-run reform impacts on grandchildren and the dynamic implications for the gender gap and child penalties. Finally, Section 7 concludes.

[8] Of the two studies that we are aware of, Del Boca et al. (2018) show that compared with children in formal childcare, children between ages 3 and 7 cared for by their grandparents are better at naming objects but perform worse in terms of non-verbal reasoning in the UK. Zhang et al. (2021) find that compared with parental care, grandparental care delays the achievement of children between ages 1 and 5 in China.

1 Institutional Setting

1.1 Early retirement schemes and the 2006 Dutch pension reform

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 aged 65 and above a flat-rate pension. 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. Contribution to the second pillar is mandatory. 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. During our sample period, the earliest possible age to claim occupational pensions is between ages 55 and 60, depending on occupational group. The third pillar consists of non-mandatory savings. See Appendix Section 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 (see Appendix A.2 and Euwals et al. (2010) for a summary and further details on the evolution of Dutch early retirement schemes). Retiring early was and still is only possible through the early pension (ER) scheme, which is part of the occupational pension scheme. Before 2006, early retirement was subsidized through the tax system, as 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). Consequently, around 80% of all workers retired at the age of 62 or younger before 2006 (Statistics Netherlands, 2009).

Since January 1, 2006, the tax benefits for early retirement schemes were eliminated. The goal was to encourage labor market participation of the elderly by speeding up the transition towards an actuarially fair early retirement system. Two types of individuals are exempted from the new bill. First, people, who were 55 years or older before January 1, 2005 are not affected by the reform. Thus, people born before January 1950 are exempted from the changes, while those born in or after January 1950 are no longer enjoying the tax advantages as of January 1, 2006. Second, people who have claimed early retirement before 2006 are not affected by the reform. 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.[9]

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

[9]The reform bill No. 29760 includes a clause to adjust fiscal policy VUT and prepension (Wet voor aanpassing fiscal behandeling VUT and pre-pension) and is sometimes referred to as the 56-plus scheme (de 56-plusregeling). For further details, see Appendix A.2 .

change being concentrated between ages 60 and 64. Therefore, the reform led to a quasi-exogenous change in early retirement incentives of the older generations, which allows us to causally estimate the impact of grandmothers' labor supply on mothers' and children's outcomes. More specifically, we employ a Regression Discontinuity Design based on grandmothers born since January 1950, as opposed to before, and compare outcomes (of grandmothers, mothers and children) when grandmothers are aged 60 to 64.

1.2 Grandparents and child care

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 the majority of 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 in 2011, according to the Survey of Income and Program Participation, 47% of children below age 15 receive grandparent-provided child care. For 93% of pre-school children grandparents are the primary child care arrangement (Laughlin (2010) and Rupert and Zanella (2018)).

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 child care than fathers. Even when conditioning 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 that they work 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, in Figure A1 we show the distribution of the different types of childcare (other than parents), in particular formal care, grandparental care, and other types of informal care. See Appendix A.4 for a detailed description of the childcare system. 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

[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 before our sample period. For more details see section B.1.

some formal care. The two most common care arrangements for children younger than 4 years old are paid formal care and a combination of 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% in the case of younger children below age 4. Lastly, the data show that maternal grandparents are more important in terms of care giving than paternal grandparents, in that more than 60% of the care-giving grandparents are maternal grandparents.

2 Data and Empirical Strategy

2.1 Data

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 us to follow families across generations and over time. Birth and marriage records enable us to link several 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 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", is referred to as *mothers* or *fathers*, and the "third generation" is referred to as *children*. We also analyze 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), as well as childcare usage, while the elderly generation is between 60 and 64.

Baseline sample: In our analysis, we start with all grandmothers born between 1948 and 1951 who have at least one daughter, since the main focus of our analysis is to estimate labor supply spillovers of grandmothers on their adult daughters. We exclude first-generation individuals who are migrants. We further drop grandmothers who are unlikely to be affected by the reform, that is, grandmothers who are self-employed or family workers, because they do not rely on early retirement schemes. We also drop grandmothers who exit the labor force before age 50, those who were never active in the labor market, those who claimed disability before the age of 55, and those who died before age 65. We are thus left with 62% of the 1948-1951 generation of women (for details on the sample construction, see Appendix B.2 and Table A12, which tests sample restrictions and shows that they are not affected by the reform).

Since we aim to investigate the impact of grandmothers' labor supply on their adult daughters'

labor supply while they have young children, the baseline sample covers adult daughters who have at least one young child when the grandmothers are aged 60 to 64.[11] 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 split this sample by the age of the youngest child. Our baseline sample consists of 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 (see Section 1.2).

Lastly, for the baseline RD analysis, we keep families with grandmothers who are within the optimal bandwidth of 8 months around January 1950 (see Section 2.2 for a discussion of the bandwidth choice). We end up with 23,497 mothers in our baseline sample. 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 deceased grandmothers, and other family members.

Summary statistics: In the baseline RD sample, grandmothers have on average 2.5 adult children and 1.7 adult daughters. The mothers 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. Our main outcome variables capture the labor supply of grandmothers and mothers, which are measured when the grandmothers are between ages 60 and 64. Grandmothers work on average 37 hours per month and earn 638 euros per month.[12] Their likelihood of employment is 42%, and 5% are employed full-time. On average, grandmothers exit the labor force at age 61 and start claiming retirement at age 63. Mothers work on average 78 hours and earn 1534 euros per month. Around 78% of them are employed, and 6% are in full-time employment. The RD sample is very comparable to the full sample of adult daughters with at least one young child of grandmothers born between 1948 and 1951 (for details see, Appendix B and Table A1).

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 in a particular age range while their grandmother was between ages 60 and 64.

Dutch primary education is for children aged 4 to 12. At the end of primary school, pupils take the Cito test, which is a high-stakes standardized test used to sort students into different secondary school tracks (such as vocational, technical, academic tracks). We merge the children in our baseline sample with the test score data, including the final Cito score, number of correct answers overall, and number of correct answers in math and verbal skills. While the Cito test is the test used in the majority of schools to determine the secondary school track for their students, schools

[11] 28.64% of adult daughters do not have any children, while grandmothers are aged 60 to 64. We provide labor supply results for this group in Section 3 and for sons-in-law, adult sons and daughters-in-law in Section 4. In Section 4.4.2 we show that the reform had no effects on adult daughters' fertility.

[12] All income measures are CPI adjusted for the year 2015.

can opt for other test types, which we do not have data on. 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 A14 presents a comparison of the children in our baseline sample with the children who took the Cito test showing that the children have similar characteristics.[13] We supplement our analysis on 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.

2.2 The Fuzzy Regression Discontinuity

In this paper, we address the research question as to whether, to what extent, and through which channels grandmothers' labor supply affects the maternal labor supply of their adult daughters and the educational performance of the grandchildren. It is difficult to causally estimate the effects for two reasons. First, unobserved variables can 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 might be reversed causality as grandmothers' retirement decisions can be affected by childcare decisions. The timing of grand-parenthood can cause a reduction in the labor supply of grandmothers (Rupert and Zanella (2018); Frimmel et al. (2020)). We address this issue by exploring a cohort-based reform in early retirement incentives in the Netherlands.

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

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

where r_i^{GM} is the running variable defined as the 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 $D = \mathbb{1}(r_i^{GM} \geq 0)$. 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 first stage estimated impact of the reform on grandmothers' labor supply outcomes. X contains demographic characteristics of the grandmother and the mother, including mother's age and migration background, number of her siblings and sisters, the age at first birth, mother's and grandmother's predetermined marital status, mother's

[13] The only notable difference between the sample of youngest children and the CITO sample of children that could be matched with the test score data is children's age. Children in the CITO sample are slightly older, because in later years further test options became available for schools (on which we do not have data), so that in the CITO sample older children receive more weight.

predetermined number of children, disability status of the grandmother's partner, predetermined 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. In the case where a grandmother may have changed her sector at some point during her employment history, we consider the one in which she was employed the longest.

In our analysis, we use a bandwidth of 8 months around the cutoff, which is the average of the mean square error optimal bandwidths generated by the Calonico et al. (2017) and Calonico et al. (2018) procedure for the different outcomes we consider, and a linear specification.[14] We present robustness results in Section 4.4.3.

Second, we investigate the middle generation: the mothers. The reform allows us to causally estimate the impact 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. The coefficient $\widehat{\alpha_1^M}$ is the the reduced form effect of the reform on outcomes of mothers.

The effect of grandmothers' labor supply on mothers' labor supply can be obtained as the ratio of the discontinuity in mothers' labor supply ($\widehat{\alpha_1^M}$) to the discontinuity in grandmothers' labor supply ($\widehat{\alpha_1^{GM}}$) (Lee and Lemieux, 2010). The following local linear regression gives us the two stage least square (2SLS) fuzzy RD estimate:

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

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 look at the youngest generation - the (grand)children. We examine the reform impact on children's educational performance. The reduced form model for children' 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. X 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

[14] In Table A11, we show the estimates using mean square error optimal bandwidths generated by Calonico et al. (2017) for each of the outcome variables.

estimated reform impact on children.[15]

2.3 Assumptions

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 A3a and b show the density plot of grandmothers 24 months and 8 months around the cutoff. The bin size is the grandmothers' birth month. Figures A3c and d show the density plot of mothers 24 months and 8 months around the cutoff. The fluctuating pattern of the density plots are 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 e and f). This suggests seasonal patterns which commonly occur in terms of birth rates (and which are not driven, for example, by 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 above described density plots.

Smoothness in covariates: Table A2 reports the estimated impact of grandmothers being born since January 1950 (as opposed to before) on a list of predetermined characteristics of grandmothers and mothers (using Equations 1 and 2). All variables are predetermined and refer to the time period when the grandmothers were aged 50 to 53. We show the estimated effects under different specifications: linear (column (1)) and quadratic (column (2)). All specifications use a bandwidth of 8 months. Covariates are smooth across the cutoff. In particular, there are no significant differences in the covariates above versus below the cutoff, with two exceptions out of 21 variables (significant at the 10 and 5 percent level, respectively), suggesting that the distribution of pre-determined characteristics is balanced around the cutoff.

Instrument validity: There are three conditions necessary to interpret the two-stage least squares estimate. First, the grandmothers' birthdate is strongly associated with the grandmothers' labor supply. We show the validity and magnitude of the first-stage relationship in Section 3.1. Second, a grandmother's birthdate only impacts her adult daughter's labor supply outcome through changes in her own labor supply. The exclusion restriction could be violated if a grandmother who was born before or since 1950 affects her adult daughters' labor supply and grandchildren's education outcomes through channels other than her own labor supply. This assumption is fundamentally

[15] We present results from regressions with clustered standard errors at primary school level. The clustering allows for correlations of test-performance within schools. These results are also robust to clustering at the mothers' level and to two-way clustering on mothers' and primary school level.

untestable. We argue that the exclusion restriction assumption is reasonable because there are no other reforms with the same grandmother birthdate cutoff. Moreover, a number of results in the mechanisms analysis (Section 4) indicate that it is the changes in grandmothers' time availability that causes the mothers to work less. In Section 4.4.1, we provide a more in-depth discussion about the exclusion restriction based on further empirical findings. 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 make grandmothers 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 readily satisfied.

3 Short-Run Spillover Effects: Direct and Indirect Effects

In this section, we first investigate the direct reform impact on grandmothers' labor supply. Then we show the indirect spillover effects on mothers with their youngest child between age 4 and 12 (baseline sample). Finally, we further divide our baseline sample by the age of the youngest child to highlight the importance of childcare responsibilities.

We hypothesize that mothers with children of primary school ages are most strongly affected by changes in the informal care provided by grandmothers. This is because the need for childcare and the compatibility of grandparental care and formal care varies according to the child's age. First, as the children grow older, the amount of care needed decreases. Children below primary school age (younger than 4 years), require more care and more intensive care, while children older than 12 years need limited or no supervision. Second, the amount of free public childcare hours increases as the children grow up. While daycare for children below age 4 is costly, from age 4 onwards children can attend primary school which is free of charge, and offers around 30 to 35 hours of care per week. According to the LISS data, grandparents provide about three half-days ($\sim$ 9 hours) per week of childcare, which – combined with the time in primary school – can free mothers from childcare responsibilities and allow them to work more during the week.

In contrast, children below age 4 do not yet attend (pre)school and thus require care full-time, which grandparents are rarely able to provide in its entirety. In fact, we find – based on own calculations from the LISS data – that for children aged 4 to 12, grandparents are the only childcare option for 20% of families, while for children below age 4, this is only the case for less than 10% of families (see discussion and further details in Section 1.2). Thus, for very young children below age 4, mothers either stay home completely or if they do work, they mostly rely on formal daycare. As a result, grandparents' time availability is less critical for this age group and hence less likely to alter mothers' labor supply decisions.

Lastly, children above age 12 are likely to require little or no supervision after their (longer)

school day. Also "mothers" (i.e., adult daughters) without children obviously require no childcare at all. We present the estimated effects for the two latter groups (children above 12 years and no children) to support our finding that the effects we find are indeed linked to time availability and the childcare provision of grandparents.

3.1 Effects on Grandmothers' Labor Supply: First-stage

Figure 1 provides graphical evidence of the first-stage reform impact on grandmothers' total hours worked. It shows the bin scatter plot of total monthly hours worked as a function of distance to the cutoff for grandmothers' birth month, which is January 1950. The solid lines are the linear fitted lines, and the shaded areas indicate the 95 percent confidence interval. We can see that grandmothers born between April 1949 to December 1949 work on average 33 hours per month between age 60 and 64, while grandmothers born between January 1950 and September 1950 work on average 41 hours between the same ages. Moreover, there is a clear jump at the cutoff from about 34.5 to 42.5 hours per month.

First-stage estimates of the pension reform are shown in Panel A of Table 1. Columns 1, 2, and 3 show the results for a local linear regression without controls, with controls, and with controls and sector fixed effects. All three specifications use a bandwidth of 8 months (as in all other tables, unless otherwise specified; see discussion in Section 2.2). Columns 4 and 5 show the robustness of the results using a bandwidth of 6 months and of 12 months. Standard errors are clustered at the grandmother level, since in the main analysis grandmothers are in the sample multiple times if they have several daughters with children in the relevant age range. The regression results are consistent with the graphical analysis. We find that the reform increases grandmothers' monthly hours worked by around 6.2 hours (equivalent to an increase of 18 percent). This effect is robust across specifications and bandwidths and is highly significant at the one percent level.

3.2 Effects on Mothers' Labor Supply: Spillover Effects

Graphical evidence in Figure 1 show that mothers' labor supply drops sharply at the cutoff. The corresponding estimates from the fuzzy regression discontinuity design are in Panel B of Table 1. Columns 1, 2, and 3 are based on a local linear regression without controls, with controls, and with controls and sector fixed effect using a bandwidth of 8 months, while Columns 4 and 5 show the robustness of the results using a bandwidth of 6 months and of 12 months. Standard errors are clustered at the grandmother level. We find that a one-hour increase in grandmothers' monthly working hours induces a decline in mothers' monthly hours worked of around 28 minutes (0.47 hours) at 5 percent significance level. When we change the bandwidth to 6 and 12 months, the estimate remains similar in terms of size and significance. In fact, reducing the bandwidth to 6

months even increases the estimated coefficient in absolute value to 0.63 at 5 percent significance; that is, a one-hour increase in grandmothers' monthly working hours decreases mothers' monthly hours worked by 38 minutes.

The patterns are similar for other measures of labor supply. When grandmothers work more, mothers are less likely to engage in formal employment (significant at the 10 percent level). The probability of working full-time decreases (significant at 10 percent level for a bandwidth of 12 months). Since a change in grandmothers' labor supply affects mothers' employment probability (and thus changes differentially who is working to the right and left 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.[16]

To understand whether and, if yes, how the change in mothers' labor supply is related to childcare responsibilities, we further divide our sample based on the childcare needs of the mother. In particular, we classify mothers (middle generation) with their youngest child below age 13 into a group *with* childcare needs (see Table 2, Columns 1 to 3) and the ones with their youngest child above age 13 or without children into a group with *little or no* childcare needs (see Columns 4 and 5 of Table 2). Moreover, we further divide mothers with the youngest child between age 0 to 12 into three different categories: 0 to 3, 4 to 7, and 8 to 12. Children aged 0 to 3 require the largest amount of care (that is, full-time care), since the child does not yet attend school. Starting at age 4, children attend school (which for the first two years is pre-school, see discussions in Section A.4) and are thus taken care of for around 6 to 7 hours per day. A mother working part-time thus requires only a few additional hours of help. Hence, the grandmothers' availability can potentially fill this gap and allow mothers to work more. The hours attended in school increase with age and at some point, children are able to spend some time unsupervised. Thus, we also present the results for children aged 8 to 12.[17]

[16]Table 1 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, 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 percentage points translate into change in total hours worked, if the entire response in hours were 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 0.47 hours, larger than 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 a larger role of the intensive margin response in terms of the overall decline in total hours.

[17]See the introduction of Section 3 for a discussion of our empirical hypotheses and detailed reasons for the age group partition. Moreover, we show in Table A13 that the probability of having the youngest child(ren) in different age groups is not affected by the reform.

Table 2 shows the estimates. As hypothesized, we find effects of the increase in grandmothers' hours worked (due to the pension reform), but only among mothers with childcare need. More specifically, we find the strongest effects on the labor supply of mothers with a primary school-aged child between 4 and 7. A one-hour increase in grandmothers' monthly working hour induces a decline in mothers' monthly hours worked of around 32 minutes (0.53 hours), which is significant at the five percent level. In contrast we do not find effects on mothers with a child aged 0 to 3, consistent with mothers either staying home or making use of full-time daycare,[18] so that a change in grandmothers' time availability makes less of a difference. Furthermore we do not find significant effects on mothers' labor supply when the youngest child is aged 8 to 12. Lastly, we analyze effects for mothers whose youngest child is above 12 years and adult daughters without children, who have little or no need of childcare/supervision, and thus act as a sort of placebo group. As expected, we find no effect on those two groups, strongly supporting our interpretation that the changes in mothers' labor supply are indeed related to the time availability of grandmothers and their childcare responsibilities. Although the F-statistics are below 10 for those two groups, Panel A of Table A5 shows that the pension reform strongly affects grandmothers of all age groups, including mothers whose youngest child is above 12 years and adult daughters without children (significant at the 5 percent level). Panel B of Table A5 shows the corresponding reduced-form estimates. Consistent with the estimates in Table 2, we show that only mothers with the youngest child between 4 and 7 respond significantly to the pension reform. The effects on hours worked for the two groups with little/no childcare need are close to zero and even of the opposite sign and thus significantly different from the effect on mothers with youngest child aged 4 to 7. In the next section, we will provide further evidence on the mechanisms underlying our results.

4 Short-Run Spillover Effects: Mechanisms and Validity

The goal of this section is to further investigate the mechanisms underlying our results discussed above. In particular, we aim to provide additional evidence that the exogenous changes in pension incentives affect mothers' labor supply through changes in the time grandmothers can devote to take care of their grandchildren. The first piece of evidence supporting the "time transfer channel" is based on the results for different age groups (younger than 12 versus older than 12/no kids), which suggests that the negative effects on mothers' labor supply is indeed related to childcare needs and thus strongly points towards a time transfer channel (instead of, say, a monetary transfer channel, since this should also affect mothers with older children or adult daughters without children).

In principle, it could be that grandparents support their adult children (who themselves have

[18] In the Netherlands, there is a strong social norm to stay at home with the children when they are below school age (Swart et al., 2019).

young children) via monetary transfers for formal child care. While we do not have direct data on monetary gift exchanges and therefore cannot directly test the “monetary transfer channel", we can test whether the gross income of grandmothers changes in the first place. However, we show in Table A3 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 any case, monetary support for formal child care by grandparents would not explain why after the reform when grandmothers work more, mothers work less.

If the time transfer channel plays a critical role, we have the following three hypotheses. First, we expect a smaller impact if grandmothers have other care responsibilities. Since those grandmothers have little time for childcare already in the absence of the reform, the reform should have little impact on grandmothers' availability for childcare. Second, we expect grandmothers who live far away to have a smaller or no impact, as they are unlikely to provide care regularly. If, on the other hand, grandmothers support childcare via monetary transfers, the residential location should be irrelevant. Third, we expect grandmothers with only one young maternal-grandchild to have a larger impact, as their time is not shared with other grandchildren and other daughters. We test these hypotheses in Table 3, as discussed in the upcoming section.

Furthermore, we test whether there is a similar effect for grandfathers on their daughters and study the impact of grandmothers' labor supply on other family members in the middle generation, such as sons, daughters-in-law and sons-in-law. We have seen in Section 1.2 that the majority of grandparents providing child care are maternal grandparents (more than 60%). Thus, if the effects we find are due to a decrease in time availability, we should observe larger effects for maternal grandparents and thus smaller effects on grandmothers' daughters-in-law.

4.1 Heterogeneity by Health, Proximity and Family Composition

Table 3 shows heterogeneous effects by health status of the grandmother's partner, the proximity of the grandmother, and by family composition. Columns 1 and 2 of Table 3 show the results for grandmothers whose partner (mostly the grandfather) is unhealthy 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, since otherwise they would not have time for child care even in the absence of the reform.

Table 3, Columns 3 and 4, show results by the distance of where the grandmother lives relative to her adult daughter. We find strong spillover effects for grandmothers living nearby, while effects are substantially smaller and insignificant when the grandmother lives in a different municipality than her daughter. In particular, one additional hour worked per month by grandmothers living

nearby causes mothers to work about 50 minutes (0.82 hours) less. In addition, the probability of employment declines significantly for grandmothers living nearby and the difference between the two subgroups is significant at the 10 percent level. Thus, the finding that grandmothers living close have large impacts on their daughters' labor supply, while those that live in a different municipality (and thus were unlikely to provide regular childcare even in the absence of the reform) do not, strongly supports the time transfer channel.

Columns 5 and 6 of Table 3 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.[19] Indeed, as hypothesized, we find that grandmothers who have exactly one maternal grandchild between ages 4 and 7 have a large and significant impact. One additional hour worked per month by grandmothers causes the mothers to work about 0.83 hours (50 minutes) less. In contrast, the impact on the other group is small and insignificant.

4.2 Grandfathers' Effect

We provide further evidence for the time transfer channel by looking at grandfathers. The reform applies to both genders, therefore grandfathers are also affected. In fact, first-stage results on grandfathers' labor supply are even stronger than on grandmothers due to their stronger attachment to the labor force (see Table A6). While both grandparents can provide childcare, previous studies show that grandmothers are more likely than grandfathers to be engaged in childcare activities (Jappens and Van Bavel, 2012; Janta, 2014). Therefore, if the mechanism behind our results is the "time transfer" channel, we expect to see grandfathers having smaller impacts on daughters' labor supply. If instead it is the "monetary transfer" channel, we expect to see a similar impact by grandfathers, as they could provide money equally well as grandmothers.

In Panel A of Table 4, we compare the effect on (adult) daughters' labor supply of grandfathers (Column 1) versus grandmothers (Column 3). We find that grandfathers have a very limited impact on daughters' labor supply. The impact on hours worked, employment probabilities and probability of full-time employment are an order of magnitude smaller than for grandmothers and only the effect in terms of employment probability is significant (but less than a third in magnitude compared to the effect of grandmothers).

In summary, even though grandfathers' labor supply is strongly affected by the pension reform, we

[19] We focus on maternal grandmothers, since they are most relevant in terms of taking care of grandchildren (see evidence based on LISS survey data discussed in Section 1.2; also see our discussion below where we compare the effects of grandmothers' labor supply on daughters and daughters-in-law and our results on grandfathers). Moreover, we focus on grandchildren aged 4 to 7, since we find the strongest effects for this subsample, and we know from the LISS data that grandparents play a particularly large role as sole childcare providers for primary school-aged children (Section 1.2).

find that grandmothers' labor supply and time availability causes significant changes in daughters' labor supply, while grandfathers have little impact. These findings provide further supportive evidence for the time transfer channel; that is, the change in grandmothers' time engagement in childcare activities is the underlying factor that drives our findings.

4.3 Impacts on Other Family Members

For a more complete understanding of the reform and labor supply spillovers on (extended) family members and to provide further evidence on mechanisms, we investigate the impact of grandmothers' labor supply not only on (adult) daughters, but also on other family members in the middle generation, such as sons-in-law, sons, and daughters-in-law. First, we expect maternal grandmothers to have a larger impact, if the underlying mechanism of our findings is a time transfer channel. Survey evidence from LISS data (as discussed in Section 1.2), as well as the existing literature show that maternal grandparents are more likely to provide childcare support than paternal grandparents (see, e.g., Danielsbacka et al. (2019) using the Survey of Health, Ageing and Retirement in Europe (SHARE) on 11 countries in Europe). Moreover, this exercise provides a more complete picture of the intergenerational spillover effect on all extended family members. Thereby it helps us understand the full scope of the unintended side effects of pension reforms, including distributional effects within families.

Panel B of Table 4 compares the impact of grandmothers' labor supply on her daughters, sons-in-law, sons, and daughters-in-law, for whom their youngest child is aged between 4 and 12. We find no significant impacts on sons and daughters-in-law. Only maternal grandmothers matter, consistent with the survey evidence discussed above.

As for the core families with the youngest child aged 4 to 12, we have shown that in response to grandmothers' labor supply increases, mothers work less. Panel B of Table 4 shows that their husbands work more, most likely to compensate for the loss in their wives' labor earnings.[20] Indeed, we find that there is no impact on overall household income. However, the fact that mothers of young children reduce their labor supply in response to the reform, while their husbands' labor supply increases, has important implications in terms of gender inequality within the household as well as in the society overall (for further evidence see Section 5.2 on the reform effect on child penalty and gender gaps).

[20] 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 1.2 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.

4.4 Threats to Validity

4.4.1 Exclusion Restriction

The exclusion restriction requires that a grandmother's birthdate only impacts her adult daughter's labor supply outcome through changes in her labor supply. In principle, the pension reform could lead to an increase in grandmothers' labor income because of delaying retirement, which in turn might have an effect on monetary transfers to her children. However, while the reform increases grandmothers' monthly labor supply, individuals born since 1950 are entitled to less generous pension benefits. Overall, we find that monthly gross income and gross household income remain unchanged (see Table A3), suggesting that changes in monetary transfers in response to the reform are unlikely.

Another potential channel through which the pension reform could affect adult daughters' labor supply outcome is a "reminder effect". Adult children of mothers affected by the pension reform might become more aware that future public pensions are less generous and thus save and work more. However, mothers with young children actually respond to the reform by working less, not more.

Moreover, neither the monetary transfer channel nor a reminder effect could explain why only mothers with young children aged 4 to 12 (who thus have childcare needs) respond with changes in their labor supply (see Table 2), but not mothers with older (or even younger) children or adult daughters without children. It would also not explain why there is only an effect on adult daughters, but not on sons or daughters-in-law, nor why there is no effect of grandfathers (see Table 4). These findings further support the validity of the exclusion restriction.

Lastly, 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. While the LCS program was introduced at the same time as the 2006 early retirement reform, both the treated and control can use this new tax-facilitated saving scheme. Moreover, if anything, the availability of the LCS plan makes our first stage estimates smaller, but does not pose a threat to the validity of our estimates.[21]

4.4.2 Fertility Effects

In our main analysis we focus on mothers with the youngest child between age 4 and 12. However, given that grandparental child care is important, it might also affect the fertility decision of their

[21] In practice, only some high-wage workers manage to retire early using the LCS plan. Lindeboom and Montizaan (2020) show that only around 15% of the 1950 cohort participated in the LCS plan, among which only 16% managed to counteract the effect of the early retirement reform and maintain their previously planned retirement dates. See Appendix A.3 for further details about the Life course savings program.

adult daughters. The limited evidence in the literature on this effect is mixed in terms of size and sign (see, for example, Battistin et al. (2014); Eibich and Siedler (2020)).

The question of whether grandparents' retirement and labor supply decisions affect their adult daughters' fertility is interesting and policy relevant. Moreover, it is important for us to test the fertility responses, since if the pension reform affects fertility we might end up with a selected sample (for example, mothers with grandmothers born since January 1950 might be less likely to have children or may time the births differently). We test this by estimating the impact of having grandmothers born since 1950 on the fertility outcomes of their adult daughters. For this fertility analysis, we use a general sample of all adult daughters of native grandmothers born 8 months around January 1950. All adult daughters with and without children are included in the analysis. Table A9 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. For 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 grandmothers turned age 55. None of these measures of fertility are affected by the reform. [22]

4.4.3 Placebo Tests and Robustness Checks

Two placebo exercises further support the credibility of our estimates. First, we use a sample of mothers with deceased grandmothers. Table A7 shows the estimated reform impacts on the labor supply of adult daughters (mothers) whose mothers (grandmothers) died before age 50. As expected, none of the estimates are significant, and the coefficients are small (compared to Table A4). 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 show the validity of our results by using placebo cutoffs up to 10-months prior and 10-months post the actual cutoff, at a bi-monthly frequency. Figure A5 shows that the F-statistics is only above 10 at the actual cutoff. Table A8 shows the reduced-form estimates and Figure A6 plots the coefficients with 95% confidence intervals. We find no significant effects of grandmothers' labor supply on maternal labor supply at these placebo cutoffs.

Lastly, our estimates are robust to a varying choice of polynomial orders (Table A10) and bandwidths (Table A11). For more details, see Appendix Section C.

[22] Ilciukas (2022) studies the impact of the same pension reform on fertility outcomes and find fertility responses. However, this paper uses a restricted sample — adult daughters of grandmothers born around the cutoff who were married or cohabiting and had no children before the reform. Our fertility sample is more general than Ilciukas (2022), since we do not make any restrictions based on marital status or fertility prior to the reform. Restricting the sample to adult daughters with children before the reform, Ilciukas (2022) does not find fertility responses.

5 Long-Run Spillover Effects

5.1 Effects on Grandchildren's Educational Achievement

We have shown that an (exogenous) increase in grandmothers' labor supply (due to the pension reform) leads to a decrease in mothers' labor supply when they have young children between ages 4 and 12. This suggests a substitution effect away from grandparental care to maternal care, raising the interesting 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-stake test taken at the end of primary school and which helps to place children into different tracks in secondary school (vocational, technical, academic). The performance on the test and the resulting track assignment has important long-run implications in terms of the likelihood of enrolling/completing university (which requires completing the academic track in secondary school), earnings and family formation outcomes (see, e.g., Dustmann et al. (2017) on the longterm effects of early track choice and Kaufmann et al. (2021) on the marriage market effects of university education).

More specifically, we examine the impact on the overall Cito score, on the number of correct answers in the math and verbal component of the test (separately and jointly) and on the likelihood of receiving the recommendation for the highest, i.e. the academic, track in secondary school. Since we find effects on mothers' labor supply that depend on the age of the *youngest* child in the family, we investigate reform spillovers on the performance of the youngest child.

Table 5 presents the RD estimates of the reform spillovers on children's test scores. Panel A shows the results for children who were aged 4 to 12 during the relevant years (i.e., when the grandmother was 60 to 64, as in our entire analysis), since this is when the grandmothers, and, as a result, the mothers change their labor supply behavior. We find that the reform has led to a slight improvement in children's Cito performance, in particular in the math component of the test.[23]

In a second step, we split the sample by age groups (see Panel B of Table 5), 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 Panel B of Table A5). Older children might at some point be able to spend some time unsupervised, which is why we divide the group of older children further. More specifically, we investigate the effect on children's educational performance for three different age groups: 4 to 7 years, 8 to 10 years, and 11 to 12 years.[24]

[23] Table A15 lends support to the smoothness condition in that covariates for the Cito sample are balanced across the cutoff.

[24] Table A16 shows that there is no selection into taking the Cito test for the different age groups.

Panel B of Table 5 shows that results in Panel A are hiding a substantial amount of heterogeneity. In particular, we find strong positive effects on children who were between 4 and 7 years old during the relevant years (see first row of Panel B), i.e. on those children whose mothers' labor supply decreased in response to the reform. Their overall Cito score increases by 17 percent of a standard deviation and the number of correct answers on the verbal and the mathematical component increase by 13 and 20 percent of a standard deviation, respectively (all significant at the one percent level). When we split the sample by gender, we find particularly large positive effects on girls suggesting that girls benefit strongly from the increased interaction with their mothers.[25]

The fourth row of Panel B in Table 5 shows regression estimates for children aged 8 to 10 during the relevant years. For this age group, the children of treated grandmothers do not perform significantly differently from children with untreated grandmothers and this is true for both boys and girls. Table A5 shows that for these children aged 8 to 10, we do not find significant effects on mothers' labor supply.

The seventh row of Panel B in Table 5 presents RD estimates for children aged 11 to 12. For this age group we find important negative reform effects. Children aged 11 to 12 perform 13 percent of a standard deviation worse in terms of overall Cito score and 13 to 19 percent of a standard deviation worse in terms of the correct number of answers on the verbal and the mathematical part of the test. 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 can have important consequences for human capital development). Our findings by gender in Panel B of Table 5 lend some support to this interpretation in that the negative effects on children aged 11 to 12 are entirely driven by boys. The differences between boys and girls are all statistically significant. It is particularly striking that for boys even the likelihood of receiving a recommendation for the highest (academic) track decreases by 7 percentage points (significant at the 5 percent level). Our findings are consistent with girls, who are generally more conscientious, studying for the high-stake test, even if unsupervised, while boys' study behavior at this age depends more strongly on adult supervision.[26] The reform thus has very negative spillover effects on boys aged 11 to 12, since it

[25]These results are in line with recent findings in the literature showing that children receiving formal childcare instead of maternal care have worse cognitive skills (e.g., Fort et al. (2020) find that one additional month in daycare decreases IQ scores by 4.7 percent of a standard deviation, in particularly for girls) and worse non-cognitive skills (e.g., Baker et al. (2008) find that a Canadian universal childcare reform –that increased the use of formal childcare and reduced maternal care time– decreased motor and social development scores by more than 10 percent of a standard deviation).

[26]As discussed in Bertrand and Pan (2013), boys are known to perform worse than girls on many noncognitive dimensions, such as in terms of conscientiousness, attention and behavioral difficulties, inhibitory control. Related to these findings, boys' educational performance is more strongly affected by negative shocks/environments, such as poor school quality, living in a single-parent household etc (see, for example, Bertrand and Pan (2013) and Autor et al. (2016)).

not only decreases their verbal and mathematical abilities, but it also decreases their likelihood of getting into the academic track in secondary school with important consequences for their long-term educational attainment, labor and marriage market outcomes.

While we know 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, mothers' labor supply did not decrease in response to an increase in grandmothers' labor supply. However, we find a strong negative reform effect on children aged 11 to 12, suggesting that there was a change in the mode of supervision in response to the likely decrease in grandmaternal care. This raises the question as to whether substitution has taken place towards formal after-school care or other types of informal care. We therefore supplement our analysis with data on whether parents applied for childcare subsidies and for which type (day care for children aged 4 to 7 or out-of-school care for children aged 4 to 12), for how many hours and for which child.

Table A17 presents the estimated reform impact on the probability to take up and the hours of daycare subsidies as well as on the probability to take up and the hours of out-of-school care. For children aged 4 to 7, we do not find significant reform effects on the likelihood of taking up daycare subsidies or after-school-care subsidies. However, the number of hours in daycare decreases somewhat (by 6 hours per month). Thus, in addition to the substitution from grandmaternal care to maternal care, we see that part of the increase in maternal care time comes from a decrease in formal daycare hours. It appears that mothers who reduce their labor supply in response to the decrease in grandmothers' availability also send their young child to daycare for slightly fewer hours.

While we do not find reform effects in terms of after-school care for children aged 8 to 10, we find a significant increase in the the probability of taking up subsidies for after-school care, as well as in the hours of after-school care (9 hours per months more) for children aged 11 to 12. 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 11 and older are at home unsupervised for a few hours in the afternoon, as hypothesized above.

Interpreting the results in terms of Cito test for 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 and formal childcare. This is consistent with the growing literature studying the relationship between childcare options/parental inputs and child development (e.g., Baker et al., 2008; Fort et al., 2020; Baker et al., 2019). For children aged 8 to 10, we find no effect in terms of test scores, which is consistent with the fact that there are no significant changes in terms of childcare modes. For children aged 11 to 12, we find negative effects on test scores which are concentrated on 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 two years prior to the high-stake test has important negative effects on the performance of boys. This is consistent with the fact that especially boys at this age need adult supervision for their human capital development (compare Aizer (2004)). Moreover, the quality of childcare in the Netherlands could be one potential reason for the test score results. Both expenditures and teacher-child ratio of both pre-primary school and out-of-school care in the Netherlands were below OECD average and at a similar level compared to the U.S. during our sample period (OECD, 2017).

5.2 Dynamic Effects on Mothers' Labor Supply and 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, 2020)). The Netherlands face a similar situation as the U.S. and other developed countries. In particular, the monthly gender wage gap in 2014 was 41.8% (women earn EUR 580 to every EUR 1,000 earned by a man), and thereby the Netherlands are among the OECD countries with a large gender gap (OECD, 2019a). In this section, we connect our empirical findings to this debate and aim to address the question as to whether grandmothers' retirement decisions affect the gender gap and child penalties. Building on the framework developed by Kleven et al. (2019a), first, we 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 grandmothers and untreated grandmothers. For details on the estimation see Appendix Section D.

The child penalty results are based on a sub-sample of parents for whom we observe labor market outcomes in all years from four years prior to seven years after birth. This yields a balanced sample of women and men with their first child born between the years 2009 and 2013.[27] We use the optimal bandwidth of seven months for the outcome total monthly hours worked and take women and men with (grand)mothers born around the January 1950 cutoff. Since we only found negative effects on daughters' labor supply, we focus on the middle generation that has a direct relation to the treatment generation, meaning only (adult) daughters and sons.

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. Such

[27]We exclude teenage births by dropping observations with first birth before age 20 and exclude late entry into parenthood after the age of 40.

effects are easiest to detect for women who have exactly one child, since additional births introduce noise and lead to confounding different factors, such as slower convergence due to treatment or due to additional births. Thus, in the main part of this analysis, we focus on women and men with only one child, which allows us to investigate the child penalty in a cleaner way.[28]

Panel (a) of Figure 2 compares the evolution of mothers' total hours worked around the birth of their child (marked in the figure as event time 0) for women with treated (blue dots) and untreated grandmothers (black triangles). Both the 90 and 95 percent confidence intervals are shown by the shaded areas. We find that women experience a sharp drop in their monthly hours worked within the first year after birth, which amounts to a 21 percent drop relative to their pre-birth work hours.

However, women with treated grandmothers recover more slowly than do women with untreated grandmothers. This difference becomes significant at 5 percent starting around four years after birth (compare Section 3) and the gap increases substantially in the following years. In other words, the reform did not only lead to short-run effects on maternal labor supply but also had dynamic effects in that it led to a slower recovery of the working hours of mothers with young children.

Panel (b) of Figure 2 compares the estimated child penalties for women (blue dots) relative to men (black triangles) separately for treated (left panel) and control groups (right panel). We also calculate the long-run relative child penalty faced by women seven years after birth and report it at the bottom of each panel. We observe that the gender gap in total hours worked starts to converge and decreases to 11 percent seven years after birth for the control group (i.e., with grandmothers not treated by the reform).[29] In contrast, for the treatment group the gap remains wide, and the long-run gap in total hours worked remains at 30 percent seven years after giving birth to their first child.[30] In Panel (b) of Figure 2, we use as men's labor market outcomes the ones of sons of the grandmothers in the relevant age range, because they are unaffected by the reform. However, if we were to use the husbands' of the women in the figure (i.e. grandmothers' sons-in-law), the gap would actually be even wider because sons-in-law increase their hours worked (see Panel B of Table 4), suggesting that the reform had even stronger effects on the gender gap within the household.

Overall, we find that the changes in grandmothers' labor supply decisions do not only affect maternal labor supply in the short run, but there are also dynamic spillover effects in the long run. The decrease in time availability of grandmothers to provide childcare, leads to a significant negative impact on mothers' long-run labor supply and to a substantial increase in the child penalty and in the gender gap within households and in the society overall.

[28] In the Appendix Figure A9, we show that also without this restriction we find dynamic treatment effects of the reform, albeit smaller.

[29] Note that this is a sample with only one child. The convergence of hours worked for the control group comes from the fact that we shut down the impacts from additional child births during the 7 years.

[30] Similarly, we also find that the reform leads to a slower recovery of monthly labor earnings. The dynamic treatment effects and the relative child penalty are smaller, but patterns are similar for labor earnings, see Appendix Figures A7 and A8.

6 Conclusion

This paper provides the first estimates of spillover effects of old age pension across generations on all family members' labor supply and show the resulting effects on children's academic performance. To estimate the effect, we exploit a cohort-based reform in early retirement incentives in the Netherlands. We show that a one-hour increase in grandmothers' hours worked causes their adult daughters with young children to work half an hour less. In contrast, the change in grandmothers' labor supply does not affect adult daughters without child or with older children, nor does it affect sons or daughters-in-law. Combined with the heterogeneity analysis, our evidence indicates the importance of intergenerational time transfers.

In addition to short-run spillover effects, we investigate long-run effects on grandchildren and dynamic effects on maternal labor market outcomes and the child penalty. We find a sizeable positive impact on the Cito test performance of children aged 4 to 7, who have experienced a substitution away from grandparental to maternal care. In particular, girls benefit from spending more time with their mothers. From a policy perspective, high-quality care provided by mothers can be made possible and shared by both parents through generous parental leave policies or by improving the quality of formal care choices. Interestingly, we also find substantial adverse effects on children aged 11 to 12, predominantly driven by boys. Our findings suggest that unsupervised time at home due to a lack of grandparental care and/or low-quality formal care can negatively affect the performance of children in high-stake tests, with decisive long-run implications. When reforming the pension system, governments should keep in mind spillovers to childcare modes.

Furthermore, our results are meaningful for recent policy discussions on gender gap and child penalty. We show that mothers with treated (grand)mothers face a much larger child penalty seven years after the first childbirth, relative to mothers whose (grand)mothers could retire earlier. We thereby provide first evidence that pension reforms can have unintended implications for the child penalty and gender gap within households and in a society overall.

Although at first glance the estimates are only applicable for 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

Adda, Jérôme, Christian Dustmann, and Katrien Stevens, "The career costs of children," *Journal of Political Economy*, 2017, *125* (2), 293–337.

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.

Angelov, Nikolay, Per Johansson, and Erica Lindahl, "Parenthood and the gender gap in pay," *Journal of Labor Economics*, 2016, *34* (3), 545–579.

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.

Bailey, Martha, 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," Technical Report apr 2020.

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 EITC and maternal time use: more time working and less time with kids?," forthcoming.

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.

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.

__ , Claudia Goldin, and Lawrence F Katz, "Dynamics of the gender gap for young professionals in the financial and corporate sectors," *American Economic Journal: Applied Economics*, 2010, *2* (3), 228–55.

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.

Brown, Kristine M. and Ron A. Laschever, "When They're Sixty-Four: Peer Effects and the Timing of Retirement," *American Economic Journal: Applied Economics*, 2012, *4* (3), 90–115.

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, "rdrobust: Software for regression-discontinuity designs," *The Stata Journal*, 2017, *17* (2), 372–404.

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.

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.

Cubas, German, Chinhui Juhn, and Pedro Silos, "Work-care balance over the day and the gender wage gap," in "AEA Papers and Proceedings," Vol. 111 2021, pp. 149–53.

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.

Daniel, Fernández-Kranz, Aitor Lacuesta, and Núria Rodríguez-Planas, "The motherhood earnings dip: Evidence from administrative records," *Journal of Human Resources*, 2013, *48* (1), 169–197.

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.

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.

Duflo, Esther and Emmanuel Saez, "Participation and investment decisions in a retirement plan: The influence of colleagues' choices," *Journal of public Economics*, 2002, *85* (1), 121–148.

__ and __ , "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment," *The Quarterly journal of economics*, 2003, *118* (3), 815–842.

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.

__ , Patrick A Puhani, and Uta Schönberg, "The long-term effects of early track choice," *The Economic Journal*, 2017, *127* (603), 1348–1380.

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.

Fadlon, Itzik and Torben Heien Nielsen, "Household labor supply and the gains from social insurance," *Journal of Public Economics*, mar 2019, *171*, 18–28.

Felfe, Christina and Rafael Lalive, "Does early child care affect children's development?," *Journal of Public Economics*, 2018, *159*, 33–53.

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

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 Donovan, "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.

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.

Goux, Dominique, Eric Maurin, and Barbara Petrongolo, "Worktime Regulations and Spousal Labor Supply," *American Economic Review*, jan 2014, *104* (1), 252–276.

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).

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*, apr 2016, *106* (4), 903–934.

_ , _ , and _ , "Long-run impacts of childhood access to the safety net," *American Economic Review*, 2016, *106* (4), 903–34.

Hurd, Michael D., "The joint retirement decision of husbands and wives," 1990.

Ilciukas, Julius, "Fertility and Parental Retirement," *SSRN Electronic Journal*, 2022.

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.

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," Technical Report, National Bureau of Economic Research 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.

Lee, David S and Thomas Lemieux, "Regression discontinuity designs in economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355.

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

Lundborg, Petter, Erik Plug, and Astrid Würtz Rasmussen, "Can women have children and a career? IV evidence from IVF treatments," *American Economic Review*, 2017, *107* (6), 1611–37.

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.

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.

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.

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.

__ , 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.

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.

Swart, Lisette, Wiljan Van den Berge, and Karen van der Wiel, "Do parents work more when children start school? Evidence from the Netherlands," *IZA Discussion Paper*, 2019, *12207*.

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.

7 Tables and Figures

Table 1: Fuzzy RD estimates - impact on mothers' labor supply

	Fuzzy RD estimates					
	(1)	(2)	(3)	(4)	(5)	Means at cutoff
Panel A:	**First-stage estimates - impact on GM's total labor supply**					
Total monthly hours worked	6.801***	6.831***	6.174***	5.947***	5.398***	34.418
	[1.493]	[1.392]	[1.340]	[1.556]	[1.086]	[47.608]
Panel B:	**Impact on mothers' labor supply**					
Total monthly hours worked	-0.469**	-0.413**	-0.465**	-0.630**	-0.405*	78.876
	[0.219]	[0.204]	[0.229]	[0.298]	[0.207]	[47.744]
Other labor supply measures						
Prob(employed)	-0.003*	-0.003*	-0.003*	-0.004*	-0.003*	0.785
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.378]
Prob(full-time employed)	-0.001	-0.001	-0.001	-0.001	-0.002*	0.066
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.214]
F-stat	20.75	24.09	21.22	14.62	24.69	
Obs. Mothers	23497	23497	23497	17930	34592	4018
Obs. Grandmothers	19548	19548	19548	14959	28739	
Bandwidth	8	8	8	6	12	
Controls	NO	YES	YES	YES	YES	
Sector FE	NO	NO	YES	YES	YES	

Notes: Table 1 shows the coefficient estimates of grandmothers' total monthly hours worked on mothers' labor supply. The top panel reports first-stage estimates and the bottom panel reports 2SLS fuzzy RD estimates. An indicator for the grandmother being born since January 1950 serves as the instrument for grandmother's total monthly hours worked. Columns 1, 2, and 3 show the results without controls, with controls, and with both controls and sector fixed effects, respectively. We use local linear regressions with a bandwidth of 8 months (optimal bandwidth, generated by the Calonico et al. (2017) and Calonico et al. (2018) procedure). Columns 4 and 5 show local linear regressions with a bandwidth of 6 and 12 months, respectively. Sample means at the cutoff (measured in the three months before the cutoff) are reported in Column 6. All outcomes are measured when the grandmothers are between ages 60 and 64. Robust standard errors clustered at grandmothers' level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table 2: Fuzzy RD estimates - by age of the youngest child

	Childcare need				
	more			no/less	
	Age of the youngest child				
	0-3	4 - 7	8-12	13 - 18	No child
	(1)	(2)	(3)	(4)	(5)
Impact on mothers' labor supply					
Total monthly hours worked	-0.004	-0.534**	-0.410	0.118	0.046
	[0.161]	[0.245]	[0.281]	[0.482]	[0.353]
Other labor supply measures					
Prob (Employed)	0.001	-0.003*	-0.004	-0.002	0.001
	[0.001]	[0.002]	[0.002]	[0.004]	[0.002]
Prob (Full-time employed)	-0.001	-0.001	-0.001	0.003	0.001
	[0.001]	[0.001]	[0.001]	[0.003]	[0.004]
F-stat	30.01	19.25	14.82	5.75	6.27
Obs. Mothers	25450	20540	11378	4983	12289
Obs. Grandmothers	20987	17519	10145	4583	10878

Notes: Table 2 shows the coefficient estimates of grandmothers' total monthly hours worked on the mother's labor supply by the age of the youngest child (Fuzzy RD estimates). An indicator for the grandmother being born since January 1950 serves as the instrument for grandmother's total monthly hours worked. Columns 1 - 3 show the results for families with childcare need based on the age of the youngest child (0-3, 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 ages 60 and 64. All specifications use local linear regression with a bandwidth of 8 months including controls and sector fixed effects. Robust standard errors clustered at grandmothers' level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table 3: Mechanisms I: Heterogeneous Effects

	Grandmother's partner is		Grandmother's residence municipality is		Number of maternal grandchildren aged 4-7	
Subgroups	unhealthy	healthy	different	same	only one	more or other age
	(1)	(2)	(3)	(4)	(5)	(6)
Impact on mothers' labor supply						
Total monthly hours worked	-0.051	-0.535**	-0.162	-0.820**	-0.829*	-0.321
	[0.499]	[0.257]	[0.251]	[0.413]	[0.459]	[0.259]
test p-value	0.386		0.170		0.335	
Other labor supply measures:						
Prob (Employed)	0.003	-0.004**	-0.000	-0.007**	-0.007*	-0.001
	[0.005]	[0.002]	[0.002]	[0.003]	[0.004]	[0.002]
test p-value	0.160		0.082		0.183	
Prob (Full-time employed)	-0.004	-0.001	-0.000	-0.002	-0.001	-0.002
	[0.003]	[0.001]	[0.001]	[0.002]	[0.002]	[0.001]
test p-value	0.438		0.281		0.818	
F-stat	3.34	17.45	15.48	9.63	7.90	13.75
Obs. Mothers	1761	21734	10385	13112	7452	16045

Notes: Table 3 shows heterogeneous effects of grandmothers' total monthly hours worked on mothers' labor supply outcomes (Fuzzy RD estimates). 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 mother and grandmother live in the same municipality. Columns 5 and 6 show the results by the number of maternal grandchildren aged 4-7. All outcomes are measured when the grandmothers are between ages 60 and 64. All specifications use local linear regression with a bandwidth of 8 months including controls and sector fixed effects. Robust standard errors clustered at grandmothers' 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.
Source: Authors' calculations from the CBS data.

Table 4: Mechanisms II: Grandfathers' vs. grandmothers' effect

Panel A:	Grandfathers'		Grandmothers'	
Family member	Daughter	Partner	Daughter	Partner
	(1)	(2)	(3)	(4)
Impact of grandparent's total monthly hours worked on family members' labor supply				
Total monthly hours worked	-0.066	0.039	-0.465**	-0.148
	[0.048]	[0.053]	[0.229]	[0.279]
Other labor supply measures:				
Prob (Employed)	-0.001**	0.000	-0.003*	-0.003
	[0.000]	[0.000]	[0.002]	[0.002]
Prob (Full-time employed)	0.000	0.000*	-0.001	0.001
	[0.000]	[0.000]	[0.001]	[0.002]
F-stat	184.31	172.10	21.22	20.40
Obs. Partners/ Daughters	23609	19840	23497	16224
Obs. Grandparents	19766	19753	19548	16182

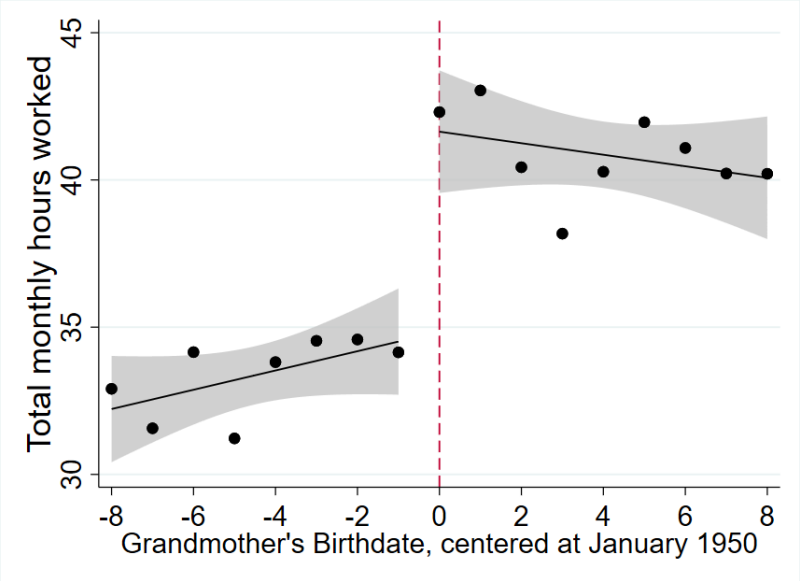
Panel B:	Grandmothers'			
Family member	Daughters	Sons-in-law	Sons	Daughters-in -law
	(1)	(2)	(3)	(4)
Impact of grandmother's total monthly hours worked on family members' labor supply				
Total monthly hours worked	-0.465**	0.648**	0.581	-0.038
	[0.229]	[0.310]	[0.407]	[0.290]
Other labor supply measures:				
Prob (Employed)	-0.003*	0.004**	0.002	-0.001
	[0.002]	[0.002]	[0.002]	[0.002]
Prob (Full-time employed)	-0.001	0.002	0.002	-0.000
	[0.001]	[0.002]	[0.003]	[0.001]
HH labor income	14.491		14.886	
	[11.685]		[14.154]	
F-stat	21.22	22.09	10.87	11.59
Obs. Family members	23497	21530	16773	16531
Obs. Grandmothers	19548	18128	14374	14192

Notes: Panel A of Table 4 shows the coefficient estimates of grandparents' total monthly hours worked on their partners' and adult daughters' (mothers') labor supply (Fuzzy RD estimates). Columns 1 - 2 refer to the grandfather, Columns 3 - 4 refer to the grandmother. Panel B of Table 4 shows the coefficient estimates of grandmothers' total monthly hours worked on other family members' labor supply, namely adult sons, daughters-in-law, and sons-in-law of the grandmothers (Fuzzy RD estimates). In both panels, 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 age 60 and 64. Household income (Panel B) 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. Both panels consider family members with a youngest (grand)child aged 4-12 when the grandparent is aged 60-64. All outcomes are measured when the grandmothers are between ages 60 and 64. All specifications use local linear regression with a bandwidth of 8 months, including controls and sector fixed effects. All income measures are CPI adjusted for the year 2015. Robust standard errors clustered at grandparent's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

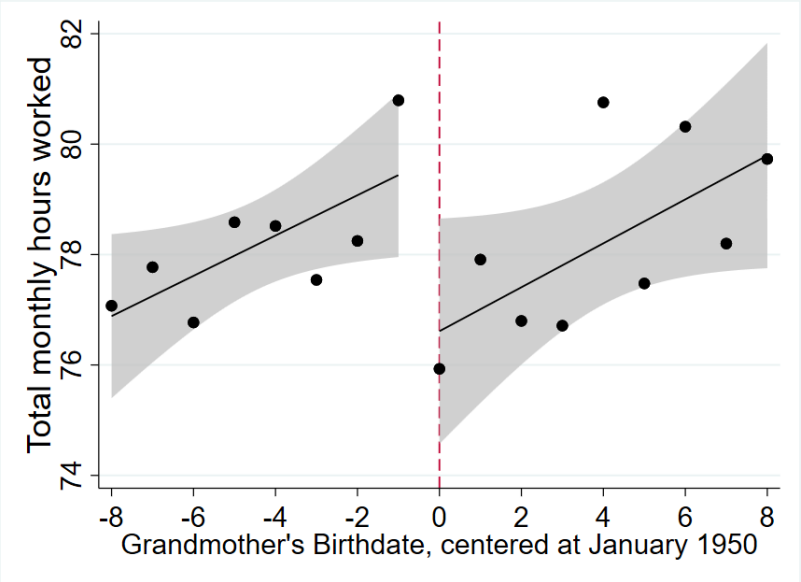
Table 5: Effects on children's educational performance (reduced-form)

RD estimates	Cito score (1)	Number of correct answers: Verbal (2)	Math (3)	Overall (4)	High track (5)	Obs. (Children)
Panel A: All youngest children						
Age between **4 - 12**	0.061	0.042	0.065*	0.072	0.007	8436
	[0.045]	[0.039]	[0.039]	[0.044]	[0.016]	
Panel B: By age groups and gender						
Age between **4 - 7**	0.171***	0.134***	0.198***	0.182***	0.032	5500
	[0.055]	[0.052]	[0.052]	[0.054]	[0.021]	
Girls between 4 - 7	0.200***	0.099	0.272***	0.206***	0.045	2785
	[0.076]	[0.069]	[0.073]	[0.074]	[0.029]	
Boys between 4 - 7	0.134	0.170**	0.113	0.153*	0.014	2715
	[0.082]	[0.078]	[0.075]	[0.080]	[0.031]	
p-value	0.103	0.030	0.135	0.055	0.658	
Age between **8 - 10**	0.075	0.066	0.050	0.082	0.012	5585
	[0.057]	[0.048]	[0.049]	[0.056]	[0.020]	
Girls between 8 - 10	0.111	0.074	0.120*	0.122	0.044	2847
	[0.078]	[0.066]	[0.069]	[0.077]	[0.028]	
Boys between 8 - 10	0.035	0.040	-0.008	0.038	-0.023	2738
	[0.082]	[0.069]	[0.069]	[0.080]	[0.029]	
p-value	0.6713	0.5680	0.9138	0.6377	0.4184	
Age between **11 - 12**	-0.131*	-0.136**	-0.193***	-0.146*	-0.037	2868
	[0.079]	[0.065]	[0.067]	[0.076]	[0.024]	
Girls between 11 - 12	-0.075	-0.083	-0.103	-0.083	-0.012	1459
	[0.109]	[0.087]	[0.095]	[0.104]	[0.035]	
Boys between 11 - 12	-0.225*	-0.277***	-0.252***	-0.247**	-0.070**	1409
	[0.116]	[0.095]	[0.098]	[0.111]	[0.036]	
p-value	0.0528	0.0034	0.0098	0.0258	0.0482	

Notes: Table 5 shows reduced-form impacts on education outcomes of children who are aged 4 - 12 when their grandmothers were between ages 60 and 64. Panel A shows results for all youngest children of a family aged 4 - 12. Panel B presents results separately for the youngest children aged 4 - 7, 8 - 10, and 11 - 12 in total and by gender. Column 1 shows the impact on the overall Cito score. Columns 2 - 4 report effects on the number of correct answers in the verbal part, mathematical part, and in the overall test, respectively. Columns 1 - 4 are based on standardized outcomes and thus measure effects in percent of the standard deviation. Column 5 shows the impact on the probability of obtaining a secondary school recommendation for the highest (academic) track (VWO). All specifications use local linear regression with a bandwidth of 8 months and include controls. Robust standard errors (clustered by the primary school the child attends) are in parenthesis.. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.



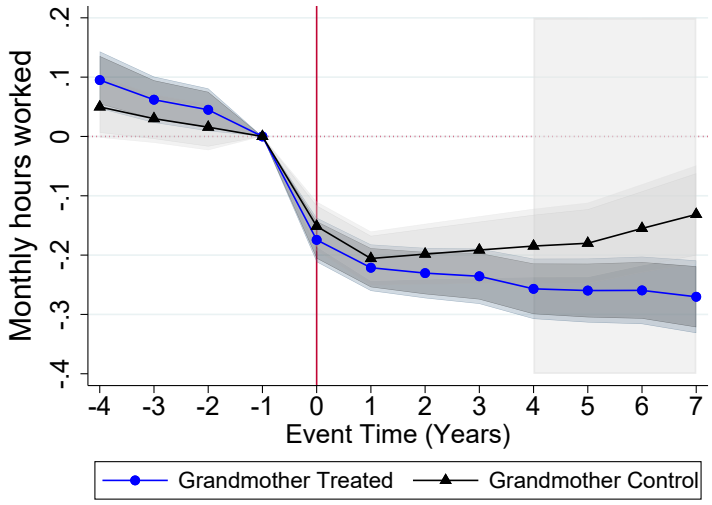
(a) Total hours worked by grandmothers (first-stage)



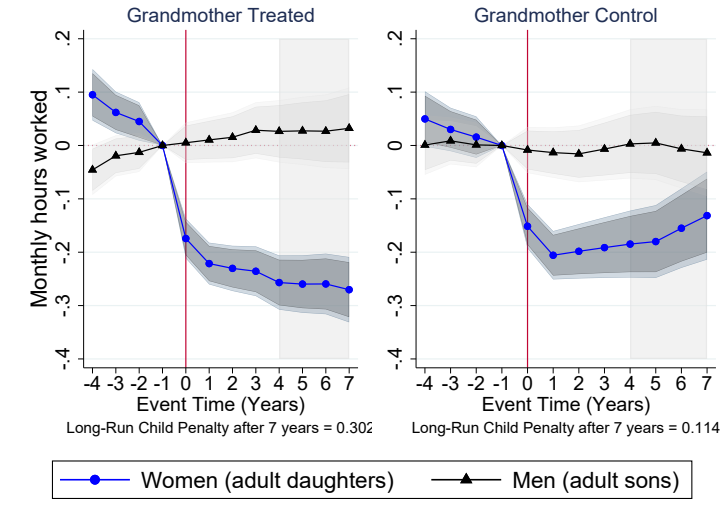
(b) Total hours worked by mothers (reduced-form)

Notes: Panel (a) of Figure 1 shows the scatter bin plot of grandmother's 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 mother's labor supply as a function of distance to the cutoff, which is grandmothers' birth month being January 1950. The solid lines are the linear fitted lines. The shaded areas indicate the 95 percent confidence interval.

Figure 1: Grandmothers' and mothers' labor supply relative to the cutoff



(a) Dynamic treatment effects on monthly working hours



(b) Relative child penalty by treatment status

Notes: Panel a of Figure 2 shows the evolution of mothers' total monthly hours worked from four years before and to seven years after they gave birth to their first child. It compares the monthly working hours of treated (blue dots) mothers, whose (grand)mothers are born since January 1950 and thus treated by the pension reform, to those of control (black triangles) mothers (with untreated (grand)mothers). Event time 0 marks the birth of the first child. Panel b of Figure 2 depicts the child penalty in total monthly working hours (including zeros) by treatment status. The left figure presents the child penalty for treated grandmothers and the right figure for control grandmothers. Blue dots document women's and black triangles indicate men's monthly working hours, the difference between which represents the child penalty.The long-run relative child penalty after 7 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. To capture the extensive and intensive margin of labor supply simultaneously, total hours worked include zeros for women not working. Different tones of shaded areas indicate the 90 and 95 percent confidence intervals.

Figure 2: Dynamic treatment effects and child penalty

For Online Publication

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

Katja M. Kaufmann
Johannes Gutenberg University Mainz, briq, CESifo, HCEO and IZA

Yasemin Özdemir
Johannes Gutenberg University Mainz

Han Ye
University of Mannheim, IZA, ZEW

List of Tables

List of Figures

A Additional Details on Institution

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 aged 65 and above a flat-rate pension.[1] AOW benefits depend on years of residence and are not related to earnings and contributions paid before retirement. They are financed by income taxes and are linked to the minimum wage (OECD, 2019b).

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

[1]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.

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. During our sample period, the earliest possible ages to claim the occupational pensions are between ages 55 and 60, depending on their occupational group.

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. These schemes were initially designed to reduce youth unemployment. 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] 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 transitioning from the generous and actuarially unfair pay-as-you-go 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, 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 even under the new more actuarially fair ER schemes. Approximately 80% of all workers retired at the age of 62 or younger before 2006 (Statistics Netherlands (2009)).

The general plan to eliminate early retirement tax advantages was announced in 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 replace the pre-pension with a new pension scheme called the

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

ABP Flexible Pension Scheme in July 2005. More generally, the reform bill no. 29760 includes a clause to adjust fiscal policy VUT and prepension (Wet voor aanpassing fiscal behandeling VUT and pre-pension) 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.

The reform bill no. 29760 includes a clause to adjust fiscal policy VUT and prepension (Wet voor aanpassing fiscal behandeling VUT and pre-pension) and is sometimes referred to as the 56-plus scheme (de 56-plusregeling). People, who were 55 years or older before January 1, 2005 are not affected by the reform.

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 Child care and primary education in the Netherlands

The amount of free public childcare increases as the child ages. In the Netherlands, children aged 0 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

[3] In the Netherlands, mothers are entitled to fully-funded maternity leave 6 weeks before and 10 weeks after childbirth. Before 2019, 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. 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).

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 A1 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 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 describe the different data sources used in the analysis. All datasets used are provided by Statistics Netherlands (CBS). Documentation for each of the files shortly described below can be found at the link provided below. Please note that these are only available in Dutch.

gpapersoontab

This file is updated yearly and provides information on the demographic background of the Dutch population that appears in the Personal Records Database (BRP) since 1 October 1994. Information includes gender, date and country of birth, migration background.

For details see: Official documentation of gpapersoontab

kindoudertab

This file links children to their legal parents and includes all registered individuals if the parents could be identified. The file refers to legal parents and thus includes biological as well as adoptive parents.

For details see: Official documentation of kindoudertab

gbaoverlijdentab

This file contains the date of death of all persons who have died since 1 October 1994 and who

were registered in the Personal Records Database (BRP) on the date of death. The date of death of persons who are not residents but were once residents of the Netherlands since 1 October 1994 and whose death information has been registered in the Register Non-Residents (RNI) are also included.

For details see: Official documentation of gbaoverlijdentab

gbahuishoudenbus

This file provides information on all individuals who appear in the Basic Registration of Persons (BRP) from 1 October 1994. Information includes their position in the household and details about the family to which they belong/ belonged, such as family composition, number of children, marital status. The data is provided in the form of spells indicating each household's start and end date and individuals belonging to the family.

For details see: Official documentation of gbahuishoudenbus

gbaadresobjectbus

This file contains encrypted information on the addresses of persons who are or have been registered since 1 January 1995.

For details see: Official documentation of gbaadresobjectbus

vslgwbtab

This file includes all objects from the Key Register of Addresses and Buildings (BAG) and all objects that were in the SSB before 1-1-2012 (historical objects). Information on the municipality in which the object is placed is provided on an annual basis.

For details see: Official documentation of vslgwbtab

secmbus

This file contains monthly data on the socioeconomic category (SECM) of individuals registered since 1 January 1995. Information on separate sources of income, on which the SECM definition is based, is indicated for each relevant period. Also, the file indicates whether a person was registered as a student in the given month.

For details see: Official documentation of secmbus

integraal persoonlijk inkomen

This file contains the annual income of individuals belonging to the population of the Netherlands on 31st December of the study year. The main data supplier are the tax authorities. All individuals that appear in the basic tax register are included. Note that this file was replaced from 2011 on by INPATAB due to the revision of the income statistics 2017.

For details see: Official documentation of integraal persoonlijk inkomen

CITOtab

This file contains information on the performance on a high-stake test taken at the end of primary school, the Cito Primary School Final Test, used to assign children to different tracks in secondary

school. Until the school year of 2013/14, only the Cito Primary School Final Test was provided. Since 2014/15 schools could choose between three compulsory final tests, of which the test designed by Cito on behalf of the Board for Tests and Exams is included in our data. The data includes information on the test results including seperate components (such as points on the verbal and mathematical part).
For details see: Official documentation of CITOtab

kinderopvang

In each year, we observe information on the type and number of childcare hours in a reporting year for each child that used childcare under the Childcare Act.
For details see: Official documentation of kinderopvang

LISS panel

The LISS panel is an online household panel. The panel consists of some 5000 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 child care 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 the following four questions based on which we calculate child care statistics.

"The following questions are about your living-at-home children born from 2004 onwards. This concerns your [CHILD/CHILDREN] [NAMES AND BIRTH YEARS CHILDREN BORN FROM 2004 ONWARDS]. For [THIS/THESE] [CHILD/CHILDREN], do you make regular use, that is at least once a week, of the following childcare options? If so, of which? More than one answer possible." Parents could indicate for each of the following options whether they make regular use of it or not:

a toddler playgroup, nursery school

b child daycare center ('kindergarten' – also half-day childcare)

c pre-school childcare

d after-school childcare

e host parent where the child goes to (arranged through a host parent agency)

f paid childsitter, where the child goes to

g paid childsitter, that comes to your home

h unpaid childsitter (for instance family/friends/neighbors/acquaintances)

i other childcare

j no, no childcare

If parents indicated to make use of an "unpaid childsitter" (option h), they were asked in addition: "Who is generally this unpaid childsitter? If more than one option applies, please indicate the person that you call on most often." The answer could be one among the following options:

i own parent (child's grandparent)

ii partner's parent (child's grandparent)

iii neighbour

iv friend/acquaintance

v brother or sister (own or partner's)

vi other own (step/adoptive/foster) child

vii someone else

Analogously, information on child care arrangement for children aged 4-12, who potentially attend primary or another type of school but do not attend secondary school yet, were collected.
"This concerns your living-at-home children born from 1995 onwards that do not attend secondary school: your [CHILD/CHILDREN] [NAMES CHILDREN BORN FROM 1995 ONWARDS THAT DO NOT ATTEND SECONDARY SCHOOL]. Do you make regular use, that is at least once a week, of the following childcare options for your children? If so, of which? More than one answer possible." Again, parents could indicate for each of the following options whether they made regular use of it or not:

a pre-school childcare

b after-school childcare

c between-school childcare

d host parent where the child goes to (arranged through a host parent agency)

e paid childsitter, where the child goes to

f paid childsitter, that comes to your home

g unpaid childcare (for instance family/friends/neighbors/acquaintances)

h other childcare

i no childcare

For those making regular use of a "unpaid childcare" (option g), a second question on the unpaid care giver followed.
"Who is generally this unpaid childsitter? If more than one option applies, please indicate the person that you call on most often." The answer could be one among the following options:

i own parent (child's grandparent)

ii partner's parent (child's grandparent)

iii neighbour

iv friend/acquaintance

v brother or sister (own or partner's)

vi other own (step/adoptive/foster) child

vii someone else

Additionally, we make use of two questions regarding parents' work and child care arrangements. In case parents report to perform paid work ("even if is it just for one or several hours per week or for a brief period"), they are asked:
"Are you currently working less hours in order to care for your children? If you have children and you are on parental leave, then consider the hours that you have additionally started working less (so in addition to the parental leave)"

For those who indicated to currently work less, a follow-up question asked about the reduction in hours:
"How many hours per week did you start working less on account of the care for your children? Do not include the hours that you have possibly taken as your parental leave." The answer could be any number of hours ranging from zero to 40.

B.2 Sample selection

In Table A12 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 eight 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 an 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 children have to take a standardized test, the *Cito test*, to determine the secondary school track they are 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 administored by Cito. It is important to note that the schools, not the parents or children, select the type of test. The data includes information on the Cito test and only for children from 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 (when the grandmother is 60-64) to mothers of our baseline sample that can be matched to their Cito test results. Table A14 compares characterictics of all children aged 4-12 (see Columns (1) and (2)), to characteristics of the youngest children aged 4-12 (Columns (3) and (4)), and to the characteristics of the children in the Cito-sample (Columns (5) and (6)). Compared to all children of our baseline mothers, children in the Cito-sample are less likely to be firstborn and somewhat more likely to have parents who are married (when the grandmother was aged 50-53). Compared to all youngest children aged 4-12, these differences disappear as we focus on the youngest child in a family. In terms of other characteristics, the Cito-sample is comparable to all children and even more to all youngest children aged 4 - 12 of the baseline mothers. Comparing the average year of birth across the three samples,

children in the Cito-sample are somewhat older than all children aged 4-12 due to the fact that since 2014/15 only part of the schools still rely on the Cito test which we have data on. Table A16 tests whether the Cito-sample restrictions and matching rates differ by treatment status. We show that among all children aged 4-12 (when the grandmothers are 60-64), 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 show that the restricted data availability of test scores does not constrain our analysis.

Table A15 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. We show results using a linear and quadratic specification with a bandwidth of 8 months around the cutoff. All covariants are smooth, so that treated and non-treated grandchildren are comparable in terms of children's covariates, such as, among other things, birth year and month, gender, birth order, and family situation.

C Placebo Tests and Robustness Checks

Two placebo exercises further support the credibility of our estimates. First, we show the reduced-form impact of having grandmothers born since 1950 using a sample of mothers with deceased grandmothers. We do not expect the pension reform to affect women with deceased mothers (grandmothers) born around the cutoff. Table A7 shows the estimated reform impacts on the labor supply of adult daughters (mothers) whose mothers (grandmothers) died before age 50. As expected, none of the estimates are significant and the coefficient size is an order of magnitude smaller (compare to Table A4). 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 show the validity of our results by using placebo cutoffs up to 10-months prior and 10-months post the actual cutoff, at a bi-monthly frequency. Figure A5 shows that the F-statistics is only above 10 at the actual cutoff. Table A8 shows the reduced-form estimates and Figure A6 plots the coefficients with 95% confidence intervals. We find no significant effects of grandmothers' labor supply on maternal labor supply at these placebo cutoffs.

We also test the robustness of the estimation results by varying the choice of polynomial orders and bandwidth. Table A10 shows the estimates in response to a one-hour increase in grandmothers' monthly working hours for a linear and a quadratic specification. The Aikake Information Criterion (AIC) and Bayesian information criterion (BIC) and AICc (AIC with a correction for small sample sizes) are also reported. According to the AIC, BIC and AICc criteria, in terms of total monthly hours worked, the linear specification fits the best.

Table A11 displays the estimated impacts and bandwidth for all relevant outcomes using the mean square error optimal bandwidths generated by the Calonico et al. (2017) and Calonico et al. (2018) procedure. We show the estimates and optimal bandwidths for both linear and quadratic specifications. Notice that the estimated impact on the total monthly hours worked is around -0.65 hours using a bandwidth of 7.25, which is higher than the baseline result. Therefore, if anything, our baseline result is rather conservative. We show results for a bandwidth of 8 to avoid switching between different bandwidths (and thus samples) for different outcomes and for different subgroups.

D Estimation of Dynamic Effects on Mothers' Labor Supply and 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 term 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 we 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 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:

$$P_t^d = \frac{\hat{\alpha}_t^{md} - \hat{\alpha}_t^{wd}}{E[\tilde{Y}_{ist}^{wd}|t]}$$

E 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, in the following, we calculate the MVPF for the policy of incentivizing later retirement, taking into account the grandmothers' and mothers' labor supply responses.

The mechanical net cost of incentivizing later retirement is the change in pension generosity. The 2006 reform reduced pension replacement rates between age 60 and 64 from 70% to 64% (Lindeboom and Montizaan (2020) Table A.1). For a typical woman with average labor earnings of 727 euro and an average pension claim duration of 17 years,[5] we calculate that the government saves about 8900 euro per person.

The behavioral costs consist of the direct impact on grandmothers and the indirect spillover effects on mothers. First, we calculate the net cost of the pension reform from the direct impact on grandmothers. The reduced form estimates show that grandmothers earn 106 euros more per month between age 60 and 64 (Table A2). Average monthly labor earnings around the cutoff are 727 euros. Using the Dutch Income Tax Calculator, we find that the government increases the amount of taxes raised by 484 euro per person during those four years.

Second, we include the spillover impacts on mothers' labor supply. The reduced form estimates show that mothers whose youngest child is between 4 and 12 years old earn 58 euros less (Table A3). Their average monthly labor earnings around the cutoff are 2064 euros. This suggests that the government loses 264 euro tax income per person during those four years because mothers work less.

[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).

We could also include the spillover impacts on grandchildren. However, we find that children treated when they are young (ages 4-7) perform better, and children treated when older (ages 11-12) perform worse. The reduced-form estimates have similar magnitudes but opposite signs. Therefore, we do not include the spillovers on grandchildren on tax revenue. Based on the early childhood development literature, we know that return to investment on younger children is higher and can have a long-term impact on their lifetime earnings. Therefore, if 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.

If we assume the government only cares about income tax revenue, the net government costs would be -224 euro. Although the increases in grandmothers' labor supply have negative impacts on mothers' labor supply, the policy of incentivizing later retirement still pays for itself if we only take into account short-run effects of (adult) daughters' labor supply.

However, Section 5.2 shows that the reform also has long-run effects on mothers' lifetime income. Assuming the government only cares about income tax revenue and the impact on maternal labor supply lasts for up to eight years, then 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 grandmother. Of course, the net government costs may remain negative if we include the tax revenue gain from grandfathers working longer and the spillover effects on sons-in-laws' labor supply.

To conclude, our MVPF exercise highlights the importance of taking into account the various types of 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" of the former, such as –say- complementing an early retirement reform with better access to high-quality childcare.

F Appendix Tables and Figures

Table A1: Summary statistics

Variables	All		Full sample		RD sample	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Grandmothers' characteristics						
Birth cohort	1949.51	[1.121]	1949.48	[1.117]	1949.53	[0.499]
Age	62.516	[0.292]	62.878	[1.110]	62.957	[1.123]
Number of adult children	2.510	[1.016]	2.475	[0.982]	2.469	[0.983]
Number of adult daughters	1.729	[0.800]	1.712	[0.785]	1.702	[0.780]
Mothers' characteristics						
Age	35.254	[4.521]	37.884	[2.938]	37.924	[2.952]
Age at first child birth	29.691	[4.200]	28.337	[3.477]	28.378	[3.518]
Age at first employment	23.589	[3.916]	24.855	[3.888]	24.726	[3.720]
Married	0.466	[0.470]	0.661	[0.460]	0.657	[0.461]
Age gap to partner	2.663	[4.395]	2.809	[3.995]	2.786	[3.987]
Number of children	1.246	[1.087]	2.015	[0.763]	2.017	[0.767]
Outcomes: grandmothers' labor supply						
Monthly hours worked	44.023	[48.245]	37.230	[48.712]	37.315	[48.628]
Prob (Employed)	0.474	[0.419]	0.415	[0.446]	0.417	[0.447]
Prob (Full-time employed)	0.067	[0.212]	0.055	[0.204]	0.053	[0.200]
Monthly labor income	803.52	[1019.71]	638.99	[947.15]	637.76	[935.73]
Monthly gross income	1635.15	[1552.12]	1419.83	[1344.81]	1395.29	[1315.91]
Monthly HH labor earnings	1725.80	[1966.94]	1307.63	[1739.54]	1287.62	[1713.83]
Age at exiting employment	61.103	[4.304]	60.812	[4.388]	60.786	[4.442]
Age at claiming pension	63.039	[3.171]	62.954	[3.230]	62.912	[3.189]
Outcomes: mothers' labor supply						
Monthly hours worked	97.255	[51.482]	78.498	[47.458]	78.176	[47.388]
Prob (Employed)	0.816	[0.334]	0.784	[0.377]	0.782	[0.378]
Prob (Full-time employed)	0.232	[0.355]	0.063	[0.209]	0.061	[0.205]
Monthly labor income	1844.71	[1219.59]	1531.13	[1188.53]	1533.70	[1193.38]
Monthly HH labor income	4280.80	[2477.81]	4525.05	[2677.43]	4546.31	[2697.09]
Obs. Mothers	147858		66252		23497	
Obs. Grandmothers	106036		55055		19548	

Notes: Table A1 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 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 to (extended) families with grandmothers with the youngest grandchild aged 4-12 when the grandmother is between 60 and 64. Columns 5 and 6 are the RD sample, which is the sample of Columns 3 and 4 restricted to families with grandmothers born within a bandwidth of 8 months before and after January 1950. Grandmothers and mothers' labor supply is measured when the grandmother is between age 60 and 64. All income measures are CPI adjusted for the year 2015.
Source: Authors' calculations from the CBS data.

Table A2: Impacts on covariates (reduced-form)

	RD estimates (1)	RD estimates (2)	Mean at cutoff
Grandmothers' characteristics			
Age	0.044	0.027	62.934
	[0.031]	[0.051]	[1.144]
Number of adult children	0.055*	0.022	2.451
	[0.032]	[0.051]	[0.892]
Number of adult daughters	0.024	-0.003	1.712
	[0.027]	[0.043]	[0.761]
Prob (Employed)	-0.006	0.024	0.793
	[0.011]	[0.017]	[0.338]
Prob (Married)	-0.008	0.012	0.835
	[0.011]	[0.018]	[0.355]
Prob (Cohabit)	0.003	0.005	0.036
	[0.006]	[0.009]	[0.166]
Prob (Partner disabled)	0.016**	0.016	0.062
	[0.008]	[0.012]	[0.235]
Birthcohort of partner	0.093	-0.169	1947.50
	[0.129]	[0.214]	[3.849]
Mothers' characteristics			
Age	0.124	0.091	37.867
	[0.081]	[0.132]	[2.899]
Native	0.007	0.019*	0.953
	[0.006]	[0.010]	[0.211]
Birth cohort	-0.074	-0.065	1974.45
	[0.091]	[0.147]	[3.211]
Prob (Married)	0.006	-0.000	0.369
	[0.012]	[0.020]	[0.434]
Prob (Employed)	-0.005	-0.008	0.772
	[0.009]	[0.015]	[0.333]
Live in same municipality as GM	0.003	0.037*	0.552
	[0.013]	[0.021]	[0.461]
Age at first child birth	0.080	0.123	28.268
	[0.010]	[0.160]	[3.609]
Age of youngest child	-0.042	0.084	2.059
	[0.071]	[0.111]	[2.022]
Age of oldest child	-0.083	-0.086	3.790
	[0.117]	[0.190]	[3.164]
Number of children	0.002	-0.025	0.842
	[0.027]	[0.044]	[0.969]
Age of first employment	0.060	0.032	24.891
	[0.105]	[0.173]	[3.809]
Obs. Mothers	23497	23497	4018
Obs. Grandmothers	19548	19548	
Polynomial	linear	quadratic	

Notes: 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 times when the grandmother was aged 50 to 53. Prob(employed) refers to formal employment only. All specifications use local linear regression with a bandwidth of 8 months. Robust standard errors clustered at grandmother's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table A3: Impacts on grandmothers' labor supply (reduced-form)

	RD estimates			Means at
	(1)	(2)	(3)	cutoff
First-Stage RD estimates				
Total monthly hours worked	6.801***	6.831***	6.174***	34.418
	[1.493]	[1.392]	[1.340]	[47.608]
Other labor supply measures:				
Prob (Employed)	0.063***	0.063***	0.054***	0.387
	[0.014]	[0.013]	[0.012]	[0.438]
Prob (Full-time employed)	0.007	0.006	0.007	0.054
	[0.006]	[0.006]	[0.006]	[0.202]
Other income measures:				
Monthly labor earnings	121.767***	123.340***	106.286***	573.065
	[27.798]	[26.076]	[24.961]	[887.669]
Monthly HH labor earnings	126.754**	142.272***	122.952**	1211.88
	[50.983]	[50.071]	[49.318]	[1647.46]
Monthly gross income	64.678	66.114*	45.148	1361.66
	[40.259]	[36.270]	[34.101]	[1304.62]
Monthly gross HH income	15.901	49.140	28.525	4082.91
	[66.725]	[64.097]	[62.482]	[2144.97]
Obs. Mothers	23497	23497	23497	4005
Obs. Grandmothers	19548	19548	19548	
Controls	NO	YES	YES	
Sector FE	NO	NO	YES	

Notes: Table A3 shows the first-stage 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. All specifications use local linear regression with a bandwidth of 8 months. Robust standard errors clustered at grandmother's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: Authors' calculations from the CBS data.

Table A4: Impacts on mothers' labor supply (reduced-form)

	RD estimates			
	(1)	(2)	(3)	Means at cutoff
Mothers' labor supply outcomes				
Total monthly hours worked	-3.193**	-2.818**	-2.871**	78.876
	[1.300]	[1.259]	[1.259]	[47.744]
Other labor supply measures:				
Prob (Employed)	-0.021**	-0.019*	-0.020*	0.785
	[0.010]	[0.010]	[0.010]	[0.378]
Prob (Full-time employed)	-0.009	-0.009	-0.009	0.066
	[0.006]	[0.006]	[0.006]	[0.214]
Obs. Mothers	23497	23497	23497	4018
Obs. Grandmothers	19548	19548	19548	
Controls	NO	YES	YES	
Sect FE	NO	NO	YES	

Notes: Table A4 shows the reduced form impacts of grandmothers being born since January 1950 on labor market outcomes on their adult daughters (mothers). Columns 1, 2, and 3 show the results without controls, with controls, and with both controls and sector fixed effects, respectively. All specifications use local linear regression with a bandwidth of 8 months. Robust standard errors clustered at grandmother's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table A5: First-stage and reduced-form effects- by age of the youngest child

	Childcare need						
	more					none/ less	
	age of youngest child						
	0-3	4-7	8-10	11-12	8-12	13-18	no child
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A:	**First-stage: reform effects on Grandmothers**						
Total monthly hours worked	7.299***	6.386***	8.823***	6.377***	7.233***	6.429**	4.667**
	[1.332]	[1.455]	[2.003]	[2.372]	[1.879]	[2.682]	[1.863]
F-stat	30.01	19.25	19.40	7.23	14.82	5.75	6.27
Panel B:	**Impact on mothers' labor supply (reduced form)**						
Total monthly hours worked	-0.026	-3.412**	-1.477	-2.497	-2.966	0.757	0.213
	[1.175]	[1.337]	[1.962]	[2.474]	[1.876]	[3.088]	[1.649]
Other labor supply measures							
Prob(Employed)	0.005	-0.022**	-0.017	-0.031	-0.025	-0.013	0.004
	[0.009]	[0.011]	[0.016]	[0.020]	[0.015]	[0.024]	[0.009]
Prob(Full-time employed)	-0.009	-0.008	0.006	-0.009	-0.006	0.019	0.006
	[0.007]	[0.006]	[0.009]	[0.011]	[0.008]	[0.0151]	[0.0162]
Obs. Mothers	25450	20540	10343	6999	11378	4984	12289
Obs. Grandmothers	20987	17519	9348	6476	10145	4584	10878

Notes: Panel A of Table A5 shows the first-stage reduced form impacts on grandmothers' total monthly hours worked by the age of the youngest (grand)child. Panel B of Table A5 shows the reduced form impacts of grandmothers being born since January 1950 on labor market outcomes on their adult daughters (mothers) by the age of the youngest child. Columns 1 - 5 show the results for families with childcare need based on the age of the youngest child (0-3, 4-7, 8-10, 11-12, 8-12, respectively). Columns 6 and 7 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 ages 60 and 64. All specifications use local linear regression with a bandwidth of 8 months including controls and sector fixed effects. Robust standard errors clustered at grandmothers' level are in parentheses.
Source: Authors' calculations from the CBS data.

Table A6: Impacts on grandfathers' labor supply (reduced-form)

	RD estimates			Means at
	(1)	(2)	(3)	cutoff
First-Stage RD estimates				
Total monthly hours worked	28.058***	28.276***	26.233***	58.951
	[2.052]	[2.025]	[1.932]	[64.226]
Other labor supply measures:				
Prob (Employed)	0.151***	0.153***	0.137***	0.430
	[0.013]	[0.013]	[0.012]	[0.422]
Prob (Full-time employed)	0.144***	0.145***	0.137***	0.267
	[0.012]	[0.012]	[0.012]	[0.373]
Other income measures:				
Hourly wage rate	4.311***	4.335***	3.967***	9.483
	[0.371]	[0.367]	[0.352]	[11.104]
Monthly labor earnings	728.209***	731.704***	683.573***	1303.44
	[54.897]	[54.178]	[52.019]	[1666.96]
Monthly gross income	388.721***	382.745***	355.081***	3826.74
	[60.231]	[59.366]	[55.937]	[1831.90]
Obs. Grandfathers	23609	23609	23609	4026
Controls	NO	YES	YES	
Sector FE	NO	NO	YES	

Notes: Table A6 shows the first-stage reduced form impacts on grandfathers' labor supply. Columns 1, 2, and 3 show the results without controls, with controls, and with both controls and sector fixed effects, respectively. All specifications use local linear regression with a bandwidth of 8 months. Robust standard errors clustered at grandfather's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table A7: Placebo test - impacts on labor supply of mothers with deceased grandmothers (reduced-form)

	RD estimates		
	(1)	(2)	Mean at cutoff
Mothers' labor supply outcomes			
Total monthly hours worked	0.415	0.161	68.467
	[4.923]	[4.827]	[49.236]
Other labor supply measures:			
Prob (Employed)	-0.001	-0.005	0.699
	[0.042]	[0.041]	[0.421]
Prob (Full-time employed)	0.006	0.005	0.055
	[0.019]	[0.019]	[0.193]
Obs. Mothers	1858	1858	312
Controls	NO	YES	

Notes: Table A7 shows the second-stage reduced form impacts on adult daughters (mothers) whose mothers (grandmothers) deceased before age 50. Columns 1 and 2 show results without and including controls, respectively. All specifications use local linear regression with a bandwidth of 8 months. Robust standard errors clustered at grandmother's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table A8: Placebo tests using other birth dates as cutoffs (reduced-form)

Distance to actual cutoff	-10	-8	-6	-4	-2	0	2	4	6	8	10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Mothers' labor supply outcomes											
Total monthly hours worked	0.774	-0.357	0.305	0.088	1.139	-2.871**	-0.84	1.670	1.581	-0.156	-0.594
	[1.243]	[1.235]	[1.234]	[1.245]	[1.253]	[1.259]	[1.247]	[1.255]	[1.266]	[1.285]	[1.284]
Other labor supply measures:											
Prob (Employed)	0.007	0.001	0.005	-0.003	0.001	-0.020*	0.002	0.021	0.005	-0.005	-0.002
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Prob (Full-time employed)	0.004	0.002	0.007	-0.001	0.006	-0.009	-0.011	-0.000	0.009	0.006	-0.003
	[0.005]	[0.006]	[0.005]	[0.006]	[0.006]	[0.006]	[0.005]	[0.005]	[0.005]	[0.006]	[0.006]
Obs. Mothers	23853	23643	23753	23861	23730	23497	23342	23196	23333	23392	23331

Notes: In Table A8 we test the validity of our results by using placebo cutoffs of 10 months back and 10 months ahead at a bi-monthly frequency. The second-stage reduced form impacts of grandmothers being born since January 1950 are reported in the table. All specifications use local linear regression with a bandwidth of 8 months including controls and sector fixed effect. Robust standard errors clustered at grandmother's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1..
Source: Authors' calculations from the CBS data.

Table A9: Impacts on mother's fertility outcomes (reduced-form)

	RD estimates		
	(1)	(2)	Mean at cutoff
Mothers' fertility outcomes			
Prob (Ever child)	0.002	0.004	0.768
	[0.006]	[0.009]	[0.422]
Prob (At least 2 children)	-0.007	-0.007	0.611
	[0.007]	[0.011]	[0.488]
Total number of children	-0.003	-0.002	1.610
	[0.016]	[0.026]	[1.147]
Age at first birth	0.018	-0.052	29.309
	[0.072]	[0.117]	[4.430]
Age at last birth	-0.052	-0.104	32.685
	[0.063]	[0.103]	[4.059]
Average age gap of children	-0.023	0.020	3.092
	[0.030]	[0.050]	[1.763]
Average age gap after GM age 55	-0.060	0.000	3.264
	[0.040]	[0.065]	[2.045]
Prob (First child after GM age 55)	-0.001	-0.004	0.470
	[0.007]	[0.011]	[0.499]
Obs. Mothers	100369	100369	16923
Obs. Grandmothers	69628	69628	
Polynomial	linear	quadratic	

Notes: Table A9 tests the impact of grandmothers being born since January 1950 on the adult daughters' (mothers') fertility outcomes. All specifications use local linear regression with a bandwidth of 8 months. Robust standard errors clustered at grandmother's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table A10: Fuzzy RD estimates by polynomial orders

Outcomes	poly	estimates	AIC	BIC	AICc	Obs.
Impact of GM's total monthly hours worked on mothers' labor supply						
Total monthly hours worked	1	-0.465**	251525	251638	251526	23497
		[0.229]				
	2	-0.850*	259250	259379	259250	23497
		[0.473]				
Other labor supply measures:						
Prob (Employed)	1	-0.003*	23495	23608	23495	23497
		[0.002]				
	2	-0.004	25453	25582	25453	23497
		[0.003]				
Prob (Full-time employed)	1	-0.001	-5550	-5437	-5550	23497
		[0.001]				
	2	-0.002	-3550	-3421	-3550	23497
		[0.002]				

Notes: Table A10 shows the 2SLS estimates of grandmothers' total monthly hours worked on mothers' labor supply outcomes by polynomial orders. An indicator for grandmothers being born since January 1950 serves as the instrument for the grandmothers total monthly hours worked. We show local linear and quadratic regressions with a bandwidth of 8 months. Robust standard errors clustered at the grandmother's level are in parentheses. We also report the AIC, its sample equivalent (AICc), and BIC criteria for each regression. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table A11: Fuzzy RD estimates by CCT optimal bandwidth and polynomial orders

Outcomes	Poly	estimates	s.e.	optimal BW	Obs.
Impact of GM's total monthly hours worked on mothers' labor supply					
Total monthly hours worked	1	-0.645**	[0.285]	7.252	20711
	2	-0.822**	[0.404]	9.986	26221
Other labor supply measures:					
Prob (Employed)	1	-0.004*	[0.002]	7.698	20711
	2	-0.004*	[0.003]	10.942	29142
Prob (Full-time employed)	1	-0.002	[0.001]	8.780	23497
	2	-0.002	[0.002]	10.091	29142
			Average of optimal BW	9.1248	
			Average of optimal BW (linear)	7.9099	
			Average of optimal BW (quadratic)	10.3398	

Notes: Table A11 shows the 2SLS estimates of grandmothers' total monthly hours worked on mothers' labor supply outcomes using the mean square error optimal bandwidths generated by the Calonico et al. (2017) and Calonico et al. (2018) procedure (the CCT bandwidths). The optimal bandwidths are generated separately for each outcome. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: Authors' calculations from the CBS data.

Table A12: Impacts on sample selection

	RD estimates (1)	RD estimates (2)	Mean at cutoff
Restrictions by Grandmothers' characteristics			
Step 1: reform relevance			
Exit labor force before age 50	-0.000	0.001	0.385
	[0.009]	[0.015]	[0.487]
Obs. Mothers	72924	72924	12307
Obs. Grandmothers	54912	54912	
Step 2: alive during treatment period			
Dead before age 65	0.001	0.002	0.023
	[0.004]	[0.006]	[0.148]
Obs. Mothers	44903	44903	7564
Obs. Grandmothers	34085	34085	
Step 3: health status/ relevance for care responsibility			
Claim disability before age 55	0.001	-0.001	0.081
	[0.007]	[0.011]	[0.272]
Obs. Mothers	43810	43810	7394
Obs. Grandmothers	33253	33253	
Restrictions by Mothers' characteristics			
Step 4: Keep by relevance of child care			
Youngest child aged 4-12	-0.014	0.003	0.595
	[0.011]	[0.017]	[0.491]
Obs. Mothers	40160	40160	6799
Obs. Grandmothers	30447	30447	
Baseline Sample			
Obs. Mothers	23497		
Obs. Grandmothers	19548		
Polynomial	linear	quadratic	

Note: Table A12 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 8 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. Columns 1 and 2 show results based on local linear and quadratic regressions with a bandwidth of 8 months, respectively. Robust standard errors clustered at grandmother's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table A13: Impacts on sub-sample selection

	RD estimates		Mean at
	(1)	(2)	cutoff
Sub-sample selection criteria			
Restrictions by age of youngest child			
Youngest aged 0-3	-0.008	-0.029*	0.647
	[0.010]	[0.017]	[0.478]
Youngest aged 4-7	-0.002	0.003	0.525
	[0.011]	[0.018]	[0.499]
Youngest aged 8-12	-0.003	0.016	0.293
	[0.010]	[0.016]	[0.455]
Youngest aged 13-18	-0.006	0.000	0.130
	[0.007]	[0.012]	[0.336]
Obs. Mothers	39293	39293	6663
Obs. Grandmothers	29921	29921	
Polynomial	linear	quadratic	

Note: Table A13 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 A12.Columns 1 and 2 show results based on local linear and quadratic regressions with a bandwidth of 8 months, respectively. Robust standard errors clustered at grandmother's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table A14: Summary statistics of children

	Baseline Sample Children					
	All		Youngest		CITO - sample	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Birth month	6.549	[3.404]	6.488	[3.386]	6.608	[3.346]
Birth year	2004.27	[3.175]	2005.27	[3.047]	2003.37	[2.460]
Girl	0.489	[0.500]	0.490	[0.500]	0.508	[0.500]
Children in Household	2.281	[0.815]	2.049	[0.775]	2.001	[0.747]
Birthorder	1.652	[0.750]	1.996	[0.737]	1.951	[0.721]
Prob (First-born child)	0.493	[0.500]	0.238	[0.426]	0.256	[0.436]
Prob (Parents married)	0.409	[0.439]	0.397	[0.438]	0.533	[0.447]
Live is same municipality as GM	0.539	[0.463]	0.551	[0.462]	0.553	[0.469]
High SES (disp. income of GM)	0.446	[0.497]	0.429	[0.495]	0.391	[0.488]
Parents' age difference	2.776	[3.084]	2.822	[3.963]	2.912	[3.865]
Age of mother (when GM aged 60-64)	38.094	[2.808]	38.058	[2.924]	38.482	[2.820]
Native mother	0.960	[0.196]	0.957	[0.204]	0.955	[0.207]
Number of aunts and uncles	2.526	[1.043]	2.463	[0.976]	2.434	[0.990]
Number of aunts	1.730	[0.809]	1.698	[0.779]	1.685	[0.783]
Mother's age first child	28.019	[3.289]	28.133	[3.455]	27.483	[3.251]
Obs. Children	41685		21443		8445	
Obs. Mothers	22828		22828		5487	

Note: Table A14 reports means and standard deviations. Columns 1 and 2 consist of all children aged 4-12 of mothers in the baseline sample (i.e., with grandmothers born 8 months before and after January 1950). Columns 3 and 4 restrict the sample in addition to the youngest children aged 4-12. Columns 5 and 6 summarize characteristics of the youngest children aged 4-12 that can be matched to their Cito outcomes (Cito-sample used for the long-run analysis). The probability of parents being married, living in the same municipality as the grandmother, and high SES are predetermined (i.e., measured when grandmothers are aged 50-53).

Source: Authors' calculations from the CBS data.

Table A15: Smoothness of children's covariates (reduced-form)

	RD estimates		Mean at
	(1)	(2)	cutoff
Birth month	0.026	0.130	6.565
	[0.156]	[0.250]	[3.375]
Birth year	-0.062	-0.003	2003.35
	[0.113]	[0.186]	[2.437]
Girl	0.028	0.008	0.500
	[0.023]	[0.037]	[0.500]
Children in Household	-0.030	-0.039	2.029
	[0.035]	[0.058]	[0.736]
Birthorder	-0.036	-0.064	1.978
	[0.033]	[0.054]	[0.707]
Prob (First-born child)	0.018	0.044	0.231
	[0.019]	[0.030]	[0.422]
Prob (Parents married)	0.003	-0.041	0.540
	[0.020]	[0.034]	[0.445]
Live is same municipality as GM	-0.002	-0.004	0.554
	[0.021]	[0.034]	[0.467]
Parents' age difference	-0.279	-0.196	2.954
	[0.183]	[0.293]	[3.794]
High SES (disp. income GM)	-0.035	-0.005	0.403
	[0.022]	[0.036]	[0.491]
Age of mother (when GM aged 60-64)	0.075	0.014	38.391
	[0.125]	[0.205]	[2.755]
Native mother	0.013	0.049***	0.945
	[0.010]	[0.017]	[0.227]
Number of aunts and uncles	0.054	-0.065	2.441
	[0.044]	[0.069]	[0.896]
Number of aunts	0.053	-0.076	1.702
	[0.037]	[0.056]	[0.769]
Mother's age first child	0.183	0.350	27.312
	[0.152]	[0.243]	[3.449]
Obs. Children	8436	8436	1521
Obs. Mothers	8221	8221	
fit	linear	quadratic	

Note: Table A15 tests the impact of grandmothers being born since January 1950 on a list of children's characteristics. Regressions are based on all children aged 4-12 when the grandmother is aged 60-64 who can be matched to their CITO results. Columns 1 and 2 show results based on local linear and quadratic regressions with a bandwidth of 8 months, respectively. Robust standard errors clustered at the mother's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

Table A16: Impacts on CITO-sample selection

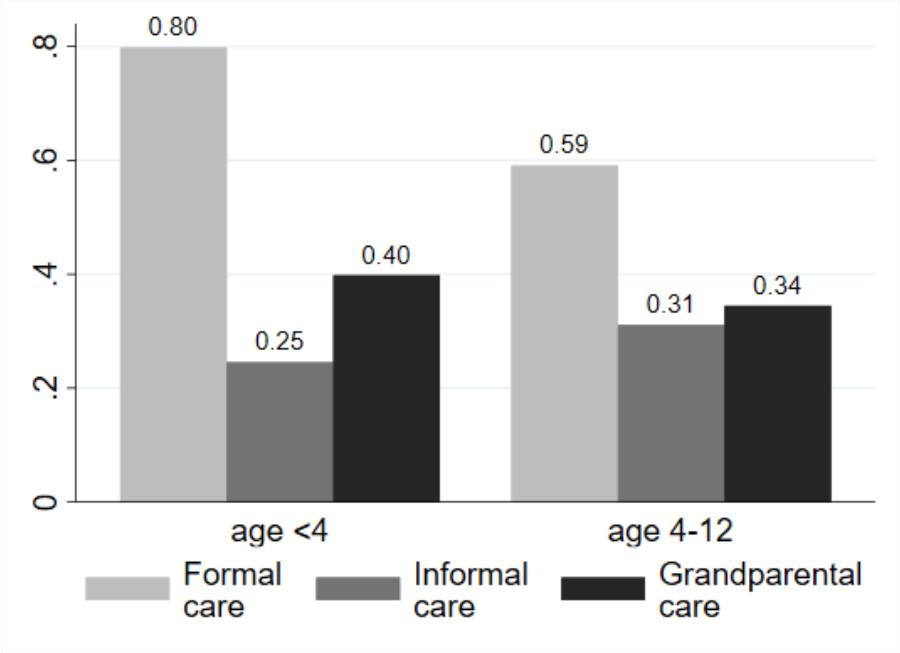
	RD estimates (1)	RD estimates (2)	Mean at cutoff
Restrictions by Age			
Relative age of child within family and test score availability			
Youngest child 4-7 and cito available	-0.013	-0.027	0.218
	[0.010]	[0.016]	[0.413]
Obs. Children	29019	29019	4887
Obs. Mother	18645	18645	
Youngest child 8-10 and cito available	-0.019	-0.018	0.267
	[0.012]	[0.019]	[0.442]
Obs. Children	21807	21807	3732
Obs. Mother	15377	15377	
Youngest child 11-12 and cito available	-0.019	0.004	0.221
	[0.014]	[0.023]	[0.415]
Obs. Children	12882	12882	2222
Obs. Mother	9758	9758	
fit	linear	quadratic	

Note: Table A16 tests the impact of grandmothers being born since January 1950 on a list of Cito-sample selection variables. Regressions are based on all children aged 4-7, 8-10, and 11-12 when the grandmother is aged 60-64, respectively. Columns 1 and 2 show results based on local linear and quadratic regressions with a bandwidth of 8 months, respectively. Robust standard errors clustered at the mother's level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.

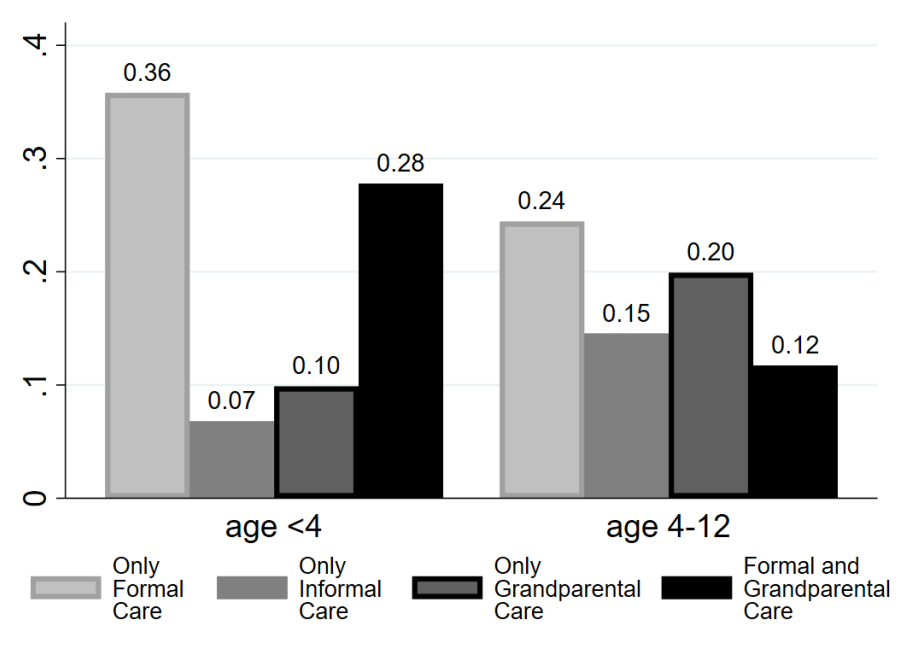
Table A17: Impacts on subsidy take-up (reduced-form)

RD estimates	Daycare		Out-of-school care		N
	Prob. (1)	Hours (2)	Prob. (3)	Hours (4)	Mothers
Age between 4-7	-0.007	-5.934*	-0.009	1.301	18683
	[0.006]	[3.051]	[0.014]	[6.807]	
Means at cutoff	0.092	38.847	0.321	122.516	
Age between 8-10	-	-	0.014	10.677	10243
			[0.015]	[7.193]	
Means at cutoff	-	-	0.180	66.916	
Age between 11-12	-	-	0.021*	8.850*	5292
			[0.011]	[4.931]	
Means at cutoff	-	-	0.048	16.114	

Notes: Table A17 shows educed-form reform impacts on childcare subsidy take-up in families with the youngest child aged 4 - 7, 8 - 10, and 11 - 12 when the grandmothers are at age 60 and 64. Subsidy take-up is shown for any child within the indicated age range, which means that a mother with multiple children in official child care will show up in multiple age groups. Columns 1 and 2 show effects on the probability of daycare take-up and the average hours of daycare usage, respectively. Columns 3 and for show effects on the probability of out-of-school care take-up and the average hours of out-of-school care usage, respectively. All specifications use local linear regression with a bandwidth of 8 months and include controls. Robust standard errors clustered by grandmother's id are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations from the CBS data.



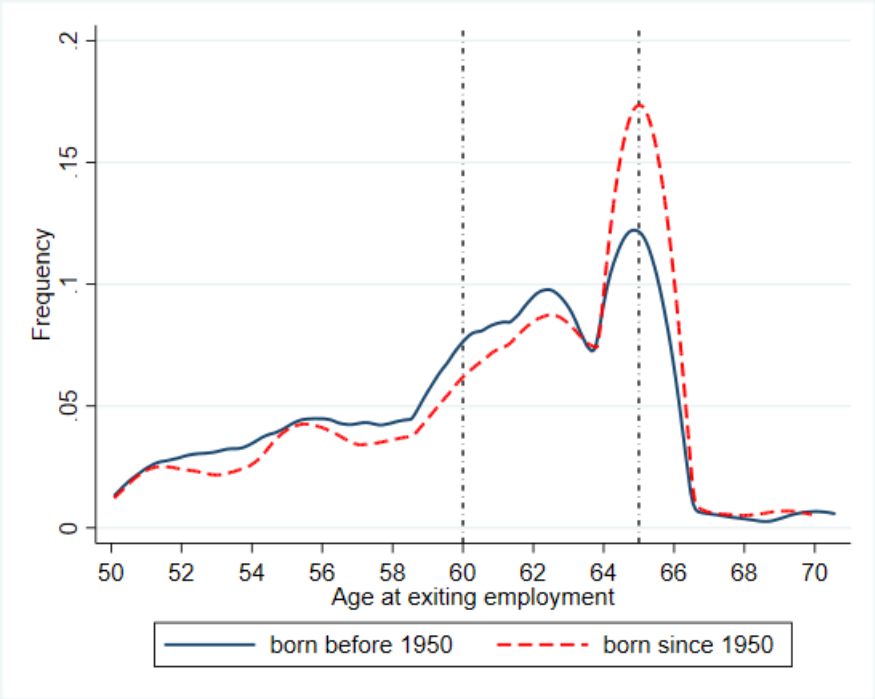
(a) Distribution of childcare modes



(b) Distribution of care mix

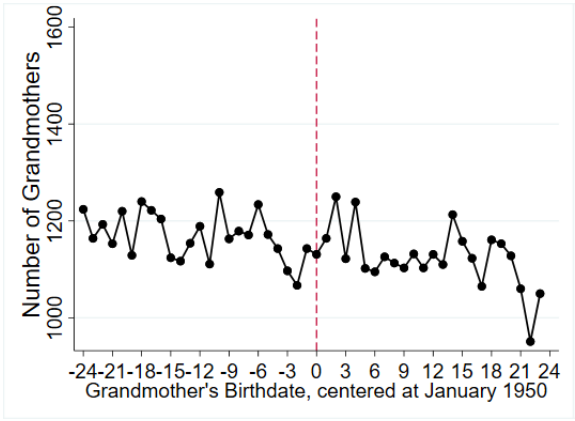
Notes: Figure A1 1 shows child care modes employed by parents in the LISS panel. Parents are asked separately for their children below 4 and between ages 4 and 12 whether they make regular use (at least once a week) of various types of child care. Panel (a) shows childcare take-up allowing for multiple answers so that the three categories are not mutually exclusive. Panel (b) shows the four most common combinations of child care modes with mutually exclusive categories.
Source: Authors' own calculations from LISS panel administered by CentERdata (Tilburg University, The Netherlands).

Figure A1: Survey evidence on childcare modes

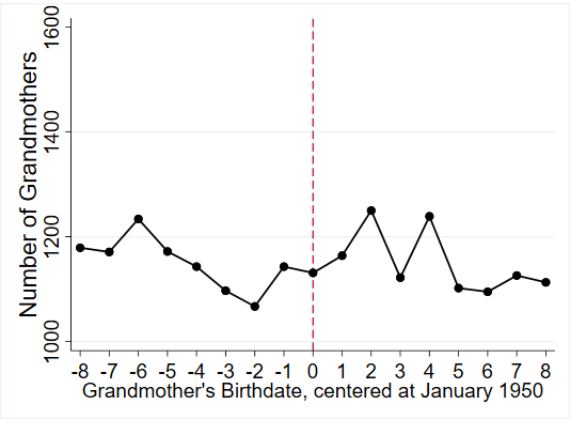


Notes: Figure A2 shows the distribution of age at exiting employment for the cohorts born before and since 1950 in baseline sample (i.e., 8 months around the cutoff). We can clearly see a shift towards later retirement for the treated cohorts.
Source: Authors' calculations from the CBS data.

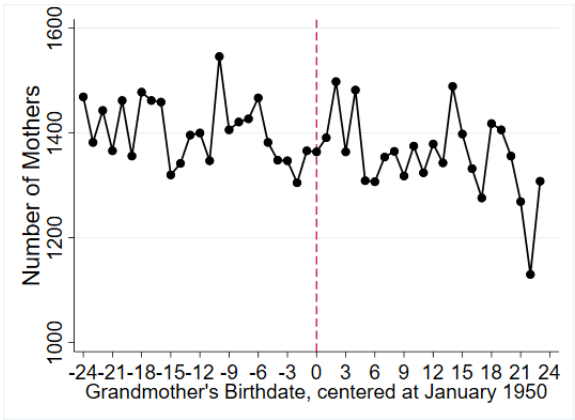
Figure A2: Distribution of age at exiting employment for women by treatment status



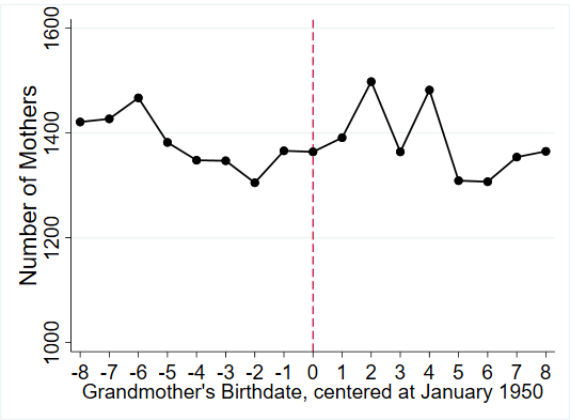
(a) Density of grandmothers (24 months)



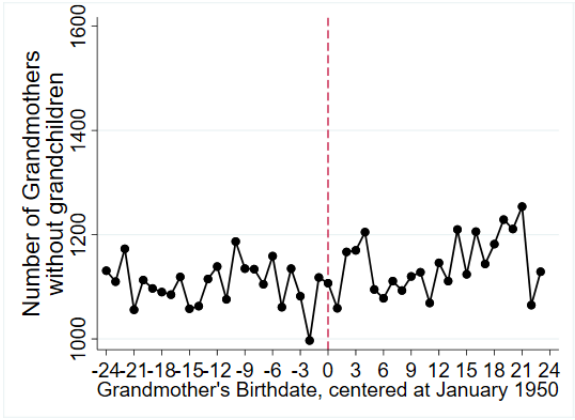
(b) Density of grandmothers (8 months)



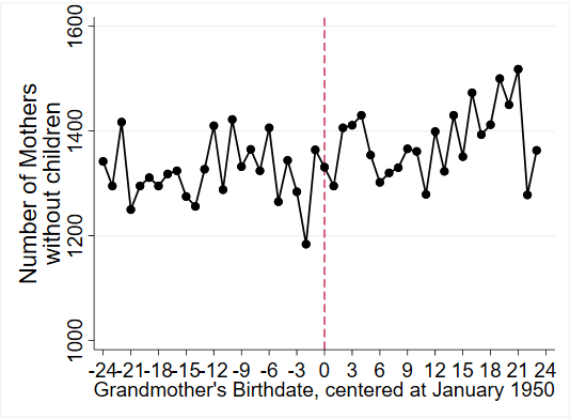
(c) Density of mothers (24 months)



(d) Density of mothers (8 months)



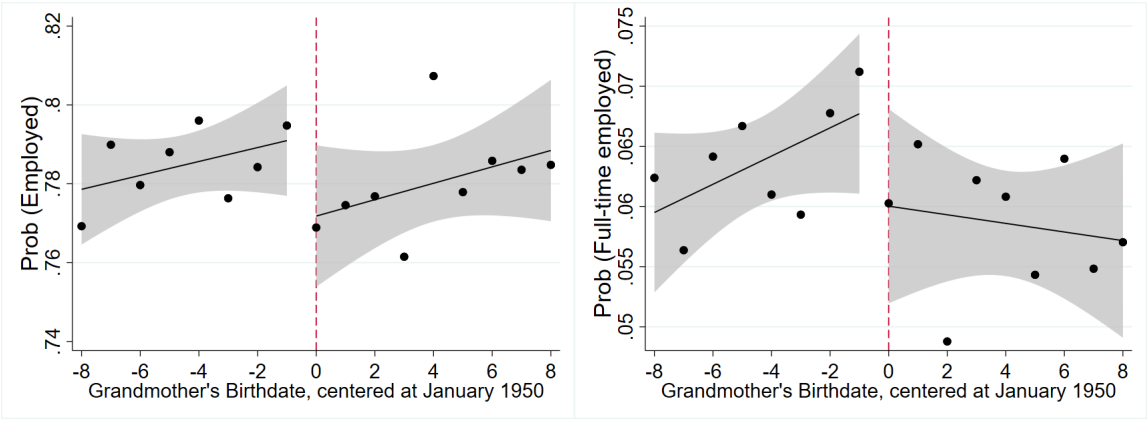
(e) Density of grandmothers without grandchildren



(f) Density of mothers without children

Notes: The (bin size/running variable) in Figure A3 is grandmothers' birth date/months. Figure A3a and A3b show the density plot of grandmothers 24 months and 8 months around the cutoff. Figure A3c and A3d show the density plot of mothers whose mothers' ('grandmother') birth month is 24 months and 8 months around the cutoff. Figure A3e and A3f 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 A3: Distribution of age at exiting employment for women by treatment status

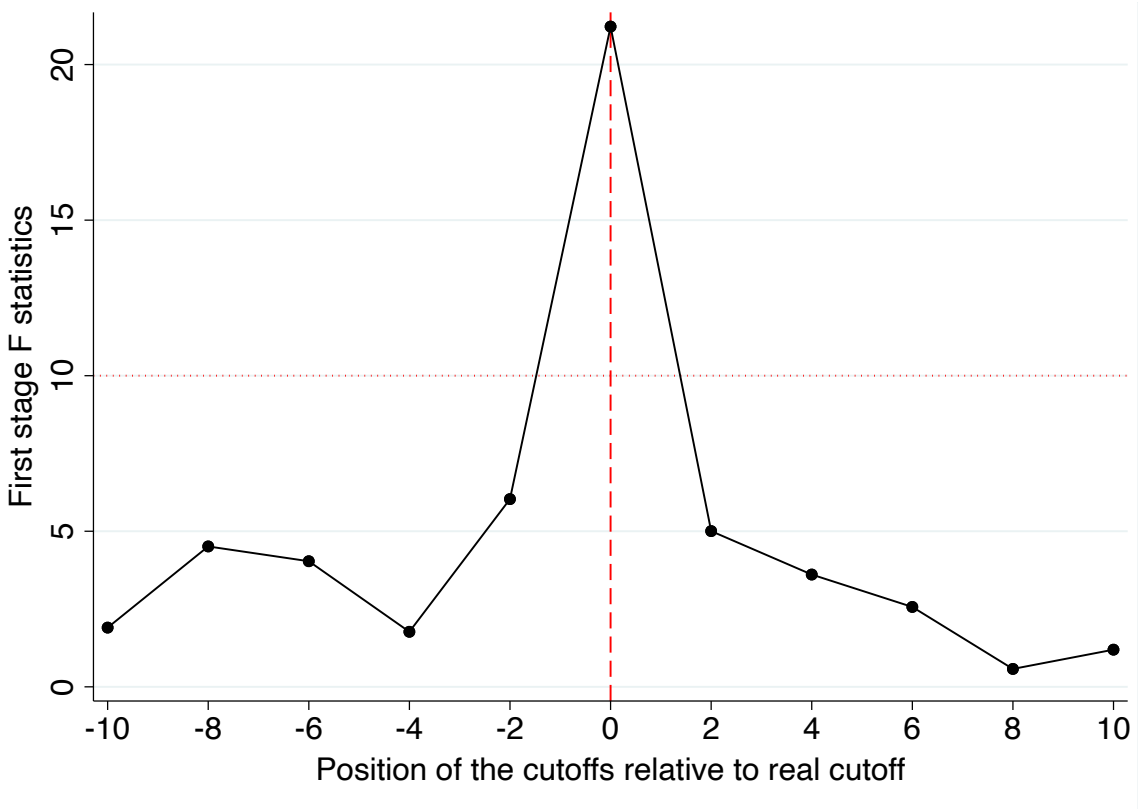


(a) Mothers' probability of employment

(b) Mothers' probability of full-time employment

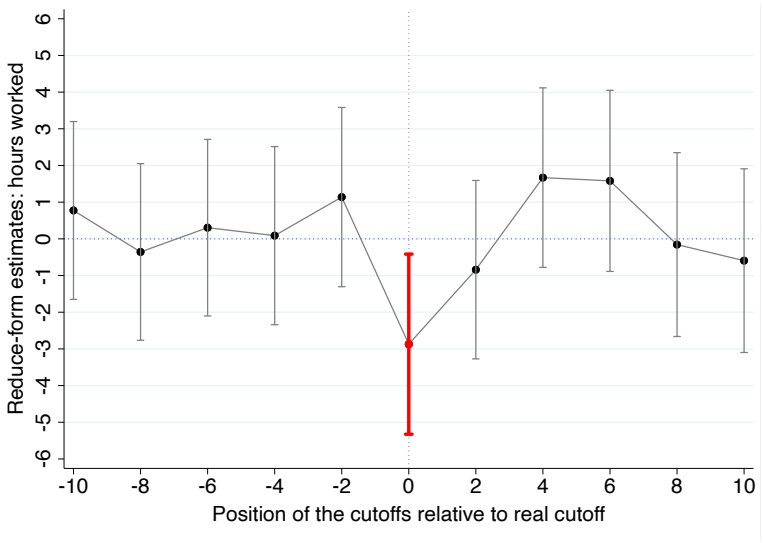
Notes: Figure A4 shows the scatter bin plots of mother's labor supply as a function of distance to the cutoff, which is grandmothers' birth month being January 1950. The solid lines are the linear fitted lines. The shaded areas indicate the 95 percent confidence interval.

Figure A4: Mothers' labor supply relative to the cutoff (reduced-form)



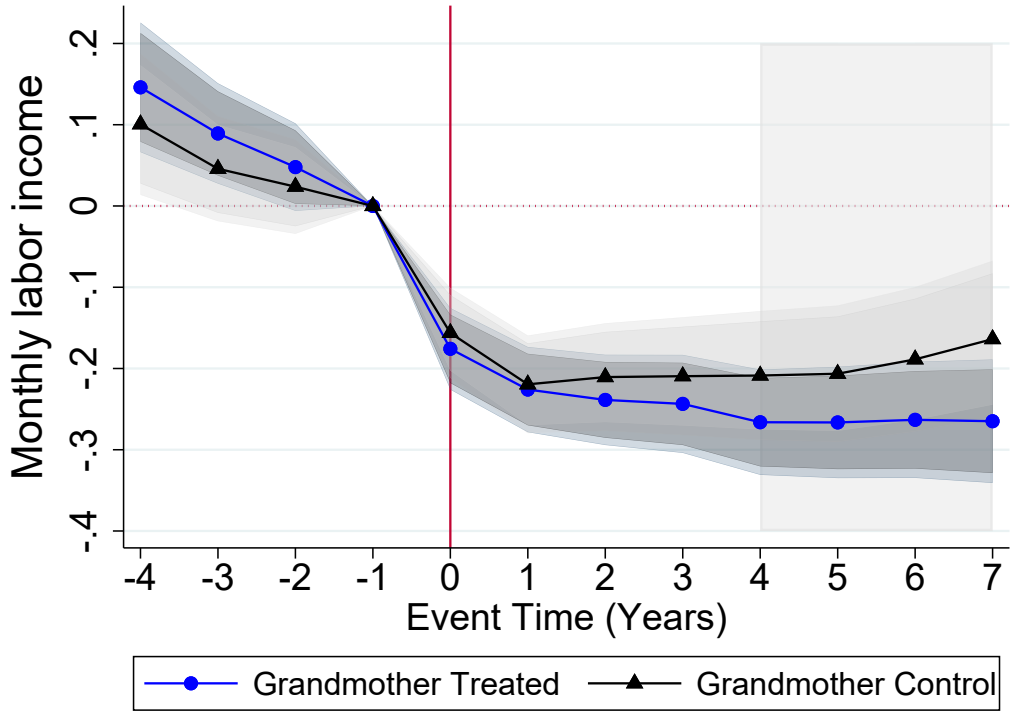
Notes: Figure A5 shows F-statistics when the cutoff is placed at "placebo" locations around the real cutoff. We estimate the first-stage effect of the reform on grandmothers' labor supply with placebo cutoffs ranging from -10 to 10.
Source: Authors' calculations from the CBS data.

Figure A5: Placebo test: F-statistics as a function of the location of cutoff



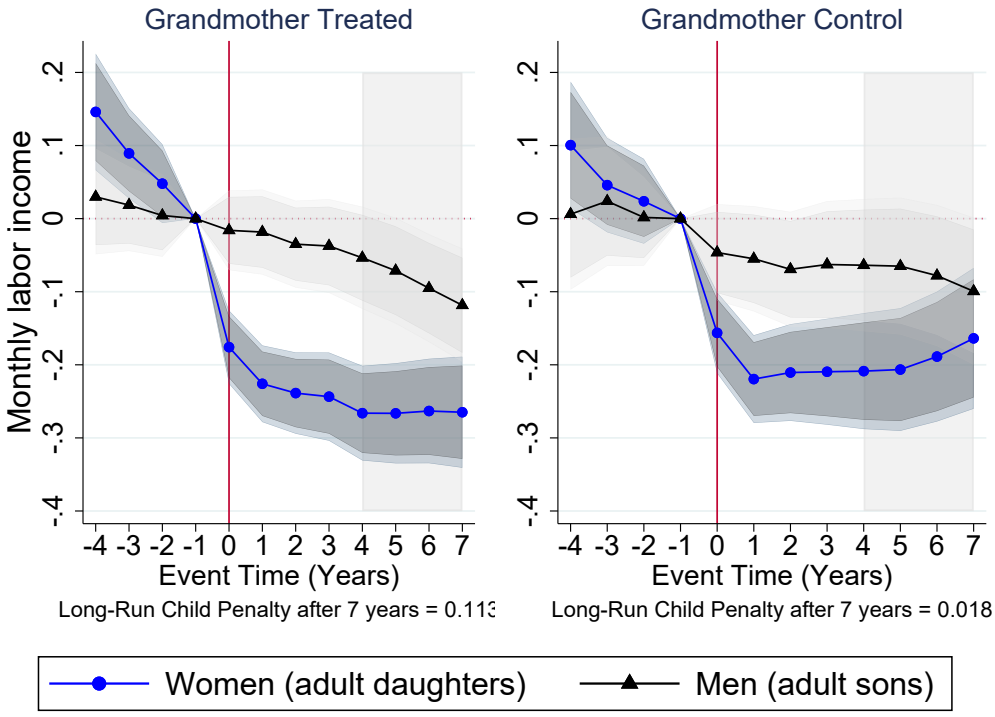
Notes: Figure A6 plots reduced-form estimates of total monthly hours worked and 95% confidence intervals by replacing the true cutoff (normalized to zero) with placebo cutoff locations ranging from -10 to 10. The red dashed line indicates the actual cutoff. Table A8 lists the regression estimates.
Source: Authors' calculations from the CBS data.

Figure A6: Placebo test: placebo estimates as a function of the location of cutoff



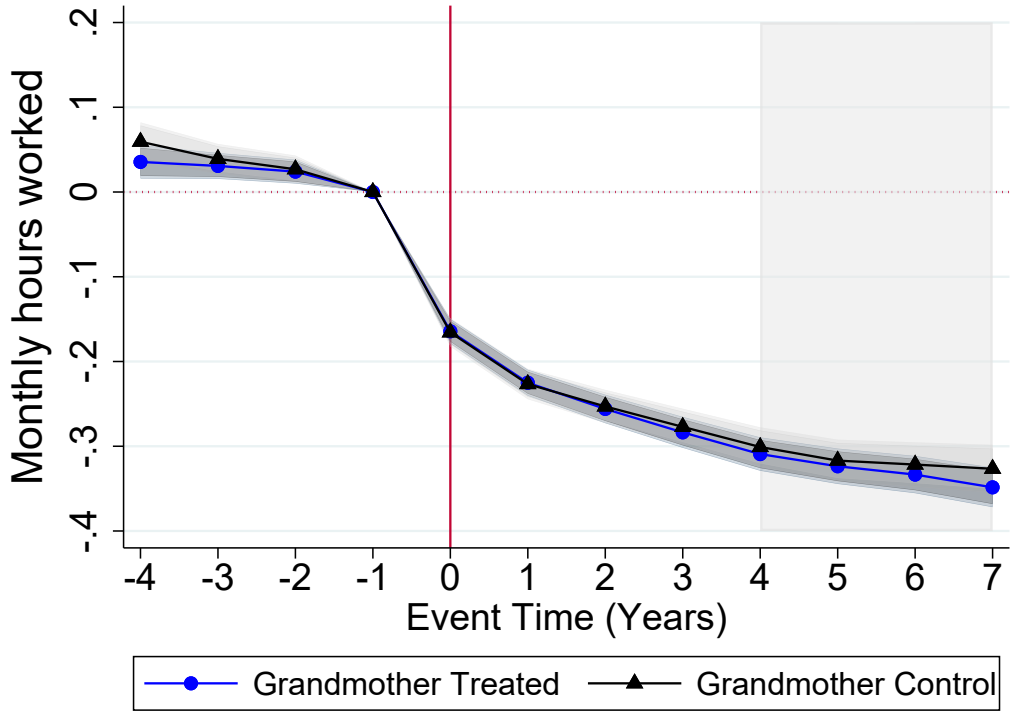
Notes: Figure A7 shows the history of monthly labor income of women for a window of four years before and up to seven years after they gave birth to their first child. The effect on monthly labor income is estimated unconditional on employment status to capture the impact on total labor supply. Monthly labor income of treated (blue dots) and control (black triangles) women are compared. Treated refers to women (mothers) with a grandmother born since January 1950 (using a bandwidth of 7 months). Event time 0 marks the birth of the first child. The shaded area between event times 4 and 7 indicates the main reform spillover window on the second generation. 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. Different tones of shaded areas indicate the 90 and 95 percent confidence intervals.

Figure A7: Dynamic treatment effects on monthly labor earnings



Notes: Figure A8 shows the child penalty in total labor earnings (including zeros) by treatment status. Blue dots document women's and the black triangles indicate men's monthly labor income. Treated refers to women/ men (i.e., adult daughters/ sons in our analysis) with a grandmother born after and including January 1950 (using a bandwidth of 7 months). Event time 0 marks the birth of the first child. The shaded area between event times 4 and 7 indicates the main reform spillover window on the second generation. 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. Different tones of shaded areas indicate the 90 and 95 percent confidence intervals. The long-run relative child penalty after 7 years (i.e., relative loss women experience compared to men) is reported below each sub-graph.

Figure A8: Relative child penalty by treatment status



Notes: Figure A9 shows the history of total monthly hours worked of women for a window of four years before and up to seven years after they gave birth to their first child. The effect on total hours worked is estimated unconditional on employment status to capture the impact on total labor supply. Monthly working hours of treated (blue dots) and control (black triangles) women are compared. Treated refers to women (mothers) with a grandmother born since January 1950 (using a bandwidth of 7 months). Event time 0 marks the birth of the first child. No restriction is made concerning additional children born between the depicted event times zero and seven. The shaded area between event times 4 and 7 indicates the main reform spillover window on the second generation. 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. Different tones of shaded areas indicate the 90 and 95 percent confidence intervals.

Figure A9: Dynamic Treatment effects on monthly working hours (incl. multiple births)