Netspar

Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

Prolonged worklife among grandfathers: Spillover effects on grandchildren's educational outcomes

Jim Been, Anne C. Gielen, Marike Knoef, Gloria Moroni

DP 05/2022-035

Prolonged worklife among grandfathers: Spillover effects on grandchildren's educational outcomes

Jim Been* Anne C. Gielen† Marike Knoef ‡ Gloria Moroni §¶

May 2022

Abstract

Recent policies aiming to prolong worklives have increased older males' labor supply. Yet, little is known about their intergenerational effects. Using unique Dutch administrative data covering three consecutive generations, this paper studies the impact of increased grandfathers' labor supply following a reform in unemployment insurance for persons aged 57.5+ on grandchildren's educational performance. We find that increased grandfathers' labor supply increases grandchildren's test scores in 6^{th} grade. The effect is driven by substitution of grandparents' informal care by formal childcare.

Keywords: Intergenerational effects, Labor supply, Unemployment insurance, Child care, Child development.

JEL Classification: J13, J14, J22, J26, J65.

*Department of Economics at Leiden University, and Netspar.

†Erasmus School of Economics, Tinbergen Institute, and IZA.

‡Department of Economics at Leiden University, Department of Econometrics OR at Tilburg University, and Netspar.

§Erasmus School of Economics and Tinbergen Institute. (Corresponding author: moroni@ese.eur.nl)

¶We thank participants at various conferences, workshops and seminars for useful comments and suggestions. Financial support from Instituut Gak and NWO (Vidi grant 452-17-007) is gratefully acknowledged.

1 Introduction

Since the early 2000s, many OECD countries have stimulated old-age labor supply by reducing the generosity of unemployment, disability, and early-retirement schemes (OECD, 2011). Although the direct labor market effects of these reforms on the targeted population have been studied extensively (e.g. Staubli and Zweimüller 2013; Atav et al. 2021), any intergenerational effects on subsequent generations have remained unaddressed. Increased labor supply of older individuals may have a strong impact on their children and grandchildren, for example, if it reduces opportunities to provide informal care to grandchildren. This paper studies the multigenerational effects of older workers' activation, focusing on grandfathers who have been most responsive to incentives to increase old-age labor supply.[1] Focusing on grandfathers is relevant because the biggest potential change in time spend with grandchildren is when grandfathers do or do not work: although they provide less care than grandmothers, on average, the amount of care is more affected by grandfathers' than grandmothers' employment status.

There is a vastly growing literature on the intergenerational transmission of labor market policies, and more broadly on the intergenerational persistence of socioeconomic outcomes. Although studies have provided evidence on correlations in socioeconomic status within families up to the fourth generation,[2] causal spillover effects on biologically linked family members have only been documented for two subsequent generations[3] (Dahl et al., 2014; Dahl and Gielen, 2021; Hartley et al., 2017; Grübl et al., 2020; Eibich and Siedler, 2020; De Haan and Schreiner, 2018) or from grandchildren to grandmothers (Truskinovsky, 2021). This paper is

[1] We show results of grandmothers too. However, for the relevant cohorts in our study, grandmothers' labor supply is very selective. As most grandmothers have left the labor market early (or never worked at all) those still active in the labor market at the cutoff age of our reform represent a non-random subsample of all grandmothers. Moreover, we find that grandmothers are much less responsive to the reform in terms of labor supply, most likely because of this selection. Therefore, we primarily focus on grandfathers, but present also evidence for grandmothers as grandparents are likely to provide care for their grandchildren together.

[2] See among others Adermon et al. 2018, 2021; Lindahl et al. 2015; Braun and Stuhler 2018; Bütikofer et al. 2018; Clark and Cummins 2015; Ferrie et al. 2021; Long and Ferrie 2018. Barone and Mocetti (2021) also provides evidence of correlation of outcomes among family ties over multiple centuries.

[3] Related studies such as Braun and Stuhler (2018) focus on intergenerational links while using surnames to link individuals from subsequent generations rather than using actual biological family links. With this approach only fathers and sons can be linked.

the first to use quasi-experimental methods to study the causal impact of grandfathers' labor market activation on outcomes of three generations (i.e. grandparents, parents, and grandchildren). Our main outcome of interest is educational performance of grandchildren by the end of primary school. Moreover, we investigate the underlying mechanisms by considering effects that arise from both direct spillovers - from the grandfathers to the grandchildren - as well as indirect spillovers - via the parents.[4]

Our setting is an unemployment insurance reform in the Netherlands aimed to activate older workers. This paper studies the causal effect of the increased labor force participation of grandfathers on their grandchildren's educational performance. The Netherlands is an interesting case because of two main reasons. First, Dutch grandparents spend a relatively large amount of time caring for their grandchildren. About 35% of grandparents look after their grandchildren every week; this is comparable to Northern European grandparents (e.g. Sweden, Denmark) but lower than Southern European grandparents (e.g. Spain, Italy, Greece) (Appendix Figure A.1). Second, formal child care is heavily subsidized and therefore widely available to parents. Our identification is based on the introduction of mandatory job search requirements for unemployment benefit recipients aged 57.5 and older, which came into effect on January 1st 2004 in the Netherlands. Earlier work by Hullegie and Van Ours (2014), Lammers et al. (2013), and Been and Knoef (2017) has shown that this reform significantly increased the active labor market participation of older men by reducing the inflow into unemployment and increasing the outflow out of unemployment to employment. The reform affected old-age labor supply of women much less than that of men. The increased labor market participation of older men is likely to reduce their time spent on non-market activities such as leisure, home production, and providing informal care to their grandchildren (Aguiar et al., 2013).[5] A reduction in informal child care provision by the grandfather can weaken or improve grandchildren's outcomes, depending on whether the quality of child care provided by the grandfathers is better or worse than the alternative, e.g. formal child

[4] As such, this paper also contributes causal evidence to the literature on grandparental child care provision and childrens' outcomes (Del Boca et al., 2018; Lei, 2019; Barschkett et al., 2021).

[5] A reduction in the grandfathers' time spent with grandchildren is likely to imply that the grandmothers' time spent with grandchildren is also reduced because of existing joint time use decisions (Stancanelli and Van Soest, 2012b).

care or parental care.

Taking advantage of the reform, we estimate the causal effect of the increased labor force participation of grandfathers (*1st generation*, or *G1*) on their grandchildren's (*3rd generation*, or *G3*) educational performance, measured by their score on a standardized test in 6^{th} grade. To account for potential endogeneity issues, we implement an IV strategy which exploits variation generated by the policy introduction date (1 January 2004), and the age eligibility criterion (57.5 and older) to instrument for grandfathers' employment. We thus compare the educational performance of grandchildren at age 12, whose grandfathers' employment statuses were and were not triggered by the reform at the time of their birth.[6] We particularly focus on grandchildren aged 0-4 around the reform as these grandchildren may be most responsive to the quality of childcare provided. In a next step, we investigate potential channels driving the multigenerational effect: we study (i) a *direct* channel - i.e. grandfathers adjusting the (time) investments in their grandchildren, and (ii) an *indirect* channel - i.e. mothers (*2nd generation*, or *G2*) responding to the reform by changing their labor supply and/or fertility, which in turn can affect grandchildren's outcomes.[7] The model is estimated using unique multigenerational administrative data on the whole Dutch population that spans three consecutive generations with grandchildren born in the years 1999-2007.

We offer three sets of results on multigenerational spillover effects. First, we find a strong positive impact of increased grandfathers' labor market participation on grandchildren's educational performance. Based on the estimated Intention-to-Treat (ITT), we find that grandchildren score 1.9% of a standard deviation higher on a standardized test in the final grade of primary school when their grandfathers are affected by the reform.[8] This effect is statistically significant and economically important. In an IV-framework, we estimate the effect of grandfathers' labor market participation on grandchildren's test score to be 37.1% of a standard deviation. This large effect reflects the local nature of our results as well as

[6]There is also a literature on the reversed relationship, documenting the impact of having a grandchild on grandparent's labor supply (see for example Rupert and Zanella 2018; Gørtz et al. 2020).

[7]In the paper we focus on mothers instead of fathers, as women's labor supply is in general more responsive to changing circumstances. Evidence for fathers is provided in the Appendix.

[8]We find that the reform increased the labor supply of grandfathers by 5.1%-points which is in line with prior studies by Hullegie and Van Ours (2014), Lammers et al. (2013), and Been and Knoef (2017).

the year-by-year accumulation of investments in grandchildren from early on in life up to age 12 (6^{th} grade). In addition, we provide evidence that the effect - although triggered by the activation of grandfathers - is likely driven by the change in the joint investment of both grandfathers and grandmothers, given that both grandparents often spend time together with their grandchildren. A set of placebo tests show that the results are not driven by any other differential group trends across different grandfather's age, other time shocks occurring before or after the reform, nor by changes in selective fertility of mothers.

Second, we find that the positive impact of the reform on grandchildren's school performance is decreasing with grandfathers' number of children and grandchildren's birth order. Assuming that potential time investments in a grandchild decline with each additional grandchild born,[9] our baseline result is likely driven by the time grandfathers invest in their grandchildren.[10] In addition, we show that the positive effect is concentrated among grandchildren living in urban areas - where informal child care is more likely to be replaced by formal child care, and among grandchildren living closer to their grandparents. Finally, the quality of grandfathers' investments may play a role in explaining our findings. Our results show that the multigenerational impact is most prevalent among grandchildren whose grandfather has a low socio-economic status (SES), suggesting that these grandchildren mostly benefit from alternative child care provisions.

Third, our results show that mothers do not stay home to replace grandparents' informal care. This suggests that the positive effect is driven by substitution of grandparents' informal care by higher quality formal care.[11] Our findings are consistent with the evidence showing significant positive effect of formal child care programs on children outcomes especially among children from disadvantaged background (e.g. Currie 2001; Cornelissen et al. 2018; Felfe and Lalive 2018; Fryer et al. 2020).

[9] Price (2008) shows that a first-born child receives more quality time by their parents than a second-born child and this explains the recent evidence on the better outcomes of first born child (Black et al., 2005).

[10] An alternative source of grandparents' investments are material investments, i.e. monetary gifts, transfers or donation. However, we show that the amount of monetary transfers from grandparents to grandchildren is fairly limited in the Netherlands.

[11] See Havnes and Mogstad (2011) for similar evidence of substitution between formal and informal child care in Norway.

The remainder of the paper is organized as follows. We first describe the institutional settings characterizing the Dutch unemployment insurance scheme and formal child care provision in Section 2. A theoretical framework is introduced in Section 3. Section 4 discusses the empirical strategy and the data are described in Section 5. We present our main results in Section 6 and provide some robustness analysis in Section 7. Finally, we conclude in Section 8.

2 Institutional setting

2.1 Dutch unemployment insurance system

Until the early 2000s, the Dutch unemployment insurance (UI) system favored older unemployed individuals with specific regulations as older workers were considered a vulnerable group in the labor market with low reemployment possibilities. Next to being entitled to extended age-based benefit duration and specific welfare policies after exhaustion of UI benefits (for details, see Appendix B), older workers were exempted from the job search requirement.

Prior to January 1st 2004, only UI recipients aged younger than 57.5 were subject to mandatory job search. UI recipients aged 57.5 and above were exempted from this obligation to actively search for a job. Effectively, this meant that older UI recipients could receive 70% of their gross prior earnings and use this to bridge the period to full retirement at the age of 65. Everything else constant, this exemption was lifted in order to stimulate older workers to exit unemployment and increase labor supply as of January 1st 2004. Hence, as of that date, older unemployed individuals also had to fulfill the job search requirement in order to retain their UI benefit.[12]

Previous studies have shown that the introduction of mandatory job search requirements for older UI recipients has increased exit from UI and decreased entry into UI (Hullegie and Van Ours, 2014; Lammers et al., 2013; Been and Knoef, 2017).[13] This paper builds upon

[12]For an overview of the exact requirements, we refer to Appendix B.

[13]Petrongolo (2009) found a similar effect of job search requirements on treated individuals, including

this earlier literature and exploits the same 2004 reform as an exogenous shock to elderly labor supply to study the impact of grandparents' labor supply on outcomes of subsequent generations.[14]

2.2 Child care in the Netherlands

2.2.1 Formal child care

Formal child care provision is provided by private organizations in the Netherlands.[15] Parents are free to choose the type of child care they prefer. Child care providers are allowed to set their own hourly price, and parents receive a child care subsidy from the central government for each hour of care used, up to a maximum price per hour beyond which parents receive no additional subsidy.[16] The total subsidy received depends on households' income and the number of children under age 13 in formal child care per household. Low-income households receive a relatively higher subsidy which can be as large as 90% when household income falls below the threshold of 20,000 euros per year.[17] When family income exceeds this threshold, the subsidy is gradually reduced to 33% for households with income equal to 90,000 euros per year or more. Subsidy rates are higher for the second than for the first child (Akgunduz et al., 2015).

As of 2005, child care subsidies have been adjusted several times. In the period 2005-

those below 57.5 years old in the UK.

[14]Note that the effect of the reform can be identified as no other policy changes affected the different groups of 57.5− and 57.5+ simultaneously, and anticipation effects are relatively small (Hullegie and Van Ours, 2014; Lammers et al., 2013; Been and Knoef, 2017).

[15]Formal child care includes centre-based daycare, centre-based playgroups (*peuterspeelzalen*), and guest-parents (*gastouders*). Till 2015, it was possible to pay grandparents for providing care to their grandchild(ren) and receive child care subsidies. The grandparent had to pay taxes over the income from child care and any social insurance benefits were cut with the amount of income from child care.

[16]This maximum price is 8.46 euros/hour and 6.49 euro/hours for daycare and guest parents in 2021, respectively. For an overview of prices, we refer to https://www.belastingdienst.nl/wps/wcm/connect/bldcontentnl/belastingdienst/prive/toeslagen/kinderopvangtoeslag/hoeveel-kinderopvangtoeslag-kan-ik-krijgen/maximaal-uurtarief-voor-de-kinderopvang.

[17]Median standardized household income is about 28,600 euro per year in 2020 according to Statistics Netherlands. To make this comparable to the threshold of 20,000 euros of non-standardized household income, we assume a two-person household with one child. This gives an equivalence factor of 1.67 and, hence, a non-standardized income of 47,762 euros per year. This is 53,768 euros if we assume two children per household (the average number of children per household is 1.57).

2008, child care subsidies were increased substantially thereby effectively cutting the net price of formal child care in half. Despite the sudden increase in the generosity of child care subsidies, these reforms only had a modest impact on maternal labor supply (Bettendorf et al., 2015). The steep rise in public spending on child care (from 1 billion euros in 2004 to 3 billion euros in 2009) that followed these reforms, initiated new reforms in which the government tried to curb the rise in public spending on child care. In 2012, the hourly subsidy reduction for parents was at least 2 to 5%-points for the first child but could exceed 10%-points for high income households. This reduced generosity led to a reduced demand for formal child care in terms of fewer children attending child care centers (Akgunduz et al., 2015). However, Been and Jongen (2017) find no statistical evidence for any maternal or paternal labor supply changes, suggesting that reduced formal care use has been replaced by informal care. The changes in subsidy generosity of formal child care are not in any way related to the age of the grandparents, and hence do not invalidate our research design. Moreover, it is important to note that most of the net hourly price changes in formal child care occurred after 2007, whereas our final sample only considers the period 1999-2007 for identification (see Section 5).

In the period under consideration in this study there were substantial increases in the use of school care (i.e. lunch at school between the morning and afternoon schedules) and guest parents (i.e. care by a person who is an official host). Data of Statistics Netherlands show an increase in school care of 214% and an increase in guest parents of 200% in the period 2001-2008. Daycare (i.e. daycare centers) is the type of formal care that is most often used. It was already on a high level and raised even further (+52%) in the period under consideration. The use of kindergarten (i.e. pre-school for kids aged 2-4) was relatively constant throughout this period (a decrease of 1%). The jump in use of formal child care observed in 2006 can be attributed to the expansion of child care subsidies (Appendix Figure A.2).

Internationally, the use of child care has a very particular structure in the Netherlands. According to OECD (2020), enrollment rates of 0-2 year olds into formal child care is about 59%, which is the third highest percentage of enrollment in the OECD. However, the average

number of hours per week of formal child care is only about 17 hours, which is the lowest number of hours in the OECD countries. This may be due to the high level of part-time work among Dutch mothers and suggests that formal child care is largely combined with informal child care in the Netherlands.

2.2.2 Informal child care by grandparents

SHARE (Survey of Health, Ageing and Retirement in Europe) 2004 data shows that only few grandparents spent time with their grandchildren on a daily basis in the Netherlands. However, about 35%, 23%, and 40% of the grandparents report spending time with their children at least every week, every month, and less often, respectively (Figure A.1 in the Appendix). This pattern of informal child care provided by the grandparent is comparable to other countries in which formal child care is broadly present (e.g. Denmark and Sweden), but very different from Southern European countries in which formal care is less accessible, and grandparents are more likely to take care of their grandchildren on a daily basis (e.g. Italy, Greece, and Spain).

Grandmothers are more involved in informal care for their grandchildren than grandfathers in the Netherlands. 45% of grandmothers look after their grandchildren for at least one day a week compared to 30% of grandfathers, although part of the time spent with the grandchildren may be spent simultaneously by the grandmother and grandfather (Appendix Figure A.3, Panel (a)). Note that this gender-difference is much less pronounced for inactive grandmothers and grandfathers (Appendix Figure A.3, Panel (b)). Employed grandfathers spent relatively little time with grandchildren. Grandmothers, on average, spent more time with grandchildren regardless of their labor market status. This may be due to the large amount of part-time work among these women. Because of shorter employment histories women were also less affected by the reform and therefore we focus on grandfathers in this analysis. These descriptives suggests that the largest potential change in time spent with grandchildren is when grandfathers do or do not work: although they provide less care than grandmothers, on average, the amount of care is mostly affected by their employment status.

In Appendix C, we show the prevalence of different forms of child care using ancillary data from the LISS panel[18], including information on child care provided by the grandparents. These data suggest that 37% (23%) of families have grandparents providing informal child care to children aged 0-4 (5-12) and 42% of families use formal daycare centers (i.e. kindergarten) which makes child care by grandparents almost equally important as kindergarten in terms of prevalence.

3 Theoretical framework

Building on Del Boca et al. (2014) and Rupert and Zanella (2018), this section provides an analytical framework in which grandparents (G1) and parents (G2) care about grandchildren's (G3) ability. In line with our empirical analysis, we assume that G1 is the grandfather and G2 the mother. Our goal is not to estimate a structural model like Del Boca et al. (2014) and Rupert and Zanella (2018), but rather to illustrate potential direct and indirect mechanisms through which grandfathers' labor supply decisions can impact grandchildren's ability. In our stylized model both G1 and G2 are altruistic with respect to the G3's (future) success in life. For simplicity, we assume that G1 is also altruistic with respect to G2, but G2 is not altruistic with respect to G1.[19] Children's ability A_t depends on various inputs, where the transformation function z_t can vary across children[20] (depending on the need of children and the quality of informal and formal investments) and over time (e.g child age):

$$A_t = z_t(A_0, A_{t-1}, I_t^{G1}, I_t^{G2}, T_t^F) \tag{1}$$

A_0 denotes the endowment at birth, which depends on the mother's age at birth (k_a) together with some unobserved factors v, i.e. $A_0(k_a, v)$. I is the investment by grandfathers (G1) and

[18]The LISS panel (Longitudinal Internet studies for the Social Sciences) surveys 5,000 representative households in the Netherlands (approximately 7,500 individuals) every month as from 2007.

[19]Intergenerational (monetary) transfers from G2 to G1 in the Netherlands is very uncommon. In part, because G1's well-being is protected relatively well by publicly provided state pensions and, if necessary, social assistance benefits. Substantial time investments from G2 to G1 are also relatively scarce due to the high use of nursing homes in the Netherlands. Recent empirical evidence from Truskinovsky (2020) suggests that G2 does not adjust labor supply in response to necessary increments in time taking care of G1.

[20]In line with earlier studies, we leave out a child-specific subscript for brevity here.

mothers (G2), respectively, which is increasing with investments in quality time (η) spent with the (grand)children and in money/gifts (ν) provided: $I_t = f(\eta_t, \nu_t)$. Time spent in formal child care is denoted by T^F.

Parents' (G2) utility function is denoted by

$$U_t^{G2} = u(C_t^{G2}, L_t^{G2}, A_{t1}, ..., A_{tK}) \tag{2}$$

where parent's utility U_t^{G2} is assumed to be additively separable in preferences for consumption C_t^{G2}, leisure L_t^{G2}, and ability of their children $A_1, ..., A_K$, with $u' > 0$ and $u'' < 0$, where K refers to G2's number of children. For simplicity, a unitary household model is assumed.[21]

Parents optimize U_t^{G2} with respect to both a budget constraint and a time constraint. The budget constraint is given by

$$C_t^{G2} + p_F \cdot T_t^F + \nu_t^{G2} = w_t^{G2} \cdot h_t^{G2} + \mu_t^{G2} \tag{3}$$

where p_F is the net price for an hour of formal child care relative to the market price of C_t[22], w is the (shadow) wage rate i.e. the price of time, h is the number of hours worked for pay, and μ any non-labor income. If there are siblings, we assume for simplicity that formal time (T^F) and informal time (η) spent on the child is spent together with the other children. The same holds for the monetary investments ν.[23]

We normalize the price of market consumption C_t to one and assume no saving or borrowing as in Del Boca et al. (2014).[24] The time constraint is given by

$$L_t^{G2} + h_t^{G2} + \eta_t^{G2} = H \tag{4}$$

where H is the total time endowment (24 hours per day). The utility function of the

[21]For comparison, Blau and Robins (1988) and Ruhm (2004) only consider mothers.

[22]We assume that the prices of C_t and ν_t are equal. Our overview of the institutional setting of formal child care in Section 2.2 indicates that p_F can be considered constant over time in our period of analysis.

[23]Note that the role of siblings is not the focus of our study, and this simplifying assumption does not influence the predictions from our model

[24]Assuming no saving or borrowing implies that households are assumed to be credit constrained.

grandfathers (G1) is given by[25]

$$U_t^{G1} = u(C_t^{G1}, L_t^{G1}, C_{t1}^{G2}, ..., C_{tJ}^{G2}, L_{t1}^{G2}, ..., L_{tJ}^{G2}, A_{t1}, ..., A_{tN}) \tag{5}$$

where J represents the number of children of the grandfather and N the total number of grandchildren. Grandfathers are altruistic with respect to both their children (G2) and grandchildren (G3), and they receive utility from their own consumption and leisure, the consumption and leisure of their children (G2), and the ability of their grandchildren. Grandfathers' budget constraint can be described by

$$C_t^{G1} + \sum_{j=1}^{J} \nu_{jt}^{G1} = w_t^{G1} \cdot h_t^{G1} + \mu_t^{G1}, \tag{6}$$

where $\sum_{j=1}^{J} \nu_{jt}^{G1}$ is the monetary gift/transfer spent on the grandchildren of the G2-children.

The time constraint for G1 is

$$L_t^{G1} + h_t^{G1} + \sum_{j=1}^{J} \eta_{jt}^{G1} + s_t^{G1} = H \tag{7}$$

where $\sum_{j=1}^{J} \eta_{jt}^{G1}$ is the time spent with the grandchildren of the G2-children. Job search time in case of unemployment is denoted by s. The reform increases job search time s and/or reduces the generosity of the unemployment insurance benefits (μ^{G1}) when not obeying to the job search requirements. This is likely to increase the labor force participation of the grandfathers (h^{G1}), and may have spillover effects on the educational outcomes of the grandchildren, either by changing the investments of grandfathers in their grandchildren (ν^{G1}, η^{G1}) and/or via an intermediate effect on G2.

This intermediate effect on G3 through G2 may exist because any reduction in G1's investments (I^{G1}) needs to be replaced by increased investment from G2 (I^{G2}) and/or from formal childcare (I^F). As a consequence, mothers' labor supply (h^{G2}) is adjusted depending on the choice for I^{G2} and I^F. A reduction in h^{G2} is likely to be evidence for choosing to invest G2's time (η^{G2}), while a status quo of increase in h^{G2} most likely resemble an increase

[25]For brevity, we leave out a grandfather-specific subscript.

in I^F.

4 Model and identification

In this section, we first describe how we identify the effect of grandfather's labor supply on grandchildren's ability. Afterwards, we explain how we investigate the potential underlying mechanisms, based on the theoretical model outlined in the previous section.

4.1 Effect of grandfathers' employment on grandchildren's ability

We estimate the causal effect of grandfathers' extensive margin labor supply on grandchildren's ability, which is measured by their educational performance on a standardized test at the age of 12. Performance on this standardized test is indicative for students' ability and is positively associated with economic outcomes later in life (Van Elk et al., 2011).

The identification of a causal effect faces potential endogeneity problems. For example, grandchildren with less affluent characteristics may have grandfathers with less affluent characteristics, who as a result may be more likely to be unemployed. In addition, grandfathers may adjust their labor market participation and time spent with grandchildren in response to grandchildren's development. To take into account potential endogeneity issues, we estimate an instrumental variable (IV) model. In the first stage, we exploit the 2004 UI reform that generated exogenous variation in grandfathers' labor supply. We estimate the following difference-in-differences (DiD) model:

$$EMP_i^{G1} = \beta_0 + \beta_1 D_i P_i + \beta_2 D_i + \beta_3 P_i + \boldsymbol{X}_i^{G1}\beta_4 + \boldsymbol{X}_i^{G3}\beta_5 + \varphi_i \tag{8}$$

where EMP_i^{G1} is the grandfathers' employment status for each grandchild i at birth.[26] D_i is a dummy for the treatment group, which are G3 whose grandfathers were 57.5 or older at the time of G3's birth. The treatment period (P_i) is equal to one when G3 is born

[26]Grandfathers are considered employed if they are either in self-employment or in paid employment

after the reform, and zero otherwise. Here, β_1 is the parameter of interest, measuring the effect of the reform on G1's employment status. Furthermore, $\mathbf{X}_i^{G1}$ is a vector containing background characteristics of the grandfather (including age and age^2, immigrant status, having a partner, and the number of G1's children at the time of G3's birth), and $\mathbf{X}_i^{G3}$ a vector containing background characteristics of the grandchild (including gender, immigrant status, and year of test fixed-effects). φ_i is a zero-mean error term.

The 2004 UI reform introducing variation in job search requirements for older unemployed individuals is a *valid* instrument because the reform has no effect on child development other than through the grandfather's labor supply decision. As shown by the theoretical model, this may be a direct effect induced by a change in grandfathers' investments in their grandchildren, and/or via generation G2. The instrument is *relevant* if the reform actually affects grandfathers' labor supply ($\beta_1 \neq 0$).[27] By using a DiD framework in the first stage, we separate the effect of the reform from other period effects. The main identifying assumption in the DiD framework is that the labor market status of the treatment (i.e. unemployment benefit recipients aged 57.5 and above) and control group (i.e. all younger unemployed) would have evolved in a similar way if the reform had not been introduced. As the job search requirement was already in place for younger unemployed (control group), our setting is a *DiD in reverse*, where the introduction of the reform makes the two groups more similar rather than more different, since the control group (individuals younger than 57.5) is always treated (Kim and Lee, 2019). The usual common trend assumption, therefore, needs to be verified *after* the treatment rather than before. Figure A.4 in the Appendix provides evidence that this is indeed the case, especially for grandfathers, as the trend in unemployment insurance between treatment and control groups becomes more similar after the treatment. [28] Moreover, earlier works studying the same reform (Hullegie and Van Ours, 2014; Lammers et al., 2013; Been and Knoef, 2017) have established that anticipation effects are small and therefore unlikely to be a threat for our identification strategy.

[27]Compared to Been and Knoef (2017) we take a binomial instead of multinomial approach and consider a static instead of dynamic model. In our setting the additional heterogeneity of a dynamic multinomial approach is not necessary and we consider the simplified version model as a reference. However, estimating a multinomial dynamic model does not alter our main conclusions (see Table A.3).

[28]In Figures A.5 and A.6 in the Appendix, we show the trends pre- and post-reform trends in mothers' labor supply and fertility decisions and grandchildren's test scores, respectively.

The second stage of our IV model is specified as follows

$$A_i^{age12} = \delta_0 + EMP_i^{G1}\delta_1 + \mathbf{X}_i^{G1}\delta_2 + \mathbf{X}_i^{G3}\delta_3 + D_i\delta_4 + P_i\delta_5 + \epsilon_i, \tag{9}$$

where A_i^{age12} is G3's score at a standardized test at the end of primary school (at age 12) and ϵ_i is a zero-mean error term, which is likely to be correlated with φ_i. The parameter of interest δ_1 measures the causal effect of grandfathers' extensive margin labor supply on grandchildren's educational outcomes. As previous literature suggests the timing of investment matters (see for example Carneiro et al. 2021), we also investigate differential effects depending on the grandchild's age.[29] The model is estimated using 2SLS.

Moreover, the theoretical framework shows us the potential pathways through which the UI reform affected grandchildren's ability. Below we explain how we will differentiate between possible direct (i.e. from G1 to G3) and indirect channels (i.e. directly from G1 to G3 through G2) in our analysis.

4.2 Potential *direct* channel

There are several mechanisms through which the reform can have spillover effects on grandchildren outcomes. Section 3 illustrates that when the reform increases employment of grandfathers (h_t^{G1}), this causes them to have less time available to provide informal care to their grandchildren (η_t^{G1}), but potentially also increasing monetary resources available that could (partly) be transferred to children and grandchildren. Ancillary data, however, show us that monetary transfers are low, so in the remainder we will focus on the time channel.[30]

Our administrative data lacks information on informal child care provision by the grand-

[29]Note that the exposure to the treatment also diminishes with age, as (pre-)school activities increase with the age of the grandchild. We can not disentangle the age and the exposure effect.

[30]In data from the Dutch LISS survey Time Use and Consumption we find that among households with the main respondent above 50 years old, donations and gifts for family, friends, charity etc. (which is broader than (grand)children only) are fairly limited and amounts to only 4% of total monthly spending in the period 2009-2019. This corresponds to about 86 euros per month of gifts and donations to family, friends, and charity. This relatively small amount also makes it unlikely that parents use money transfers from grandparents to pay for child care. Especially, since (gross) child care costs are about 68 euros per day (given the hourly price of 8.46 euros/hour in footnote 16).

fathers (η_t^{G1}). However, in line with the theoretical model, ancillary data from SHARE confirms that informal childcare is higher for non-employed grandfathers than for employed grandfathers (Appendix figure A.3, Panel (b)). When children receive less informal care from grandfathers this may be positive or negative for grandchildren's outcomes, depending on the quality of the alternative (formal care or parental care).[31]

To corroborate the direct time channel we investigate whether the intergenerational effect is larger for grandchildren who were likely to receive a larger investment from their grandfathers in absence of the reform. Therefore, first, we consider heterogeneity in the multigenerational effect of the reform by the total number of children (G2) for each grandfather. As shown by Equation (7), grandfathers' total number of children (G2) influence the direct time investment that grandfathers can devote to their grandchildren. When grandfathers' time available to provide informal care to their grandchildren is reduced due to the reform, the impact is likely largest for those who have only few G2-children.[32] Second, grandfathers are likely to spend more time with their first grandchild as compared to later born grandchildren, as time available has to be distributed among more grandchildren once more are born. Therefore, we also estimate heterogeneous effect with regard to the grandchild being the first grandchild. If the time channel is important, we expect to find largest effects for first-born grandchildren.

Third, there is evidence of higher availability of childcare in urban areas as compared to rural areas, due to relocation of private child care providers in more profitable urban areas (Noailly and Visser, 2009). As a result, a reduction in informal care is more likely to be replaced by formal child care in urban areas. To further support the direct time channel we estimate our baseline model by grandchildren's residential area, i.e. urban or rural. Note that the quality of formal child care is similar in Dutch urban and rural areas (Slot et al., 2018). So, if the effect is driven by substitution of informal grandparental care by formal care, we expect to find that effects are concentrated among grandchildren living in urban

[31]Other forms of informal care do not seem to be widely available/accessible, see Table C.1 in the Appendix.

[32]We investigate heterogeneous effects regarding the number of G2 children instead of the number of G3 children. That is because the number of G3 children may be endogenous and because time may be spent on a grandchild and his/her brothers and/or sister together.

areas.[33] Finally, we test whether effects vary across socio-economic status of the grandfather. A higher G1 socio-economic status may be associated with higher quality of investments in G3. Hence, if the time channel is important, we expect a larger positive effect on G3 school performance among grandchildren whose grandfathers have a low socio-economic status, as proxied by low G1 pre-reform income and G1 immigrant status.

One potential concern is that the reform may have reduced stigma for grandfathers who become employed thanks to the reform but would have been unemployed without the reform. This could increase the well-being of the grandfather and may increase the quality of the time investment made by the grandfather to the grandchildren (function z in Equation (1)). In Figure A.14, we show how the reform affected mental health (ITT) and conclude that the reform significantly decreased the use of mental health drugs, both in the short- and long-run. In Table A.16 we show the incidence of mental health drug use, which suggests that the positive effects on mental health can be considered sizable. *Ceteris paribus*, we would expect that the quality of the time investment made by the grandfather to the grandchildren would increase because of this. Even with this increased quality of grandfathers' time investments, we find that grandchildren's development is better after the reform when grandfather's care is replaced by formal childcare.

4.3 Potential *indirect* channel: via G2 mothers

Section 3 illustrates that, in addition to formal care (T_t^F), informal child care provision of grandfathers (η_t^{G1}) may also be substituted by care provided by the mother (η_t^{G2}). In this case her labor supply is likely to decline.

To investigate this empirically, we estimate the effect of the reform on (G2) mothers' labor supply as follows:

$$h_{t,j}^{G2} = \gamma_0 + D_{t,i}P_{t,j}\gamma_1 + D_{t,i}\gamma_2 + P_{t,j}\gamma_3 + \mathbf{X}_{t,j}^{G1}\gamma_4 + \mathbf{X}_{t,j}^{G2}\gamma_5 + \pi_j + \varepsilon_{t,j} \tag{10}$$

[33] Maternal labor force participation is also higher in urban areas than in rural areas (Polman and van der Elst, 2008), therefore the need to replace a reduction in informal care by formal child care may also be higher in urban areas.

where $h_{t,j}^{G2}$ measures the labor supply of the mother j, as measured in terms of both extensive (employment status) and intensive (full time factor) margins as well as by income from paid employment. Here, γ_1 measures the ITT effect of less attractive UI benefits for G1 on G2's labor supply. We also include mother's j fixed effect π_j to control for unobserved time invariant characteristics.

Second, when grandfathers are more often employed (and hence informal care provided by the grandfather is expected to be lower) this might change a mothers' fertility decision, whereby mothers might decide to delay fertility up to the grandfathers' retirement (Haan and Wrohlich, 2010; Bauernschuster et al., 2016; Bick, 2016; Adda et al., 2013). For simplicity, we assumed fertility to be exogenous in the theoretical framework, but we will study fertility empirically, by estimating the ITT effect of the reform on mothers' fertility choices

$$k_{a,j}^{G2} = \theta_0 + D_i P_j \theta_1 + D_i \theta_2 + P_j \theta_3 + \mathbf{X}_j^{G1} \theta_4 + \mathbf{X}_j^{G2} \theta_5 + \phi_j \tag{11}$$

where $k_{a,j}$ is the fertility decision of mother j, measured by the total number of children and the mother's age at first birth.

5 Data

We use administrative panel data from Statistics Netherlands which contain all Dutch registered inhabitants. In these data, individuals can be linked to their parents based on a unique identification number.[34] This allows us to construct a linked dataset containing information on three subsequent generations: grandparents (G1), parents (G2), and grandchildren (G3). Our sample consist of grandchildren whose grandfathers were between the ages of 50 and 63 years in the period around the reform, i.e. 1999-2007.[35] Figure A.8 provides an overview of the timeline of the data. Information on demographic characteristics includes gender, age, number of children, marital status, and country of origin. Furthermore, we merge reg-

[34] Access to these data can be obtained via a remote access facility after a confidentiality agreement with Statistics Netherlands has been signed.

[35] This is the same age range as considered in Been and Knoef (2017).

ister based social statistics, e.g. from the tax office and from the national institution of educational testing, gathered by Statistics Netherlands.

We define the labor market status of the grandfathers by their main income source. Information on the various income sources is available from 1999 to 2016, and includes earnings from paid employment, income from self-employment, pensions, and income from various social security and welfare schemes. Based on the amount of income from either source, we define the main source of income, and determine whether an individual is employed (that is, in paid employment or self-employed). Only, for the few cases where the income from pension is higher than the income from employment, we still consider these individuals as active in the labor market. This is motivated by the fact that every individual who is employed for the past 26 of the 36 weeks is entitled to UI benefits (regardless of their amount of income).[36]

Our outcome of interest is the test score of grandchildren (G3) on a standardized CITO test in 6^{th} grade at the end of primary school. The CITO score is an important input for decisions related to the level of the secondary school, and proxies students' ability. We have test score information for overall test performance, as well as for language and math performance separately. The test scores are standardized with mean zero and standard deviation one within cohorts, in the full population of grandchildren.

5.1 Descriptive statistics

Grandfathers - G1: We focus here on the labor force participation of grandfathers as for the cohorts under consideration only a minority of women was participating in the labor market and often working only on a part-time basis. Because of shorter employment histories they have reduced (or even none) UI coverage and are less affected by the reform. The descriptive statistics for women are presented in Appendix Table A.1. In addition, we show in Appendix D that grandparents are often spending time together with their grandchildren,

[36]In addition, we classify individuals as inactive when they report zero income. Because this is administrative data, not having a source of income implies being out of the labor market. Robustness for excluding these individuals (2%) do not change the results. These results are available upon request.

so the activation of grandfathers by the reform might lead to a reduction in informal care provision by grandmothers as well. In the remainder of this study we will therefore refer to the decreased informal care provided by both grandparents.

Table 1 provides descriptive statistics for the grandfathers (G1), separately for the control group (aged 50-57.5 years old) and the treatment group (57.5-63 years old), in the control period (2000-2003) and the treatment period (2004-2010). More than 90% of the grandfathers have a partner, and grandfathers in the treatment group have on average more grandchildren compared to grandfathers in the control group, probably due to their higher age. Considering labor market status, we find that the incidence of paid employment and self employment is higher in the control group than in the treatment group. Between the control and the treatment period labor market activity of the control group declined somewhat, while we find a small increase in the treatment group. Unemployment doubled in the control group before and after 2004, whereas it decreased in the treatment group, indicating an effect of the 2004 UI reform.

Mothers - G2: Moving one generation forward, Table 2 provides descriptive statistics for G2 mothers (whose youngest child is between 0 and 4) by grandfather's treatment and control status. Virtually all mothers (94-97%) have a partner. The most notable differences between control and treatment group is that the average number of children is slightly lower in the control group as compared to the treatment group. Similarly, mothers in the treatment group are, on average, slightly older at first birth.[37] Both employment, wage, and the fulltime factor[38] are somewhat higher in the treatment period compared to the control period.

Grandchildren - G3: The distribution of the test scores at 6^{th} grade by gender reveals that girls outperform boys in language, and boys outperform girls in math. The gender difference is masked in the total score, containing both language and math scores (Figure A.9 in the Appendix). In Table 3, we provide descriptive statistics for G3 grandchildren by G1's

[37]We empirically test for any fertility effects of the reform in the remainder of the paper.

[38]The fulltime factor takes values from 0 to 1 and should be interpreted as the higher the value, the closer to a full time job the mother is.

treatment and control period.[39] Comparing the mean of the test score (total, language and math) in the treatment period and treatment group, we can already find suggestive evidence for a positive effect of the unemployment insurance reform. In addition to the standardized test at age 12 (6^{th} grade) we also look at alternative outcomes such as G3 non-cognitive skills at age 12. We distinguish between internalizing and externalizing behavior of grandchildren, which are often used as proxy for non-cognitive skills of children.[40] We use information on individual drugs prescription included in the Dutch administrative data (ATC4 code) to proxy for the incidence of internalizing and externalizing behavior of grandchildren.[41]

In our sample 50% of grandchildren is female, the average year test is 2013 and 2017 for the control and treatment period, respectively. Table 3 also shows descriptives for grandchild birth order and the number of G1's children, as well as G1's level of income, the grandchild living area (urban or rural)[42] and the distance in kilometers between the G1 and G3 living areas.[43]

6 Results

This section describes our main results. First, we replicate the results obtained by Been and Knoef (2017) on our sample of grandfathers and confirm that the reform increased the paid employment of older individuals at the expenses of unemployment benefits and inactivity (first stage of the IV). Second, we estimate the spillover effect of the reform on grandchildren's

[39] Because the main analysis will be based on grandfathers, G3's treatment status is based on grandfather's treatment

[40] See for example Moroni et al. (2019); Attanasio et al. (2020a,b).

[41] These two traits of non-cognitive skills are increasingly used in the child development literature, see for example Moroni et al. (2019); Attanasio et al. (2020a,b). A child with internalizing problems is typically withdrawn, anxious or depressed and more likely to be diagnosed of with Autism Spectrum Disorder (ASD). A child with externalizing problems is often disruptive, hyperactive and aggressive and more likely to develop Attention Deficit Hyperactivity Disorder (ADHD).

[42] Grandchild living area is defined on the basis of the municipality of grandchildren before the reform. Urban areas have population higher or equal than 50000. Rural areas have population lower than 50000. If the child is born after the reform we consider the mother's municipality before the reform.

[43] Distance is measured as Euclidean distance (equivalent to 2/3 of the geographical distance) between the 4-digit-postcodes of grandchildren and grandfathers before the reform. If the child is born after the reform we consider the mother's postcode before the reform. Figure A.10 shows that the majority of grandchildren live within 5km distance from their grandfathers with an average distance of 21.3km.

educational outcomes, both as an Intention-to-Treat (ITT) and as the second stage of the IV-framework. Third, we explore potential *direct* and *indirect* mechanisms that can drive this relationship.

6.1 Labor market participation of grandfathers

Table 4, Column 1 presents the estimated impact of the reform on grandfather's labor market participation (Equation 8). Our results show that the reform has significantly increased the labor supply of G1 grandfathers.[44] In particular, extending search requirements in UI to individuals older than 57.5 has increased labor supply of older males by 5.1%-points.[45] This effect is both economically and statistically significant. The accompanied F-statistic on the excluded instrument is 632 which is bigger than the critical value of 104.7 as argued by Lee et al. (2021). Hence, the estimation results suggest that the instrument is sufficiently strong and *relevant*. For women we do not find an effect (Appendix Table A.4, Column (1)), which aligns with earlier findings by Hullegie and Van Ours (2014), Lammers et al. (2013), and Been and Knoef (2017).

6.2 Intergenerational spillover on grandchildren

The reduced form estimation results in Column (2) of Table 4, show that the reform has increased G3's test score by 1.9% of a standard deviation. This coefficient should be interpreted as the Intention-to-Treat (ITT), i.e. the average treatment effect of the reform on grandchildren's outcomes regardless of whether the reform has actually triggered a change in the activity status of grandfathers.

[44]Our model is a simplified version of the model in Been and Knoef (2017). In Tables A.2 and A.3 in the Appendix we show the corresponding results when using OLS, FE or multinomial logit models in the full sample of grandfathers. Results from multinomial logit models show that, as a result of the reform, the probability of being in paid employment or self-employed increases, whereas the probability of being on unemployment benefits decreases as compared of being inactive (where inactive in this case includes retired or receiving other benefits, e.g. disability benefits.)

[45]In Table A.5 in the Appendix we show that the reform also increased grandfathers' income by about 1.3 percent. On the other hand, employed grandfathers tend to reduce the hours of work as shown in Column (4) by the effect of the reform on grandfathers full time factor.

Column (3) presents the IV results, where we exploit the exogenous variation in grandfather's labor market status triggered by the unemployment insurance reform. Having a working grandfather increases G3's total test score by 37.1% of a standard deviation.[46] In Table A.6 in the Appendix we present the naive OLS estimates assuming that there is no endogeneity issue related to grandparent's activity. For grandfathers, we find that labor market activation increases G3's total test score by 12.6% of a standard deviation. This estimated effect exceeds the ITT-estimate but is smaller than the IV-estimate which suggests that endogeneity is likely an issue and biases the estimates downward. Hence, it is most likely that less affluent grandfathers have stopped working and invest in their grandchildren.

Taking the IV-estimates as our baseline results, we find no differential effect between performance in language and math, nor between grandchild's gender (Appendix Figure A.11).[47] The effects that we find are sizable, but in line with the evidence provided for early childhood programs (means tested), such as the US *Head Start* program. For example, Elango et al. (2016) provide a thorough comparison of different early childhood programs and find that their effect range from 0.39 to 0.54 of a standard deviation on achievement test score at age 5-10.

There are several explanations for the relatively large effect of grandfather's activity on grandchildren's test scores. First, our results have to be interpreted as local average treatment effect, i.e. the effect on the grandchildren whose grandfathers changed their activity status as a response to the reform. If compliers' characteristics are different from those of the total population then our result reflects the effect for a selective part of the population. Recently, Blandhol et al. (2022) have argued that IV-estimates cannot be interpreted as LATE per se. We investigate this further in Section 7.3 and conclude that our results reflect the local nature of the estimates.

An alternative explanation for the relatively large effects is the 'skills beget skills' prin-

[46]The F-statistic (= 632.446) indicates that the instrument, i.e. the 2004 UI reform, is a strong instrument of grandfather's activity.

[47]The benchmark results do not distinguish between maternal and paternal grandfathers. When we estimate separate models between the two, we find that the results are larger for paternal grandfathers as compared to maternal grandfathers, although with overlapping confidence intervals (Table A.7).

ciple (Cunha and Heckman, 2007). Being exposed to a grandfather who is employed since child's birth (i.e. the time when we measure the grandfather's treatment status) might have cumulative effects over time. We therefore explore whether G3-age at treatment impacts the results. We estimate the benchmark model in Equation (9) by considering the grandfathers' treatment at different grandchild ages, ranging from birth (benchmark model) up to age 5.[48] Figure 1 shows that the impact of grandfather's treatment varies by the age of the grandchild.[49] The effect becomes statistically insignificant from age 3 onwards, which suggests that the effect of having a grandfather active in the labor market can accumulate over the years but has the most significant impact in the first three years of life.[50] Note that compulsory schooling starts at 4 years old in the Netherlands, which means that after the age of 4 the informal child care provided by the grandfathers is likely to become less important.

Third, and related to the second point, is the accumulation of effects caused by the persistence of formal child care. Using ancillary data from the LISS panel, we show that the persistence of using formal child care from one year to the following is quite substantial (0.62-0.75, depending on the specification; Table C.2 in the Appendix). This means that there is a high state-dependence in the choice for formal child care: choosing formal child care in the current period substantially increases the probability of using formal child care in the following years. Such persistence is much smaller when looking at informal child care, especially when provided by grandfathers (see Table C.3 in the Appendix). This implies that if parents start using formal child care since birth (e.g. due to the lower availability of grandparents), they are more likely to continue using it in the years thereafter.

[48]For example, considering the treatment status when grandchild is one year old instead of at birth, implies to include in the treated both grandchildren that were already treated since birth and grandchildren that start to be treated when they turn one. These grandchildren are older when treated and less exposed to the treatment.

[49]Note however, that because the standardized test is taken at age 12 for every grandchild we cannot disentangle exposure from age effects.

[50]The corresponding estimation results are shown in Table A.9 in the Appendix, with the IV results shown in Panel A, ITT results shown in Panel B and OLS results shown in Panel C. The pattern of the ITT results fully confirms the pattern in the IV results. The decreasing impact of grandfather's activity on grandchildren outcome does not appear in the OLS results. This is reassuring as, on the one hand it shows the bias reflected in the OLS estimates, on the other hand it shows that the reform has triggered an exogenous change in the grandfathers' activity status which however has decreasing effects over grandchild's age.

Fourth, the large effect that we find on grandchildren's educational outcomes may be partly explained by the effect of the reform on alternative grandchildren's outcomes such as grandchildren's behavior - outcomes who are often found to be cross-productive (Cunha and Heckman, 2008) during child development. However, we show in Table A.8 in the Appendix that the reform did not have any effect on grandchildren's behavioral problems as measured by internalizing and externalizing behavior at age 12.[51]

Finally, the expansion in child care subsidies in the period 2005-2009 (with increasing female labor supply as the primary objective) has most likely increased the quality of formal child care as well (Akgunduz et al., 2015). Increased child care subsidies may also have caused increased child care use, but not differently for those with grandfathers younger or older than 57.5 years (the cutoff used in the identification strategy). Although the expansion in child care subsidies does not render our methods and causal inference invalid, our results should be interpreted against this background in which substitution between informal child care by grandfathers and formal (subsidized) child care became easier for parents over time.

Overall, the positive spillover effect suggests that if the increased labor supply of grandfathers induces a reduction of the informal child care provided to their grandchildren, this reduction is likely compensated by higher quality alternative child care for the grandchildren (either formal care or care provided by the child's parents (G2)). This will be further explored in the next section.

6.2.1 *Direct* channel

This section explores whether the increased labor market participation of grandfathers has a direct effect on their grandchildren's educational performance at the end of primary school via reduced time investments of grandfathers in their grandchildren.

Figure 2 shows that the intergenerational impact on G3 school performance is larger for grandchildren whose grandfathers have fewer G2 children (Panel (a)) and among first born

[51]We also estimate the effect of the reform on grandchildren's behavioral problems as measured by internalizing and externalizing behavior at age 7 and also find no effect. The results of these estimates are available upon request.

grandchildren (Panel (b)). Potential time available for informal care investments is likely larger for these grandchildren than for those born in larger families or for grandchildren of higher birth order. Hence, the positive spillover is concentrated among the former group as grandfathers were likely to invest more time in these grandchildren in absence of the reform.

To corroborate the direct channel further, Figure 3, Panel (a) shows that the positive effect of the reform on G3's test scores is relatively large in urban areas. Moreover, G3's test scores in urban areas increased most when the grandfather was living close to the grandchild and was thus more likely to spent more time with the grandchild (Figure 3, Panel (b)).[52] As formal care use is relatively high among grandchildren in urban areas and affected by the reform (Figure A.12 in the Appendix) this suggests that the positive effect on G3's test scores may be driven by a replacement of grandparental informal care by formal care.[53]

Figure 4 considers the potential impact of the quality of grandfathers' investments (relative to that of formal care) and shows that the positive impact on G3 school performance is concentrated among families with low socio-economic stats (SES) grandfathers. These results are consistent with existing evidence on the heterogeneous impact of child care on children outcomes, where the return to child care is shown to be higher for children coming from a more disadvantaged background (e.g. Cornelissen et al. 2018; Felfe and Lalive 2018)[54], suggesting that replacing the informal care by high quality formal care benefits the grandchild.

Finally, although our results are triggered by a change in grandfathers' labor market activation - which in turn reduces their informal child care provision - we provide evidence that the change in the labor market activity of grandfathers affects not only his own allocation of time but also grandmothers' allocation of time. In Section D in the Appendix we use the LISS panel data (2007-2019) and its time use and consumption module to estimate the effect

[52]Distance is measured using the 4-digit-postcode of both grandfathers and grandchildren prior to the reform, therefore avoiding the distance to be an outcome of the reform itself.

[53]We don't find such effects for grandchildren residing in rural areas, which aligns with the fact that the effect of the reform is concentrated in urban areas. See results in Figure A.13 of the Appendix.

[54]Depending on the institutional context, i.e. the availability and quality of formal child care, the effect of child care on children outcomes may also be negative as shown by Fort et al. (2020) who finds that in Italy parental child care is of a better quality than the formal one.

of retirement on time use, in particular on time spent in work, leisure and helping other family members, e.g. child care. We consider the couples' joint time use decision by estimating a Simultaneous Equation Model (SEM) to infer a causal effect of retirement induced by the statutory retirement age of 65 in a RDD framework.

Results reported in Table D.1 in the Appendix show evident cross-effects of retirement: the retirement of either the male or the female in the couple increases the leisure time of both the individual who retires and his/her partner, implying that couples value joint leisure. In addition, we find that retirement only significantly increases time spent with grandchildren for males and that male retirement also increases female partner's time spent with grandchildren. Hence, grandfathers are most likely to spend time with their grandchildren together with grandmothers. As such, these results provide evidence that the multigenerational effect that we find can be ascribed to the change in the informal child care provided by both grandfathers and grandmothers, although only triggered by grandfathers' labor market activation.[55]

All in all, our investigation of the direct mechanisms suggest a substitution between formal and informal child care whereby we observe larger effects among grandchildren who require and/or receive the largest investments from their grandfathers, and for those who likely benefit most from high quality formal child care.

6.2.2 *Indirect* channel

So far we have interpreted the positive spillover effect as a result of the reduction in grandfathers' informal child care likely being replaced by formal child care. Alternatively, the reduction in grandfathers' informal care induced by the reform could also impact care provided by the G2-mother, and as such *indirectly* impact children's outcomes. We study this potential indirect channel by investigating changes in G2-mothers' choices regarding their labor supply and fertility.

[55]Note also that, because females increase the time with their grandchildren when the male partners retire (Table D.1 in the Appendix), we can exclude the possibility that grandmothers are spending more time with their grandchildren when grandfathers are working; so there is no evidence of competition in grandmothers' time between grandfathers and grandchildren.

Table 5 shows the ITT of the reform on mothers' (whose youngest child is between 0 and 4) outcomes regarding their labor supply and fertility. We focus on various indicators for mother's labor supply - employment incidence (dummy) (Column 1), working full time (index) (Column 2), and log annual earnings (Column 3) - and find that mothers do not reduce their labor supply as a response to the reform.[56] In Appendix Table A.18 we show that it is primarily lower educated mothers that increase labor supply, which corroborates the idea that children from families with a lower socioeconomic status benefit relatively more from formal child care.[57]

Maternal fertility decisions are proxied by the number of children (Column 4) and the age at which they first gave birth (Column 5). We find that the reform only has a very small negative impact on the number of children. Stronger effects are found for mothers postponing fertility decisions, i.e. mothers tend to delay their first child as a response to the reform by about 6 months. Note that the positive effect on maternal employment can partly be the result of the negative effects on fertility.[58]

All in all, our results are consistent with the existing evidence that Dutch female labor supply is typically not responsive to grandparental possibilities to provide informal care (Rellstab et al., 2020). In line with Havnes and Mogstad (2011), there appears to be substitution between formal care and informal child care provided by grandfathers, without affecting mothers' labor supply. This is also consistent with the result that mothers do not decrease their labor supply following increases in the costs of formal child care (Bettendorf et al., 2015; Been and Jongen, 2017). Our findings that the change in grandfathers' labor market activity mostly affects the timing of G2-mother's fertility rather than total fertility are in line with earlier work by Eibich and Siedler (2020).

By comparing the findings for the direct and indirect mechanisms, our results suggest

[56] In line with the labor supply literature we find no decreases in labor supply for fathers (whose youngest child is between age 0 and 4). See Table A.17.

[57] We have to be cautious in interpreting results based on mothers' education as this information is only available for a selective sample as shown in Appendix Table A.19. This is the reason why we don't use this information in our main analysis.

[58] In our robustness section (Section 7, Table A.24) we rule out that the effect is driven by those G3 conceived after the reform, which suggests that our findings are not driven by selection in fertility.

that the inputs from formal child care into child development are of better quality than the inputs from grandfathers, on average. First, the indirect mechanism indicates that mothers do not reduce their labor supply to provide higher child care. Hence, mothers replace the grandparental time input by formal child care. Second, this substitution leads to better outcomes for grandchildren as measured by the CITO-score at age 12, on average. This interpretation is strengthened by the heterogeneous effects by SES.

7 Robustness checks

7.1 Placebo tests

In this section we provide evidence on the robustness of our results by performing placebo tests to (i) rule out differential group trends and (ii) rule out the presence of other time shocks.

To test for the presence of differential group trends we assign two different placebo thresholds for the grandfather's age, while keeping the actual treatment period (2004-2007). First, we restrict the analysis to grandchildren whose grandfathers were younger than 56, and assign grandchildren whose grandfathers were 50-52 to the untreated group and grandchildren whose grandfathers were 53-55 to the treated group. Reassuringly, results reported in Table A.20 in the Appendix indicate that neither the ITT results (Column 1), nor the IV results (Column 3) show any significant effect. Unsurprisingly, the OLS results (Column 2) - which show the effect of having an employed grandfather on grandchildren educational outcomes - are unchanged, as they remain endogenous in this test. In this placebo test none of the grandfathers were actually treated. Second, we restrict the analysis to grandchildren whose grandfathers were older than 58 - therefore all belonging to the treatment group according to the UI reform - and use placebo treatment group whereby grandchildren whose grandfathers were 59-60 are assigned to the untreated group and grandchildren whose grandfathers were 61-63 to the treated group. Results of this test are presented in Table A.21 in the Appendix and once again show that there is no effect when looking at the ITT (Column 1) or when

estimating the IV model (Column 3). Taken together, these placebo tests indicate that there are no differential group trends that affect our results, but that the age threshold that matters is the UI reform threshold of age 57.5. If there had been other shocks that affected younger or older group differentially, the estimated effect of the placebo tests would be been significantly different from zero.

To test for the presence of other time shocks we assign two different placebo thresholds for the treatment period, while keeping the actual treatment group (57.5+ years old). First, we restrict the analysis to grandchildren born prior to 2004, so none of the grandfathers were actually treated. We consider grandchild births in 2000-2001 as the untreated period and those in 2002-2003 as the treated period. Results reported in Table A.23 in the Appendix show no effect, neither when looking at the ITT results (Column 1), nor when looking at the IV results (Column 3). Similarly to the previous tests, the OLS results remain unchanged. Alternatively, we restrict the analysis to grandchildren born after 2003 - therefore all belonging to the treatment period according to the UI reform - and consider grandchild births in 2004-2005 as the untreated period and those in 2006-2007 as the treated period. Results of this test are presented in Table A.22 in the Appendix and show some significant effect of the placebo reform (Column 1), which however is fully driven by second born children (Column 2), due to the within-family spillover effect of the reform across siblings and/or cousins. Once we restrict the analysis on first births only (Column 3), the effect of the placebo reform completely disappears as expected, both in the ITT (Column 3) and in the IV model estimates (Column 5). It is important to point out that in the benchmark model the effects are larger among first born (see Figure 2, Panel (b)), so we would have had larger effects in the placebo test as well if the placebo effects were not fully driven by the spillover effect coming from cousins and siblings. Taken together, the lack of significant effects from the placebo tests show that there were no other time shocks which differentially impacted the treatment and control group that can explain the difference between treated and untreated grandchildren apart from the UI reform that we study here.

7.2 Selective fertility

In Table 5 we documented that one of the indirect mechanisms at play is through delayed fertility by G2 mothers. Although the delay of about 6 months is relatively small (column 5), selection into fertility can still influence our findings, e.g. when only affluent mothers' decide to give birth after the reform and this correlates with the G3-outcomes later on in life. In Table A.24 in the Appendix we explore whether selection into fertility influences our results. In order to show this we use the month of conception, to compare the benchmark ITT with the ITT of a model where the treatment period is replaced by a dummy taking value one if the child has been conceived after the reform and zero otherwise. We also restrict the sample of grandchildren born between 2003-2005 in order to focus around the date of the reform to be able to isolate the month of conception from the month of birth. Column (1) shows the results of the benchmark ITT in the restricted sample. Column (2) shows the results considering the month of conception to define the treatment period. The ITT when using the month of conception is not statistically significant, therefore implying that the effect that we find is fully driven by grandchildren who have been conceived before the reform and ended up being treated after the reform, therefore indicating that positive selection into fertility does not drive our results.

7.3 Compliers analysis

Our benchmark results (Table 4) indicated that having an employed grandfather at the grandchild's birth increases the grandchild's standardized test score by 37% of a standard deviation. This result should be interpreted as local average treatment effect. In this section we describe the compliers' characteristics and compare them to the characteristics of the total population.

We characterize the compliers by looking at different subgroups of the population, defined by the grandfather's pre-reform characteristics such as having a partner, immigrant status, number of children, income, and living area.[59] Table A.25 in the Appendix shows that there

[59]In our case the fraction of compliers represents 5.1% of the sample. However, among the treated group,

are some statistically significant difference between the compliers and the total sample. Compliers are more likely to have a partner, less likely to be immigrant, more likely to have few children, less likely to have high income and less likely to live in rural area. Considering the results in Section 6.2.1, we found that the effect of the reform is larger among grandchildren whose grandfathers have fewer children (Figure 2, Panel (a)), grandchildren living in urban areas (Figure 3, Panel (a)), and grandchildren whose grandfathers have lower income (Figure 4, Panel (a)), characteristics that are more likely to be found among the compliers than in the total sample. Therefore we confirm that the large effect that we estimate reflects the local nature of our results. Keeping this in mind, the external validity of our results may not be compromised given that – although statistically different – the means of characteristics between the compliers and the total population are very similar in magnitude.

8 Conclusions

This paper studies the effect of grandfathers' labor market activation on grandchildren's educational outcomes. We make two important contributions to the growing literature on intergenerational persistence of socioeconomic outcomes as well as on the literature on the intergenerational spillovers of welfare reforms. First, we are the first to use quasi-experimental methods to study causal effects of activation policies on three subsequent generations in population-wide administrative data from the Netherlands. Second, we provide detailed analyses of a potential direct (i.e. substitution between informal grandparental care and formal child care) and indirect mechanism (i.e. substitution between parental care and formal child care) driving our results which adds to the scarce literature on the impact of informal child care provided by the grandfathers on grandchildren's outcomes.

To infer causal estimates of the effect of grandfathers' labor supply on grandchildren's outcomes, we use an IV-approach exploiting a reform in unemployment insurances targeted at older workers. The reform introduced mandatory job search requirements for unemployment

compliers make up 18.5% of the sample. The size of compliers among treated is estimated as the first stage times the fraction of people who are treated and divided by the fraction of people that are employed.

of 57.5 years and older on the 1st of January 2004. Prior studies have shown that the reform met its goal in increasing labor supply of older workers and men, especially (Hullegie and Van Ours, 2014; Lammers et al., 2013; Been and Knoef, 2017). This is confirmed by our study. Since the effect is especially strong among grandfathers, our analyses and conclusions are largely based on grandfathers. Grandmothers are less responsive to the reform, work less prior to the reform, and already provide substantial amount of care for the grandchildren during working life.

We find substantial positive effects of grandfathers' labor supply on grandchildren's educational outcomes, measured by test scores at the age of 12. Our Intention-to-Treat (ITT) estimates suggest that grandfathers' treatment by the reform increases the standardized test score by 1.9% of a standard deviation. In an IV-framework, this implies an effect as large as about 37% of a standard deviation which is economically large. This large effect is likely due to the local nature of the results as well as the year-to-year accumulation from birth to age 12 when the test is taken. We find that grandchildren aged 0-4 around the reform are most substantially affected, consistent with the idea that children are most responsive to the quality of childcare in this age range. The effect is not driven by reduced maternal labor supply (and, hence, not by replacing grandparental care by parental care) nor by the selection of women choosing to have children (i.e. fertility). We find that this effect is primarily driven by grandfathers who are most likely to have spent more time with their grandchildren. Therefore, our results suggest that informal child care is primarily substituted by formal child care with formal child care being a higher-quality investment in children. This is especially true for children from families with a lower socioeconomic background. Our results are robust to many sensitivity checks and heterogeneity analyses. The conclusion adds to the literature on the positive effect of formal child care on child development (Currie, 2001; Cornelissen et al., 2018; Felfe and Lalive, 2018; Fryer et al., 2020). The sign and size of the effect may depend strongly on the institutional setting, including the (financial) availability and quality of formal child care.

Note that our conclusions are valid for grandfathers who otherwise (without the reform) would have used unemployment as an early retirement pathway. Although using social

insurances as an exit route was highly common in the Netherlands until the early 2000s, the question arises to what extent our results also hold for the labor force participation of grandparents (including grandmothers) in general. To get more insights on the external validity of our results, it would be interesting to investigate other activation reforms (e.g. in disability insurance or early retirement schemes) for future research.

A policy response based on these results could be to stimulate the use of formal child care, because of the potential positive effects on both maternal labor supply and child investments especially for children from families with a lower socioeconomic status. This may also have important long-run effects as stimulating the use of formal child care among families with a lower socioeconomic status increases the equality of opportunity of children. Especially, as children from disadvantaged families are less likely to use formal child care even though they might benefit the most (Cornelissen et al., 2018). Regardless of the quality argument, our results suggest the existence of important multigenerational spillover effects to grandchildren of reforms aimed at grandparents. Such effects should be taken into account in designing future social policy as the costs and benefits of social policy may exceed the targeted population.

Table 1: G1 Men - Descriptive Statistics: Household and individual characteristics by treatment

	(1999-2003) Control Period				(2004-2007) Treatment Period			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	Age 50-57.5 Control Group		Age 57.5-63 Treatment Group		Age 50-57.5 Control Group		Age 57.5-63 Treatment Group	
Age	53.803	2.048	59.890	1.836	54.599	1.844	60.288	1.842
Year of birth	1946.986	2.309	1941.307	2.193	1950.546	2.033	1945.177	2.071
Immigrant	0.135	0.342	0.141	0.348	0.158	0.364	0.129	0.335
Partner	0.921	0.269	0.922	0.267	0.905	0.293	0.915	0.278
Children in the hh	0.472	0.499	0.236	0.424	0.412	0.492	0.196	0.397
Number of children	2.112	1.014	2.241	1.130	2.174	1.061	2.110	1.016
Number of grandchildren	4.743	3.194	5.159	3.718	4.954	3.362	4.740	3.224
Paid employment	0.698	0.459	0.417	0.493	0.668	0.471	0.428	0.495
Self-employment	0.086	0.280	0.073	0.259	0.082	0.274	0.069	0.253
Unemployment benefits	0.012	0.109	0.048	0.213	0.024	0.152	0.041	0.198
Inactivity	0.204	0.403	0.463	0.499	0.227	0.419	0.462	0.499
Observations	1233516		1093603		416134		1041550	

Source: SSB Dutch administrative data

Notes: The table shows the mean and standard deviation of all the demographic and labor information for grandfathers by treatment and control period (1999-2003 and 2004-2007) and treatment and control group (50-57.5 and 57.5-63) (*generation 1*). The sample includes information between 1999-2007 as detailed in Figure A.8 and individuals between 50-63 years old. Because the sample considers only grandfathers, the number of grandchildren reported in the table is the average number of grandchildren among individuals with at least one grandchild.

Table 2: G2 Women - Descriptive Statistics: G2 characteristics - By treatment of G2's fathers.

	(2000-2003) Control Period				(2004-2007) Treatment Period			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	Age 50-57.5 Control Group		Age 57.5-63 Treatment Group		Age 50-57.5 Control Group		Age 57.5-63 Treatment Group	
Mother age	28.811	3.023	31.961	3.045	28.789	3.188	32.420	3.181
Year of birth	1972.244	3.307	1969.149	3.274	1976.347	3.373	1972.835	3.322
Immigrant	0.127	0.333	0.114	0.318	0.139	0.346	0.100	0.300
Partner	0.955	0.208	0.970	0.170	0.943	0.233	0.967	0.179
Number of children	1.696	0.774	1.916	0.875	1.747	0.783	1.950	0.855
Mother's age at first birth	26.139	3.177	28.503	3.386	25.836	3.244	28.698	3.404
Employment	0.706	0.455	0.724	0.447	0.748	0.434	0.777	0.416
Wage	17012.199	14727.487	20004.601	14211.913	18212.488	12065.199	22655.059	15262.449
Full-time factor	0.548	0.279	0.566	0.265	0.556	0.263	0.585	0.246
Observations	505449		725136		261133		519873	

Source: SSB Dutch administrative data

Notes: The table shows the mean and standard deviation of all the demographic and labor information for mothers (*generation 2*) whose youngest child is between 0 and 4 by treatment and control period (1999-2003 and 2004-2007) and treatment and control group (50-57.5 and 57.5-63) of the mothers' father (*generation 1*). The sample includes information between 1999-2007 as detailed in Figure A.8. Because the sample considers only mothers, the number of children reported in the table is the average number of children among women with at least one child.

Table 3: G3 - Descriptive Statistics: Final sample - By treatment of G3's grandfathers.

	(1999-2003) Control Period				(2004-2007) Treatment Period			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	Age 50-57.5 Control Group		Age 57.5-63 Treatment Group		Age 50-57.5 Control Group		Age 57.5-63 Treatment Group	
Total	-0.073	0.987	0.071	0.964	0.025	0.954	0.170	0.918
Language	-0.060	0.991	0.072	0.969	0.011	0.968	0.148	0.930
Math	-0.077	0.999	0.055	0.975	0.003	0.970	0.141	0.938
Externalizing behavior	0.060	0.238	0.051	0.220	0.058	0.235	0.051	0.220
Internalizing behavior	0.001	0.035	0.001	0.035	0.001	0.035	0.001	0.031
Female	0.502	0.500	0.499	0.500	0.510	0.500	0.507	0.500
Year test	2013	1.449	2013	1.451	2017	1.174	2017	1.204
First birth	0.399	0.490	0.224	0.417	0.393	0.488	0.238	0.426
Few G1 children	0.670	0.470	0.615	0.487	0.664	0.473	0.681	0.466
G1 High income	0.746	0.435	0.674	0.469	0.788	0.409	0.778	0.416
Urban	0.441	0.496	0.428	0.495	0.460	0.498	0.482	0.500
G1 - G3 distance below 2 km	0.386	0.487	0.343	0.475	0.385	0.487	0.321	0.467
G1 - G3 distance 2-10 km	0.299	0.458	0.285	0.451	0.290	0.454	0.271	0.444
G1 - G3 distance 10+ km	0.316	0.465	0.372	0.483	0.325	0.468	0.409	0.492
Observations	265089		359387		91513		162488	

Source: SSB Dutch administrative data.

Notes: The table shows the mean and standard deviation of all the demographic information for grandchildren (*generation 3*) by treatment and control period (1999-2003 and 2004-2007) and treatment and control group (50-57.5 and 57.5-63) of the grandfathers (*generation 1*). First birth is a dummy variable taking value 1 if the grandchild is the first G1's grandchild and zero otherwise. Few G1 children is a dummy variable taking value 1 if G1 has up to two children and zero if G1 has more than two children. G1 High income is a dummy variable taking value one if the G1's income before the reform is above the median income and zero otherwise. Urban is a dummy variable that takes value one if the grandchild lives in urban areas (municipality with a population above 50000) before the reform. G1-G3 distance measures the kilometer distance between grandchild and grandfather before the reform. The sample includes information between 1999-2007 as detailed in Figure A.8 at grandchild's birth.

Table 4: The impact of G1 employment on G3 educational outcomes

	Grandfather		
	(1)	(2)	(3)
	First stage	ITT	IV
Treat Group	0.021***	0.002	-0.006
	(0.002)	(0.004)	(0.005)
Treat Period	0.002	0.514***	0.513***
	(0.003)	(0.008)	(0.008)
Treat Group × Treat Period	0.051***	0.019***	
	(0.002)	(0.005)	
Employed			0.371***
			(0.094)
Observations	878837	878837	878837
R squared	0.187	0.027	
Endogeneity test			0.000
F-stat			632.446

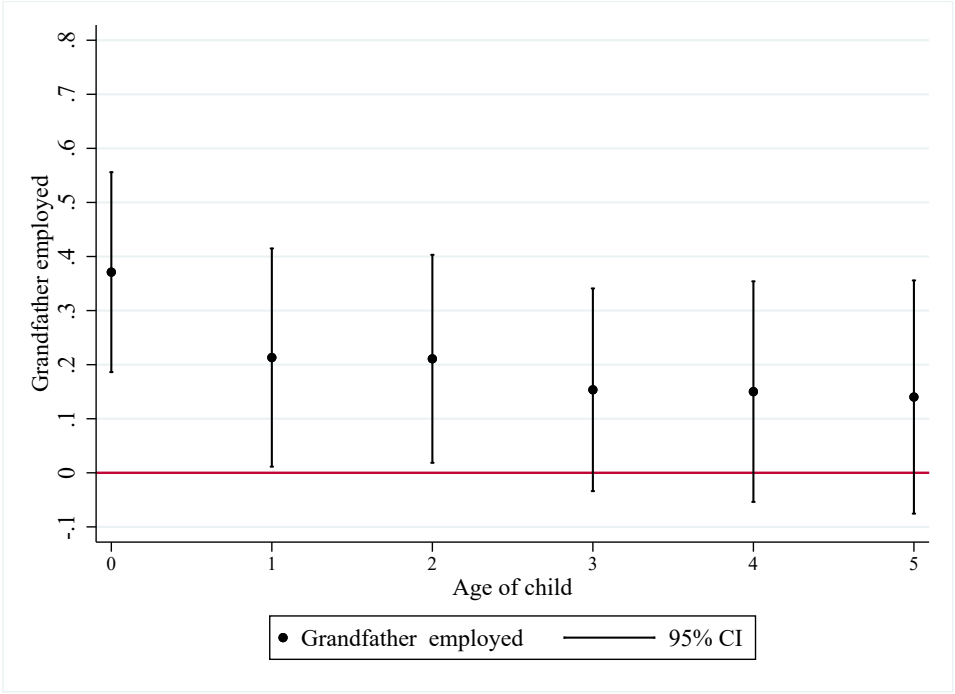
Source: SSB Dutch administrative data.
Notes: The dependent variable is grandfather's employment in column (1) and grandchildren's standardized total test score at age 12 in columns (2) and (3). Column (2) shows the ITT results. Column (3) shows estimates of Equation (9) for grandfathers. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors at grandchild's level in parenthesis.

Table 5: The impact of the reform on G2-mother's labor supply and fertility

	(1)	(2)	(3)	(4)	(5)
	Employment	Full time	Earnings	Num Children	Age at 1st birth
Treat group	-0.005***	-0.005***	-0.018***	0.031***	-0.028
	(0.001)	(0.001)	(0.003)	(0.001)	(0.022)
Treat Period (2004-2007)	-0.018***	-0.015***	-0.063***	0.063***	-0.288***
	(0.001)	(0.001)	(0.003)	(0.001)	(0.018)
Treat Group x Treat Period	0.013***	0.006***	0.031***	-0.030***	0.524***
	(0.002)	(0.001)	(0.004)	(0.002)	(0.026)
Observations	1197901	732507	907390	3379472	259770
R squared	0.017	0.065	0.029	0.545	0.284

Source: SSB Dutch administrative data.
Notes: The dependent variables are G2-mothers' labor supply when the youngest child is between age 0 and 4 (Column 1-3) and fertility (Column 4-5) decisions. Employment (Column 1) is a dummy variable that indicates whether the mothers is on paid employment. Full time (Column 2) is a full time index going from zero to one, where one indicate a full time job. Earnings (Column 3) are income from paid employment measured in logarithm. Number of children (Column 4) counts the total fertility. Mother's age at first birth (Column 5) is expressed in years. The table shows the estimates of Equation (10) and (11) considering the treatment status of her G1 father. All estimates control for grandparent's background characteristics and mother's background characteristics (i.e. age and age squared (except for the last column), immigrant status, having a partner, and the number of G2's children (only in column 1-3)) and mother's fixed effects. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: The impact of G1 employment on G3 educational outcomes - By grandchildren's age at treatment

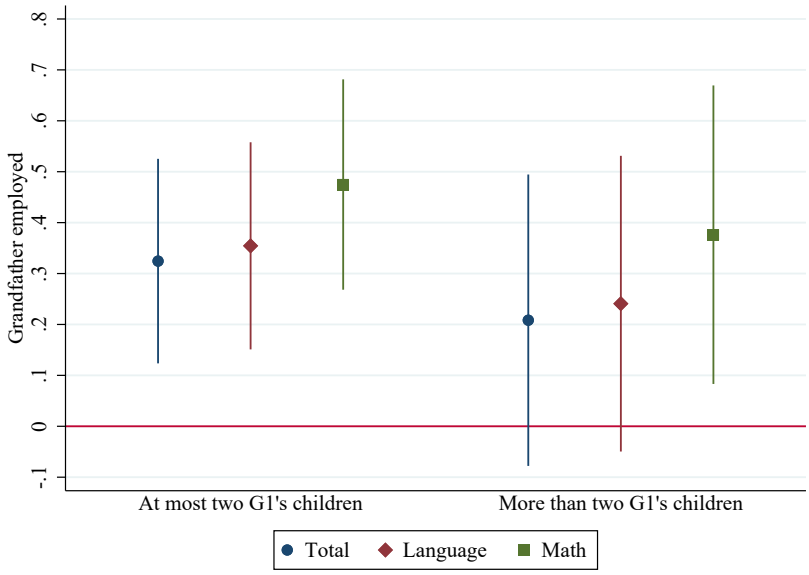


Source: SSB, Dutch Administrative Data

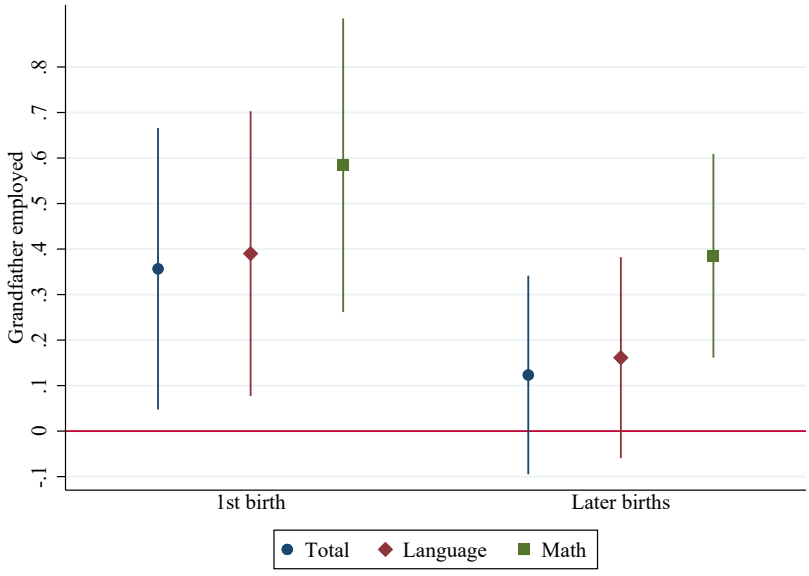
Notes: The dependent variable is grandchildren's standardized test score at age 12. The figure shows the estimates of Equation (9) by grandchildren's age at treatment from birth till age 5. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Corresponding full set of results are reported in Table A.9.

Figure 2: The impact of G1 employment on G3 educational outcomes

(a) By number of G1's children



(b) By birth order

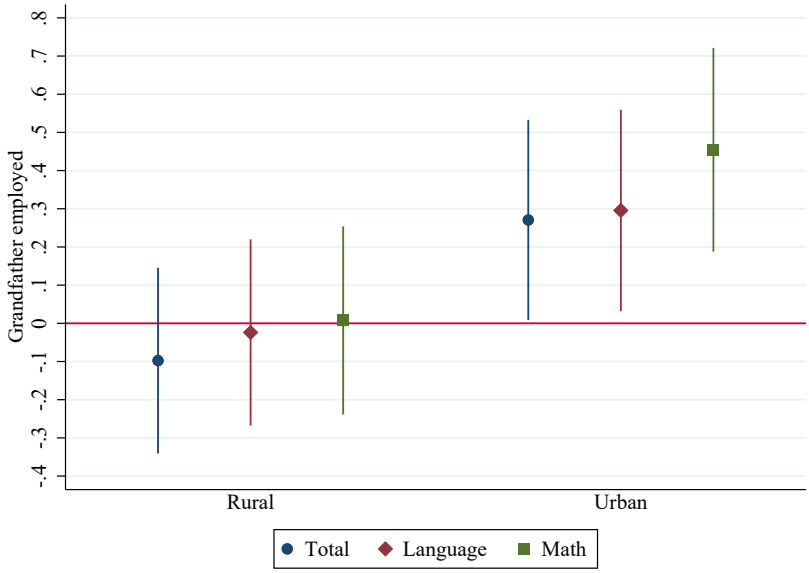


Source: SSB, Dutch Administrative Data.

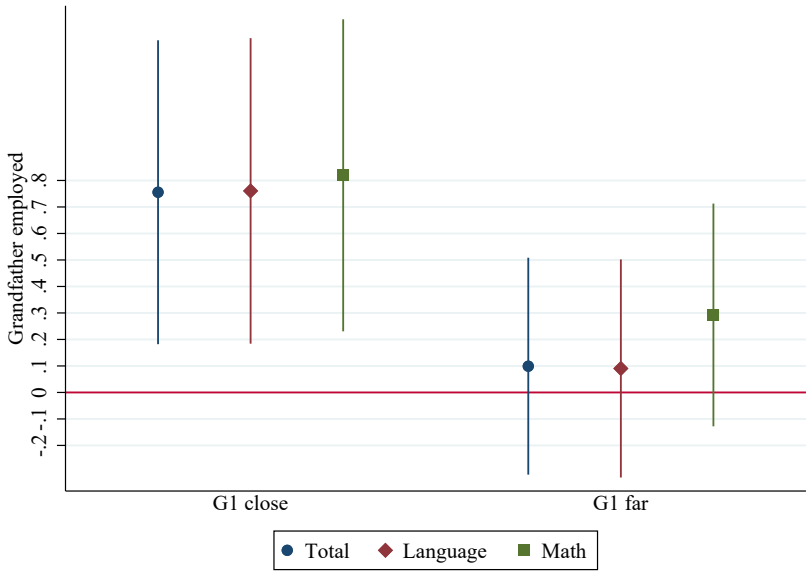
Notes: The dependent variable is grandchildren's standardized test score at age 12. The figure shows the estimates of Equation (9) for different subjects, i.e. total score, language and math and by the number of G1's children in Panel (a) and by grandchild's birth order in Panel (b). All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). See Table A.10 and A.11 in the Appendix for the full set of results including ITT and OLS estimates.

Figure 3: The impact of G1 employment on G3 outcomes - By rural or urban area

(a) By rural or urban area



(b) By distance between grandfathers and grandchildren in urban area

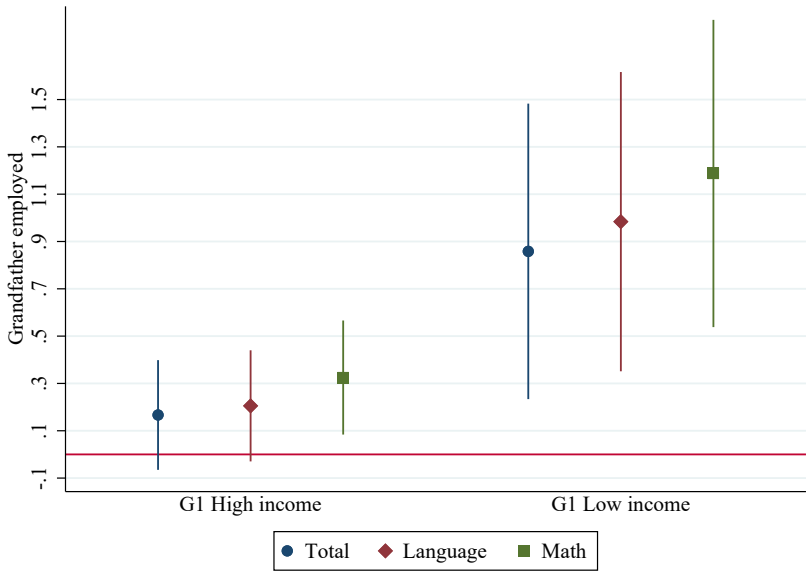


Source: SSB, Dutch Administrative Data.

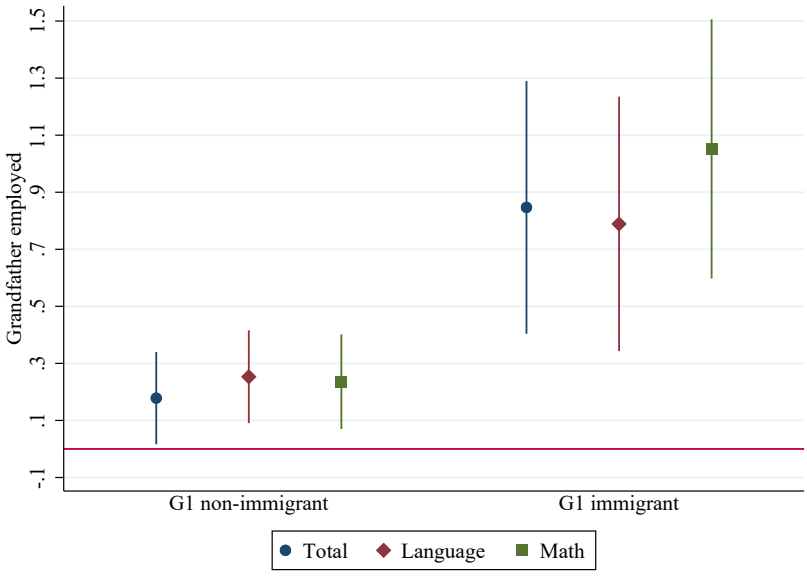
Notes: The dependent variable is grandchildren's standardized test score at age 12. Panel (a) of the figure shows the estimates of Equation (9) for different subjects, i.e. total score, language and math and by the type of living area of grandchildren. Panel (b) of the figure shows the estimates of Equation (9) for different subjects, i.e. total score, language and math and by distance between grandfathers and grandchildren within urban areas. If grandchildren and grandfathers live within 2 km distance they are considered to live close; if they live more than 10 km away they are considered to live far. Observations of individuals living between 2-10 km are excluded to be able to compare the distance more clearly. The distribution of distance is shown in Figure A.10 in the Appendix. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). See Table A.12 and A.13 in the Appendix for the full set of results including ITT and OLS estimates.

Figure 4: The impact of G1 employment on G3 educational outcomes

(a) By Pre-reform G1's level of income



(b) By G1's immigrant status



Source: SSB, Dutch Administrative Data.

Notes: The dependent variable is grandchildren's standardized test score at age 12. The figure shows the estimates of Equation (9) for different subjects, i.e. total score, language and math and by pre-reform G1's level of income in Panel (a) and by G1's immigrant status in Panel (b). All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). See Table A.14 and A.15 in the Appendix for the full set of results including ITT and OLS estimates.

References

Adda, J., Dustmann, C., and Stevens, K. (2013). The career costs of children. *Journal of Political Economy*, 125(2):293–337.

Adermon, A., Lindahl, M., and Palme, M. (2021). Dynastic human capital, inequality, and intergenerational mobility. *American Economic Review*, 111(5):1523–48.

Adermon, A., Lindahl, M., and Waldenstrom, D. (2018). Intergenerational wealth mobility and the role of inheritance: Evidence from multiple generations. *Economic Journal*, 128(612):F482–F513.

Aguiar, M., Hurst, E., and Karabarbounis, L. (2013). Time use during the Great Recession. *American Economic Review*, 103(5):1664–1696.

Akgunduz, E., Jongen, E., Leseman, P. P., Plantenga, J., et al. (2015). Quasi-experimental evidence on the relation between child care subsidies and child care quality. Technical report, CPB Netherlands Bureau for Economic Policy Analysis.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics*. Princeton university press.

Atav, T., Jongen, E. L., and Rabate, S. (2021). Increasing the effective retirement age: Key factors and interaction effects. *IZA Discussion Paper 14150*.

Attanasio, O., Blundell, R., Conti, G., and Mason, G. (2020a). Inequality in socio-emotional skills: A cross-cohort comparison. *Journal of Public Economics*, 191:104171.

Attanasio, O., De Paula, Á., and Toppeta, A. (2020b). The persistence of socio-emotional skills: Life cycle and intergenerational evidence. *NBER Working Paper 27823*.

Barone, G. and Mocetti, S. (2021). Intergenerational mobility in the very long run: Florence 1427–2011. *The Review of Economic Studies*, 88(4):1863–1891.

Barschkett, M., Spiess, C. K., and Ziege, E. (2021). Does grandparenting pay off for the next generations? Intergenerational effects of grandparental care. *IZA DP No. 14795*.

Bauernschuster, S., Hener, T., and Rainder, H. (2016). Children of a (policy) revolution: the introduction of universal child care and its effect on fertility. *Journal of the European Economic Association*, 14(4):975–1005.

Been, J. and Goudswaard, K. (2021). Intertemporal and intratemporal consumption smoothing at retirement: Micro evidence from detailed spending and time use data. *Journal of Pension Economics and Finance*, DOI: https://doi.org/10.1017/S1474747221000330.

Been, J. and Jongen, E. L. (2017). The effect of child care subsidies on the time allocation and well-being of parents. Technical report.

Been, J. and Knoef, M. (2017). Job-search requirements for unemployed at the end of working life effects on unemployment dynamics and self-employment probabilities. *Journal of Human Resources*, 52(2):491–530.

Bettendorf, L. J., Jongen, E. L., and Muller, P. (2015). Childcare subsidies and labour supply — evidence from a large Dutch reform. *Labour Economics*, 36:112–123.

Bick, A. (2016). Time use during the Great Recession. *Journal of the European Economic Association*, 14(3):639–668.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, 120(2):669–700.

Blandhol, C., Bonney, J., Mogstad, M., and Torgovitsky, A. (2022). When is tsls actually late? *NBER Working Paper 29709*.

Blau, F. and Robins, P. (1988). Child-care costs and family labor supply. *The Review of Economics and Statistics*, 70:374–381.

Braun, S. T. and Stuhler, J. (2018). The transmission of inequality across multiple generations: testing recent theories with evidence from Germany. *The Economic Journal*, 128(609):576–611.

Bütikofer, A., Dalla-Zuanna, A., Salvanes, K. G., et al. (2018). Breaking the links: Natural resource booms and intergenerational mobility. *NHH Dept. of Economics Discussion Paper*, 19.

Carneiro, P., García, I. L., Salvanes, K. G., and Tominey, E. (2021). Intergenerational mobility and the timing of parental income. *Journal of Political Economy*, 129(3):757–788.

CBS (2011). Welzijnswerk en kinderopvang; personeel en productie, 2001-2008. https://opendata.cbs.nl/statline/CBS/nl/dataset/71470ned/table?ts=1623748313491.

CBS (2021). Arbeidsdeelname en werkloosheid per maand. https://opendata.cbs.nl/statline/CBS/nl/dataset/80590ned/table?ts=1635350141349.

Clark, G. and Cummins, N. (2015). Intergenerational wealth mobility in England, 1858–2012: surnames and social mobility. *The Economic Journal*, 125(582):61–85.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2018). Who benefits from universal child care? Estimating marginal returns to early child care attendance. *Journal of Political Economy*, 126(6):2356–2409.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2):31–47.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782.

Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives*, 15(2):213–238.

Dahl, G. B. and Gielen, A. C. (2021). Intergenerational spillovers in disability insurance. *American Economic Journal - Applied Economics*, 13(2):116–50.

Dahl, G. B., Løken, K. V., and Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, 104(7):2049–74.

De Haan, M. and Schreiner, R. C. (2018). The intergenerational transmission of welfare dependency. *CESifo Working Paper Series*.

Del Boca, D., Flinn, C., and Wiswall, M. (2014). Household choices and child development. *Review of Economic Studies*, 81(1):137–185.

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

Duggan, M., Singleton, P., and Song, J. (2007). Aching to retire? The rise in in the full retirement age and its impact on social security disability rolls. *Journal of Public Economics*, 91:1327–1350.

Eibich, P. and Siedler, T. (2020). Retirement, intergenerational time transfers, and fertility. *European Economic Review*, 124:103392.

Elango, S., García, J. L., Heckman, J. J., and Hojman, A. (2016). *4. Early Childhood Education*. University of Chicago Press.

Felfe, C. and Lalive, R. (2018). Does early child care affect children's development? *Journal of Public Economics*, 159:33–53.

Ferrie, J., Massey, C., and Rothbaum, J. (2021). Do grandparents matter? Multigenerational mobility in the United States, 1940–2015. *Journal of Labor Economics*, 39(3):000–000.

Fort, M., Ichino, A., and Zanella, G. (2020). Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families. *Journal of Political Economy*, 128(1):158–205.

Fryer, R. G., Levitt, S. D., List, J. A., and Samek, A. (2020). Introducing cogx: A new preschool education program combining parent and child interventions. NBER working paper27913.

Gørtz, M., Sander, S., and Sevilla, A. (2020). Does the child penalty strike twice. *CEBI Working paper series, 30/20*.

Grübl, J., Lackner, M., and Winter-Ebmer, R. (2020). Intergenerational transmission of unemployment: Causal evidence from Austria. *IZA Discussion Paper No. 13068*.

Haan, P. and Wrohlich, K. (2010). Can child care policy encourage employment and fertility?: Evidence from a structural model. *Labour Economics*, 18:498–512.

Hartley, R. P., Lamarche, C., and Ziliak, J. P. (2017). Welfare reform and the intergenerational transmission of dependence. *Journal of Political Economy, forthcoming*.

Havnes, T. and Mogstad, M. (2011). Money for nothing? Universal child care and maternal employment. *Journal of Public Economics*, 95(11-12):1455–1465.

Hullegie, P. and Van Ours, J. (2014). Seek and ye shall find: How search requirements affect job finding rates of older workers. *De Economist*, 162(4):377–395.

Kalwij, A., Kapteyn, A., and De Vos, K. (2018). Work capacity at older ages in the Netherlands. *NBER Working Paper, No. 24636.*

Kim, K. and Lee, M.-j. (2019). Difference in differences in reverse. *Empirical Economics*, 57(3):705–725.

Knoef, M., Been, J., Alessie, R., Caminada, K., Goudswaard, K., and Kalwij, A. (2016). Measuring retirement savings adequacy: A multi-pillar approach in the Netherlands. *Journal of Pension Economics and Finance*, 15(1):55–89.

Lammers, M., Bloemen, H., and Hochguertel, S. (2013). Job search requirements for older unemployed: Transitions to employment, early retirement and disability benefits. *European Economic Review*, 58:31–57.

Lee, D., McCrary, J., Moreira, M., and Porter, J. (2021). Valid t-ratio inference for iv. *NBER Working Paper 29124*.

Lei, L. (2019). Childcare choice and long-run human capital outcomes in China. *Available at SSRN 3474026*.

Lindahl, M., Palme, M., Massih, S. S., and Sjögren, A. (2015). Long-term intergenerational persistence of human capital an empirical analysis of four generations. *Journal of Human Resources*, 50(1):1–33.

Long, J. and Ferrie, J. (2018). Grandfathers matter (ed): occupational mobility across three generations in the US and Britain, 1850–1911. *The Economic Journal*, 128(612):F422–F445.

Moroni, G., Nicoletti, C., and Tominey, E. (2019). Child socio-emotional skills: The role of parental inputs. *HCEO Working paper 2019-038*.

Nagore Garcia, A., Rossi, M., and Van Soest, A. (2021). Retirement of the self-employed in the Netherlands. *Small Business Economics*, 56(1):385–402.

Noailly, J. and Visser, S. (2009). The impact of market forces on child care provision: Insights from the 2005 child care act in the Netherlands. *Journal of Social Policy*, 38(3):477–498.

OECD (2011). Pensions at a glance 2011: Retirement-income systems in OECD and G20 countries. OECD publishing, www.oecd.org/els/social/pensions/PAG.

OECD (2019). Benefits and wages: Net replacement rates in unemployment (edition 2019), oecd social and welfare statistics (database). https://doi.org/10.1787/68e3dbd7-en.

OECD (2020). Education at a glance 2020. OECD publishing, https://www.oecd.org/education/education-at-a-glance.

Petrongolo, B. (2009). The long-term effects of job search requirements: Evidence from the UK JSA reform. *Journal of Public Economics*, 93:1234–1253.

Polman, N. and van der Elst, M. (2008). Verschillen tussen stad en platteland in arbeidsparticipatie van vrouwen. Een analyse van ondernemerschap en werken in loondienst. *Rapport 4.08.01; ISBN/EAN: 978-90-8615-206-3*.

Price, J. (2008). Parent-child quality time does birth order matter? *Journal of Human Resources*, 43(1):240–265.

Rellstab, S., Bakx, P., García-Gómez, P., and Van Doorslaer, E. (2020). The kids are alright-labour market effects of unexpected parental hospitalisations in the Netherlands. *Journal of Health Economics*, 69:102275.

Ruhm, C. (2004). Parental employment and child cognitive development. *The Journal of Human Resources*, 39:155–192.

Rupert, P. and Zanella, G. (2018). Grandchildren and their grandparents' labor supply. *Journal of Public Economics*, 159:89–103.

Slot, P., Jepma, I., Muller, P., Romijn, B., and Leseman, P. (2018). Quality of Dutch daycare, toddler care, out-of-school care and childminder care: Combined measurements 2017-2018.

Stancanelli, E. and Van Soest, A. (2012a). Joint leisure before and after retirement: A double regression discontinuity approach. *IZA Discussion Paper, No. 6698*.

Stancanelli, E. and Van Soest, A. (2012b). Retirement and home production: A regression discontinuity approach. *American Economic Review*, 102(2):600–605.

Staubli, S. and Zweimüller, J. (2013). Does raising the early retirement age increase employment of older workers? *Journal of Public Economics*, 108:17–32.

Truskinovsky, Y. (2020). The kids are alright - labour market effects of unexpected parental hospitalisations in the Netherlands. *Journal of Health Economics*, 69(102275).

Truskinovsky, Y. (2021). The unintended consequences of informal child care subsidies for older women's retirement security. *Journal of Human Resources*, doi: 10.3368/jhr.59.1.0818-9715R2.

Van Elk, R., Van Der Steeg, M., and Webbink, D. (2011). Does the timing of tracking affect higher education completion? *Economics of Education Review*, 30(5):1009–1021.

Table A.1: G1 Women - Descriptive Statistics: Household and individual characteristics by treatment

	(1999-2003) Control Period				(2004-2007) Treatment Period			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	Age 50-57.5 Control Group		Age 57.5-63 Treatment Group		Age 50-57.5 Control Group		Age 57.5-63 Treatment Group	
Age	53.611	2.074	59.839	1.837	54.425	1.897	60.157	1.848
Year of birth	1947.254	2.342	1941.390	2.210	1950.773	2.069	1945.353	2.084
Immigrant	0.136	0.343	0.132	0.338	0.164	0.370	0.126	0.332
Partner	0.866	0.340	0.835	0.371	0.853	0.355	0.832	0.374
Children in the hh	0.386	0.487	0.179	0.383	0.333	0.471	0.144	0.351
Number of children	2.113	1.030	2.293	1.156	2.172	1.082	2.124	1.034
Number of grandchildren	4.760	3.306	5.309	3.824	4.954	3.462	4.790	3.328
Paid employment	0.388	0.487	0.188	0.390	0.418	0.493	0.225	0.417
Self-employment	0.053	0.225	0.036	0.186	0.046	0.209	0.035	0.183
Unemployment benefits	0.010	0.100	0.018	0.133	0.017	0.131	0.019	0.135
Inactivity	0.549	0.498	0.759	0.428	0.519	0.500	0.722	0.448
Observations	1522609		1103908		620913		1176993	

Source: SSB Dutch administrative data

Notes: The table shows the mean and standard deviation of all the demographic and labor information for grandmothers by treatment and control period (1999-2003 and 2004-2007) and treatment and control group (50-57.5 and 57.5-63) (*generation 1*). The sample includes information between 1999-2007 as detailed in Figure A.8 and individuals between 50-63 years old. Because the sample considers only grandparents, the number of grandchildren reported in the table is the average number of grandchildren among individuals with at least one grandchild.

Table A.2: The impact of the reform on G1 employment status - Full sample - By gender

	OLS		FE	
	(1)	(2)	(3)	(4)
	Men	Women	Men	Women
Treat Group	0.017***	0.004***	0.044***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)
Treat Period	-0.005***	0.051***	-0.018***	-0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
Treat Group x Treat Period	0.056***	0.001	0.004***	0.001
	(0.002)	(0.002)	(0.001)	(0.001)
Observations	3599117	4281544	3599117	4281544
R squared	0.178	0.078	0.187	0.072

Source: SSB Dutch administrative data.

Notes: The table shows the estimates by gender for the impact of the reform on grandparents' employment status using OLS and fixed effect. Here we consider the full sample of grandparents whose labor market status can be observed over time from 1999 to 2007. This is different from the final sample reported in Table 4 and Table A.4 where treatment status for the grandparents (G1) is measured in the year G3 was born. Additional controls include grandparents' characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis.

Table A.3: The impact of the reform on G1 employment status - By gender - Multinomial Logit - Reference category Inactive

	(1) Men	(2) Women
Paid employment		
Treat Group	0.100***	0.139***
	(0.011)	(0.011)
Treat Period	-0.171***	-0.088***
	(0.013)	(0.012)
Treat Group x Treat Period	0.064***	-0.017
	(0.012)	(0.011)
Self-employed		
Treat Group	-0.018	-0.047**
	(0.019)	(0.023)
Treat Period	0.034	0.062**
	(0.023)	(0.027)
Treat Group x Treat Period	0.117***	0.022
	(0.020)	(0.024)
Unemployment benefits		
Treat Group	0.796***	0.633***
	(0.023)	(0.026)
Treat Period	0.537***	0.519***
	(0.027)	(0.029)
Treat Group x Treat Period	-0.526***	-0.464***
	(0.022)	(0.025)
Observations	3069621	3664641

Source: SSB Dutch administrative data.

Notes: The table shows the estimates by gender for the impact of the reform on grandparents' employment status defined as employed, self-employed, on unemployment benefits or inactive. The reference category is inactive. The effect is estimated using multinomial logit. Here we consider the full sample of grandparents whose labor market status can be observed over time from 1999 to 2007. This is different from the final sample reported in Table 4 and Table A.4 where treatment status for the grandparents (G1) is measured in the year G3 was born. Additional controls include grandparents' characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children) and lagged employment status. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis.

Table A.4: The impact of G1 employment on G3 educational outcomes

	Grandmother		
	(1)	(2)	(3)
	First Stage	ITT	IV
Treat Group	0.003*	0.003	0.060
	(0.002)	(0.004)	(0.126)
Treat Period	0.039***	0.509***	1.218
	(0.003)	(0.007)	(1.639)
Treat Group × Treat Period	-0.001	0.016***	
	(0.002)	(0.004)	
Employed			-18.339
			(42.958)
Observations	1050988	1050988	1050988
R squared	0.078	0.026	
Endogeneity test			0.000
F-stat			0.187

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandmother's employment in column (1) and grandchildren's standardized total test score at age 12 in columns (2) and (3). Column (2) shows the ITT results. Column (3) shows estimates of Equation (9) for grandmothers. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors at grandchild's level in parenthesis.

Table A.5: The impact of the reform on G1 labor market outcomes of grandfathers.

	Income (ln)		Full time	
	(1)	(2)	(3)	(4)
	OLS	FE	OLS	FE
Treat Group	-0.000	0.020***	0.038***	0.058***
	(0.001)	(0.001)	(0.001)	(0.001)
Treat Period	-0.029***	-0.025***	-0.008***	-0.005***
	(0.002)	(0.001)	(0.001)	(0.001)
Treat Group x Treat Period	0.071***	0.013***	0.011***	-0.011***
	(0.002)	(0.001)	(0.001)	(0.001)
Observations	3521990	3521990	1593188	1593188
R squared	0.036	0.003	0.118	0.158

Source: SSB Dutch administrative data.

Notes: The table shows the estimates for the impact of the reform on grandfathers income expressed in logarithm (Columns 1 and 2) and full time factor (Columns 3 and 4). Columns (1) and (3) estimates OLS models and control for grandparents' characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children). Columns (2) and (4) additionally control for individual fixed effect to control for potential time invariant unobserved heterogeneity. Here we consider the full sample of grandfathers whose labor market status and income can be observed over time from 1999 to 2007. This is different from the final sample reported in Table 4 where treatment status for the grandfathers (G1) is measured in the year G3 was born. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis.

Table A.6: The impact of G1 employment on G3 outcomes - grandfathers and grandmothers - OLS results

	(1) Grandfather	(2) Grandmother
Treat Group	0.003	0.007**
	(0.004)	(0.003)
Treat Period	0.522***	0.514***
	(0.007)	(0.007)
Employed	0.125***	0.101***
	(0.002)	(0.002
Observations	878837	1050988
R squared	0.027	0.028

Source: SSB Dutch administrative data.

Notes: The table shows the corresponding OLS estimates for the impact of having a grandparent active in the labor market (Corresponding IV estimates shown in Table 4 and A.4 respectively for grandfathers and grandmothers.). Additional controls include grandparents' characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children) and grandchildren's characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Cluster standard errors at grandchild's level in parenthesis.

Table A.7: The impact of G1 employment on G3 outcomes - grandfathers from mother and from fathers

	Grandfathers from fathers			Grandfathers from mothers		
	(1)	(2)	(3)	(4)	(5)	(6)
	ITT	OLS	IV	ITT	OLS	IV
Treat Group	-0.006	-0.003	-0.017**	0.008	0.008	0.003
	(0.006)	(0.006)	(0.008)	(0.005)	(0.005)	(0.007)
Treat Period	0.518***	0.529***	0.514***	0.511***	0.515***	0.512***
	(0.010)	(0.009)	(0.011)	(0.009)	(0.008)	(0.009)
Treat Group x Treat Period	0.023***			0.014**		
	(0.007)			(0.006)		
Employed		0.126***	0.514***		0.124***	0.255**
		(0.003)	(0.159)		(0.003)	(0.110)
Observations	389979	389979	389979	488858	488858	488858
R squared	0.027	0.030		0.027	0.030	
Endogeneity test			0.013			0.234
F-stat			211.887			423.707

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. The table shows the estimates of Equation (9) separately for grandfathers from fathers (Column 3) and grandfathers from mothers (Column 6). The correspondent ITT and OLS results are shown respectively in Column (1 and 2) and (4 and 5). All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis.

Table A.8: The impact of the reform on G3 behavioral problems at age 12.

	(1) Externalizing	(2) Internalizing	(3) Intern or Extern
Treat Group	-0.003***	-0.000	-0.003***
	(0.001)	(0.000)	(0.001)
Treat Period	-0.028***	0.000	-0.028***
	(0.002)	(0.000)	(0.002)
Treat Group x Treat Period	0.001	-0.000	0.001
	(0.001)	(0.000)	(0.001)
Observations	878837	878837	878837
R squared	0.015	0.000	0.015

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's internalizing and externalizing problems at age 12. The table shows the estimates of the corresponding ITT of Equation (9) for a different outcome. We use information on individual drugs prescription included in the Dutch administrative data to proxy for the incidence of internalizing and externalizing behavior of grandchildren. The ATC4 code reported in the data is used to identify the prescription code for medication for ADHD as a proxy for externalizing behavior, and the prescription of antidepressant as a proxy for internalizing behavior. The incidence of internalizing problems is 0.1% and of externalizing problems is 5.4%. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors at grandchild's level in parenthesis.

Table A.9: The impact of G1 employment on G3 outcomes - Exposure and Age effect

	(1) At birth	(2) Age=1	(3) Age=2	(4) Age=3	(5) Age=4	(6) Age=5
Panel A: IV						
Treat Group	-0.006	-0.001	0.000	-0.001	0.004	0.002
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Treat Period	0.513***	0.494***	0.474***	0.451***	0.428***	0.408***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
Employed	0.371***	0.213**	0.211**	0.154	0.150	0.140
	(0.094)	(0.103)	(0.098)	(0.096)	(0.104)	(0.110)
Observations	878837	868474	867231	806510	715751	615405
Endogeneity test	0.000	0.000	0.000	0.000	0.000	0.000
F-stat	632.446	523.231	575.179	598.008	499.562	440.945
Panel B: ITT						
Treat Group	0.002	0.002	0.004	0.001	0.005	0.003
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)
Treat Period	0.514***	0.494***	0.473***	0.450***	0.427***	0.407***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)
Treat Group x Treat Period	0.019***	0.010**	0.010**	0.008	0.008	0.008
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Observations	878837	868474	867231	806510	715751	615405
R-squared	0.027	0.031	0.029	0.032	0.034	0.035
Panel C: OLS						
Treat group	0.003	0.001	0.004	0.001	0.005	0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Treat Period	0.522***	0.497***	0.477***	0.452***	0.429***	0.409***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
Employed	0.125***	0.125***	0.122***	0.119***	0.121***	0.122***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	878837	868474	867231	806510	715751	615405
R-squared	0.030	0.035	0.032	0.035	0.037	0.038

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized test score at age 12. The table shows the estimates of Equation (9) by grandchildren's age at treatment from birth till age 5 (corresponding Figure 1). Panel A shows the IV estimates. Panel B shows the ITT estimates. Panel C shows the OLS estimates. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Cluster standard errors at grandchild's level in parenthesis.

Table A.10: The impact of G1 employment on G3 outcomes - Number of G1's children

	(1) ITT	(2) OLS	(3) IV
Treated x At most two G1's children	0.018***		
	(0.006)		
Treated x More than two G1's children	0.008		
	(0.008)		
Employed x At most two G1's children		0.087***	0.324***
		(0.003)	(0.102)
Employed x More than two G1's children		0.193***	0.208
		(0.004)	(0.146)
Treat Group x At most two G1's children	0.027***	0.019***	0.036**
	(0.004)	(0.004)	(0.016)
Treat Group x More than two G1's children	-0.048***	-0.034***	-0.084***
	(0.005)	(0.005)	(0.028)
Treat Period x At most two G1's children	0.495***	0.504***	0.496***
	(0.008)	(0.007)	(0.008)
Treat Period x More than two G1's children	0.555***	0.549***	0.553***
	(0.010)	(0.008)	(0.012)
At most two G1's children	0.006	0.080***	-0.095
	(0.004)	(0.005)	(0.116)
Observations	878837	878837	878837
R squared	0.027	0.031	
Endogeneity test			0.056
F-stat			198.489

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. The table shows the estimates of Equation (9) for grandfathers, with heterogeneous effect by G1's number of children (Column 3) (corresponding to Figure 2, Panel (a)). The correspondent ITT results and OLS are shown respectively in Columns (1) and (2). All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Cluster standard errors at grandchild's level in parenthesis.

Table A.11: The impact of G1 employment on G3 outcomes - by birth order

	(1) ITT	(2) OLS	(3) IV
Treated x 1st Birth	0.018**		
	(0.008)		
Treated x Later births	0.005		
	(0.006)		
Employed x 1st Birth		0.096***	0.357**
		(0.004)	(0.158)
Employed x Later births		0.126***	0.123
		(0.003)	(0.111)
Treat Group x 1st Birth	0.019***	0.016***	0.065*
	(0.005)	(0.005)	(0.034)
Treat Group x Later births	0.001	-0.000	-0.021
	(0.004)	(0.004)	(0.013)
Treat Period x 1st Birth	0.480***	0.488***	0.479***
	(0.009)	(0.008)	(0.010)
Treat Period x Later births	0.532***	0.531***	0.533***
	(0.008)	(0.007)	(0.009)
Later births=1	-0.179***	-0.195***	0.018
	(0.004)	(0.006)	(0.137)
Observations	878837	878837	878837
R squared	0.033	0.036	
Endogeneity test			0.231
F-stat			274.047

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. The table shows the estimates of Equation (9) for grandfathers, with heterogeneous effect by G3's gender (Column 3) (corresponding to Figure 2, Panel (b)). The correspondent ITT results and OLS are shown respectively in Columns (1) and (2). All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Cluster standard errors at grandchild's level in parenthesis.

Table A.12: The impact of G1 employment on G3 outcomes - By G3 living area

	(1) ITT	(2) OLS	(3) IV
Treated x Rural	-0.007		
	(0.007)		
Treated x Urban	0.019**		
	(0.008)		
Employed x Rural		0.091***	-0.093
		(0.003)	(0.124)
Employed x Urban		0.164***	0.307**
		(0.004)	(0.133)
Treat Group x Rural	-0.024***	-0.038***	-0.077***
	(0.005)	(0.004)	(0.020)
Treat Group x Urban	0.046***	0.056***	0.110***
	(0.005)	(0.005)	(0.026)
Treat Period x Rural	0.527***	0.521***	0.529***
	(0.009)	(0.008)	(0.009)
Treat Period x Urban	0.547***	0.551***	0.540***
	(0.009)	(0.008)	(0.011)
Rural=1	0.007	0.063***	0.318***
	(0.004)	(0.005)	(0.121)
Observations	837546	837546	837546
R squared	0.029	0.032	
Endogeneity test			0.100
F-stat			265.417

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. The table shows the estimates of Equation (9) for grandfathers, with heterogeneous effect by G3's living area (Column 3) (corresponding to Figure 3, Panel (a)). The correspondent ITT results and OLS are shown respectively in Columns (1) and (2). All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Cluster standard errors at grandchild's level in parenthesis.

Table A.13: The impact of G1 employment on G3 outcomes - By distance between grandchildren and grandfathers in urban areas

	(1) ITT	(2) OLS	(3) IV
Treated x G1 ≤ 2km	0.037**		
	(0.015)		
Treated x G1 > 10km	-0.001		
	(0.011)		
Employed x G1 ≤2km		0.166***	0.756**
		(0.007)	(0.311)
Employed x G1 > 10km		0.121***	0.099
		(0.005)	(0.200)
Treat Group x G1 ≤2km	-0.051***	-0.039***	0.057
	(0.010)	(0.009)	(0.049)
Treat Group x G1 > 10km	0.042***	0.034***	-0.031
	(0.008)	(0.008)	(0.032)
Treat Period x G1 ≤2km	0.539***	0.552***	0.515***
	(0.016)	(0.013)	(0.022)
Treat Period x G1 >10km	0.514***	0.510***	0.517***
	(0.015)	(0.012)	(0.013)
G1 ≤2km	-0.334***	-0.361***	-0.795***
	(0.007)	(0.009)	(0.212)
Observations	256902	256902	256902
R squared	0.084	0.088	
Endogeneity test			0.090
F-stat			40.395

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. The table shows the estimates of Equation (9) for grandfathers and by distance between grandfathers and grandchildren within urban areas (corresponding to Figure 3, Panel (b)). If grandchildren and grandfathers live within 2 km distance they are considered to live close; if they live more than 10 km away they are considered to live far. Observations of individuals living between 2-10 km are excluded to be able to compare the distance more clearly. The distribution of distance is shown in Figure A.10. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects).

Table A.14: The impact of G1 employment on G3 outcomes - by pre-reform G1's level of income

	(1) ITT	(2) OLS	(3) IV
Treated x G1 Low income	0.027**		
	(0.011)		
Treated x G1 High income	0.005		
	(0.006)		
Employed x G1 Low income		0.139***	0.858**
		(0.004)	(0.346)
Employed x G1 High income		0.029***	0.167
		(0.003)	(0.119)
Treat Group x G1 Low income	-0.035***	-0.021***	0.023
	(0.006)	(0.006)	(0.042)
Treat Group x G1 High income	0.014***	0.011***	-0.013
	(0.004)	(0.004)	(0.014)
Treat Period x G1 Low income	0.513***	0.523***	0.486***
	(0.011)	(0.008)	(0.021)
Treat Period x G1 High income	0.499***	0.502***	0.505***
	(0.008)	(0.007)	(0.008)
G1 High income=1	0.189***	0.223***	0.398***
	(0.005)	(0.005)	(0.146)
Observations	848276	848276	848276
R squared	0.036	0.037	
Endogeneity test			0.083
F-stat			22.213

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. The table shows the estimates of Equation (9) for grandfathers, with heterogeneous effect by G1's pre-reform level of income (Column 3) (corresponding to Figure 4, Panel (a)). The correspondent ITT results and OLS are shown respectively in Columns (1) and (2). All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Cluster standard errors at grandchild's level in parenthesis.

Table A.15: The impact of G1 employment on G3 outcomes - by G1's immigrant status

	(1) ITT	(2) OLS	(3) IV
Treated x G1 immigrant	0.046***		
	(0.013)		
Treated x G1 non immigrant	0.011**		
	(0.005)		
Employed x G1 immigrant		0.225***	0.847***
		(0.006)	(0.248)
Employed x G1 non immigrant		0.082***	0.178**
		(0.002)	(0.084)
Treat Group x G1 immigrant	0.126***	0.190***	0.333***
	(0.007)	(0.006)	(0.057)
Treat Group x G1 non immigrant	0.146***	0.173***	0.201***
	(0.003)	(0.002)	(0.025)
Treat Period x G1 immigrant	0.539***	0.558***	0.531***
	(0.012)	(0.009)	(0.015)
Treat Period x G1 non immigrant	0.515***	0.519***	0.514***
	(0.008)	(0.007)	(0.008)
G1 non immigrant	0.223***	0.286***	0.548***
	(0.006)	(0.006)	(0.149)
Observations	920619	920619	920619
R squared	0.024	0.027	
Endogeneity test			0.017
F-stat			38.059

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. The table shows the estimates of Equation (9) for grandfathers, with heterogeneous effect by G1's immigrant status (Column 3) (corresponding to Figure 4, Panel (b)). The correspondent ITT results and OLS are shown respectively in Columns (1) and (2). All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Cluster standard errors at grandchild's level in parenthesis.

Table A.16: Grandfathers' use of of drugs

	Mental health drugs	
	Mean	SD
2006	0.144	0.351
2007	0.147	0.354
2008	0.147	0.354
2009	0.070	0.256
2010	0.071	0.257
2011	0.073	0.260
2012	0.072	0.259
2013	0.072	0.259
2014	0.073	0.260
2015	0.074	0.261
2016	0.075	0.263
Observations	878837	

Source: SSB Dutch administrative data.
Notes: The variable mental health drugs is a dummy indicator for whether the grandfather within the year of observation is prescribed of any of the mental health drugs in ACT4 code.

Table A.17: The impact of the reform on G2 outcomes - Fathers - Fixed effect models - Married or Cohabiting

	(1) Employment	(2) Full time	(3) Earnings
Treat group	0.000	-0.002*	-0.008***
	(0.001)	(0.001)	(0.002)
Treat Period (2004-2007)	-0.002*	-0.006***	-0.046***
	(0.001)	(0.001)	(0.002)
Treat Group x Treat Period	0.002	0.003***	0.021***
	(0.001)	(0.001)	(0.003)
Observations	826495	607140	734246
R squared	0.004	0.003	0.017

Source: SSB Dutch administrative data.

Notes: The dependent variables are G2-fathers' labor supply when the youngest child is between age 0 and 4 (Columns 1-3) decisions. Employment (Column 1) is a dummy variable that indicates whether the mothers is on paid employment. Full time (Column 2) is a full time index going from zero to one, where one indicate a full time job. Earnings (Column 3) are income from paid employment measured in logarithm. The table shows the estimates of Equation (10) considering the treatment status of his G1 father. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and mother's background characteristics (i.e. age and age squared (except for the last column), immigrant status, having a partner, and the number of G2's children (only in columns 1-3)) and father's fixed effects. The sample is restricted to married or cohabiting fathers. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Cluster standard errors at grandchild's level in parenthesis.

Table A.18: The impact of the reform on G2 outcomes - By mother's education - FE

	(1) Employment	(2) Full time	(3) Earnings	(4) Num Children	(5) Age at 1st birth
Treat group	-0.028***	0.011	-0.032	-0.036***	-0.032
	(0.009)	(0.013)	(0.039)	(0.006)	(0.134)
Treat Period (2004-2007)	-0.049***	-0.017*	-0.105***	-0.073***	-0.239
	(0.009)	(0.009)	(0.031)	(0.007)	(0.169)
Treat group × Treat Period (2004-2007)	0.033***	0.015	0.074*	-0.069***	0.288
	(0.011)	(0.013)	(0.042)	(0.008)	(0.312)
Treat group × Treat Period (2004-2007) × Education=2	-0.009	0.004	-0.012	-0.020**	0.129
	(0.013)	(0.014)	(0.047)	(0.009)	(0.342)
Treat group × Treat Period (2004-2007) × Education=3	-0.020*	-0.009	-0.042	-0.013	0.215
	(0.012)	(0.013)	(0.043)	(0.009)	(0.317)
Treat group × Treat Period (2004-2007) × Education=4	-0.025**	-0.010	-0.055	0.092***	0.185
	(0.012)	(0.013)	(0.043)	(0.009)	(0.316)
Treat group × Treat Period (2004-2007) × Education=5	-0.022*	-0.013	-0.063	0.196***	0.066
	(0.012)	(0.014)	(0.044)	(0.011)	(0.320)
Treat group × Treat Period (2004-2007) × Education=6	-0.044**	-0.017	-0.075	0.268***	-0.338
	(0.019)	(0.020)	(0.059)	(0.025)	(0.370)
Observations	741781	471661	572157	2109872	168874
R squared	0.013	0.056	0.020	0.554	0.335

Source: SSB Dutch administrative data.

Notes: The dependent variables are G2-mothers' labor supply when the youngest child is between age 0 and 4 (Columns 1-3) and fertility (Column 4-5) decisions. Employment (Column 1) is a dummy variable that indicates whether the mothers is on paid employment. Full time (Column 2) is a full time index going from zero to one, where one indicate a full time job. Earnings (Column 3) are income from paid employment measured in logarithm. Number of children (Column 4) counts the total fertility. Mother's age at first birth (Column 5) is expressed in years. The table shows the estimates of Equation (10) and (11) considering the treatment status of her G1 father, heterogenous by mother's level of education. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and mother's background characteristics (i.e. age and age squared (except for the last column), immigrant status, having a partner, and the number of G2's children (only in columns 1-3)) and mother's fixed effects. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis.

Table A.19: Selection of the G2 sample - when including G2 education

	Total sample		Education sample		T-test
	Mean	SD	Mean	SD	p-value diff
G2 Employment	0.752	0.432	0.765	0.424	0.000
G2 Full-time	0.581	0.263	0.590	0.267	0.000
G2 Earnings	20070.035	13635.701	21227.141	14623.780	0.000
G2 Num of Children	1.692	0.775	1.667	0.758	0.000
G2 Age at first child	27.706	3.634	27.649	3.756	0.000
G2 age	30.715	3.588	30.565	3.688	0.000
G2 immigrant	0.122	0.327	0.141	0.348	0.000
G2 partner	0.954	0.210	0.940	0.237	0.000
G1 age	57.811	3.538	57.768	3.557	0.000
G1 immigrant	0.145	0.352	0.160	0.367	0.000
G1 partner	0.917	0.276	0.912	0.283	0.000
G1 active	0.586	0.493	0.584	0.493	0.005
G1 Treated Group	0.608	0.488	0.605	0.489	0.000
G1 Treated Period	0.424	0.494	0.453	0.498	0.000
Observations	1268016.000	.	787612.000	.	.

Source: SSB Dutch administrative data.
Notes: The table shows the mean comparison of the characteristics between the full sample of mothers and the sample of mothers with non-missing educational information.

Table A.20: The impact of the reform on G3 educational outcomes - Placebo threshold for the grandfather's age: 50-52 vs 53-55

	(1) ITT	(2) OLS	(3) IV
Treat Group (age 53-55)	-0.025***	-0.022***	-0.094
	(0.009)	(0.009)	(0.661)
Treat Period	0.507***	0.513***	-0.319
	(0.014)	(0.013)	(7.522)
Treat Group x Treat Period	0.015		
	(0.010)		
Employed		0.166***	35.061
		(0.005)	(315.457)
Observations	247398	247398	247398
R squared	0.021	0.026	
Endogeneity test			0.000
F-stat			0.012

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. Sample is restricted to grandchildren whose grandfather's is younger than 56. The table shows the estimates of Equation (9) but assigning a placebo treatment group: grandchildren whose grandfathers are 50-52 years old are in the untreated group and grandchildren whose grandfathers are 53-55 years old are in the treated group. The correspondent ITT and OLS results are shown respectively in Columns (1) and (2). All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors at grandchild's level in parenthesis.

Table A.21: The impact of the reform on G3 educational outcomes - Placebo threshold for the grandfather's age: 59-60 VS 61-63

	(1) ITT	(2) OLS	(3) IV
Treat Group (age 61-63)	0.009	0.017**	0.118
	(0.007)	(0.007)	(0.100)
Treat Period	0.524***	0.526***	0.450***
	(0.010)	(0.010)	(0.076)
Treat Group x Treat Period	0.010		
	(0.006)		
Employed		0.103***	2.146
		(0.003)	(2.022)
Observations	400263	400263	400263
R squared	0.019	0.022	
Endogeneity test			0.000
F-stat			1.995

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. Sample is restricted to grandchildren whose grandfather's is older than 58. The table shows the estimates of Equation (9) but assigning a placebo treatment group: grandchildren whose grandfathers are 59-60 years old are in the untreated group and grandchildren whose grandfathers are 61-63 years old are in the treated group. The correspondent ITT and OLS results are shown respectively in Columns (1) and (2). All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors at grandchild's level in parenthesis.

Table A.22: The impact of the reform on G3 educational outcomes - Placebo threshold for treatment period: 2000 2001 vs 2002 2003

	(1) ITT	(2) OLS	(3) IV
Treat Group	0.014**	0.008	0.016**
	(0.006)	(0.005)	(0.008)
Treat Period (2002-2003)	0.485***	0.477***	0.488***
	(0.007)	(0.007)	(0.010)
Treat Group x Treat Period	-0.004		
	(0.006)		
Employed		0.123***	-0.124
		(0.003)	(0.175)
Observations	509029	509029	509029
R squared	0.045	0.048	
Endogeneity test			0.000
F-stat			177.492

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. Sample is restricted to grandchildren born before 2004. The table shows the estimates of Equation (9) but assigning a placebo treatment period: grandchildren born in 2000-2001 belong to the untreated period and grandchildren born in 2002-2003 belong to the treated period. Column (1) shows the results for the all births. Column (2) restricts the sample to later births. Column (3-5) restrict the sample to first birth. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors at grandchild's level in parenthesis.

Table A.23: The impact of the reform on G3 educational outcomes - Placebo threshold for treatment period: 2004 2005 vs 2006-2007

	(1) ITT All births	(2) ITT Later births	(3) ITT First birth	(4) OLS First birth	(5) IV First birth
Treat Group	-0.003	-0.003	0.014	0.011	0.010
	(0.007)	(0.009)	(0.013)	(0.012)	(0.037)
Treat Period (2006-2007)	0.787***	0.789***	0.777***	0.777***	0.776***
	(0.018)	(0.021)	(0.029)	(0.028)	(0.031)
Treat Group x Treat Period	0.034***	0.040***	0.002		
	(0.009)	(0.012)	(0.016)		
Employed				0.074***	0.102
				(0.008)	(0.706)
Observations	254361	179733	74628	74628	74628
R squared	0.033	0.034	0.039	0.041	
Endogeneity test					0.000
F-stat					11.071

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. Sample is restricted to grandchildren born between 2003-2005. The table shows the estimates of the corresponding ITT of Equation (9) in Column (1). Column (2) shows the corresponding results replacing the treatment period dummy with a dummy taking value one if the child has been conceived after the reform and zero otherwise. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors at grandchild's level in parenthesis.

Table A.24: The impact of the reform on G3 educational outcomes - By G3 month of conception.

	(1) ITT	(2) ITT
Treat Group	0.005	0.022***
	(0.007)	(0.007)
Treat Period	0.486***	
	(0.009)	
Treat Group × Treat Period	0.016**	
	(0.007)	
Conceived after reform		0.603***
		(0.009)
Treat Group × Conceived after reform		-0.009
		(0.008)
Observations	310770	310770
R squared	0.047	0.051

Source: SSB Dutch administrative data.

Notes: The dependent variable is grandchildren's standardized total test score at age 12. Sample is restricted to grandchildren born between 2003-2005. The table shows the estimates of the corresponding ITT of Equation (9) in Column (1). Column (2) shows the corresponding results replacing the treatment period dummy with a dummy taking value one if the child has been conceived after the reform and zero otherwise. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects). Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors at grandchild's level in parenthesis.

Table A.25: Description of compliers

	(1)	(2)	(3)	(4)	(5)	(6)
	First stage	P(Employed)	P(Treated)	Sample means	Means of compliers	Difference
Compliers	0.051	0.624	0.185			
G1 Partner	0.051	0.635	0.186	0.927	0.928	-0.001***
G1 Immigrant	0.044	0.449	0.165	0.136	0.117	0.019***
G1 Few children	0.056	0.643	0.194	0.649	0.706	-0.057***
G1 High income	0.049	0.753	0.201	0.726	0.692	0.034***
G1 Rural	0.049	0.640	0.164	0.617	0.609	0.008***
Observations	878837					

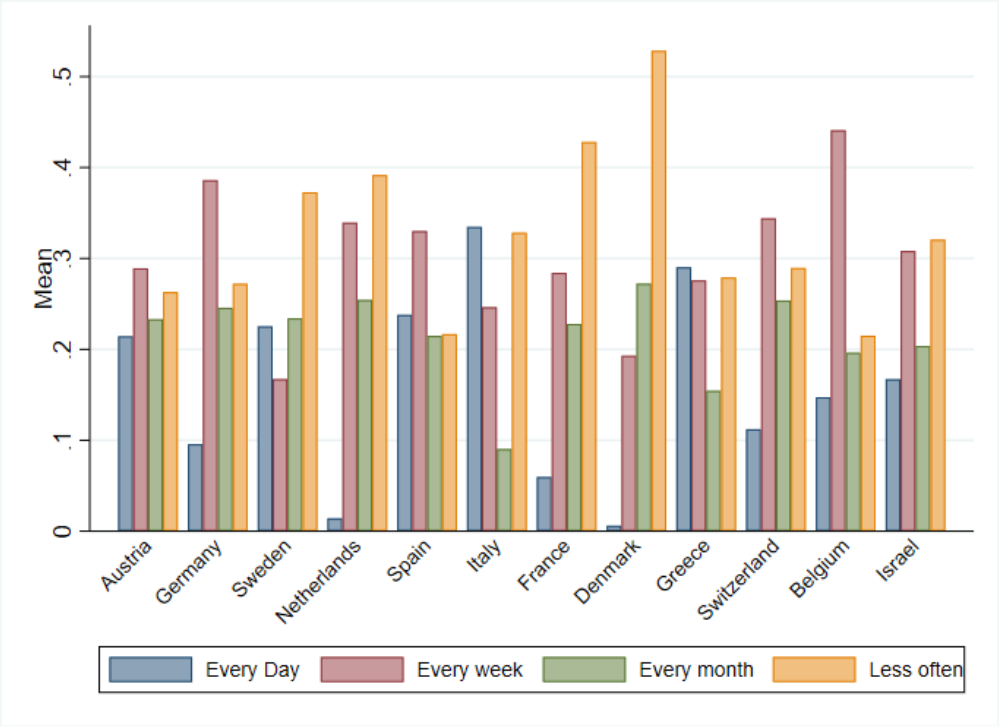
Source: SSB Dutch administrative data.

Notes: Column (1) shows the probability of compliance (first stage) for different subgroups identified by the grandfather's observable characteristics. Columns (2) and (3) show the correspondent probability of compliance among the Employed and Treated respectively. Columns (4) and (5) report respectively the sample mean and the means of compliers for each subgroup. Column (6) reports the difference between the two means and the statistical significant difference based on bootstrapped standard errors with 1000 replications. The mean of these characteristics in the group of compliers is estimated as $P(X = 1|complier)$, and the overall sample mean is $P(X = 1)$. Notice that $P(X = 1|complier) = P[X = 1|D_{1i} > D_{0i}]$ can be estimated as such:

$$P[X = 1|D_{1i} > D_{0i}] = P[X = 1]\frac{E[D_i|Z_i = 1, X = 1] - E[D_i|Z_i = 0, X = 1]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

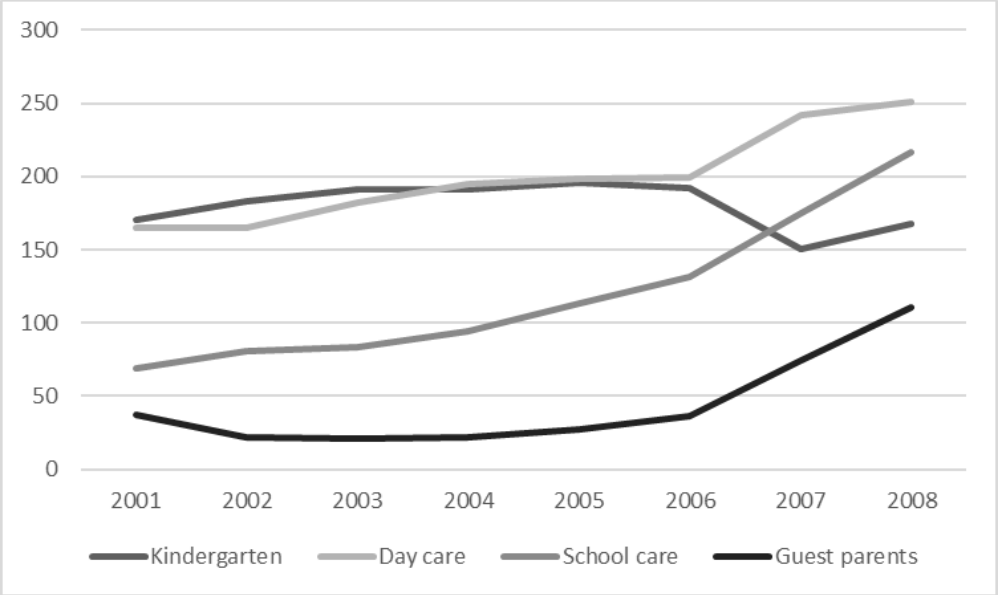
where, looking at the second term, the denominator is the first stage coefficient in the full sample, whilst the numerator is the first stage in the subsample of the population with $X = 1$. See Angrist and Pischke (2008) for a similar application.

Figure A.1: Grandparents time with grandchildren across countries



Source: SHARE (Survey of Health, Ageing and Retirement in Europe) 2004.
Notes: The figure shows the mean grandparents reply to the question "How often do you look after your grandchildren?", with answers going from "Every day" to "Less often", by countries.

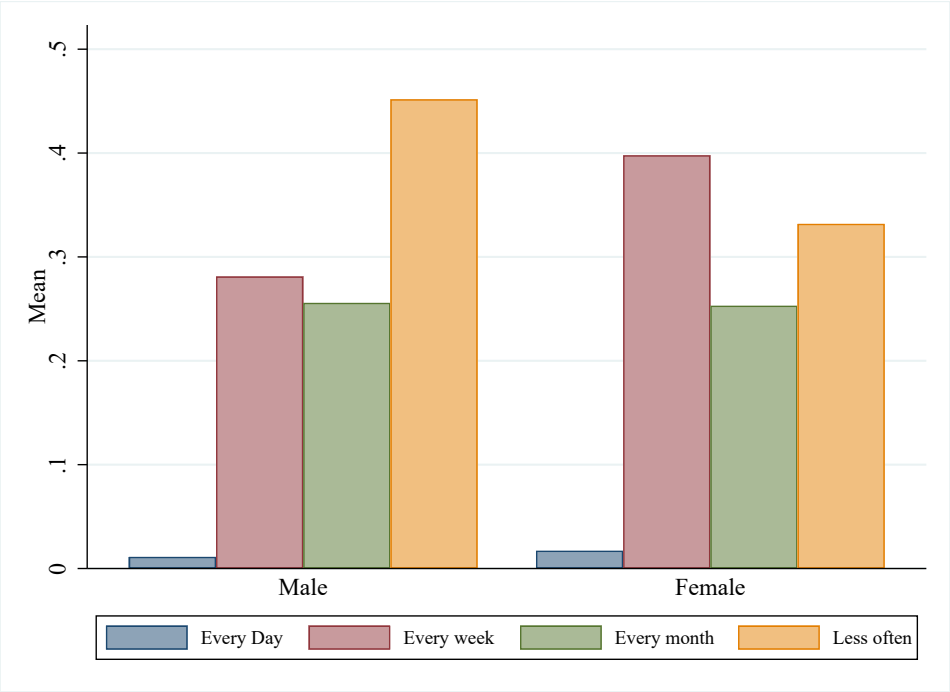
Figure A.2: Use of types of child care (1,000s)



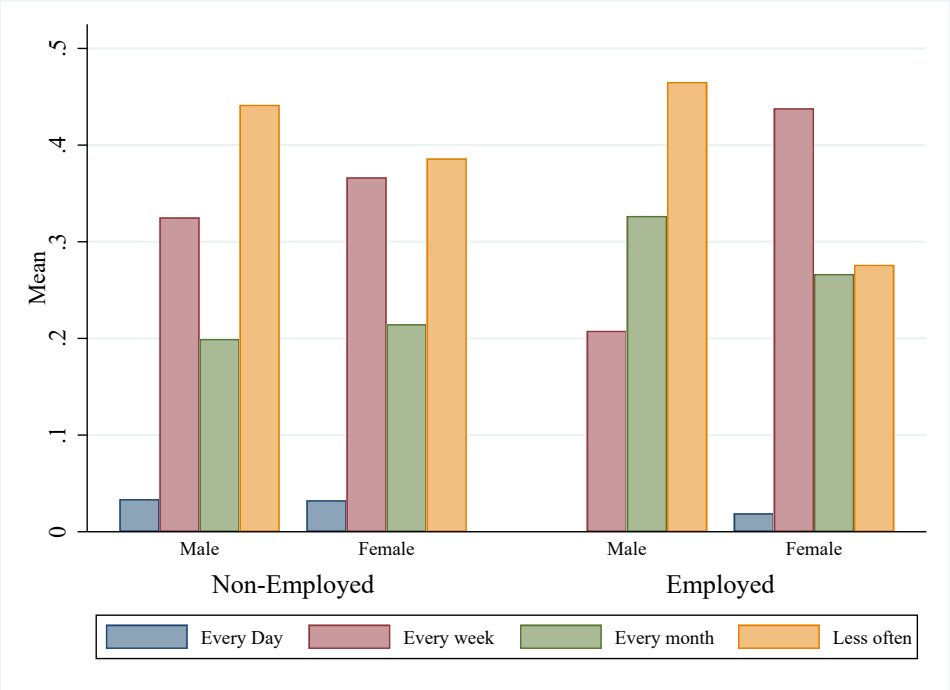
Source: CBS (2011).
Notes: Yearly use of formal child care for children aged 0-12 in 2001-2008, differentiated by four types of formal child care. Kindergarden is pre-school for kids aged 2-4; School care is lunch at school between the morning and afternoon schedules; Day care is daycare centers and guest parents is care by a person who is an official host. Statistics are not available before 2001. The strong increase in 2006 is due to the increased generosity in child care subsidies introduced in 2005.

Figure A.3: Grandparents time with grandchildren in the Netherlands

(a) By gender of grandparents

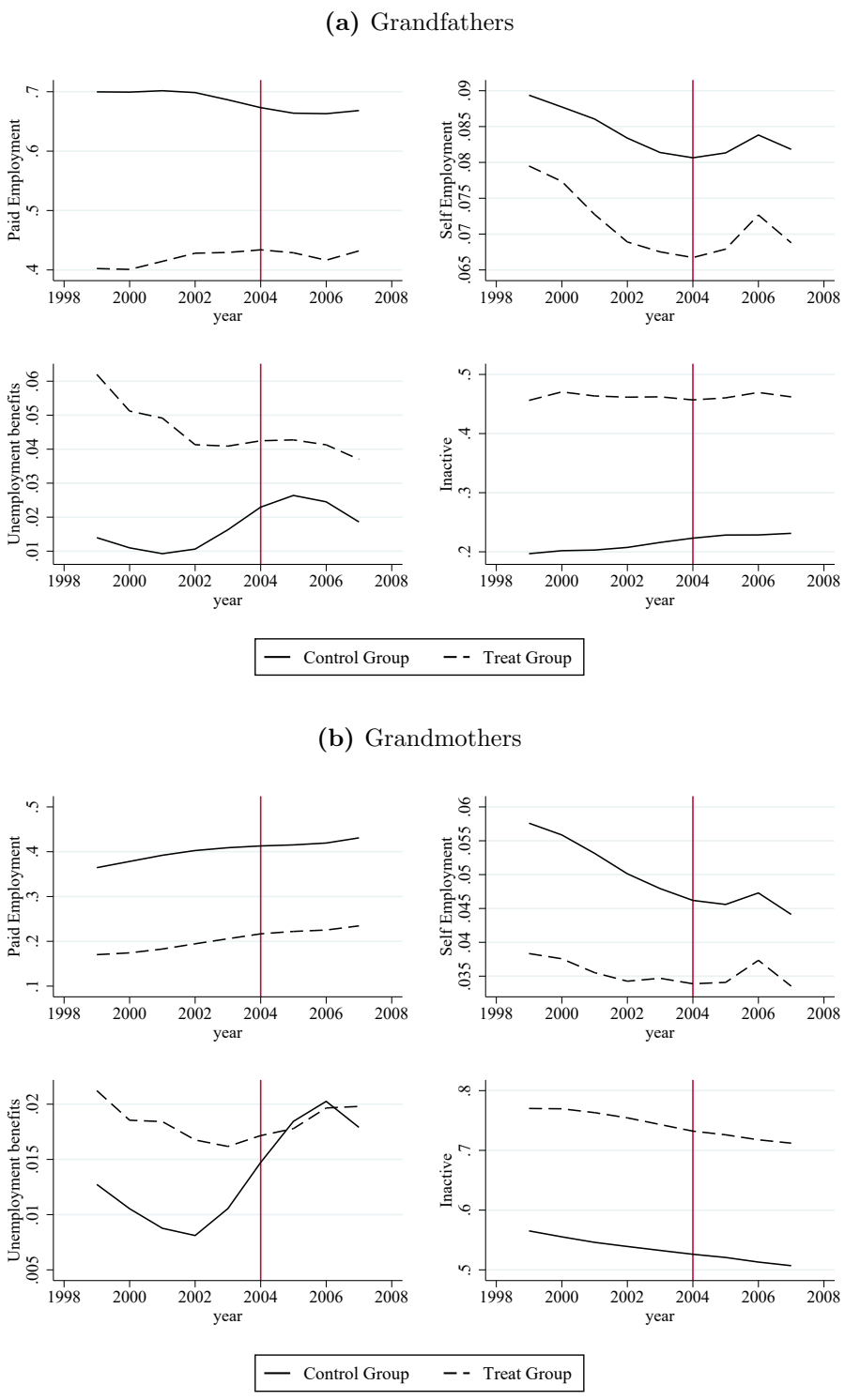


(b) By gender and employment status of grandparents



Source: SHARE (Survey of Health, Ageing and Retirement in Europe) 2004.
Notes: The figure shows the mean grandparents reply to the question "How often do you look after your grandchildren?", with answers going from "Every day" to "Less often", by grandparents' gender.

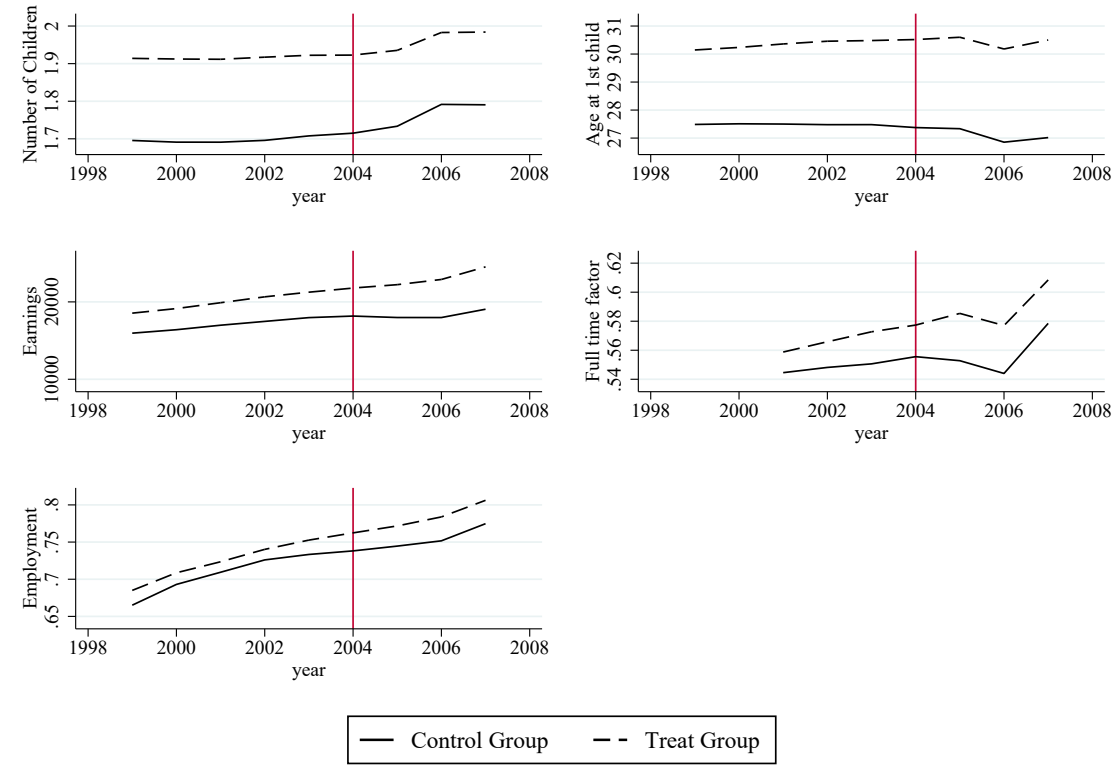
Figure A.4: Average labor market status of grandparents over time



Source: SSB, Dutch Administrative Data.

Notes: The figure shows the average labor market status of grandfathers (Panel A) and grandmothers (Panel B) over time (1999-2007) differently for the control (age 50-57.5) and treated group (age 57.5 - 63).

Figure A.5: Average G2 mothers' labor supply and fertility decisions over time

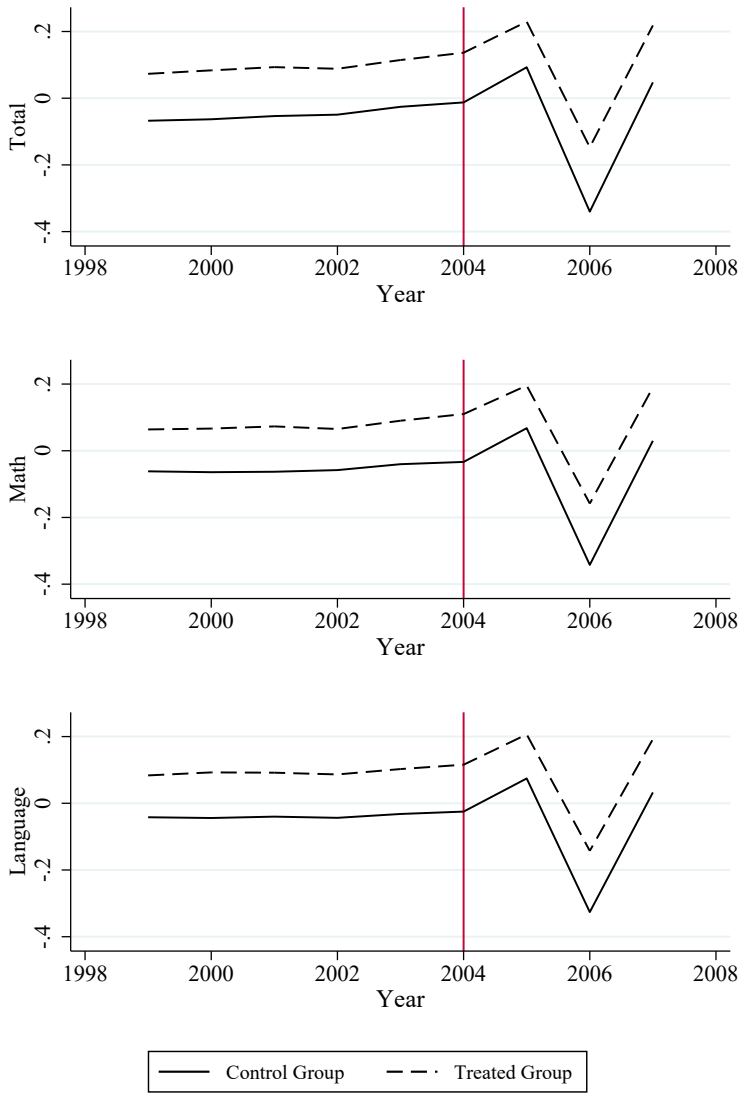


Num obs:505449

Source: SSB, Dutch Administrative Data.

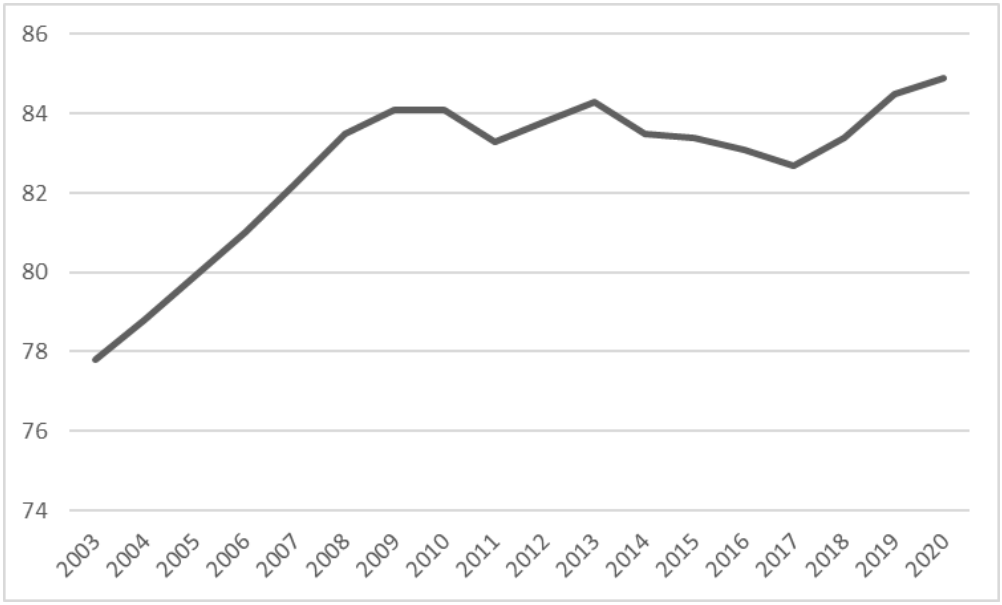
Notes: The figure shows the G2 mothers' labor supply and fertility decisions over time (1999-2007) differently for the control (age 50-57.5) and treated group (age 57.5 - 63) defined on the basis of G2's father.

Figure A.6: Trend of grandchildren educational outcomes pre and post reform.



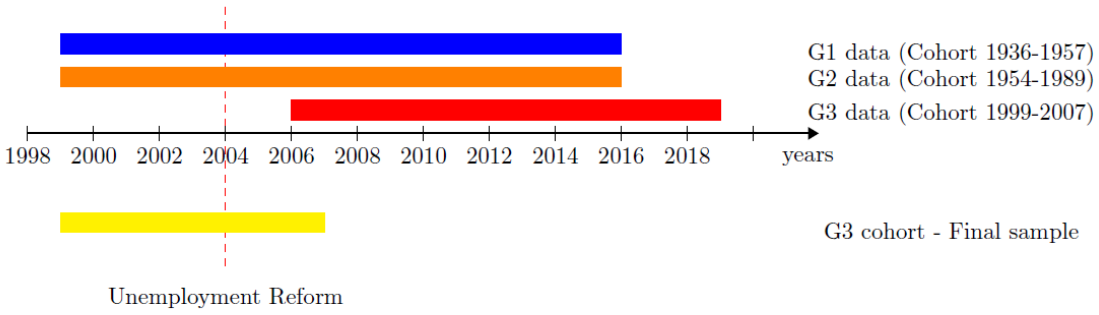
Source: SSB, Dutch Administrative Data.
Notes: The figure shows the G3 test score over time (1999-2007) differently for the control (age 50-57.5) and treated group (age 57.5 - 63) defined on the basis of G3's grandfathers treatment status at G3's birth.

Figure A.7: Female (age 25-45) labor force participation (%)



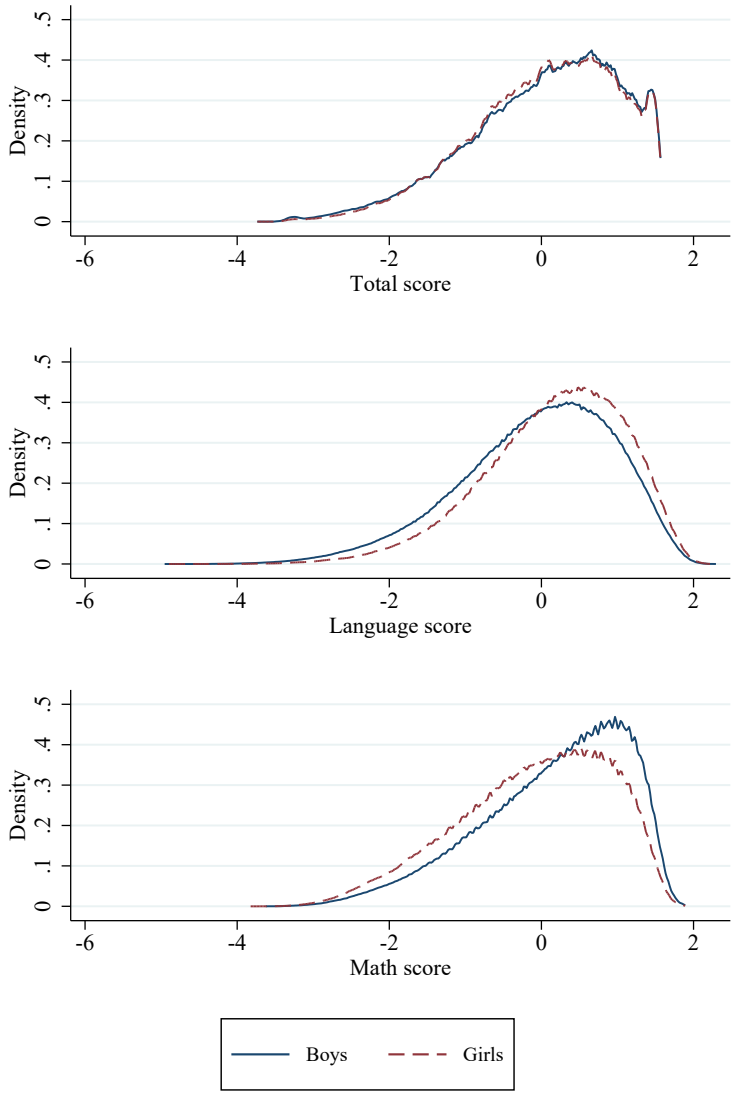
Source: CBS (2021).
Notes: Labor force participation of women aged 25-45 over the period 2003-2020. Statistics are not available before 2003.

Figure A.8: Timeline of the data



Notes: The figure illustrates the availability of data for G1, G2 and G3 cohorts from SSB Dutch administrative data. Labor market information of G1 is available for the years 1999-2016 (blue bar). Labor market information of G2 is available for the years 1999-2016 (orange bar). The test score of G3 is available for the years 2006-2019 (red bar). Since G3's test scores are measured at the age of 12 and we restrict our sample to those G1's aged 50-63 to identify the effect of the 2004 reform, we are left with a final sample that spans the years 1999-2007 (yellow bar).

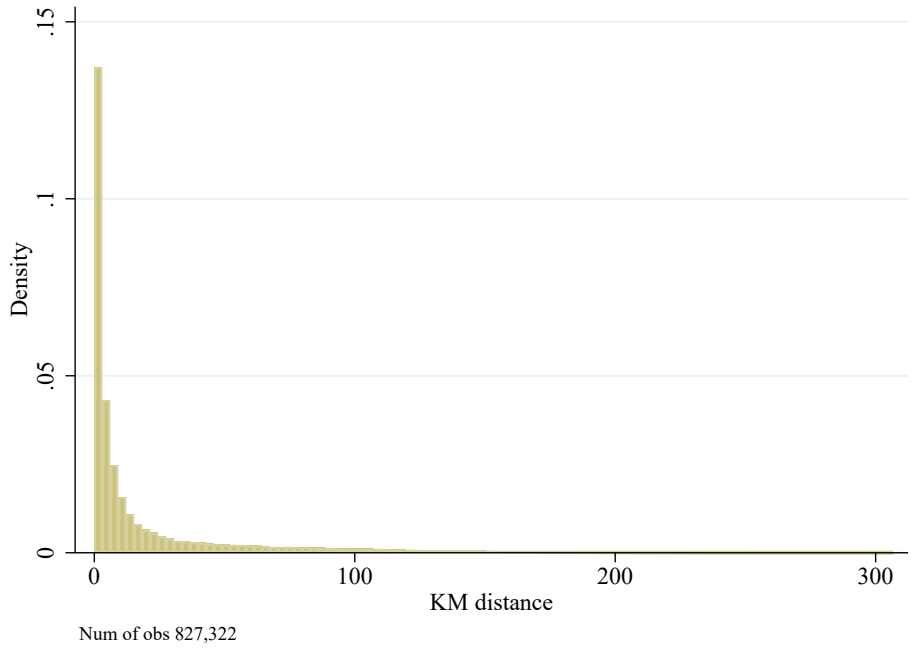
Figure A.9: G3 Grandchildren - Distribution of CITO test in 6^{th} grade - by gender



Source: SSB Dutch administrative data.

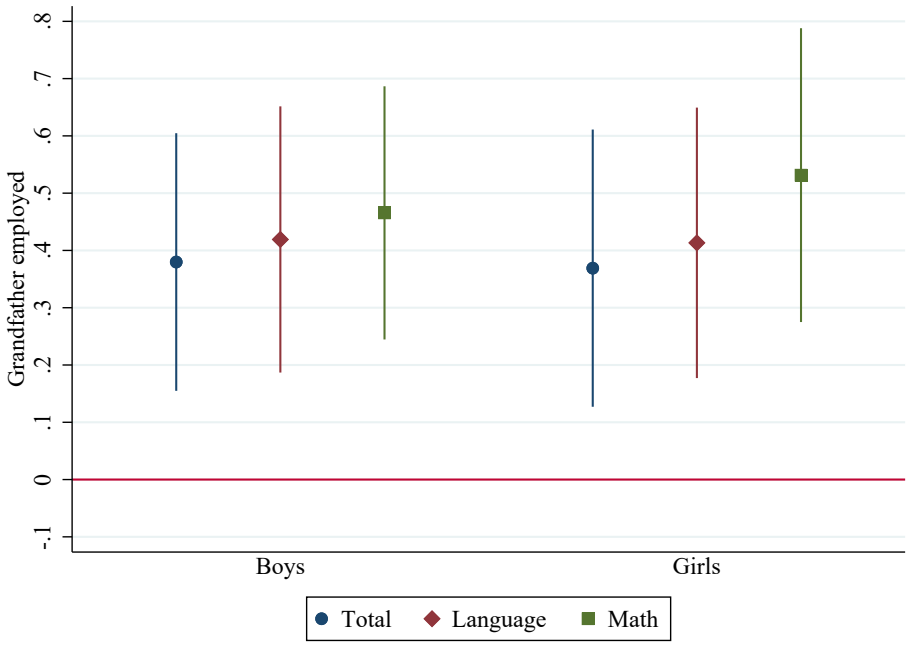
Notes: The figure shows the distribution separately for boys and girls of the CITO test in 6^{th} grade standardized with mean zero and standard deviation 1 within cohorts, in the full population of grandchildren.

Figure A.10: Description of distance between grandfathers and grandchildren



Source: SSB, Dutch Administrative Data.
Notes: Distance is measured as Euclidean distance (equivalent to 2/3 of the geographical distance) between the 4-digit-postcodes of grandchildren and grandfathers before the reform. If the child is born after the reform we consider the mother's postcode before the reform.

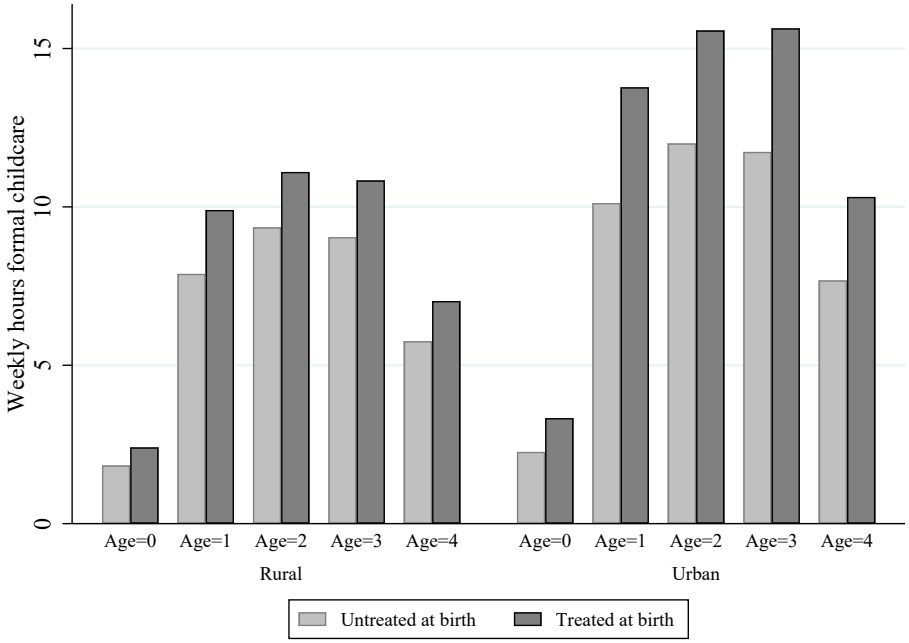
Figure A.11: The impact of G1 employment on G3 educational outcomes - by subject and gender



Source: SSB, Dutch Administrative Data.

Notes: The dependent variable is grandchildren's standardized test score at age 12. The figure shows the estimates of Equation (9) for different subjects, i.e. total score, language and math and by child's gender. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects).

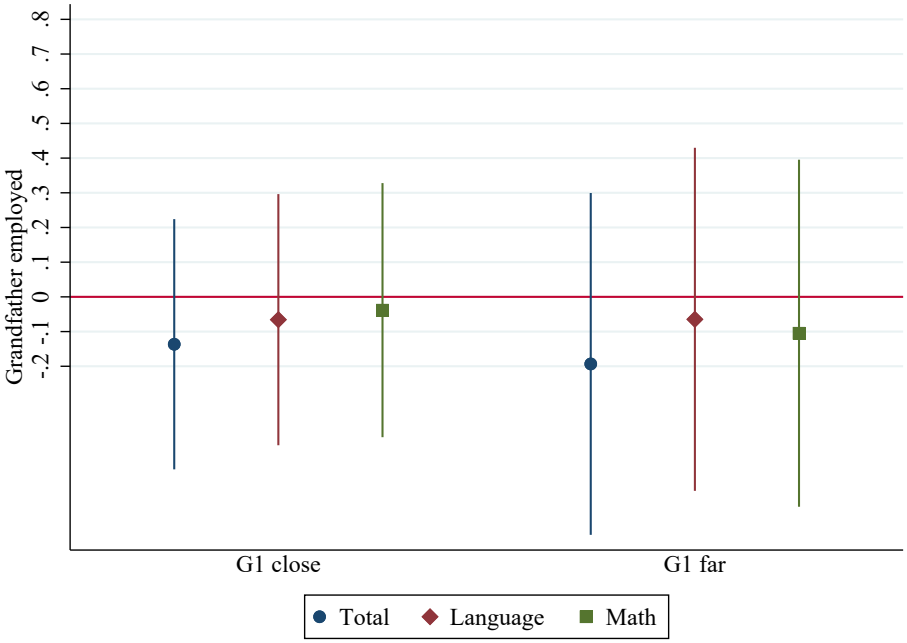
Figure A.12: Use of childcare (2007-2011) at different child's age - by treatment status at birth and by living area.



Source: SSB, Dutch Administrative Data.

Notes: The figure shows the average weekly use of formal childcare by grandchildren treatment status at birth, from 2007 till 2011, separately for grandchildren living in urban or rural areas. Because the information is available only from 2007 onwards (treatment period), the treatment status of grandchildren is defined only based on the age of grandfathers.

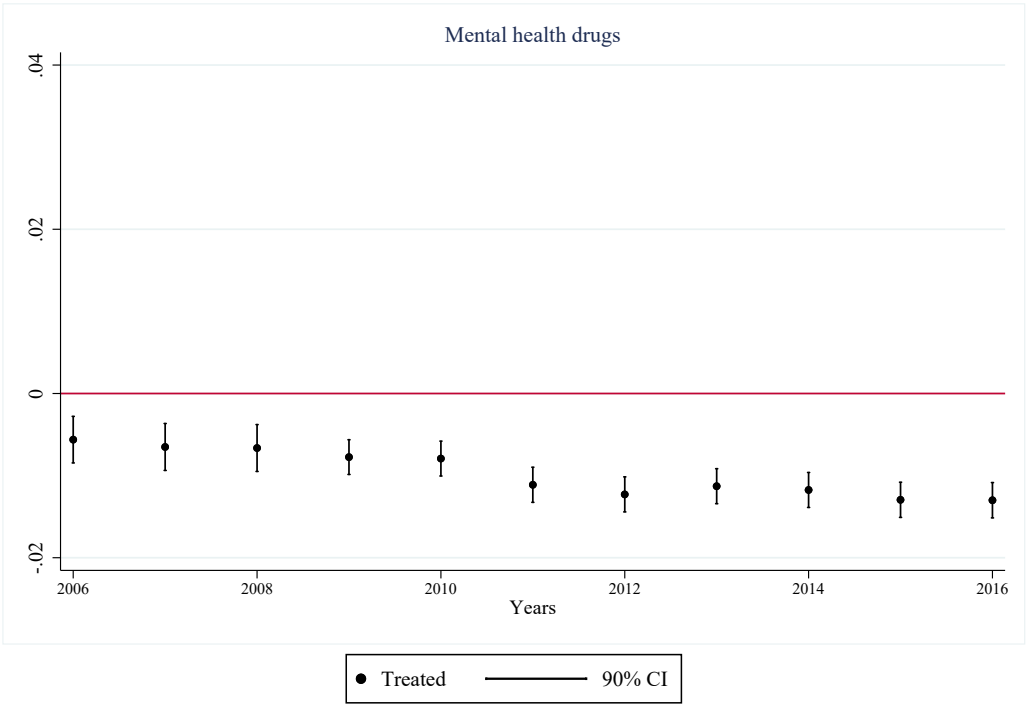
Figure A.13: The impact of G1 employment on G3 outcomes - By distance between grandfathers and grandchildren within rural areas



Source: SSB, Dutch Administrative Data.

Notes: The dependent variable is grandchildren's standardized test score at age 12. The figure shows the estimates of Equation (9) for different subjects, i.e. total score, language and math and by distance between grandfathers and grandchildren within rural areas. If grandchildren and grandfathers live within 2 km distance they are considered to live close; if they live more than 10 km away they are considered to live far. Observations of individuals living between 2-10 km are excluded to be able to compare the distance more clearly. The distribution of distance is shown in Figure A.10. All estimates control for grandparent's background characteristics (i.e. age and age squared, immigrant status, having a partner, and the number of G1's children at the time of G3's birth) and grandchild's background characteristics (i.e. gender, immigrant status, and year of test fixed-effects).

Figure A.14: The impact of the reform on grandfather's health



Source: SSB, Dutch Administrative Data.
Notes: The figure shows the first stage estimates on grandfathers health outcomes in the long run. The outcome variables, mental health drugs is a dummy indicators for whether the grandfather within the year of observation is prescribed any mental health drugs. Effects estimated are in percentage points.

B Institutional Setting: UI in the Netherlands

B.1 Eligibility, replacement rate, and duration

Employees in paid employment who loose their job are entitled to unemployment insurance (UI) benefits in the Netherlands provided that they have worked at least 26 of the last 36 weeks and that the job loss is not culpable to the employee.[60] The replacement rate is about 70% of last earnings,[61] which is fairly generous compared to the situation in other European countries.[62] Maximum benefit duration ranges between 2 to 5 years, depending on year of enrollments, age and work experience.[63] When UI benefits are exhausted, people can apply for asset- and income-based means-tested welfare benefits that guarantee a minimum standard of living.

Entitlement to a UI benefit comes with a mandatory job search requirement, which aims to increase the probability of finding new employment.[64] Not abiding the mandatory job search requirements can have severe consequences ranging from financial sanctions to losing the entitlement to UI benefits altogether. The job search requirement involves the unemployed to (i) have an intake meeting at the unemployment office, where individual criteria are made regarding the expected activities undertaken during unemployment; (ii) have the obligation to accept suitable job-offers, where suitable job offers are defined by the educational level and the time spent in unemployment; (iii) have to make at least a pre-determined number of applications (either by letter, e-mail, phone call or nuncupative contact with a company, registering at an agency, having a job interview and doing an assessment); (iv) have to participate in educational programs and job search assistance when not being able to find work within six months; 5) have regular report meetings every 4-6 weeks in addition to the mandatory intake meeting and the follow-up to explain the further procedures.

[60] Culpable reasons to become unemployed are mostly instant dismissals by the employer and voluntary quits.

[61] Prior to 2016, replacement rates were 70% for the total duration of UI benefits. As of 2016, the replacement rate increased from 70% to 75% of the last earnings for the first two months with an absolute maximum of about 3,100 euros. From the third month onwards, the replacement rate is reduced to 70% with an absolute maximum of about 2,900 euros.

[62] The net replacement rate in the Netherlands for the first 2 months of job loss is one of the highest in the OECD and is about 30, 40, and 15 percentage points higher than in the US, UK, and Germany, respectively (OECD, 2019).

[63] During the period under study the entitlement period has changed. Prior to October 2006, the maximum entitlement to UI benefits was age-dependent and amounted to a maximum of 42, 48 and 60 months for persons aged 50-54, 55-59 and 60-64 respectively. Since October 2006, the duration of benefits depends on work history with a minimum duration of three months. This is extended by one month for every year worked up to a maximum of 38 months for those who worked at least 4 out of the last 5 years. As of 2016, the maximum has been reduced to 24 months. In addition, the accumulation of months has been made less generous: one month for every year of work in the first 10 years of work history, and half a month for every additional year of work beyond the first 10 years.

[64] Exemptions are made for those who are within one year of their statutory retirement age, informal caretakers, voluntary workers (under some conditions), and starting entrepreneurs.

B.2 Specific regulations for older individuals

As older workers often experience difficulties with finding new employment, the Dutch unemployment insurance scheme contains some specific policies and exemptions targeted to older unemployed. As a result of these measures, the Dutch unemployment insurance has long served as an attractive early retirement pathway for older individuals. Since the early 2000s, various reforms in UI have taken place to limit possibilities to retire early and to increase elderly labor market participation. An important measure to increase labor supply among older UI recipients was to make job search mandatory to the 57.5 and older as of January 1st 2004.

As part of the efforts to stimulate older workers active labor market participation, there have also been changes in the extended UI benefits available to older individuals to which they were entitled after exhaustion of regular UI benefits. Until August 2003, individuals under UI could extend the maximum benefit duration up to the age of 65 by applying for extended UI benefits. These extended UI benefits amounted 70% of minimum wage. From August 2003, extended UI benefits were abolished.[65] Note, however, that this change does not differentially affect those under/above age 57.5 and hence this policy reform does not invalidate our research approach.

B.3 Other reforms

In our period of analysis, there have been several reforms that have affected the labor supply decisions of older workers in the Netherlands,[66] such as the abolishment of early retirement schemes (*VUT & prepensioen*) and simultaneous introduction of the life course arrangement (*Levensloopregeling*) in 2005,[67] the decreased eligibility and generosity in disability insurance (*WIA*) in 2006,[68] and the decreased duration in unemployment insurance benefits. However, none of these reforms coincide with the treated and control groups in the 2004 UI reform,

[65] At the same time, the *IOAW* was introduced to provide a minimum income provision for elderly born before January 1st 1965, who become unemployed after the age of 50, and whose UI benefit had expired. These benefits complement household income up to the subsistence level for those households that fall below this level. Hence, eligibility is means-tested based on household income, but assets are not taken into account (which is the main difference compared to welfare benefits). In 2009, another supplementary income scheme was introduced: the *IOW*. The *IOW* provided benefits of at most 70% of the minimum wage, depending on the level of income before unemployment, to individuals who become unemployed after the age of 60, and whose UI benefits expired. Compared to *IOAW* benefits, *IOW* benefits do not take into account household income, but only personal income. *IOW* was initially introduced as a temporary arrangement to alleviate job finding difficulties among older unemployed during the Great Recession. However, in 2014 and 2019 the arrangement has been extended for four years.

[66] For an overview of reforms from 1995-2016 aimed at the elderly in the Netherlands, see Kalwij et al. (2018).

[67] The 2005 abolishment of early retirement schemes only affected those born after January 1st 1950. Those affected by this reform could start saving more in the tax-beneficial life course arrangement, although take-up rates are very low.

[68] The 2006 decrease in eligibility and generosity of disability insurance is not targeted at older workers and generally applies to all ages.

and hence these reforms do not invalidate our methodology.

C Prevalence of different forms of (in)formal childcare

Using ancillary data from the LISS Social Integration and Leisure survey for 2008-2017, we are able to get additional insights in the use of different forms of childcare, both formal and informal. The Social Integration and Leisure is a module that is part of the annual core questionnaires of the LISS panel (Longitudinal Internet Studies for the Social sciences) administered by CentERdata at Tilburg University. The annual core waves consist of 5,000 representative households (about 8,000 individuals) in the Netherlands. These households are followed as of 2008. The LISS module includes the following question to infer information regarding the prevalence of using different types of (in)formal childcare:

The following questions are about your living-at-home children. For [THIS/THESE] [CHILD(REN)], 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.

In Table C.1, we present the average use of different types of childcare. These descriptive statistics suggest that about 42% of mothers aged 20-59 use daycare centers (i.e. kindergarten) for their children aged 0-4.[69] Unpaid (i.e. informal) childcare is almost equally important with 41% which is consistent with the statistics presented by OECD (2020). In almost 90% of the cases, the unpaid child sitter of children aged 0-4 is a grandparent. For children aged 5-12, this is about 83% of the cases. Hence, 37% of mothers use grandparents for informal childcare of children aged 0-4. This is only 23% of mothers in case of children aged 5-12. In general, we observe that mothers use less formal and informal care for children aged 5-12 than for children aged 0-4. For this age group of children, mothers mostly rely on between-school (i.e. lunch) and after-school care.

The descriptive statistics in Table C.1 indicate that grandparents are an important source of care for their grandchildren and that this option if frequently combined with formal childcare in the form of kindergarten. However, mothers appear to have particular persistent preferences for kindergarten if they use kindergarten. In Table C.2 we show that the state-dependence of using kindergarten is about 0.60-0.75. This means that if kindergarten is used at time t, there is a probability of 0.60-0.75 that kindergarten is also used at time $t+1$. In relation to our main results in the paper, this implies that the positive effects of replacing informal childcare with formal childcare are likely to accumulate because of persistence in choices for formal childcare.

Similarly, we show the estimated persistence in using unpaid child sitters, focusing on maternal and paternal grandparents, in Table C.3. We find that the state-dependence in using unpaid child sitters is much lower than for using kindergarten. Furthermore, we find

[69] For the group of children aged 5-12, the kindergarten option no longer formally exists.

that the persistence of using the maternal grandparents is smaller than using the paternal grandparents, although maternal grandparents are more frequently used as the unpaid child sitter. This suggests that mothers especially substitute away from childcare by grandparents if other options, such as formal childcare, tend to be available.

Table C.1: Prevalence of (in)formal child care for children aged 0-4 and 5-12.

	Age 0-4		Age 5-12	
	Mean	SD	Mean	SD
No care	0.136	0.343	0.380	0.486
Playgroup	0.279	0.449	n/a	n/a
Daycare center	0.420	0.494	n/a	n/a
Pre-school	0.006	0.078	0.035	0.184
Between-school	n/a	n/a	0.296	0.457
After-school	0.031	0.174	0.191	0.393
Guestparent	0.100	0.300	0.041	0.180
Paid childsitter (away)	0.027	0.161	0.033	0.180
Paid childsitter (home)	0.038	0.190	0.043	0.202
Unpaid care ...	0.412	0.492	0.277	0.448
..by maternal grandparents	0.628	0.483	0.602	0.490
..by paternal grandparents	0.267	0.443	0.227	0.419

Source: LISS Social Integration & Leisure, 2008-2017.
Notes: Selection of mothers aged 20-49. N = 2,468 for children aged 0-4. N = 4,276 for children aged 5-12. Respondents are asked about whether they make regular use of the type of child care (1 - yes, 0 - no). The mean can be interpreted as the percentage of mothers using the particular type of child care.

Table C.2: Persistence of using daycare centers

	(1)	(2)	(3)	(4)
Daycare_{t-1}	0.753***	0.696***	0.616***	0.650***
	(0.019	(0.024)	(0.028)	(0.050)
Daycare_{t-2}				0.097 **
				(0.047)
Observations	1367	1367	1367	781
Demographics	No	Yes	Yes	Yes
Year dummies	No	Yes	Yes	Yes
Random effects	No	No	Yes	Yes

Source: LISS Social Integration & Leisure, 2008-2017.
Notes: The dependent variable is a dummy that is 1 if the household makes use of daycare centers and 0 is the household does not. Hence, the coefficient of $Daycare_{t-1}$ can be interpreted as the probability that a households used daycare at time t given that it used daycare at time $t-1$. Demographic control variables include: age, age squared, low-educated, high-educated, partner, urban, number of kids 0-4, number of kids 5-12, and paid work. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors at the personal level in parentheses.

Table C.3: Persistence of using unpaid childcare

		Age 0-4		Age 5-12	
	Total unpaid	Grandmother	Grandfather	Grandmother	Grandfather
Y_{t-1}	0.420***	0.169***	0.289***	0.250***	0.480 ***
	(0.029)	(0.045)	(0.058)	(0.063)	(0.101)
Observations	1367	550	550	237	237
Demographics	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
Random effects	Yes	Yes	Yes	Yes	Yes

Source: LISS Social Integration & Leisure, 2008-2017.
Notes: The dependent variable is a dummy that is 1 if the household makes use of an unpaid child sitter and 0 is the household does not. Hence, the coefficient of Y_{t-1} can be interpreted as the probability that a households used unpaid childcare at time t given that it used unpaid childcare at time $t-1$. Total unpaid can include anyone providing the care. Unpaid care by the grandmother or grandfather is conditional on the unpaid child sitter being a grandparent. Demographic control variables include: age, age squared, low-educated, high-educated, partner, urban, number of kids 0-4, number of kids 5-12, and paid work. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Clustered standard errors at the personal level in parentheses.

D The effect of retirement on joint time spending

D.1 Empirical model

To estimate the causal effect of retirement on time use, we use a fuzzy Regression Discontinuity Design (RDD). The RDD exploits a jump in the probability to retire at age 65 that is greater than zero but less than one. This discontinuity is induced by the statutory retirement age in the Dutch pension system. For details, we refer to the next section. Following Stancanelli and Van Soest (2012b) and Stancanelli and Van Soest (2012a), we estimate a Simultaneous Equation Model (SEM) to allow for couples' joint decisions in time use. To infer a causal effect of retirement, we use the RDD in the retirement equations. The RDD is likely to affect the choice of retirement and affect time use only through the retirement decision.

The SEM exists of the following system of equations in which we define P_{itm} as male time spend on paid work of person i at time t. P_{itf} is female time spend on paid work of person i at time t. L_{itm} and L_{itf} as time spent in leisure activities by men and women in a household respectively, H_{itm} and H_{itf} as time spent in helping others, including tending to the grandchildren, by men and women in a household respectively and R_{itm} and R_{itf} the dummies indicating whether the male or female is retired respectively. We estimate all time use categories in levels.

$$P_{itm} = Z_{itm}\beta^{pmm} + Z_{itf}\beta^{pfm} + f(A_{itm})\delta^{pmm} + f(A_{itf})\delta^{pfm} + R_{itm}\gamma^{pmm} + R_{itf}\gamma^{pfm} + \epsilon^{pm}_{itm} \quad (12)$$

$$P_{itf} = Z_{itm}\beta^{pmm} + Z_{itf}\beta^{pfm} + f(A_{itm})\delta^{pmm} + f(A_{itf})\delta^{pfm} + R_{itm}\gamma^{pmm} + R_{itf}\gamma^{pfm} + \epsilon^{pm}_{itf} \quad (13)$$

$$L_{itm} = Z_{itm}\beta^{lmm} + Z_{itf}\beta^{lfm} + f(A_{itm})\delta^{lmm} + f(A_{itf})\delta^{lfm} + R_{itm}\gamma^{lmm} + R_{itf}\gamma^{lfm} + \epsilon^{lm}_{itm} \quad (14)$$

$$L_{itf} = Z_{itm}\beta^{lmm} + Z_{itf}\beta^{lfm} + f(A_{itm})\delta^{lmm} + f(A_{itf})\delta^{lfm} + R_{itm}\gamma^{lmm} + R_{itf}\gamma^{lfm} + \epsilon^{lm}_{itf} \quad (15)$$

$$H_{itm} = Z_{itm}\beta^{hmm} + Z_{itf}\beta^{hfm} + f(A_{itm})\delta^{hmm} + f(A_{itf})\delta^{hfm} + R_{itm}\gamma^{hmm} + R_{itf}\gamma^{hfm} + \epsilon^{hm}_{itm} \quad (16)$$

$$H_{itf} = Z_{itm}\beta^{hmm} + Z_{itf}\beta^{hfm} + f(A_{itm})\delta^{hmm} + f(A_{itf})\delta^{hfm} + R_{itm}\gamma^{hmm} + R_{itf}\gamma^{hfm} + \epsilon^{hm}_{itf} \quad (17)$$

$$R_{itm} = Z_{itm}\beta^{rmm} + Z_{itf}\beta^{rfm} + D_{itm}\theta^{rmm} + f(A_{itm})\delta^{rmm} + D_{itm}\cdot f(A_{itm})\phi^{rmm} + D_{itf}\theta^{rmf} + f(A_{itf})\delta^{rmf} + D_{itf}\cdot f(A_{itf})\phi^{rmf} + \epsilon^{rm}_{itm} \quad (18)$$

$$R_{itf} = Z_{itm}\beta^{rmf} + Z_{itf}\beta^{rff} + D_{itm}\theta^{rfm} + f(A_{itm})\delta^{rfm} + D_{itm}\cdot f(A_{itm})\phi^{rfm} + D_{itf}\theta^{rff} + f(A_{itf})\delta^{rff} + D_{itf}\cdot f(A_{itf})\phi^{rff} + \epsilon^{rf}_{itf} \quad (19)$$

with $f(\cdot)$ being a polynomial and

$$A_{itk} = (age_{itk} - 65) \quad (20)$$

$$D_{itk} = 1(age_{itk} \geq 65) \quad (21)$$

Here, we assume that $f(\cdot)$ follows a linear effect.[70] Z is a vector of characteristics of the individual, such as educational level dummies, and a dummy for having a migration background.

Equations 12-26 are jointly estimated using (Full Information) Maximum Likelihood (FIML), following Stancanelli and Van Soest (2012a) and Stancanelli and Van Soest (2012b), which explicitly accounts for contemporaneous correlation that might exist between the error terms of Equations 19-26.[71] The estimation of each equation in isolation ignores the information about the covariances between the equations and information about the exclusion restrictions on all other equations. The error terms of paid work, leisure, time spend with grandchildren, and retirement are assumed to be iid normally distributed with mean zero which implies a linear estimation for all equations.

The main coefficients of interest are γ^{pkk}, γ^{lkk}, and γ^{hkk}. These coefficients measure the spouses' (cross-)effects of retirement on paid work, leisure, and time spend with grandchildren respectively. Causal identification of the effects of retirement on time use comes from the

[70] Assuming $f(\cdot)$ to be quadratic does not alter our main conclusions.

[71] When the model is correctly specified, then full information estimation is more efficient than estimators that do not take into account the covariances between the equations.

discontinuities in the probability to retire. θ^{rkk} measures the jump in the probability to retire at age 65.

D.2 Discontinuities in the Dutch pension system

In the Netherlands, households receive relatively generous public pensions at retirement. The public pension is a pay-as-you-go system and involves a flat-rate public pension benefit for all residents as from the statutory retirement age. Because of the rapid ageing of the population, the statutory retirement age, which has been 65 since the introduction of the public pension scheme in 1957, has been gradually raised in steps as of January 2013. Flexible take-up, such as in the United States, is not possible. The public pension benefits are a substantial component of total income in retirement and, depending on other sources, about 1/3 to 1/2 of total income in retirement (Knoef et al., 2016).

The fixed statutory retirement age combined with substantial benefits cause a discontinuous jump in the probability of retirement at age 65. Prior evidence has shown that households are very responsive to the (change in) institutional retirement ages which causes discontinuous retirement probabilities right before and after reaching the retirement age (see, among others, Duggan et al. (2007)). Such evidence of responsiveness to the statutory retirement age is also found for the Netherlands in recent papers by Atav et al. (2021) and Nagore Garcia et al. (2021). For a more extensive discussion on the discontinuity, we refer to Been and Goudswaard (2021).

D.3 LISS Time Use data

We use data from the LISS panel (Longitudinal Internet Studies for the Social sciences) administered by CentERdata at Tilburg University. The annual core waves consist of 5,000 representative households (about 8,000 individuals) in the Netherlands. These households are followed over the time period 2008-2019. The LISS Background Variables was supplemented with an additional module on Time Use and Consumption in the 2009, 2010, 2012, 2015, 2017, and 2019 waves. Time use categories are either available in all waves (like paid work) or are missing in 2015 and 2017 waves (like leisure time). For a more extensive overview of the LISS data for analyzing retirement and time use decisions, we refer to Been and Goudswaard (2021). Compared to Been and Goudswaard (2021), we analyze couples' joint decisions as the LISS allows us to study both the individual and household level and, hence, any interactions within the household.

To analyze time spend on paid work, leisure, and time spend with grandchildren, we use the following questions from LISS.

How much time did you spend in the last seven days on paid work (in employment or as self-employed; do NOT include the time spent traveling to and from work, but DO count overhours)?

How much time did you spend in the last seven days on leisure time activities (such as watching TV, reading, sports activities, hobbies, computer as hobby, visiting friends or family, traveling, going out, etc.)?

How much time did you spend in the last seven days on helping other family members (for instance assistance with administrative chores, washing, dressing, seeing the doctor, ***tending to the grandchildren****, etc.)*

For the analyses, we select heterosexual couples aged 55-75. Hence, we select those 10 years prior to the statutory retirement age of 65 and those 10 years after. We only select those households in which the female does not report the housekeeping or social benefits as primary occupation. This is to make sure that retirement actually measures going from an active paid labor state to retirement. When we do not make this selection we would make cross-sectional comparisons between women that work at an old age and retire and women who have not been active at all. However, these groups are not comparable as there tends to be selection in working at older ages.

This issue also explains why our first-stage results for grandmothers are not significant in the main analysis: there is only a small selective group of women that work at older ages in 2004. Most older women of the cohorts involved do not (re)consider paid work as many have already left the labor market.

In Figure D.1 , we show that both men and women show a discontinuous jump in the probability to retire. In Figures D.2, D.3 and D.4 we show the discontinuities in the dependent variables. We find no discontinuous jumps in our control variables (not shown here).

D.4 Estimation results

In Table D.1, we show the estimations results for the most important parameters of the SEM shown in Equations 1-8. To estimate the model, we only consider the waves prior to 2013 for two reasons: 1) leisure time is only available for the waves 2009, 2010, 2012 and 2) using the statutory retirement age of 65 is only valid before 2013 as the statutory retirement aged increased step-wise as of January 1st 2013.

To identify these estimates as causal, we use the discontinuity in the statutory retirement age. In Table D.2, we show the estimations results for this discontinuity. Upon reaching the statutory retirement age of 65, males show a 9 percentage points bigger probability of retiring (Column 1). This is 22 percentage points for females (Column 2). For males we find that the distance to the cutoff matters more than for females with about 8 percentage points per year from the cutoff versus 4 percentage points. We find no evidence of cross-effects of reaching the retirement age on retirement of the spouse. [72]

Table D.1 shows that male retirement decreases paid work by about 40 hours per week,

[72]The F-statistic on the excluded instruments for males is 295.69. The F-statistic is 33.11 for females.

increases leisure time by about 11 hours per week, and increases time spend with family members among which grandchildren with about 3 hours per week. Male retirement tends to have cross-effects with the female spouse. Male retirement decreases paid work by the spouse by about 10 hours per week, increases leisure of the spouse by about 7 hours per week, and increase time spend with grandchildren by about 4 hours per week.

Female retirement reduces paid work by about 14 hours per week suggesting part-time work prior to retirement, increases leisure by about 11 hours per week, and does not affect time spend with grandchildren. Contrasting male retirement, female retirement only has cross-effects with the male spouse in the case of leisure time. Female retirement increases leisure of the male spouse by about 8 hours per week.

These estimation results show us two interesting patterns. Firstly, cross-effects of retirement suggest that couples value joint leisure time. Secondly, while women spend more time with their grandchildren than men (See Figures 7 and 8), retirement only significantly increase time spend with grandchildren for males. The result that male retirement also increases grandmothers' time spend with grandchildren (by about the same number of hours) suggests that grandfathers are most likely to spend time with their grandchildren together with the grandmother.

This additional analysis supports our main analysis in two ways. Firstly, it corroborates the selectivity of grandmothers' employment at older ages. Secondly, it gives us an idea of the processes underlying retirement and time spend with grandchildren. In particular, we show that grandfathers' retirement tends to increase time spend with grandchildren of both the grandfather and grandmother. Before and after retirement, grandmothers spend more time with the grandchildren than grandfathers. Much of the grandfathers' time spend with the grandchild is most likely jointly spend with the grandmother. This is also in line with the joint leisure time of couples we find in this ancillary analysis.

Table D.1: The impact of retirement on time use

	(1) P_{itm}	(2) P_{itf}	(3) L_{itm}	(4) L_{itf}	(5) H_{itf}	(6) H_{itf}
γ_1^{jm} (Male)	-40.51***	-9.57***	10.71**	6.85*	2.91***	3.80***
	(3.49)	(2.84)	(4.51)	(3.85)	(1.07)	(1.46)
γ_1^{jf} (Female)	8.33	-14.24**	7.83**	11.16**	-0.90	-2.72
	(7.59)	(6.70)	(4.59)	(5.64)	(1.71)	(3.05)

Source: LISS Panel, 2008-2019.
Notes: * Significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level. Based on 2,209 observations. Robust standard errors are reported. The dependent variables in columns (1)-(6) are measured in hours per week. Control variables include education (low, high) and migration dummies for both spouses.

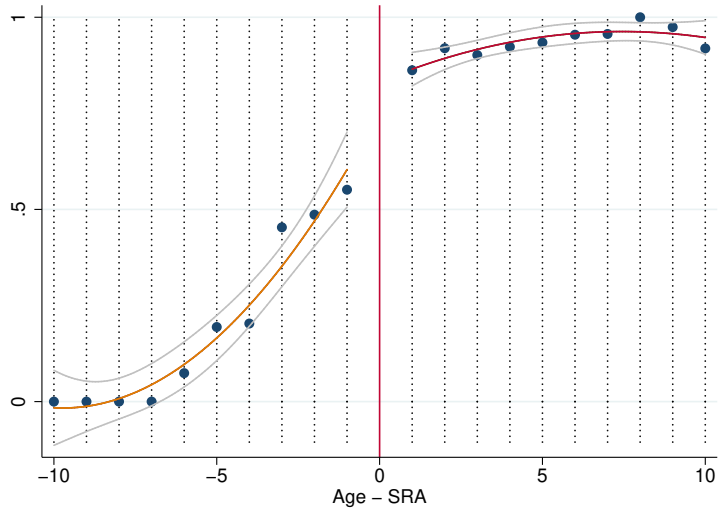
Table D.2: Estimation results retirement.

	(1) Male R_{itm}	(2) Female R_{itf}
$\theta_1^{jkm}\ D_{itm}$	0.09**	0.07
	(0.04)	(0.08)
$\theta_1^{jkf}\ D_{itf}$	0.02	0.22***
	(0.04)	(0.06)
$\phi_1^{jkm}\ D_{itm} \cdot A_{itm}$	-0.08***	0.00
	(0.00)	(0.01)
$\phi_1^{jkm}\ D_{itm} \cdot A_{itm}$	-0.01**	-0.04 ***
	(0.00)	(0.01)

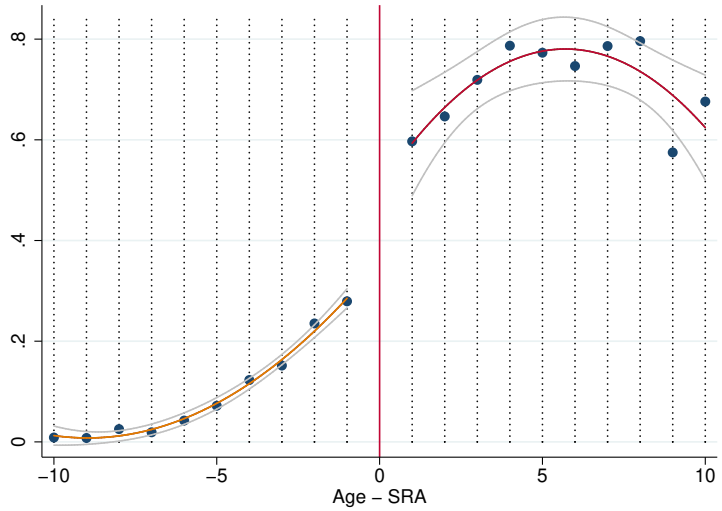
Source: LISS Panel, 2008-2019.
Notes: The dependent variables in columns (1)-(2) are dummies for retirement. Control variables include education (low, high) and migration dummies for both spouses. Based on 2,209 observations. Robust standard errors are reported in parethesis. * Significant at the 0.10 level; ** at the 0.05 level; *** at the 0.01 level.

Figure D.1: Percentage retired by distance to the SRA

(a) Male



(b) Female

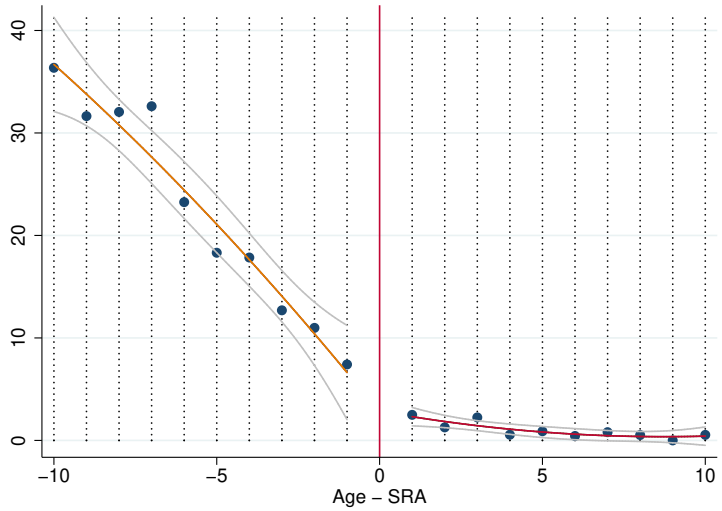


Source: LISS Panel, 2008-2019.

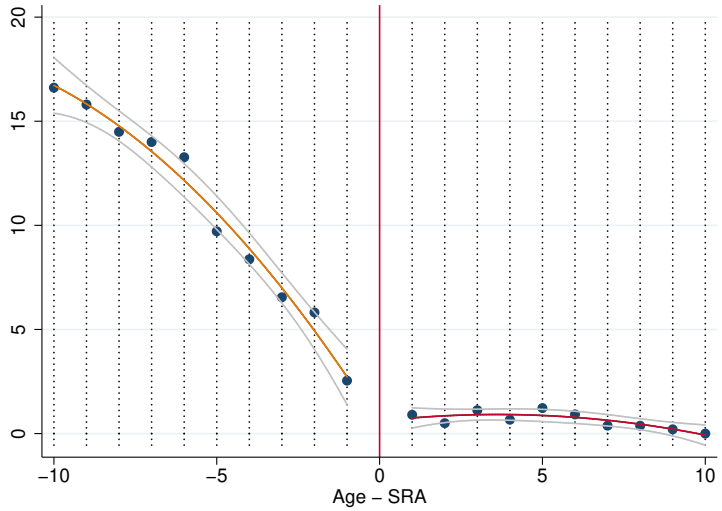
Notes: Fraction of retirees (dots), smooth quadratic function (red line) and 95% Confidence Intervals (grey lines).

Figure D.2: Hours work (p/w) by distance to the SRA

(a) Male



(b) Female

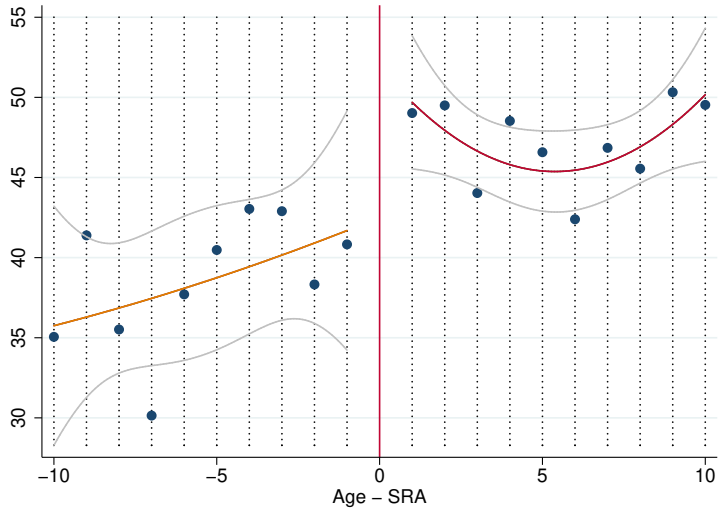


Source: LISS Panel, 2008-2019.

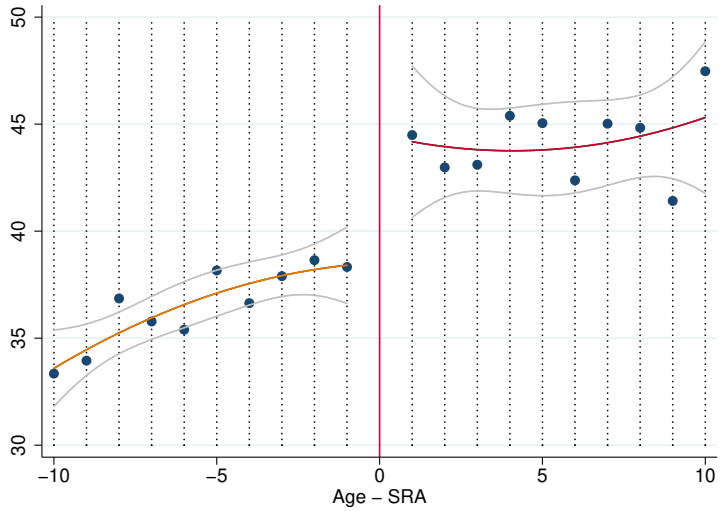
Notes: Fraction of retirees (dots), smooth quadratic function (red line) and 95% Confidence Intervals (grey lines).

Figure D.3: Hours leisure (p/w) by distance to the SRA

(a) Male



(b) Female

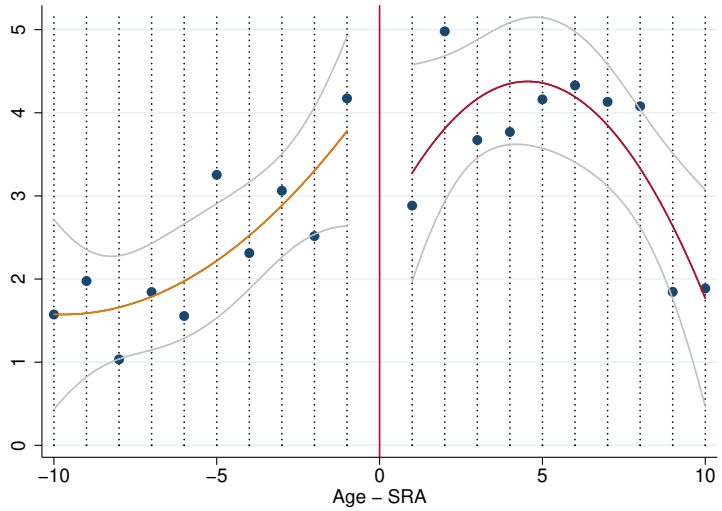


Source: LISS Panel, 2008-2019.

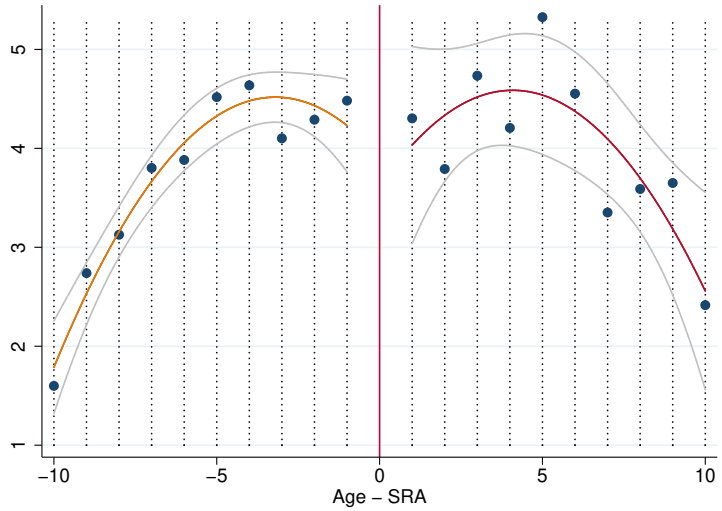
Notes: Fraction of retirees (dots), smooth quadratic function (red line) and 95% Confidence Intervals (grey lines).

Figure D.4: Hours care (p/w) by distance to the SRA

(a) Male



(b) Female



Source: LISS Panel, 2008-2019.

Notes: Fraction of retirees (dots), smooth quadratic function (red line) and 95% Confidence Intervals (grey lines).