Are the grandparents alright? The health consequences of grandparental childcare provision*

Peter Eibich† and Xianhua Zai‡

Abstract

March 2023

This paper examines the causal effect of childcare provision on grandparents' health in the US. We propose the sex ratio among older adults' children as a novel instrument for grandparental childcare provision. Our instrument is rooted in the demographic literature on grandparenthood and exploits that parents of daughters transition to grandparenthood earlier and invest more in their grandchildren than parents of sons. We estimate 2SLS regressions using data from the Health and Retirement Study. The results suggest that childcare provision is not beneficial for grandparents' health and may even be detrimental for physical functioning and subjective health.

Keywords: Grandparents, Childcare Provision, Instrumental Variables, Health

JEL classification I10, J13, J14, C26

*We would like to thank Benedicte Apouey and Eric Bonsang as well as participants of the workshop "Behaviours, Health and Ageing" at Université Paris Dauphine for their helpful comments.

†Université Paris Dauphine-PSL, Place du Maréchal de Lattre de Tassigny, 75016 Paris, France. peter.eibich@dauphine.psl.eu

‡Max Planck Institute for Demographic Research, Konrad-Zuse-Str. 1, 18057, Rostock, Germany. zai@demogr.mpg.de

1 Introduction

Grandparents across the globe play an important role in raising their grandchildren. For example, the US Census Bureau estimates that in 2011 4.8 million children under 5 (24%) received care from their grandparents (Laughlin 2013). In the UK, around 40% of grandparents provide regular care for their grandchildren, and 89% of these provide care at least once a week (Age UK 2017). In the EU, 21% of children under 3 received some childcare from sources other than their parents or formal childcare in 2020 with substantial variation across countries (Eurostat 2022).[1] Grandparental childcare provision can reduce the cost of childrearing for young parents by substituting for formal care or own childcare provision, in particular in contexts with strong family ties (Battistin et al. 2014; Chen et al. 2000). Even in contexts where formal childcare is both available and affordable, grandparents often make important contributions by offering a flexible alternative source of childcare, e.g., in case of illness or during school holidays.

While grandparenting is highly beneficial to parents (Compton 2015), the consequences for grandparents themselves are less clear. Looking after grandchildren might provide grandparents with physical and mental stimulation, thereby helping to maintain their health in old age. This would imply that grandparenting can be considered as "active ageing", i.e., an activity that benefits both older individuals and the wider society. Yet, keeping up with young children can also be physically strenuous and stressful. The negative health effects of informal care provision by older parents or spouses have been documented extensively in the literature (Bom et al. 2018; Bom and Stöckel 2021; Heger 2017; Schmitz and Westphal 2015). It seems possible that grandparents find caring for young children similarly demanding. Therefore, the overall effect of grandchild care provision on health of grandparents is ambiguous. In this paper, we empirically estimate the health effects of grandparenting for grandparents.

Previous studies on grandchild care provision and grandparents' health report

[1] Formal childcare here includes grandparents, other household members, and professional child carers.

contradictory findings. Several studies report that grandparents caring for their grandchildren are in better health, have fewer mobility limitations and fewer depressive symptoms (Danielsbacka et al. 2019; Di Gessa et al. 2016; Ku et al. 2013; Wang et al. 2020; Zeng et al. 2021). Yet, interpreting these estimates as causal is challenging. First, the transition to grandparenthood is not random. Lai et al. (2021) report that older adults expecting to become grandparents in the future are healthier than those who do not expect this transition. A possible explanation is that healthier individuals are, ceteris paribus, able to have more children than individuals in poor health, which in turn means that they are more likely to have grandchildren. Moreover, healthier parents are more likely to survive until they become grandparents and their lifespan overlaps longer with their grandchildren's lifespan (Margolis and Verdery 2019). Second, health is an important precondition for all activities in old age such that grandparents in poor health are less capable to provide grandchild care. Taken together, these arguments suggest that grandparents providing childcare are positively selected on health.

A few previous studies address the endogeneity of grandchild care provision. Ates (2017) finds that the positive association between grandparenting and health in Germany disappears when introducing individual-fixed effects. While fixed effects can resolve bias from selection on time-invariant unobservable characteristics (e.g., long-term health conditions or family size), it does not address potential reverse causality introduced by an unexpected health shock that reduces a grandparent's capacity to provide childcare. Brunello and Rocco (2019) and Ku et al. (2012) use instrumental variables (IVs) to address such endogeneity. Brunello and Rocco (2019) use data on European grandparents from the Survey of Health, Ageing and Retirement in Europe (SHARE). Their IV strategy exploits variation in the propensity of grandchild care provision due to the random timing of the survey and changes in the likelihood of grandchild care provision by the age of grandchildren. They find a sizable increase in depressive symptoms for grandparents providing childcare. Ku et al. (2012) examine Taiwanese grandparents and use marital

status of parents and the number of grandchildren as IVs. Their findings indicate that even after addressing the endogeneity, grandchild care provision is beneficial for Taiwanese grandparents' health.

This study examines the causal effect of grandparental childcare provision on grandparents' health in the US. We propose the sex ratio (defined as the number of daughters relative to the total number of children) as a novel instrument to address the endogeneity of grandchild care provision. Our IV is motivated by two insights from the demographic literature on grandparenthood. First, parents of daughters transition to grandparenthood earlier than parents of sons. Second, grandparents are more likely to provide childcare for grandchildren born to a daughter than for those born to a son. We use data from the Health and Retirement Study (HRS) with detailed health information on the number of functional limitations, self-rated health status, and depressive symptoms. Our results indicate that the effects of grandparenting on health are not significant and, in some instances, negative. We therefore conclude that grandparental childcare provision should not be considered as "active ageing". Similar to informal care provision, grandparents' care for their grandchildren can be detrimental to their own health.

This study contributes to the literature by estimating a credibly identified causal effect of grandchild care provision on grandparents' health. We propose a novel instrument motivated by the demographic literature on grandparenthood, and we conduct a battery of tests and falsification exercises that suggest that the IV assumptions likely hold. Our paper is also the first study to provide causal evidence in the US context. The contradictory findings by Brunello and Rocco (2019) and Ku et al. (2012) suggest that the health effects of grandparenting might be context-dependent. The US is a particularly interesting context characterized by both expensive formal childcare compared to some of the European countries examined by Brunello and Rocco (2019) and weaker family ties compared to East Asian societies (Ku et al. 2012).

The remainder of the paper is organized as follows. Section 2 describes the data, key

variables, and sample statistics. Section 3 first motivates our novel instrument, then describes our estimation strategy, and discusses the assumptions for our IV model. Section 4 presents our main results as well as several robustness checks. Section 5 discusses our findings and concludes.

2 Data

2.1 Sample Description

We use data from the HRS, a nationally representative longitudinal study of Americans aged 51 and above. Respondents are surveyed every other year since 1992. The survey includes different birth cohorts who enter the study as they become eligible. The core cohort, the HRS cohort, has been followed and interviewed since 1992. Since 1993, the HRS has included the Study of Assets and Health Dynamics Among the Oldest Old (AHEAD) cohort of individuals born before 1924; the Children of the Depression Age (CODA) cohort of people born between 1924 and 1930; and the War Babies cohort (WB) of individuals born between 1942 and 1947. An additional Early Baby Boomers (EBB) cohort of people born between 1948 and 1953 was added to the sample in 2004, and the Mid-Baby Boomers cohort of individuals born between 1954 and 1959 was added in 2010. Liu and Zai (2022) document the details on how respondents enter each survey wave and the number of unique individuals in each wave.

The HRS asks respondents (including cohabiting spouses) detailed information about their own demographic characteristics, health, employment, financial situation, and intergenerational transfers as well as demographic information about their family members such as children and parents. To explore the effect of grandparenting on grandparents' health, we restrict our working sample to HRS respondents aged between 40 to 70 who have at least one child. On the one hand, we aim to include as many potential grandparents as possible to maximize the sample size. On the other hand, we are concerned about the validity of our instrument if we include individuals older than 70.

This older population is more likely to be frail and dependent and thus not able to provide grandchild care. We check the sensitivity of our results to this age restriction in section 4.2 using a sample without age limits. The distribution of age of respondents in Appendix Figure A1 is almost symmetric around 70.

Our study sample includes 84,483 observations (20,508 unique individuals) and covers the period from 1996 (wave 3) to 2014 (wave 12) in which the HRS asks respondents consistent questions on grandparenting.

2.2 Dependent Variables

The HRS includes detailed information on the health outcomes of respondents. We mainly focus on three dimensions of health: self-reported health status, physical health, and mental health.

First, the HRS asks respondents to self-report their general health status. Possible answers range from 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". We create an indicator for poor self-reported health, which is one if self-reported health is "fair" or "poor", and zero otherwise. Similar indicators of fair or poor health are also employed in Dave et al. (2006) using the HRS data, in Eibich (2015) using the German Socio-Economic Panel Study (SOEP), and in Kuka (2020) using the Survey of Income and Program Participation (SIPP). We also consider alternative cut-off points to show robustness in section 4.2. While self-reported health is subjective and might be affected by reporting heterogeneity, it is a good predictor of mortality (Idler and Benyamini 1997; DeSalvo et al. 2006; Kuka 2020).

Second, we use more objective measures about physical health conditions. The HRS provides indices of functional limitations, such as limitations in Activities of Daily Living (ADLs) and limitations in Instrumental Activities of Daily Living (IADLs). The ADLs include items such as bathing, eating, dressing, getting in or out of bed, and walking across a room and the IADLs assess difficulties in using the phone, managing money, taking

medications, shopping for groceries, and preparing hot meals.[2] All these indices range from 0 to 5. An index with a value of 5 means that an individual has difficulties with all activities considered, while a value of zero means that the individual has no limitations. We create dichotomous indicators which equal one if an individual reports limitations for four or more items, and zero otherwise. We consider other operationalizations as robustness checks in section 4.2.

Third, we further use information about respondents' mental health. The HRS asks respondents about their mental health using the Center for Epidemiologic Studies Depression (CES-D) score. The CES-D score captures the number of adverse sentiments a respondent experienced all or most of the time in the past two years, including whether an individual was depressed, felt alone, felt sad, had restless sleep, felt everything was an effort, could not get going, felt unhappy, and did not enjoy life. The CES-D scale has been validated as an instrument to identify major depression in older adults (Irwin et al. 1999). The main results in section 4 reports the effect of grandparenting on a binary indicator constructed using five or more depressive symptoms as the cutoff (8 items in total). Other alternative measures are considered as robustness checks in section 4.2.

2.3 Treatment Indicators

Our treatment variable of interest is whether individuals provide grandchild care or not. There are two relevant grandparenting questions in the data. First, the HRS asks respondents whether they and their spouse spent 100 or more hours taking care of their grandchildren or great-grandchildren since the last wave.[3] If the answer is yes, respondents are asked to which child they provided grandchild care. This question was not asked in waves 1 and 2.[4]

[2]More details on the construction of these measures can be found in Chien et al. (2015).

[3]The question asked in the HRS is "Did you or your husband/wife/partner or your late husband/late wife/late partner spend 100 or more hours in total in the last two years taking care of great-grandchildren/grandchildren?"

[4]In wave 2 of the HRS, the AHEAD cohort was asked whether grandchild care was provided for a year or longer. This question is no longer asked from wave 3. The question is "Which of your children is the parent of those grandchildren (or great-grandchildren)?"

Second, the HRS asks respondents to estimate their childcare hours provided in the last two years. This question is asked separately for the respondent and the spouse.[5] For those who cannot remember the hours or do not know the exact hours or refuse to give the number of hours on grandparenting, the HRS further asks the minimum and maximum values of hours of grandchild care provided.[6] There are both advantages and disadvantages of using each question to construct our treatment indicator of grandparenting. The first question does not distinguish between grandchild care provided by respondents or their spouses, which would introduce measurement errors since we are interested in estimating the health effects on those who are actually grandparenting. On the other hand, the question only requires respondents to answer "yes" or "no" and might thus be less affected by recall bias than asking for the exact number of grandchild care hours provided over the last two years.

For our main analysis, we use the self-reported number of hours of grandchild care by respondents. Among those who are grandparenting, the majority of grandparents provide less than 1,000 hours over two years. Appendix Figure A2 shows the distribution of grandchild care hours reported by HRS respondents for these grandparents. We construct a binary indicator of grandparenting status, which indicates whether the respondent reported 100 or more hours of grandchild care over the last two years. If the number of hours is missing and the minimum and maximum values are above 100, we assume that the respondent is grandparenting. To examine the potential measurement errors in the treatment variable, we use the first question on childcare provided by the respondent and their partner to construct an alternative treatment indicator, which is 1 if the answer is "yes", i.e., the respondent and their partner provided at least 100 hours of childcare since the last wave. We also explore other cutoffs for the self-reported number of childcare hours as robustness checks in section 4.2.

[5]The grandchild care hour question in the HRS is "Roughly how many hours altogether did you spend since the last wave?" for the respondent and "Roughly how many hours altogether did your husband/wife/partner spend since the last wave?" for respondent's spouse.

[6]The quote in HRS is "Did it amount to a total of less than MAX BREAKPOINT, more than MIN BREAKPOINT, or what?" The MIN BREAKPOINTs are 0, 200, 201, 500, and 501. The MAX BEAKPOINTs are 199, 200, 499, 500, and 5,000.

2.4 Sample Statistics

Table 1 presents the summary statistics of the working sample of HRS respondents who are between 40 to 70 in each survey year. The average age of the sample is around 61. About 60 percent of the sample are female. The average educational attainment of the sample is around 12 years. On average, each respondent has between three to four children. The oldest child is on average about 40 years old and the youngest child is on average about 31 years old. About half of the respondents' children are daughters. The majority of the sample is married or living with a partner and white. Approximately 33 to 39 percent of respondents provide some grandchild care according to the different definitions discussed earlier.

The average self-reported health status of respondents is good and about 30 percent of respondents report fair or poor health status. The average ADL score and IADL score are both close to zero, which indicates that our sample is relatively healthy. The share of respondents with more than three functional limitations in ADL or IADL is about 2 percent. The average CESD depression score is 1.5 out of 8 and 12 percent of individuals have four or more depressive symptoms. Detailed definitions of these variables are provided in Appendix Table A1.

3 Methods

In this section, we review findings from the demographic literature on grandparenthood and grandparenting to motivate the sex ratio as our valid IV for providing grandchild care. Then we discuss the estimation strategy and provide evidence for the assumptions required for a causal interpretation in the IV framework.

3.1 Sex Ratio as an IV for Grandparenting

The transition to grandparenthood as well as the decision to provide grandchild care are endogenous choices, which depend on many factors that are plausibly related to health.

Table 1: Summary Statistics of the Sample

Variable	Mean	S.D.	Obs.
Demographics			
Age	60.89	5.98	84,483
Female	0.60	0.49	84,483
Education (years)	12.24	3.13	84,268
Number of children	3.72	2.02	84,483
Age of youngest child	30.54	8.37	84,188
Age of oldest child	39.75	6.99	84,188
Income from earnings ($1,000)	16.41	33.40	84,483
Income from pensions ($1,000)	3.91	20.68	84,483
Marital status			
Married/partnered	0.76	0.43	84,403
Separated/divorced	0.13	0.34	84,403
Widowed	0.10	0.29	84,403
Never married	0.02	0.12	84,403
Race/ethnicity			
White	0.75	0.43	84,348
Black/African	0.18	0.39	84,348
Other	0.07	0.25	84,348
Instrumental variable			
Sex ratio	0.50	0.29	84,354
Grandparenting			
Grandparenting for at least one child (Q1)	0.39	0.49	84,382
Grandparenting for at least 100 hours (Q2)	0.34	0.47	84,483
Heath variables			
Self-report health	2.82	1.12	84,444
Self-report fair or poor health	0.27	0.45	84,444
ADL limitations	0.26	0.80	84,426
Percentage of having ADL limitations (3+)	0.02	0.13	84,426
IADL limitations	0.19	0.67	84,420
Percentage of having IADL limitations (3+)	0.01	0.11	84,420
CESD score	1.54	2.05	79,477
Percentage of having depressive emotions (4+)	0.12	0.32	79,477

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. The definitions of these variables can be found in Appendix Table A1.

For example, parents with larger families are more likely to become grandparents (Margolis and Verdery 2019), and parents who give birth earlier in life are more likely to become grandparents at younger ages. Family size and age at the first birth have been linked to health and mortality of mothers in particular (Mirowsky 2005; Wu and Li 2012), but they are also related to socioeconomic status (Adserà 2017). Beyond the transition to grandparenthood, grandparents' capacity to provide grandchild care depends, among other factors, on the proximity between grandparents and their adult children (Compton 2015), as well as their health (Eibich and Siedler 2020).

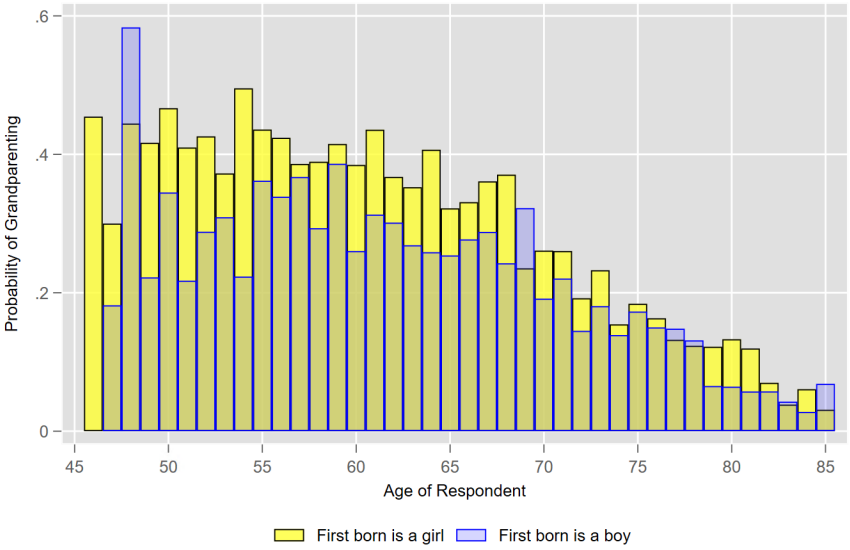
In this study, we address the endogeneity of grandparenting by using the sex ratio, defined as the number of daughters divided by the total number of children of a respondent, as an instrument for grandparenting.[7] Our choice of this instrument is motivated by a few stylized facts derived from the demographic literature on fertility and grandparenthood.

It is well-documented that women tend to give birth earlier than men (Margolis and Verdery 2019). This implies, *ceteris paribus*, that parents of daughters will transition to grandparenthood earlier than parents of sons. The gender of a child can be considered as good as randomly determined, thus the gender of a person's first-born child may serve as a suitable instrument that predicts the transition into grandparenthood (and subsequently grandparenting) (Rupert and Zanella 2018). This is also borne out in our data: Fig. 1 shows the share of individuals who are grandparenting by age for individuals with a first-born daughter and those with a first-born son, respectively. At most younger ages, older adults with a first-born daughter are much more likely to provide grandchild care than those with a first-born son. This gap narrows substantially with age and mostly disappears beyond age 70.

Second, older adults with a higher number of daughters are overall more likely to become

[7]The total number of children is defined very broadly and potentially includes deceased as well as non-biological children (e.g., adopted or step-children). The number of living children is arguably a more relevant predictor of grandparenting, however, selective mortality among male children might bias our results. Reassuringly, our results remain robust using the total number of living children to define the sex ratio instrument (results available on request).

Figure 1: Sex of the first-born child and grandparenting

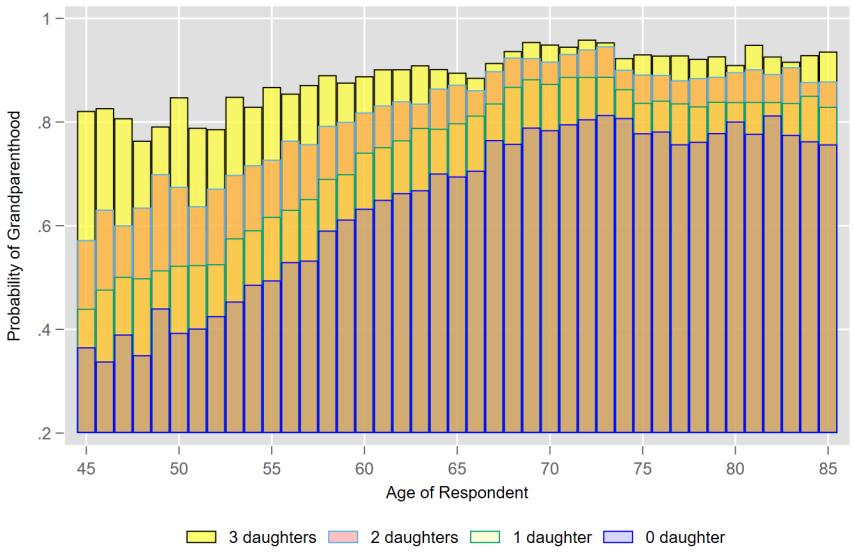


Notes: The data is the HRS from 1996 to 2014. The sample is limited to individuals with only one child. This graph draws the share of individuals who provide some grandchild care by age for individuals with a first-born daughter and those with a first-born son, respectively. Grandparenting is defined as an indicator that is 1 if the estimated grandchild care hours reported by respondents are at least 100 hours.

grandparents and they transition into grandparenthood at younger ages (see Fig. 2). This is in part because the likelihood of becoming a grandparent increases with the number of children (Margolis and Verdery 2019). Some children may remain childless (by choice or involuntarily). Even if the probability of remaining childless is correlated between siblings, we would expect that the likelihood of all children remaining childless is smaller for an older parent with three children than for an older parent with one child.

Third, women show less variation in age at first birth than men (Margolis and Verdery 2019), thus older parents with a first-born son and a second-born daughter might still on average transition into grandparenthood earlier than older parents with two sons. Fig. 3 shows the likelihood of grandparenthood for older adults with two children. Until about age 60, the likelihood of becoming a grandparent for parents with two daughters is considerably higher than for all other groups. The likelihood is slightly higher for parents with a first-born daughter and a second-born son than for parents with a first-born son and a second-born daughter, although differences between these two groups are relatively small. The likelihood

Figure 2: Grandparenthood and the number of daughters

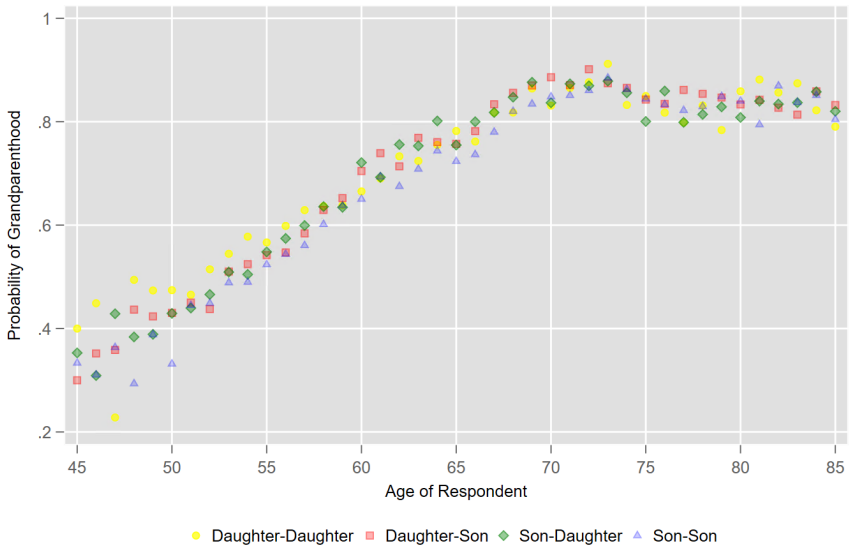


Notes: The data is the HRS from 1996 to 2014. This graph draws the share of individuals who are grandparents by age for all individuals with at least one child. Grandparenthood is defined as an indicator that is 1 if the number of grandchildren reported by the respondent is at least one.

of becoming a grandparent tends to be the lowest for parents of two sons, although the differences between groups largely vanish from age 70 onward.

In addition, the sex of a child not only affects their parents' likelihood to become grandparents, but also the extent of their involvement with the grandchild. Maternal grandparents invest more time into grandchild care than paternal grandparents (Compton and Pollak 2011; Danielsbacka et al. 2011). Several explanations have been proposed for this phenomenon – uncertainty around paternal kinship (Danielsbacka et al. 2011), stronger bonds between mothers and daughters (Somary and Strieker 1998), and a longer shared lifetime between maternal grandparents and their grandchildren (Margolis and Verdery 2019). Somary and Strieker (1998) find few differences in grandparenting behavior across lineage, but note that they control for proximity between grandparents and grandchildren, which plays an important role for grandchild care investments (Compton 2015). We find a similar pattern in our data. While there is considerable variation for younger grandparents, from age 60 onwards until about age 75 we observe that older parents with two daughters

Figure 3: Grandparenthood for older parents with two children

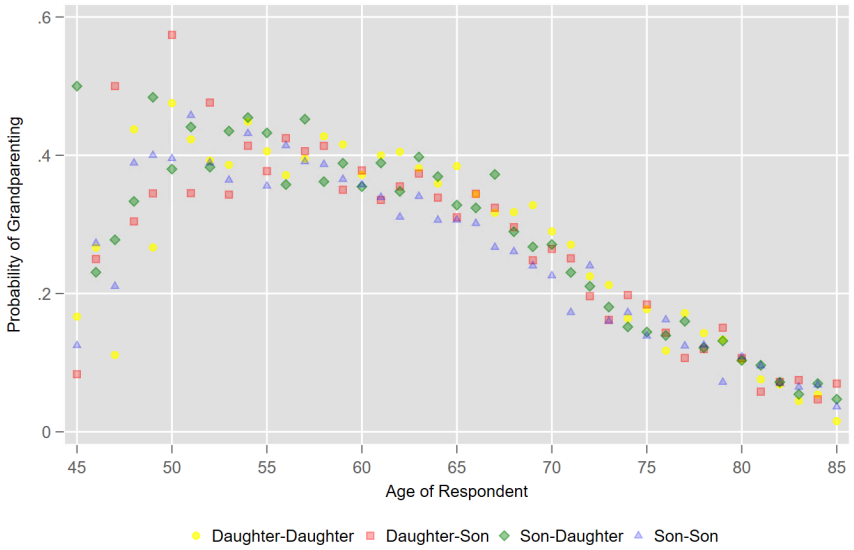


Notes: The data is the HRS from 1996 to 2014 with the respondents limited to having two children. This graph draws the share of individuals who are grandparents by age for individuals with two children. Grandparenthood is defined as an indicator that is 1 if the number of grandchildren reported by the respondent is at least one.

are most likely to provide grandchild care, whereas grandparents with two sons tend to be the least likely group to provide grandchild care (Fig. 4).

In summary, the literature suggests that the number of children, the sex of children as well as their birth order are related to the likelihood of becoming a grandparent, the age at the transition into grandparenthood as well as grandparental investment into their children. The number of children itself is endogenous and might be related to health of the older parents (Wu and Li 2012). Similarly, the absolute number of daughters is clearly correlated with family size and therefore not a suitable instrument. On the other hand, the sex of the first-born child ("birth order instrument") is plausibly exogenous, but exploits very limited variation. We therefore propose to use the sex ratio as an instrument. In contrast to the absolute number of daughters, the sex ratio is not correlated with family size ($\rho = -0.0029$ in our sample), because we essentially condition on the total number of children. The sex ratio does not draw on variation in the birth order of children, but the instrument accounts for variation stemming from parents of daughters becoming grandparents earlier on average

Figure 4: Grandparenting for older parents with two children



Notes: The data is the HRS from 1996 to 2014 with respondents limited to having two children. This graph draws the share of individuals who are grandparents by age for individuals with two children. Grandparenting is defined as an indicator that is 1 if the estimated grandchild care hours reported by respondents are at least 100 hours.

than parents of sons and being more likely to provide grandchild care, regardless of the number of children they have. For example, the birth order instrument and the sex ratio instrument exploit the same variation – parents of a daughter as compared to parents of a son – for parents with one child. For parents with two children, the birth order instrument only distinguishes between parents with a first-born daughter and parents with a first-born son, regardless of the sex of the second child. The sex ratio instrument distinguishes parents with two sons from parents with one daughter and those with two daughters, under the assumption that parents with two daughters are most likely to provide grandchild care and parents with two sons are least likely to provide grandchild care. Since 88% of the HRS sample has two or more children, the sex ratio instrument provides much more variation than the birth order instrument. Therefore we expect that the sex ratio instrument should be stronger than the birth order instrument.

3.2 Model Specification

Using the sex ratio as an instrument for grandparenting, we estimate the first stage of our IV model as follows:

$$Grandparent_{it} = \delta Sexratio_{it} + X_{it}^{'}\beta + \epsilon_{it} \tag{1}$$

where $Grandparent_{it}$ is the grandparenting status of an individual i in year t. $Sexratio_{it}$ is the ratio of the number of daughters to the number of children of an individual i in year t. X_{it} is a vector of covariates. In our preferred specification, we control for individual demographic characteristics such as age (quadratic polynomial), race/ethnicity, religion, gender, birth place and census region fixed effects as well as year-fixed effects, fixed effects for the year of birth of the first-born child, age of the youngest child, cohort fixed effects of the individual, family size and control variables for the economic situation such as the annual income from earnings[8] and the annual income from all pensions and annuities. ϵ_{it} are the standard errors clustered at the individual level which allow for correlation within individuals across waves.

We estimate the effect of grandparenting on health in the second stage of the model as follows:

$$Y_{it} = \alpha Grandparent_{it} + X_{it}^{'}\eta + \mu_{it} \tag{2}$$

where Y_{it} is an indicator of the health status of individual i in year t. The other controls X_{it} are the same as in equation 1. We estimate our IV model using linear two-stage least squares estimation (2SLS). Although we use longitudinal data, the model does not include individual fixed effects for two reasons: (i) If the required IV assumptions hold, the inclusion of individual fixed effects is not necessary for causal identification, (ii) the sex ratio instrument is time-invariant and would therefore drop out of the model when applying the fixed effects transformation.

[8]This includes salaries, bonuses, overtime pay, commissions or tips, second job earnings, and professional practice or trade income.

3.3 IV Assumptions

The interpretation of our IV estimates as causal effects requires three assumptions: (i) reliability, i.e., the sex ratio should be correlated with grandparenting, (ii) validity, i.e., the sex ratio should be as good as randomly assigned and should not affect health through any other mechanisms than through grandchild care provision, and (iii) monotonicity, i.e., the sex ratio should affect the likelihood of grandparenting in the same direction (non-negative in this case) for all observations in our sample. In this section, we will discuss the plausibility of these assumptions in details.

3.3.1 Reliability

Table 2 shows estimates of the first-stage using equation 1. In column 1, we regress our indicator of grandparenting only on the sex ratio instrument. The estimate suggests that *ceteris paribus* older adults that only have daughters (i.e., a sex ratio of 1) are 6.4 percentage points more likely to provide grandchild care than older adults with only sons (i.e., a sex ratio of 0). For parents with two children, this would imply that every daughter increases the likelihood of grandparenting by 3.2 percentage points. In columns 2-6, we successively add control variables to account for standard demographic characteristics as well as socioeconomic differences of grandparents. The point estimate of the sex ratio instrument is barely affected by the introduction of these controls. The Kleibergen-Paap F-statistic on the strength of the excluded instrument is larger than 53 in all models, which exceeds thresholds that have traditionally been used as a rule-of-thumb. This suggests that the sex ratio is indeed a sufficiently strong predictor of grandparenting, i.e., the reliability assumption holds.

3.3.2 Validity

The validity assumption consist of two parts - exogeneity of the instrument and the exclusion restriction. Exogeneity of the instrument implies that the instrument should be as good as

Table 2: First stage estimates

Dependent variable: Grandparenting						
Model	1	2	3	4	5	6
Sex ratio	0.064***	0.064***	0.064***	0.063***	0.063***	0.063***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Year FE + Birth year FE of first-born		Y	Y	Y	Y	Y
Cohort FE + Birth year FE of youngest-born			Y	Y	Y	Y
Demographics				Y	Y	Y
Familiy size					Y	Y
Socioeconomic controls						Y
Mean of dependent variable	0.34	0.34	0.34	0.34	0.34	0.34
Observations	84,354	84,276	84,187	83,709	83,709	83,709
1st stage Kleibergen-Paap F-statistic	53.4	53.6	54.3	53.9	53.9	53.9

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell reports estimates from a separate specification using equation 1. The sex ratio is defined as the number of daughters divided by the total number of children of an individual. Grandparenting is defined as an indicator that is 1 if the estimated grandchild care hours reported by respondents are at least 100 hours. Column 1 reports estimates without any controls. Column 2 adds year-fixed effects and fixed effects for the year of birth of the first-born child of an individual. Column 3 adds age of the youngest child and cohort fixed effects of the individual. Column 4 includes individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects. Column 5 further controls for the number of children of individuals. Column 6 adds socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

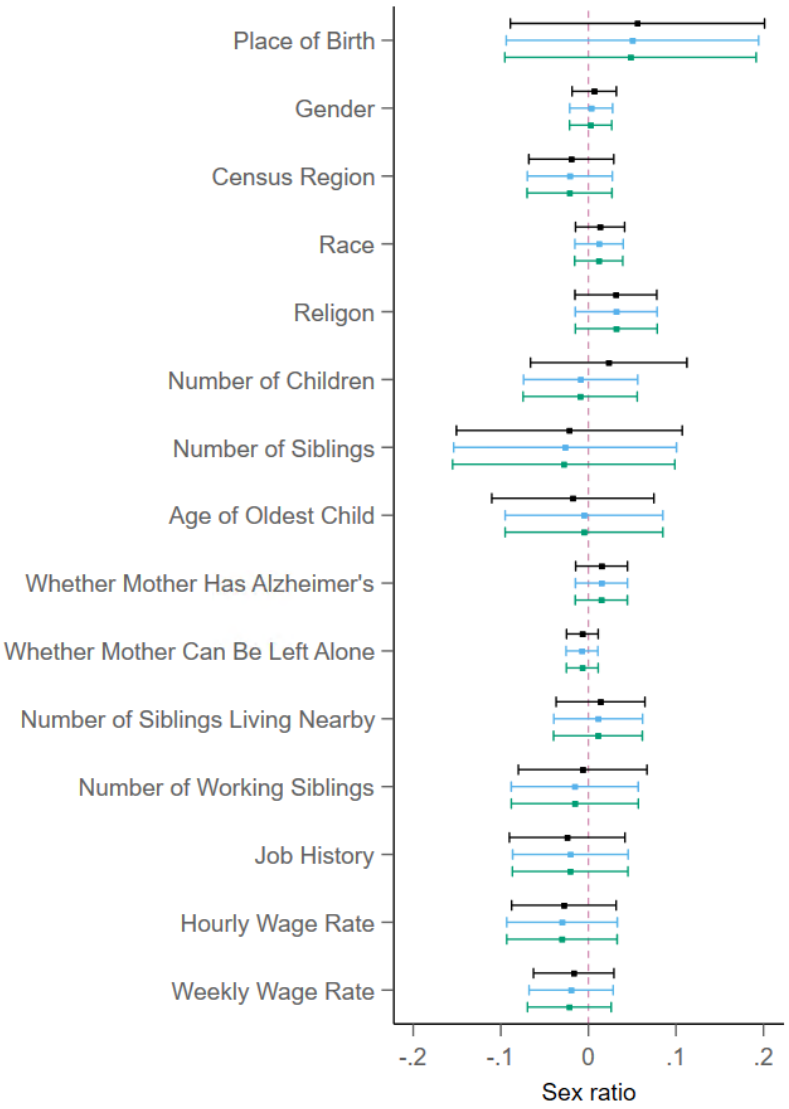
randomly assigned. This does not require that the probability of giving birth to a son is the same as the probability of giving birth to a daughter, rather it means that the probability of having a daughter should not be correlated with any characteristics of the parents. We argue that this assumption is highly plausible. To examine potential violations of this assumption, we check for covariate balancing. If the sex ratio is as good as randomly assigned, we would expect that the distribution of covariates that are not affected by the treatment should be similar across the different values of the instrument, or put differently, there is no significant correlation between such covariates and our instrument. If we would find a significant correlation, this would imply that the instrument might be affected by selection. We regress a battery of covariates as dependent variables on our sex ratio IV controlling for *(i)* year and first-born fixed effects (Model 2 of Table 2), *(ii)* additional cohort fixed effects for the individual, age fixed effects of the youngest child as well as individual demographic

characteristics (Model 4 of Table 2), and *(iii)* a full specification including also annual pension income and earnings (Model 6 of Table 2).

Figure 5 shows the point estimates and confidence intervals of the regression coefficients on the sex ratio for each dependent variable listed on the vertical axis in three specifications. For all covariates, estimates remain similar across models and none of these estimates is significant at the 95% confidence level, i.e., there is no evidence for a potential violation of exogeneity of the instrument.

The exclusion restriction requires that the sex ratio should not affect older parents' health through any other pathway than its effect on grandparenting. Previous studies have documented that parents of daughters differ in several aspects from parents of boys, e.g., parents of daughters are at a higher risk of divorce (Kabátek and Ribar 2021), they are more likely to vote for left-wing political parties (Oswald and Powdthavee 2010), and daughters provide more informal care (Dahlberg et al. 2018) . We argue that despite these findings the exclusion restriction should hold in our case. First, we note that the exclusion restriction does not imply that the instrument should not affect any other variables than our treatment. A violation of this restriction would only occur if these other outcomes are in turn related to health status at holder ages. It is not clear, e.g., why a temporary increase in the risk of divorce or voting preferences should have a causal effect on older adults' health (Oswald and Powdthavee 2010; Kabátek and Ribar 2021). While receipt of informal care is plausibly related to health at older ages, we would expect that this primarily occurs later in life, i.e., any provision of grandchild care would occur before the grandparents require care themselves. We also conduct another falsification exercise following Angrist et al. (2010) to detect potential violations of the exclusion restriction. We estimate the reduced form regression for a subsample of HRS respondents without grandchildren. In this subsample, the sex ratio instrument is not supposed to predict grandparenting (i.e., the treatment) since none of the respondents have grandchildren. The exclusion restriction requires that the instrument is associated with the outcome only through its effect on the treatment. This

Figure 5: Covariate balance for the sex ratio instrument



Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. The graph shows the point estimates and confidence intervals of the regression coefficients on the sex ratio for each covariate as the dependent variable listed on the vertical axis in models 2, 4, 6 of Table 2. The sex ratio is defined as the number of daughters divided by the total number of children of an individual. Grandparenting is defined as an indicator that is 1 if the estimated grandchild care hours reported by respondents are at least 100 hours.

implies that there should be no significant relationships between the sex ratio instrument and the health outcomes in the reduced form regression for individuals without grandchildren, because there is no valid first stage in this subsample. A significant reduced form estimate signals a violation of the IV validity because it would suggest that the sex ratio is related to

health through pathways other than grandparenting.

Table 3 reports the reduced form estimates from our preferred model specification for the subsample of respondents who do not have grandchildren. None of the estimates is statistically significant, and the point estimates are very small and close to zero (the reduced form estimates for our working sample are shown in Appendix Table A2 for comparison).

Table 3: Falsification exercise: Reduced form regression

Dependent variable	ADL	IADL	Poor subjective health	Depressed
Sex ratio	0.002	-0.004	0.017	-0.009
	(0.006)	(0.005)	(0.019)	(0.015)
Mean of dependent variable	0.021	0.014	0.302	0.121
Number of clusters	5,904	5,902	5,907	5,460
Observations	6,393	6,391	6,395	5,915

Notes: The data used are from the HRS 1996 to 2014 of the subsample of respondents who do not have grandchildren. All models control for year-fixed effects, fixed effects for the year of birth of the first-born child of an individual, age of the youngest child, cohort fixed effects of the individual, individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects, the number of children of individuals, and socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

The subsample of individuals without grandchildren consists of individuals that will become grandparents later in life as well as individuals that will never become grandparents. The latter group is likely highly selected and it may be possible that we fail to detect any violations of the exclusion restriction due to this selection bias. We therefore repeat the falsification exercise using only observations from individuals that are not yet grandparents, but who are observed to become grandparents at a later point in the panel. The results in appendix Table A3 are in line with our earlier findings, i.e., we fail to detect any violations of the exclusion restriction.

None of these falsification exercises provides evidence that the validity assumption is violated. While we cannot rule out such violations, the results improve our confidence that the sex ratio is a valid instrument for grandchild care provision.

3.3.3 Monotonicity Assumption

We test the monotonicity assumption by re-estimating the first stage regression of equation 1 for different subgroups within our working sample. The monotonicity assumption is satisfied if our instrument affects treatment in the same direction for all observations in the sample. This implies that the estimated effect of the sex ratio on grandparenting should be positive or zero for any arbitrarily defined subsample within our working sample. A significant negative effect in the first stage would imply a violation of monotonicity.[9]

Table 4 shows estimates of the first-stage regression for ten different subsamples defined by demographic characteristics. We note that the size of the first-stage estimate varies considerably – between 3.3 percentage points for men and 8.9 percentage points for older adults with more than four siblings. This suggests that the compliers differ from the overall population, but it does not suggest any violation of monotonicity. For all subsamples, the sex ratio instrument increases the likelihood of grandparenting and estimates are strongly significant.

In summary, these tests and falsification exercises suggest that the three assumptions required for a causal interpretation of our IV estimates are likely to hold.

4 Results

4.1 The effect of grandparenting on health

Before estimating the causal effect of grandparenting on health using the sex ratio instrument, we examine the relationship between grandparenting and health using ordinary least squares regression (OLS). Comparing OLS and 2SLS estimates will provide an indication of the size and direction of the bias caused by the endogeneity of grandparenting. Table 5 shows the results for our preferred model specification for all four health outcomes. Note that all health

[9] A negative but insignificant point estimate might either reflect a true zero effect (which does not violate monotonicity) or a violation of monotonicity.

Table 4: Falsification exercise: Monotonicity

Dependent variable: Grandparenting										
Subsample	Male	Female	$\leq$ 4 Children	$>$ 4 Children	$\leq$ 4 Siblings	$>$ 4 Siblings	Married	Not Married	White	Not White
Sex ratio	0.033***	0.082***	0.063***	0.058**	0.054***	0.089***	0.058***	0.084***	0.055***	0.087***
	(0.013)	(0.011)	(0.009)	(0.023)	(0.010)	(0.016)	(0.010)	(0.014)	(0.010)	(0.016)
Mean of dependent variable	0.285	0.377	0.349	0.318	0.341	0.339	0.349	0.313	0.343	0.332
Number of clusters	8,664	11,600	15,361	6,014	15,494	6,020	16,176	6,570	14,633	5,631
Observations	33,570	50,139	60,830	22,879	61,675	21,538	62,751	20,879	62,832	20,877
Kleibergen-Paap F-statistic	6.71	52.89	47.45	6.44	29.56	29.49	33.60	33.92	30.70	28.85

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell shows first stage estimates from our preferred model specification in equation 1 in each subsample. All models control for year-fixed effects, fixed effects for the year of birth of the first-born child of an individual, age of the youngest child, cohort fixed effects of the individual, individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects, the number of children of individuals, and socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

indicators are binary measures where a value of 1 represents worse health outcomes.

Table 5: OLS estimates

Dependent variable	ADL	IADL	Poor subjective health	Depressed
Grandparenting	-0.011***	-0.009***	-0.039***	-0.019***
	(0.001)	(0.001)	(0.004)	(0.003)
Mean of dependent variable	0.018	0.011	0.274	0.115
Number of clusters	20,255	20,255	20,261	19,575
Observations	83,655	83,650	83,673	78,783

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell shows OLS estimates from our preferred model specification with full controls for each dependent variable. ADL and IADL are dichotomous indicators which equal one if an individual reports limitations for more than three items, and zero otherwise. Poor subjective health is an indicator for poor self-reported health, which is one if self-reported health is "fair" or "poor", and is zero otherwise. Depressed is a binary indicator constructed using 4 items of depressive symptoms as the cutoff (8 items in total). All models control for year-fixed effects, fixed effects for the year of birth of the first-born child of an individual, age of the youngest child, cohort fixed effects of the individual, individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects, the number of children of individuals, and socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Overall, the results of Table 5 show that grandparenting is associated with better health across all dimensions of health considered here. The estimates suggest that grandparents caring for their grandchildren have fewer limitations in ADL and IADL, they rate their subjective health as better, and they show fewer depressive symptoms. Although these results seemingly support the notion that active ageing is beneficial for older adults' health, they should not be interpreted as causal effects. It is plausible that grandparents in good health are more likely to provide childcare than grandparents with poor health, which implies that these estimates may reflect reverse causality rather than the causal effect of grandparenting on health.

The results from our 2SLS regressions in Table 6 confirm the presence of such reverse causality. The point estimates for all four health outcomes are positive, suggesting that grandparenting leads to worse health in the form of more functional limitations, more depressive symptoms, and worse self-reported health. However, only the effects on IADLs

and subjective health are statistically significant at the 10% confidence level. Tables A4-A7 in the online appendix show that these results are overall robust to the inclusion or exclusion of covariates. The increase in functional limitations of IADL corresponds to change of 0.47 standard deviations, whereas the change in subjective health corresponds to a change of 0.60 standard deviations. These are substantial effect sizes, which suggest that childcare provision can be a strenuous activity for grandparents.

Table 6: 2SLS estimates

Dependent variable	ADL	IADL	Poor subjective health	Depressed
Grandparenting	0.032	0.052*	0.269*	0.097
	(0.039)	(0.028)	(0.147)	(0.091)
Mean of dependent variable	0.018	0.011	0.274	0.115
Number of clusters	20,255	20,255	20,261	19,575
Observations	83,655	83,650	83,673	78,783
Kleibergen-Paap F-stat.	53.66	53.61	54.12	54.56
AR F statistic	0.667	3.945	3.674	1.16
AR p-value	0.414	0.047	0.055	0.281

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell shows 2SLS estimates from our preferred model specification with full controls in equation 2 for each dependent variable. All models control for year-fixed effects, fixed effects for the year of birth of the first-born child of an individual, age of the youngest child, cohort fixed effects of the individual, individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects, the number of children of individuals, and socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

4.2 Robustness

We assess the sensitivity of our results to the restrictions imposed on our sample and the definition of grandparenting in a series of robustness check. First, we re-estimate our models using the full sample of HRS respondents regardless of age. The results (Panel A in appendix Table A8) are qualitatively similar to those from our main specification in Table 6. Second, we exclude respondents without grandchildren from the working sample. Our sex ratio

instrument identifies variation in both the timing of grandparenthood and the likelihood to provide grandchild care for existing grandchildren. This implies that the control group in our main specification consists of both older adults that are not (yet) grandparents as well as grandparents that do not provide childcare for their grandchildren. Excluding older adults without grandchildren from the sample means that both treatment and control group consist exclusively of grandparents and our instrument identifies variation in the likelihood to provide grandchild care. The results in Panel B of appendix Table A8 are both quantitatively and qualitatively similar to our main results in Table 6.

Appendix Table A9 reports the estimates using several alternative definitions of grandparenting to assess whether our results might be affected by measurement errors in the treatment indicator. The definition used in our main specification is a binary indicator whether respondents report providing at least 100 hours of grandchild care over the past two years. The corresponding estimates (Table 6) are repeated in the first row of appendix Table A9 to facilitate comparisons. We consider five alternative definitions: *(i)* an alternative indicator for whether the respondent and their spouse or partner provided at least 100 hours of grandchild care over the past two year as discussed in section 2.3, *(ii)* a binary indicator that is based on the same information as our main specification but excludes observations who do not know the number of care hours and reported a maximum number of hours of childcare provision,[10] *(iii)* a binary indicator for whether respondents provided any grandchild care over the past two years, *(iv)* a binary indicator for whether respondents provided at least 50 hours of grandchild care over the past two years, and *(v)* a binary indicator for whether respondents provided at least 200 hours of grandchild care over the past two years. The results in appendix Table A9 confirm that our results are robust to these changes across different treatment definitions. Both the magnitude as well as the significance of the estimated effects on IADL and subjective health are very similar

[10] As discussed in section 2.3, for respondents who report missing grandchild hours, the HRS further asks the minimum and maximum of the bracket range of care hours. We exclude 3,644 respondents who reported 0 minimum care hours and 199 maximum care hours (3,339) and 200 maximum care hours (305) since the actual care hours for this sample are ambiguous.

across specifications.

In addition, we check the sensitivity of our estimates to different outcome definitions. Appendix Table A10 reports the results for the health indicators defined with different cutoffs. The baseline column shows our main estimates in Table 6. The sign of the estimates on different health indicators is consistent with the baseline model while the magnitude and significance varies across health outcome definitions. For ADLs, grandparenting significantly increases the likelihood of reporting at least one or two limitations and the magnitude of these coefficients is much higher (22 percentage points for at least one limitation and 13 percentage points for at least two limitations) than that of the baseline estimate. However, the estimates on at least three and five ADL limitations are much smaller and not statistically significant. For IADLs, grandparenting is consistently harmful across different definitions. The magnitude of the coefficients with at least two or three IADL limitations is similar to that of the baseline estimate. Similar to the pattern of ADL outcomes, the magnitude of the coefficient wit at least one IADL limitation is much higher (16 percentage points) and the magnitude shrinks to two percentage points for the IADL measure with five limitations. For subjective health status, grandparenting significantly increases individuals' likelihood to self-report their health as poor by 15 percentage points (compared to 27 percentage points in our baseline specification). For mental health, none of the estimates is statistically significant, although the sign indicates an increase in depressive symptoms across specifications. Overall, our estimates of the effects of grandchild care provision on health are robust across different outcome definitions.

5 Discussion

This study examines the effect of childcare provision on grandparents' health in the U.S.. We propose a novel instrument, the sex ratio, for grandparental childcare provision, drawing on insights from the demographic literature on grandparenthood. Our sex ratio instrument

measures the share of daughters among all children born to a person, which captures that parents of daughters transition on average earlier into grandparenthood and grandparents are more likely to provide care for grandchildren born to their daughters than to grandchildren born to their sons. We conduct several tests and falsification exercises that support the sex ratio to be a valid instrument for grandparenting.

Our OLS results are in line with earlier studies suggesting that grandchild care provision is indeed positively associated with grandparents' health, but this association is likely driven by reverse causality. Once we address such endogeneity using the sex ratio as an IV, we find that effects of grandparenting on health are insignificant or even negative. These negative effects are not always precisely estimated, and thus a more conservative conclusion is that – contrary to most previous studies (Choi and Zhang 2021; Di Gessa et al. 2016; Ku et al. 2013; Wang et al. 2020; Zeng et al. 2021) – grandparenting does not improve grandparents' health.

Our findings suggest that previously reported positive associations between grandparenting and health are biased, likely due to reverse causality. Yet, it is possible that the context of the study also matters and that findings may differ based on, e.g., the role of the family and the strength of family ties. The U.S. is an interesting setting with neither particularly strong family ties (compared to, e.g., East Asia) nor with extensive subsidized formal childcare places (e.g., as in Northern Europe). It seems possible that health effects in this setting are very different from those reported, e.g., for China (Choi and Zhang 2021; Wang et al. 2020).

We also acknowledge a few limitations of our study. For example, it seems plausible that the health effects of grandparenting will differ based on its intensity (Zeng et al. 2021). Unfortunately, information in the HRS on the number of grandparenting hours is only reported retrospectively and the distribution suggests that it is subject to recall bias. Therefore, here we only consider effects of grandparenting at the extensive margin, but future work should examine differences in intensity, e.g., based on time use diaries.

Similarly, it seems plausible that the effects of grandparenting may differ based on the tasks taken over by grandparents. For example, taking care of infants during the day or when parents are close-by may be much less demanding than looking after these children overnight. Unfortunately, such data is not available in the HRS. Last, a larger sample (or stronger instruments) are needed to examine treatment effect heterogeneity across relevant subgroups.

In summary, our results show that grandparenting does not improve the health of grandparents, rather it may even be detrimental to grandparents' health. Instead, it seems that good health is an important precondition for grandparents to provide childcare and this reverse causality causes the frequently observed positive associations documented in the literature. This implies that grandparenting should not be considered as "active ageing" – a socially desirable activity that preserves or improves older adults' health. Instead, similar to informal care provision grandparenting appears to be an activity that older adults engage in to help their family even though it may be detrimental to their own health or well-being.

References

Adserà, A. (2017). Education and fertility in the context of rising inequality. *Vienna Yearbook of Population Research*, 15:63–92.

Age UK (2017). 5 million grandparents take on childcare responsibilities.

Angrist, J., Lavy, V., and Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4):773–824.

Ates, M. (2017). Does grandchild care influence grandparents' self-rated health? Evidence from a fixed effects approach. *Social Science & Medicine*, 190(C):67–74. Publisher: Elsevier.

Battistin, E., De Nadai, M., and Padula, M. (2014). Roadblocks on the road to grandma's house: Fertility consequences of delayed retirement. IZA Discussion Paper 8071, Institute for the Study of Labor (IZA).

Bom, J., Bakx, P., Schut, F., and van Doorslaer, E. (2018). The impact of informal caregiving for older adults on the health of various types of caregivers: A systematic review. *The Gerontologist*, 59(5):e629–e642.

Bom, J. and Stöckel, J. (2021). Is the grass greener on the other side? The health impact of providing informal care in the UK and the Netherlands. *Social Science & Medicine*, 269:113562.

Brunello, G. and Rocco, L. (2019). Grandparents in the blues. The effect of childcare on grandparents' depression. *Review of Economics of the Household*, 17(2):587–613. Publisher: Springer.

Chen, F., Short, S. E., and Entwisle, B. (2000). The impact of grandparental proximity on maternal childcare in China. *Population Research and Policy Review*, 19(6):571–590.

Chien, S., Campbell, N., Hayden, O., Hurd, M., Main, R., Mallett, J., Martin, C., Meijer, E., Miu, A., Moldoff, M., et al. (2015). RAND HRS data documentation, version N. *RAND Corporation*.

Choi, S.-w. E. and Zhang, Z. (2021). Caring as curing: Grandparenting and depressive symptoms in China. *Social Science & Medicine*, 289:114452.

Compton, J. (2015). Family proximity and the labor force status of women in Canada. *Review of Economics of the Household*, 13(2):323–358.

Compton, J. and Pollak, R. A. (2011). Family Proximity, Childcare, and Women's Labor Force Attachment. *National Bureau of Economic Research Working Paper Series*, No. 17678. featured in NBER digest on 2012-03-30.

Dahlberg, L., Berndt, H., Lennartsson, C., and Schön, P. (2018). Receipt of formal and informal help with specific care tasks among older people living in their own home. national trends over two decades. *Social Policy & Administration*, 52(1):91–110.

Danielsbacka, M., Tanskanen, A. O., Coall, D. A., and Jokela, M. (2019). Grandparental childcare, health and well-being in Europe: A within-individual investigation of longitudinal data. *Social Science & Medicine*.

Danielsbacka, M., Tanskanen, A. O., Jokela, M., and Rotkirch, A. (2011). Grandparental Child Care in Europe: Evidence for Preferential Investment in More Certain Kin. *Evolutionary Psychology*, 9(1):147470491100900102. Publisher: SAGE Publications Inc.

Dave, D., Rashad, I., and Spasojevic, J. (2006). The effects of retirement on physical and mental health outcomes. Technical report, National Bureau of Economic Research.

DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., and Muntner, P. (2006). Mortality prediction with a single general self-rated health question. *Journal of General Internal Medicine*, 21(3):267–275.

Di Gessa, G., Glaser, K., and Tinker, A. (2016). The health impact of intensive and nonintensive grandchild care in Europe: New evidence from SHARE. *The Journals of Gerontology: Series B*, 71(5):867–879.

Eibich, P. (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. *Journal of Health Economics*, 43:1–12.

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

Eurostat (2022). Living conditions in Europe - childcare arrangements.

Heger, D. (2017). The mental health of children providing care to their elderly parent. *Health Economics*, 26(12):1617–1629. Publisher: John Wiley & Sons, Ltd.

Idler, E. L. and Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior*, pages 21–37.

Irwin, M., Artin, K. H., and Oxman, M. N. (1999). Screening for depression in the older adult: Criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CESD). *Archives of Internal Medicine*, 159(15):1701–1704.

Kabátek, J. and Ribar, D. C. (2021). Daughters and divorce. *The Economic Journal*, 131(637):2144–2170.

Ku, L.-J., Stearns, S., Houtven, C., and Holmes, G. (2012). The health effects of caregiving by grandparents in Taiwan: An instrumental variable estimation. *Review of Economics of the Household*, 10(4):521–540. Publisher: Springer.

Ku, L.-J. E., Stearns, S. C., Van Houtven, C. H., Lee, S.-Y. D., Dilworth-Anderson, P., and Konrad, T. R. (2013). Impact of Caring for Grandchildren on the Health of Grandparents in Taiwan. *The Journals of Gerontology: Series B*, 68(6):1009–1021.

Kuka, E. (2020). Quantifying the benefits of social insurance: unemployment insurance and health. *Review of Economics and Statistics*, 102(3):490–505.

Lai, D. W. L., Li, J., and Bai, X. (2021). To be or not to be: relationship between grandparent status and health and wellbeing. *BMC Geriatrics*, 21(1):204.

Laughlin, L. (2013). Who's minding the kids? child care arrangements: Spring 2011. Technical Report P70-135, US Census Bureau.

Liu, Y. and Zai, X. (2022). The unintended effect of Medicaid aging waivers on informal caregiving. *Working Paper*.

Margolis, R. and Verdery, A. M. (2019). A Cohort Perspective on the Demography of Grandparenthood: Past, Present, and Future Changes in Race and Sex Disparities in the United States. *Demography*, 56(4):1495–1518.

Mirowsky, J. (2005). Age at First Birth, Health, and Mortality. *Journal of Health and Social Behavior*, 46(1):32–50. Publisher: SAGE Publications Inc.

Oswald, A. J. and Powdthavee, N. (2010). Daughters and left-wing voting. *The Review of Economics and Statistics*, 92(2):213–227.

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

Schmitz, H. and Westphal, M. (2015). Short- and medium-term effects of informal care provision on female caregivers' health. *Journal of Health Economics*, 42:174–185.

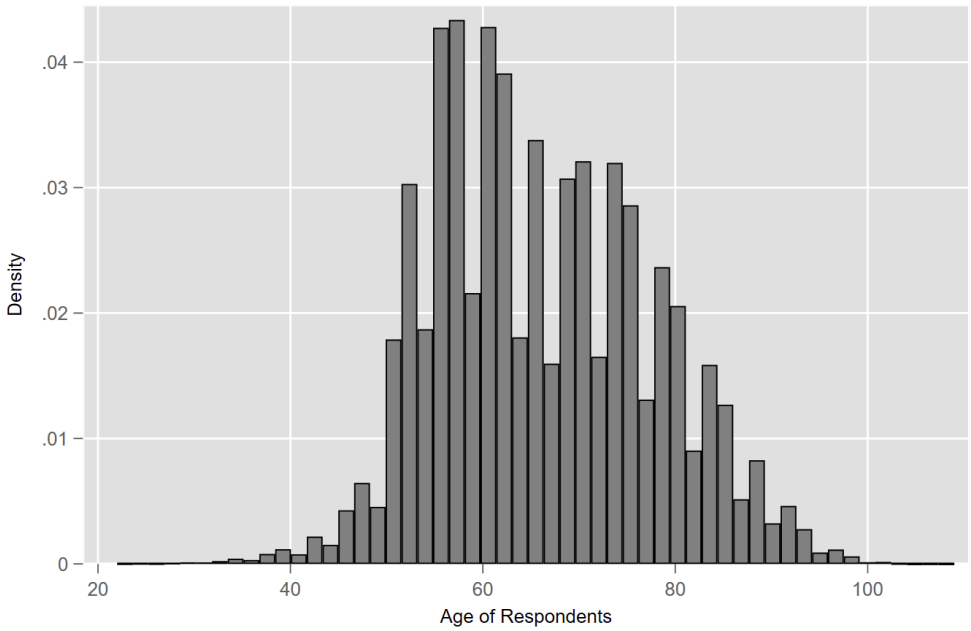
Somary, K. and Strieker, G. (1998). Becoming a Grandparent: A Longitudinal Study of Expectations and Early Experiences as a Function of Sex and Lineage1. *The Gerontologist*, 38(1):53–61.

Wang, H., Fidrmuc, J., and Luo, Q. (2020). Grandparenting and well-being of the elderly in China. Technical report, Bank of Finland, Institute for Economies in Transition. Issue: 18/2020.

Wu, X. and Li, L. (2012). Family size and maternal health: evidence from the One-Child policy in China. *Journal of Population Economics*, 25(4):1341–1364.

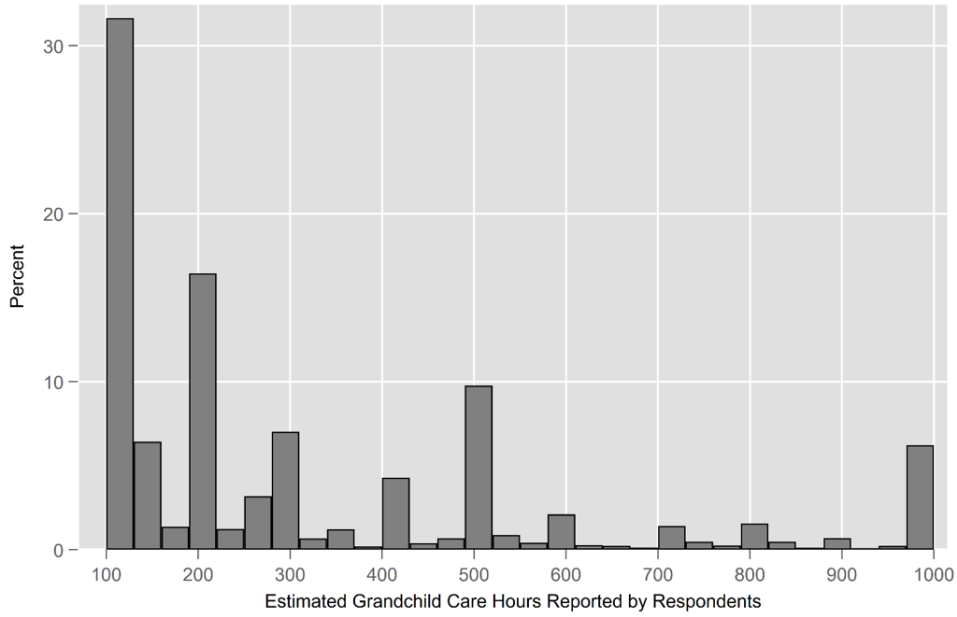
Zeng, Y., Chen, Y.-C., and Lum, T. Y. S. (2021). Longitudinal impacts of grandparent caregiving on cognitive, mental, and physical health in China. *Aging & Mental Health*, 25(11):2053–2060. Publisher: Routledge.

Figure A1: Distribution of Age of HRS Respondents



Notes: The data is the HRS from 1992 to 2014. This graphs draws the distribution of age of HRS respondents. The vertical axis shows the density of age. The age eligibility of respondents is above 50 and the age of their spouses can be any age as shown in the plot.

Figure A2: Distribution of grandchild care hours



Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. This graphs draws the distribution of grandchild care hours for respondents who are grandparenting and provide less than 1,000 hours over the last two years. Respondents providing less than 100 hours of grandchild care (defined as not grandparenting in our study) are omitted for clarity.

Table A1: Definitions of variables

Variable	Definition
Demographics	
Age	Age of respondents in years
Female	Dichotomous indicator of respondents being female (female=1, 0=male)
Education	Years in school of respondents
Marital status	
Married/partnered	Dichotomous indicator of respondents being married or having a partner living together
Separated/divorced	Dichotomous indicator of respondents being divorced or separated from marriage
Widowed	Dichotomous indicator of respondents having spouses or partners dead
Never married	Dichotomous indicator of respondents never getting married
Race/ethnicity	
White	Dichotomous indicator of respondents being white
Black/African	Dichotomous indicator of respondents being black or Hispanic
Other	Dichotomous indicator of respondents' race other than white or black or African
Instrumental variable	
Sex ratio	The ratio between the number of daughters to all children
Grandparenting	
Grandparenting for at least one child (Q1)	Q1: whether the respondent and spouse spent 100 or more hours taking care of their grandchildren or great-grandchildren since the last wave
Grandparenting for at least 100 hours (Q2)	Q2: how many estimate childcare hours provided in the last two years for the respondent and spouse, separately
Heath variables	
Subjective health	Binary indicator for poor subjective health based on respondent's self-reported general health status, 1 for "fair" or "poor", and 0 for "excellent", "very good" or "good".
ADL	Binary indicator for respondents that report 3 or more difficulties (out of 5 possible) on the index of difficulties in Activities of Daily Living (ADL), indicating respondents having any problem in bathing, eating, getting dressed, getting in/out of bed, and walking across a room
IADL	Binary indicator for respondents that report 3 or more difficulties (out of 5 possible) on the index of difficulties in Instrumental Activities of Daily Living (IADL), indicating respondents having any problem in using the phone, managing money, taking medications, shopping for groceries, and preparing hot meals
Mental health	Binary indicator for poor mental health based on scoring 4 or more points on the Center for Epidemiological Studies Depression (CESD) scale: sum of five negative indicators minus two positive indicators. The negative indicators measure sentiments all or most of the time: depression, everything is an effort, restless sleep, felt alone, sad, and could not get going. The positive indicators measure whether respondents felt happy and enjoyed life.

Table A2: Reduced form regression of the working sample

Dependent variable	ADL	IADL	Poor subjective health	Depressed
Sex ratio	0.002	0.003**	0.017*	0.006
	(0.002)	(0.002)	(0.009)	(0.006)
Mean of dependent variable	0.018	0.011	0.274	0.115
Number of clusters	20,255	20,255	20,261	19,575
Observations	83,655	83,650	83,673	78,783

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell shows reduced form estimates from our preferred model specification with full controls. All models control for year-fixed effects, fixed effects for the year of birth of the first-born child of an individual, age of the youngest child, cohort fixed effects of the individual, individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects, the number of children of individuals, and socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A3: Falsification exercise: Reduced form regression

Dependent variable	ADL	IADL	Poor subjective health	Depressed
Sex ratio	0.002	-0.004	0.011	-0.002
	(0.006)	(0.005)	(0.020)	(0.015)
Mean of dependent variable	0.019	0.013	0.300	0.118
Number of clusters	5,641	5,639	5,644	5,216
Observations	6,019	6,017	6,022	5,569

Notes: The data used are from the HRS 1996 to 2014 of the subsample of respondents who do not have grandchildren now but have later. All models control for year-fixed effects, fixed effects for the year of birth of the first-born child of an individual, age of the youngest child, cohort fixed effects of the individual, individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects, the number of children of individuals, and socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A4: Robustness of 2SLS estimates on ADL

Dependent variable: ADL						
Model	1	2	3	4	5	6
Grandparenting	0.034	0.029	0.030	0.032	0.032	0.032
	(0.038)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
Year FE + Birth year FE of first born		Y	Y	Y	Y	Y
Cohort FE + Birth year FE of youngest born			Y	Y	Y	Y
Demographics of R				Y	Y	Y
Family size					Y	Y
Socioeconomic controls						Y
Mean of dependent variable	0.019	0.019	0.019	0.018	0.018	0.018
Number of clusters	20,473	20,416	20,391	20,255	20,255	20,255
Observations	84,297	84,220	84,131	83,655	83,655	83,655
Kleibergen-Paap F-stat.	53.29	53.41	54.17	53.73	53.74	53.66
AR F statistic	0.809	0.567	0.607	0.691	0.693	0.667
AR p-value	0.368	0.451	0.436	0.406	0.405	0.414

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell reports estimates from a separate specification using equation 2 for the dependent variable. Column 1 reports estimates without any controls. Column 2 adds year-fixed effects and fixed effects for the year of birth of the first-born child of an individual. Column 3 adds age of the youngest child and cohort fixed effects of the individual. Column 4 includes individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects. Column 5 further controls for the number of children of individuals. Column 6 adds into socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A5: Robustness of 2SLS estimates on IADL

Dependent variable: IADL						
Model	1	2	3	4	5	6
Grandparenting	0.049*	0.046*	0.046*	0.053*	0.053*	0.052*
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Year FE + Birth year FE of first born		Y	Y	Y	Y	Y
Cohort FE + Birth year FE of youngest born			Y	Y	Y	Y
Demographics of R				Y	Y	Y
Family size					Y	Y
Socioeconomic controls						Y
Mean of dependent variable	0.011	0.011	0.011	0.011	0.011	0.011
Number of clusters	20,472	20,415	20,390	20,255	20,255	20,255
Observations	84,291	84,214	84,125	83,650	83,650	83,650
Kleibergen-Paap F-stat.	53.22	53.34	54.08	53.67	53.68	53.61
AR F statistic	3.375	2.960	3.065	3.971	3.989	3.945
AR p-value	0.066	0.085	0.080	0.046	0.046	0.047

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell reports estimates from a separate specification using equation 2 for the dependent variable. Column 1 reports estimates without any controls. Column 2 adds year-fixed effects and fixed effects for the year of birth of the first-born child of an individual. Column 3 adds age of the youngest child and cohort fixed effects of the individual. Column 4 includes individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects. Column 5 further controls for the number of children of individuals. Column 6 adds into socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A6: Robustness of 2SLS estimates on poor subjective health

Dependent variable: Poor subjective health						
Model	1	2	3	4	5	6
Grandparenting	0.309**	0.311**	0.303**	0.273*	0.275*	0.269*
	(0.153)	(0.154)	(0.153)	(0.151)	(0.151)	(0.147)
Year FE + Birth year FE of first born		Y	Y	Y	Y	Y
Cohort FE + Birth year FE of youngest born			Y	Y	Y	Y
Demographics of R				Y	Y	Y
Family size					Y	Y
Socioeconomic controls						Y
Mean of dependent variable	0.275	0.275	0.275	0.274	0.274	0.274
Number of clusters	20,480	20,422	20,397	20,261	20,261	20,261
Observations	84,316	84,238	84,149	83,673	83,673	83,673
Kleibergen-Paap F-stat.	53.65	53.84	54.59	54.19	54.20	54.12
AR F statistic	4.596	4.574	4.375	3.577	3.624	3.674
AR p-value	0.032	0.033	0.037	0.059	0.057	0.055

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell reports estimates from a separate specification using equation 2 for the dependent variable. Column 1 reports estimates without any controls. Column 2 adds year-fixed effects and fixed effects for the year of birth of the first-born child of an individual. Column 3 adds age of the youngest child and cohort fixed effects of the individual. Column 4 includes individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects. Column 5 further controls for the number of children of individuals. Column 6 adds into socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A7: Robustness of 2SLS estimates on depressed

Dependent variable: Depressed						
Model	1	2	3	4	5	6
Grandparenting	0.157*	0.127	0.118	0.101	0.101	0.097
	(0.094)	(0.093)	(0.093)	(0.092)	(0.092)	(0.091)
Year FE + Birth year FE of first born		Y	Y	Y	Y	Y
Cohort FE + Birth year FE of youngest born			Y	Y	Y	Y
Demographics of R				Y	Y	Y
Family size					Y	Y
Socioeconomic controls						Y
Mean of dependent variable	0.115	0.115	0.115	0.115	0.115	0.115
Number of clusters	19,779	19,726	19,698	19,575	19,575	19,575
Observations	79,356	79,284	79,198	78,783	78,783	78,783
Kleibergen-Paap F-stat.	53.51	54.43	55.22	54.68	54.68	54.56
AR F statistic	2.979	1.956	1.710	1.230	1.233	1.160
AR p-value	0.084	0.162	0.191	0.267	0.267	0.281

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell reports estimates from a separate specification using equation 2 for the dependent variable. Column 1 reports estimates without any controls. Column 2 adds year-fixed effects and fixed effects for the year of birth of the first-born child of an individual. Column 3 adds age of the youngest child and cohort fixed effects of the individual. Column 4 includes individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects. Column 5 further controls for the number of children of individuals. Column 6 adds into socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A8: Robustness of estimates to sample change

Dependent variable	ADL	IADL	Poor subjective health	Depressed
A. Sample without age restrictions				
Grandparenting	0.067	0.150**	0.487**	0.052
	(0.062)	(0.064)	(0.195)	(0.101)
Mean of dependent variable	0.037	0.037	0.306	0.107
Number of clusters	27,565	27,566	27,574	26,227
Observations	143,392	143,381	143,386	131,643
Kleibergen-Paap F-stat.	47.87	47.79	48.42	53.32
AR F statistic	1.244	6.617	7.582	0.272
AR p-value	0.265	0.010	0.006	0.602
B. Sample restricted to grandparents				
Grandparenting	0.033	0.065**	0.274*	0.124
	(0.042)	(0.031)	(0.159)	(0.100)
Mean of dependent variable	0.018	0.011	0.272	0.114
Number of clusters	19,991	19,991	19,998	19,301
Observations	77,262	77,259	77,278	72,868
Kleibergen-Paap F-stat.	45.23	45.2	45.62	45.69
AR F statistic	0.613	5.132	3.249	1.599
AR p-value	0.434	0.024	0.072	0.206

Notes: Each cell shows 2SLS estimates from our preferred model specification with full controls in equation 2 for each sub-sample. The sample in Panel A uses data from the HRS 1996 to 2014 of all individuals without age limits. The sample in Panel B uses data from the HRS 1996 to 2014 of individuals who are 40 to 70 and excludes those who report no grandchildren. All models control for year-fixed effects, fixed effects for the year of birth of the first-born child of an individual, age of the youngest child, cohort fixed effects of the individual, individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects, the number of children of individuals, and socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A9: Robustness of estimates to treatment definitions

Dependent variable	ADL	IADL	Poor subjective health	Depressed
Main specification	0.032	0.052*	0.269*	0.097
	(0.039)	(0.028)	(0.147)	(0.091)
Mean of dependent variable	0.018	0.011	0.274	0.115
Kleibergen-Paap F-stat.	53.66	53.61	54.12	54.56
AR F statistic	0.667	3.945	3.674	1.160
AR p-value	0.414	0.047	0.055	0.281
Couples' hours	0.030	0.051*	0.260*	0.096
	(0.038)	(0.027)	(0.143)	(0.091)
Mean of dependent variable	0.018	0.011	0.274	0.115
Kleibergen-Paap F-stat.	52.15	52.11	52.46	50.7
AR F statistic	0.667	3.945	3.674	1.160
AR p-value	0.414	0.047	0.055	0.281
Excluding reported maximum hours	0.042	0.062**	0.318**	0.100
	(0.043)	(0.031)	(0.162)	(0.098)
Mean of dependent variable	0.019	0.012	0.275	0.115
Kleibergen-Paap F-stat.	47.03	46.98	47.35	48.41
AR F statistic	0.958	4.383	4.385	1.065
AR p-value	0.328	0.036	0.036	0.302
$>$ 0 hours	0.031	0.052*	0.266*	0.096
	(0.039)	(0.028)	(0.146)	(0.091)
Mean of dependent variable	0.018	0.011	0.274	0.115
Kleibergen-Paap F-stat.	49.73	49.69	50.05	49.95
AR F statistic	0.667	3.945	3.674	1.160
AR p-value	0.414	0.047	0.055	0.281
$>$ 50 hours	0.032	0.054*	0.277*	0.099
	(0.04)	(0.029)	(0.153)	(0.094)
Mean of dependent variable	0.018	0.011	0.274	0.115
Kleibergen-Paap F-stat.	47.99	47.95	48.3	48.81
AR F statistic	0.667	3.945	3.674	1.160
AR p-value	0.414	0.047	0.055	0.281
$>$ 200 hours	0.045	0.068**	0.337*	0.111
	(0.048)	(0.034)	(0.177)	(0.111)
Mean of dependent variable	0.019	0.012	0.274	0.115
Kleibergen-Paap F-stat.	45.82	45.85	45.99	45.13
AR F statistic	0.914	4.375	4.027	1.033
AR p-value	0.339	0.037	0.045	0.309

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell shows 2SLS estimates from our preferred model specification with full controls in equation 2 for each dependent variable. All models control for year-fixed effects, fixed effects for the year of birth of the first-born child of an individual, age of the youngest child, cohort fixed effects of the individual, individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects, the number of children of individuals, and socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A10: Robustness of estimate to outcome definitions

		A. ADL			
Definition	Baseline (4 limitations)	*1*	*2*	*3*	*5*
Grandparenting	0.032	0.217**	0.127*	0.059	0.007
	(0.039)	(0.110)	(0.077)	(0.054)	(0.024)
		B. IADL			
Definition	Baseline (4 limitations)	*1*	*2*	*3*	*5*
Grandparenting	0.052*	0.156	0.056	0.079*	0.017
	(0.028)	(0.096)	(0.062)	(0.043)	(0.016)
		C. Poor subjective health			
Definition	*Baseline ("Poor" or "Fair")*	*"Poor"*			
Grandparenting	0.269*	0.152*			
	(0.147)	(0.085)			
		D. Depressed			
Definition	Baseline (5 symptoms)	*2*	*3*	*4*	*6*
Grandparenting	0.097	0.068	0.081	0.135	0.048
	(0.091)	(0.139)	(0.123)	(0.108)	(0.074)

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 40 to 70. Each cell shows 2SLS estimates from our preferred model specification with full controls in equation 2 for each dependent variable. The baseline is the outcome definition used in the main tables of health. The other cutoffs are used as alternative definitions to create the health indicators. All models control for year-fixed effects, fixed effects for the year of birth of the first-born child of an individual, age of the youngest child, cohort fixed effects of the individual, individual demographics such as age (quadratic polynomial), race, religion, gender, birth place and census region fixed effects, the number of children of individuals, and socio-economic controls such as the annual earnings from wages or salary, bonuses, second job income, and professional practice income as well as the annual sum of pensions and annuities. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.