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

Peter Eibich† and Xianhua Zai‡

Abstract

March 2024

This paper examines the causal effect of childcare provision on grandparents' health in the US. We use the sex ratio among older adults' children as an instrument for grandparental childcare provision. Our instrument 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 detrimental for physical functioning and subjective health. We show that these effects increase with the intensity of grandchild care provision, and the effects are driven primarily by grandmothers.

Keywords: Grandparents, Childcare Provision, Instrumental Variables, Health

JEL classification I10, J13, J14, C26

*We would like to thank editor Xi Chen and three anonymous reviewers for helpful comments and suggestions that have improved the paper. We would like to thank Benedicte Apouey, Pedro Pita Barros, Judith Bom, Eric Bonsang, Katharina Janke, Vladimir Kozlov, Tim Riffe, and Jeroen Spijker, Hansel Teo, Christine Valente as well as participants of the EuHEA semniar, the workshop "Behaviours, Health and Ageing" at Université Paris Dauphine and EAPS Health, Morbidity and Mortality Working Group in Budapest and of Society for Longitudinal and Lifecourse Studies conference (SLLS) in Munich for their helpful comments. Carlotta Dietrich provided excellent research assistance.

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

‡Corresponding author, Max Planck Institute for Demographic Research, 1 Konrad-Zuse-Str., 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 ($\sim$ 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 childcare provided by grandparents is highly beneficial to parents (Dimova and Wolff 2011; Compton 2015; Bratti et al. 2018), 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 childcare provision can be considered as "active ageing", i.e., an activity that benefits both older individuals and 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 grandparental childcare provision 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, because 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). Health is also 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 grandparents' childcare provision 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 use the sex ratio (defined as the number of daughters relative to the total number of children) as an instrument to address the endogeneity of grandchild care provision. Our IV is motivated by two insights from the demographic literature on grandparenthood - *(i)* parents of daughters transition to grandparenthood earlier than parents of sons, and *(ii)*, grandparents are more likely to provide childcare for grandchildren born to a daughter than for those born to a son. Previous studies on the labour market consequences of grandparenthood have used the sex of the first-born child as an instrument for becoming a grandparent (Rupert and Zanella 2018). Using the sex ratio as an instrument for grandchild care provision (previously used, e.g., by Salm et al. (2021)) follows similar considerations as the sex of the first-born child,[2] but exploits more variation in the data.[3] We conduct extensive diagnostic checks of the IV assumptions. While our instrument is strongly related to grandparental childcare provision, we also find weaker associations with marital status (Kabátek and Ribar 2021) and informal care receipt (Van Houtven and Norton 2004), which raise concerns about the exclusion restriction. We therefore implement the "plausibly exogenous" approach proposed by Conley et al. (2012) to derive bounds for our estimates that allow for reasonable violations of the exclusion restriction.

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 grandchild care provision on health are negative, implying that (similar to informal caregivers) grandparents looking

[2]The sex of a child is determined randomly at conception and daughters have grandchildren earlier than sons.

[3]For individuals with one child, both instruments are identical. For individuals with two or more children, the sex of the first-born child only distinguishes two groups in the data (first-born daughter vs. first-born son), whereas the sex ratio distinguishes between three or more groups (two daughters, one daughter, no daughters for families with two children).

after their grandchildren do so despite the impact it has on their health. We therefore conclude that grandparental childcare provision should not be considered as "active ageing".

Considering potential mechanisms, we show that the effects increase with the intensity of grandchild care provision. We find no evidence for activity substitution, i.e., grandparents caring for their grandchildren do not reduce their engagement in activities that are beneficial for their health (e.g., exercise) . Heterogeneity analyses suggest that these effects are driven primarily by grandmothers, which likely reflects their higher rates of grandchild care provision.

This study contributes to the literature by estimating a credibly identified causal effect of grandchild care provision on grandparents' health. We use an established instrument motivated by the demographic literature on grandparenthood, we conduct a battery of tests and falsification exercises of the IV assumptions, and we derive treatment bounds that allow for possible violations of the exclusion restriction. 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, dependent and treatment variables, and sample statistics. Section 3 first motivates our instrument, describes the estimation strategy, discusses the assumptions for our IV model, and address potential violations of IV assumptions. Section 4 presents our main results, addresses IV validity issues, examines the robustness of our estimates, explores potential mechanisms and heterogeneity, and discusses the external validity of our results. 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.

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 grandchild care provision on grandparents' health, we restrict our working sample to HRS respondents aged between 50 to 80 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 80. 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 4.3 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 120,066 observations (25,300 unique individuals) and covers the period from 1996 (wave 3) to 2014 (wave 12) in which the HRS asks respondents

consistent questions on grandparents' childcare provision.

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 functioning, 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". 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.[4] 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.

Third, we use information about respondents' mental health. The HRS asks respondents about their mental health using the Center for Epidemiologic Studies Depression (CESD) score. The CESD 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 CESD scale has been validated as an instrument to identify major depression in older adults (Irwin et al. 1999). For all health measures, we

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

consider other operationalizations as robustness checks in section 4.3. We consider the four health measures described here as our main outcomes, because these measures are frequently used in academic studies as well as in clinical practice. We report results for alternative health measures based on limitations in gross motor skills, mobility, large muscle use, and fine motor skills as well as a measure of cognitive functioning in the appendix.

2.3 Treatment Indicators

Our treatment variable of interest is whether individuals provide grandchild care or not. There are two relevant 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.[5] 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.[6] 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.[7] For those who cannot remember the hours or do not know the exact hours or refuse to give the number of hours, the HRS further asks the minimum and maximum values of hours of grandchild care provided.[8] There are both advantages and disadvantages of using each question to construct our treatment indicator of grandchild care provision. 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 looking after the grandchildren. On the other

[5]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?"

[6]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)?"

[7]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.

[8]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.

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 providing childcare, 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 grandparental childcare provision, 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 providing childcare. 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 4.3.

2.4 Sample Statistics

Table 1 presents the summary statistics of the working sample of HRS respondents who are between 50 to 80 in each survey year. The average age of the sample is around 66. About 58 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 44 years old and the youngest child is on average about 35 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 28 to 33 percent of respondents provide some grandchild care according to the different definitions discussed earlier.

The average self-reported health status of respondents is good. The average ADL score

Table 1: Summary Statistics of the Sample

Variable	Mean	S.D.	Obs.
Demographics			
Age	65.84	7.93	120,204
Female	0.58	0.49	120,204
Education (years)	12.14	3.19	119,989
Number of children	3.72	2.07	120,204
Age of oldest child	43.76	8.53	119,881
Age of youngest child	34.75	9.63	119,881
Marital status			
Married/partnered	0.71	0.45	120,105
Separated/divorced	0.11	0.32	120,105
Widowed	0.15	0.36	120,105
Never married	0.01	0.11	120,105
Race/ethnicity			
White	0.78	0.41	120,070
Black/African	0.16	0.37	120,070
Other	0.05	0.22	120,070
Instrumental variable			
Sex ratio	0.49	0.29	120,066
Grandparenting			
Grandparenting for at least one child (Q1)	0.33	0.47	120,101
Grandparenting for at least 100 hours (Q2)	0.28	0.45	120,204
Heath variables			
Self-report health status (1-5)	2.88	1.12	120,142
ADL limitations (0-5)	0.30	0.87	120,130
IADL limitations (0-5)	0.25	0.79	120,129
CESD score (0-8)	1.50	1.99	112,246

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. The definitions of these variables can be found in Appendix Table A1.

and IADL score are both close to zero, which indicates that our sample is relatively healthy. The average CESD depression score is 1.5 out of 8. 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 grandchild care to motivate the sex ratio as our 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 Grandparental Childcare Provision

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. 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 grandchild care provision by using the sex ratio, defined as the number of daughters divided by the total number of children of a respondent, as an instrument.[9] The instrument relies on two distinct mechanisms that link

[9]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

the sex ratio to grandparents' childcare provision - (i) parents of daughters transition to grandparenthood earlier than parents of sons, and (ii) parents of daughters invest more in their offspring than parents of sons.

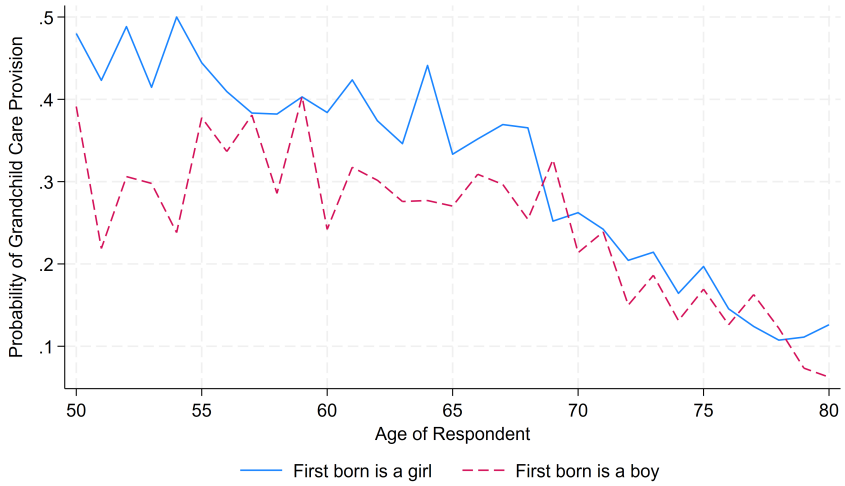
It is well-documented that women tend to give birth earlier than men (Margolis and Verdery 2019). This implies, *ceteris paribus*, that parents of daughters will transition to grandparenthood earlier than parents of sons. The gender of a child can be considered as good as randomly determined, thus the gender of a person's first-born child might serve as a suitable instrument that predicts the transition into grandparenthood (and subsequently grandchild care provision) (Rupert and Zanella 2018). This is also borne out in our data: Figure 1 shows the share of individuals who are providing grandchild care by age for individuals with a first-born daughter and those with a first-born son, respectively, for older adults with at least one grandchild. 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.

While the sex of the first-born child is a plausibly exogenous instrument, it also relies on very limited variation. Our sex ratio instrument exploits that parents of daughters (on average) transition to grandparenthood earlier than parents of sons, regardless of birth order. Figure 2 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 very similar for parents with one daughter and one son, regardless of the birth order. The likelihood 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. This pattern is likely driven by the lower variation in age at first birth among women than among men (Margolis and Verdery 2019),[10] which implies that, conditional on family size, the number of daughters is predictive of an earlier

sex ratio instrument (Appendix Table A2).

[10]The lower variation in age at first birth among women implies that parents with a first-born son and a second-born daughter will in many cases become grandparents due to the first birth of their daughter.

Figure 1: Sex of the first-born child and grandchild care provision



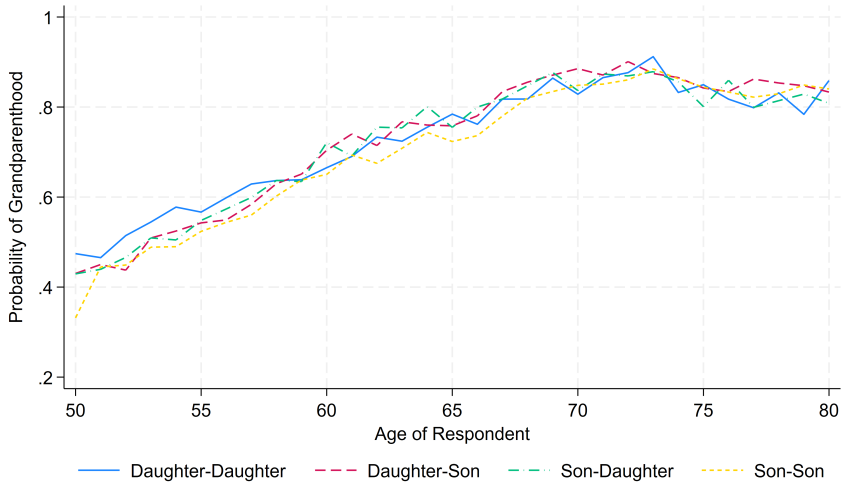
Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. The sample is limited to individuals with only one child and at least one grandchild. This graph draws the share of individuals who provide some grandchild care by age for individuals with a first-born daughter and those with a first-born son, respectively. Grandchild care is defined as an indicator that is 1 if the estimated grandchild care hours reported by respondents are at least 100 hours.

transition to grandparenthood.[11]

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 tend to invest more time into grandchild care than paternal grandparents (Compton and Pollak 2011; Danielsbacka et al. 2011). The literature has proposed three possible explanations for this difference: First, maternal grandparents share a longer lifetime with their grandchildren (Margolis and Verdery 2019). As discussed above, they tend to become grandparents earlier in life, and are consequently younger and on average healthier than paternal grandparents. Hence, they can invest more into their grandchildren than paternal grandparents. Second, from an evolutionary perspective grandparents invest into their grandchildren to ensure the survival of their kin. Since there is more uncertainty around paternal kinship, grandparents will invest preferentially into their daughters' offspring

[11] While family size is in itself predictive of grandparenthood, it is also endogenous and we therefore condition on family size.

Figure 2: Grandparenthood for older parents with two children



Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80 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.

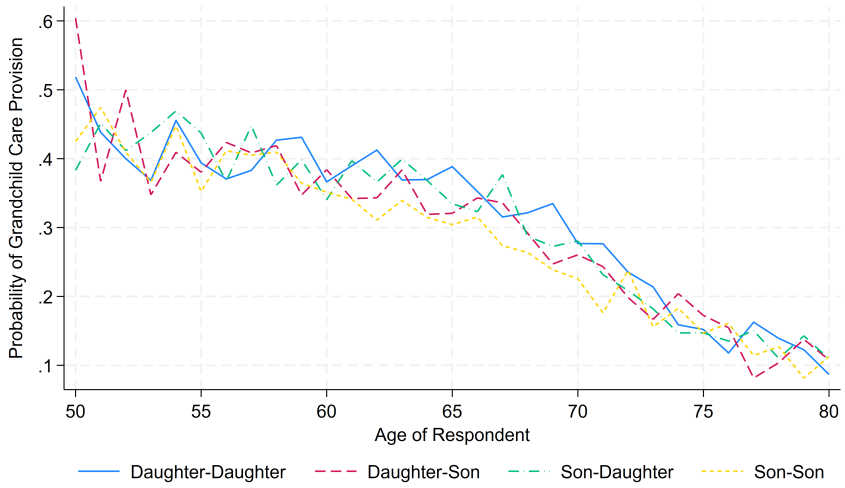
rather than into their sons' (Danielsbacka et al. 2011). Third, mothers tend to share stronger bonds with their daughters (Somary and Strieker 1998) and might therefore invest more into their daughters' children. While the first mechanism suggests that any differences in grandchild care provision are driven by differences in the timing of the transition to grandparenthood, the second and third explanations imply that children's sex is also predictive of grandparental childcare provision conditional on the timing of grandparenthood. Figure 3 shows that from age 60 onwards until about age 75, grandparents with two daughters are most likely to provide grandchild care, whereas grandparents with two sons tend to be the least likely group to provide grandchild care.[12]

The third mechanism (emotional bonds between mothers and daughters) also raises concerns about instrument validity, since these bonds might either affect other outcomes or are in turn affected by unobserved confounders. For example, Somary and Strieker (1998) report few differences in grandparents' behavior across lineage, but note that they control

[12]Figure 2 shows the probability of grandparenthood, whereas Figure 3 shows the probability of grandparental childcare provision conditional on grandparenthood.

for proximity between grandparents and grandchildren, which plays an important role for grandchild care investments (Compton 2015). It seems plausible that the proximity between grandparents and their children might also affect grandparents' health through other mechanisms than the provision of grandchild care. We discuss such concerns in more detail in section 3.3.2

Figure 3: Grandparental childcare for older parents with two children



Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80 with respondents limited to having two children. This graph draws the share of individuals who provide grandchild care by age for individuals with two children conditional on grandparenthood. Grandparental childcare is defined as an indicator that is 1 if the estimated grandchild care hours reported by respondents are at least 100 hours.

In summary, the literature suggests that maternal grandparents invest more into their grandchildren, because *(i)* they become grandparents earlier in life, and *(ii)* grandparents prefer to invest in their daughters' offspring. Importantly, it is not just the first child that matters - having one or more daughters is predictive of grandchild care provision regardless of birth order. However, family size (and consequently the absolute number of daughters) is endogenous and might be correlated with, e.g., the health of the older parents (Wu and Li 2012). We therefore use the sex ratio as our instrument to exploit random variation in the number of daughters born to a respondent conditional on family size.[13] Compared to the

[13]In contrast to the absolute number of daughters, the sex ratio is not correlated with family size ($\rho = -0.0007$ in our sample).

sex of the first-born child ("birth order instrument") used in previous studies, the sex ratio instrument draws on more variation for respondents with two or more children. For example, for parents with one child, 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 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.[14]

3.2 Model Specification

We estimate the first stage of our IV model as follows:

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

where $GCcare_{it}$ is a binary indicator for individuals providing grandchild care 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, and family size. These covariates can all be considered as predetermined

[14] We note that both instruments require the assumption that becoming a grandparent should not have a direct effect on health for older adults that do not provide care for their grandchildren. We discuss this assumption in more details later when considering the validity of the exclusion restriction, but here we note that we are generally comfortable with this assumption.

and they capture important demographic differences in fertility and grandparenthood. While the validity of the instrument does not depend on the inclusion of these covariates, they might help to improve the precision of our estimates.[15] ϵ_{it} are the standard errors clustered at the individual level which allow for correlation within individuals across waves.

We estimate the effect of grandchild care provision on health in the second stage of the model as follows:

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

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)* we aim to exploit variation in the sex ratio between individuals arising from the random assignment of sex at conception of the child rather than variation in the instrument over time for individuals that occurs due to new births, deaths of children, or misreporting.[16]

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 grandparental childcare provision, *(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 providing grandchild care

[15] We include fixed effects for the year of birth of the first-born and age of the youngest child, because our sample also includes individuals with only one child. Table A3 shows that our results are robust to including fixed effects for year of birth for the oldest and the youngest child.

[16] In our sample, 16.5% of individuals show variation in the sex ratio over time. Of these, 14% (2.4% of all individuals in the sample) experience the loss of a child, 28% (4.7% of individuals in our sample) appear to misreport the number of children in at least one wave, and the remaining 58% are new births or adoptions. Our results remain robust if we exclude individuals with changes in the sex ratio over time (Table A4) or if we use an alternative measure based on the number of living children (Table A2).

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 grandparental childcare provision only on the sex ratio instrument. The estimate suggests that - in line with the demographic literature on grandparenthood - *ceteris paribus* older adults that only have daughters (i.e., a sex ratio of 1) are 5 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 grandchild care provision by 2.5 percentage points. In columns 2-5, we successively add control variables to account for standard demographic characteristics. 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 46 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 grandchild care provision, 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 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. The sex of a child is randomly determined at conception, and sex-selective abortion or miscarriage rates are unlikely to play a major role in the context of this study. The exclusion restriction requires that the

Table 2: First stage estimates

Dependent variable: Grandparenting					
Model	1	2	3	4	5
Sex ratio	0.050***	0.050***	0.050***	0.048***	0.048***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Year FE + Birth year FE of first-born		Y	Y	Y	Y
Cohort FE + Birth year FE of youngest-born			Y	Y	Y
Demographics				Y	Y
Family size					Y
Mean of dependent variable	0.28	0.28	0.28	0.29	0.29
Number of clusters	25,300	25,227	25,202	25,055	25,055
Observations	120,066	119,968	119,880	119,293	119,293
1st stage Kleibergen-Paap F-statistic	46.82	49.48	49.94	47.16	47.10

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. 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. Grandparental childcare provision 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. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

sex ratio should not affect older parents' health through any other pathway than its effect on grandparental childcare provision. As noted earlier, one concern is that our instrument identifies variation in grandchild care provision that arises because older parents of daughters transition to grandparenthood earlier than older parents of sons. If the transition to grandparenthood itself affects health in other ways than through childcare provision, the exclusion restriction would be violated. We are not very concerned about such direct effects of grandparenthood on health. While previous studies report effects of grandparenthood on labor market outcomes (Frimmel et al. 2022), changes in labor force participation are likely driven by (anticipated) grandchild care provision. It is possible that becoming a grandparent might have other effects on older adults, but we would expect that such effects are either unrelated to health (e.g., financial transfers to support their children) or have limited, positive effects (e.g., increasing life satisfaction or improvements

in health behavior), in which case our estimated negative health effects can be regarded as a conservative upper bound of the true effect of grandchild care provision.

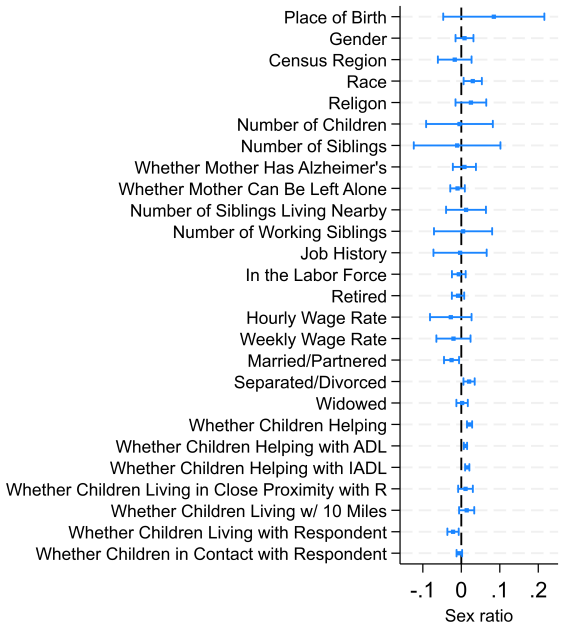
The exclusion restriction will also be violated if the sex of a child affects individuals' lives beyond their transition to grandparenthood. Previous studies have documented that parents of daughters differ in several aspects from parents of sons, e.g., parents of daughters are at a higher risk of divorce (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). It is plausible that some of these differences (such as divorce or informal care receipt) might have consequences for health in later life. In particular, we would expect that a previous divorce is related to worse health and health behavior, whereas the health effects of informal care receipt may depend on the counterfactual.[17] This means that it is not clear in which direction such violations of the exclusion restriction would affect our results.

We conduct two diagnostic checks for the validity of our instrument. First, we check for covariate balancing. If the validity assumption holds, 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. Note that failure to detect such correlations does not in turn imply that the validity assumption holds, since unobserved confounders remain a concern. We regress a battery of covariates as dependent variables on our sex ratio IV controlling for year and first-born fixed effects, cohort fixed effects for the individual, age fixed effects of the youngest child as well as individual demographic characteristics. Figure 4 shows the point estimates and confidence intervals of the regression coefficients on the sex ratio for each dependent variable listed on the vertical axis. The sex ratio appears to be uncorrelated with demographic characteristics of the respondent, implying that it is as good as randomly assigned. However, the sex ratio is negatively associated with being married and positively

[17] Informal care receipt would likely improve health if the alternative is an unmet need for care, whereas the health effects are less clear if the alternative is formal care provided by a paid professional.

related to divorce, implying that (in line with the literature) more daughters increase the risk of divorce. Similarly, older adults with more daughters are also more likely to receive informal care in our data. Interestingly, older adults with more daughters are less likely to live with any of their children. There are no significant differences in the proximity to their children. Nevertheless, correlations between the sex ratio and marital status as well as informal care receipt raise concerns about the validity of our instrument. We note, however, that both the conditional and unconditional correlations between our instrument and these potential confounders (see Table A5) are considerably smaller than our first-stage estimates (Table 2) and our reduced form estimates (Table A6). This suggests that these mechanisms lead to no or at worst minor violations of instrument validity. We address such potential violations in more detail in section 3.3.4.

Figure 4: Covariate balance for the sex ratio instrument



Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. The graph shows the point estimates and confidence intervals of the coefficients on the sex ratio using the regression model with controls including 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, and individual demographics such as age (quadratic polynomial). The sex ratio is defined as the number of daughters divided by the total number of children of an individual.

We 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 grandchild care provision (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 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 IV validity because it would suggest that the sex ratio is related to health through pathways other than grandchild care. 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 A6 for comparison).

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 A7 are in line with our earlier findings, i.e., we fail to detect any violations of the exclusion restriction.

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

Table 3: Falsification exercise: Reduced form regression

Dependent variable	ADL	IADL	Self-reported health	CESD Score
Sex ratio	0.000	0.006	0.047	-0.042
	(0.035)	(0.031)	(0.040)	(0.078)
Mean of dependent variable	0.34	0.28	2.99	1.68
Number of clusters	8,402	8,400	8,405	7,682
Observations	9,194	9,192	9,195	8,405

Notes: The data used are from the HRS 1996 to 2014 of the subsample of respondents who are 50 to 80 and who do not have grandchildren. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. 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, and the number of children of individuals. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

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 grandparental childcare provision 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.[18]

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. For all subsamples, the sex ratio instrument increases the likelihood of grandparental childcare provision and (with one exception) estimates are strongly significant.

Additionally, we also examine the robustness of the first-stage estimate across random subsamples from our working sample. We repeatedly draw a random 25% subsample from

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

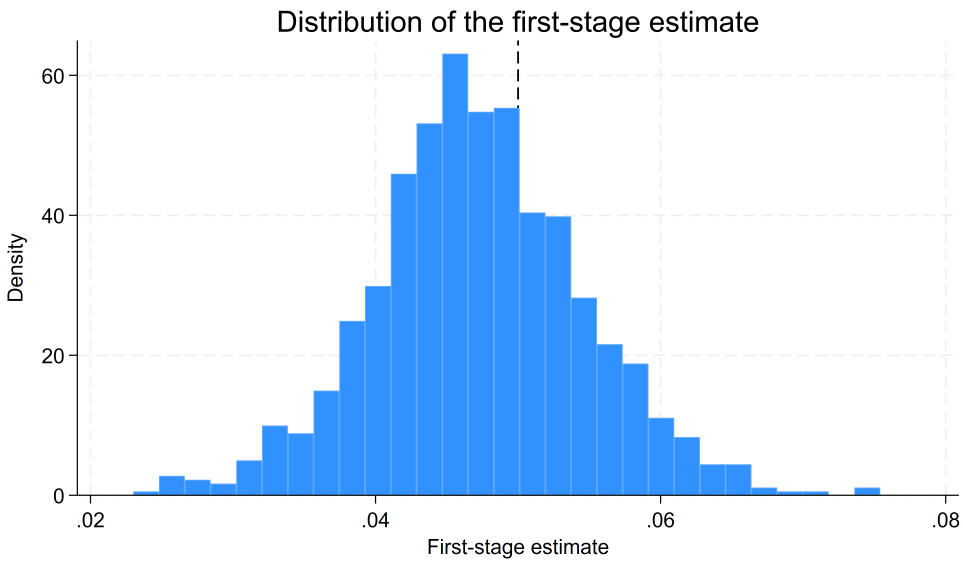
Table 4: Falsification exercise: Monotonicity

	Dependent variable: Grandparenting									
Subsample	Male	Female	$\leq$ 4 Children	$>$ 4 Children	$\leq$ 4 Siblings	$>$ 4 Siblings	Married	Not Married	White	Not White
Sex ratio	0.027**	0.063***	0.050***	0.031	0.040***	0.079***	0.045***	0.056***	0.041***	0.072***
	(0.010)	(0.009)	(0.007)	(0.019)	(0.008)	(0.015)	(0.009)	(0.011)	(0.008)	(0.015)
Mean of dependent variable	0.25	0.31	0.29	0.27	0.28	0.30	0.30	0.24	0.28	0.30
Number of clusters	10,873	14,182	19,091	7,266	20,032	6,639	19,049	9,939	18,812	6,243
Observations	49,665	69,628	86,829	32,461	92,318	26,326	84,183	35,012	93,429	25,864
1st stage Kleibergen-PaapF-statistic	6.45	45.68	45.96	2.50	25.89	28.33	28.19	26.37	27.14	24.34

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. 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, and the number of children of individuals. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

our working sample and re-estimate the first-stage regression in this subsample. Figure 5 shows the distribution of the point estimates across 1,000 random draws. While there is considerably variation in the magnitude of the effect, the point estimate is positive across all subsamples.

Figure 5: Monotonicity assumption: distribution of first stage estimates



Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. This graph draws the distribution of the point estimates of the first stage using equation (1) across 1,000 random subsamples of the working sample.

In summary, these tests and falsification exercises suggest that the reliability and monotonicity assumptions hold, but the exclusion restriction may be violated because the sex ratio also affects other characteristics that are plausibly linked to health.

3.3.4 Addressing violations of the exclusion restriction

We address these potential violations of the exclusion restriction using the "plausibly exogenous" approach proposed by Conley et al. (2012). In their framework, violations of the exclusion restriction are represented by a direct effect of the instrument on the outcome. They propose four different methods that can be used to construct a confidence interval around an estimated treatment effect which remains valid even in the presence of

(small) violations of the exclusion restriction. The methods differ in their assumptions about the plausible values of the direct effects of the instrument.

Here, we implement the "union of confidence intervals" approach, which imposes the fewest assumptions and yields the largest interval. This approach only requires that we specify the support for the direct effect of the instrument on the outcome. We consider the following second stage regression:

$$Y_{it} = \alpha GCcare_{it} + X_{it}^{'}\eta + \gamma Sexratio_{it} + \mu_{it} \tag{3}$$

In equation (3), γ represents a direct effect of the instrument on the outcome, which violates the exclusion restriction. If the true value of γ was known as γ_0, we could simply estimate

$$(Y_{it} - \gamma_0 Sexratio_{it}) = \alpha GCcare_{it} + X_{it}^{'}\eta + \mu_{it} \tag{4}$$

to conduct valid inference on α . If γ_0 is unknown, but the support of γ, G, is limited and known, then we can construct valid confidence intervals for each element of G by assuming that $\gamma = \gamma_0$ and estimating equation (4). The union of these confidence intervals then forms a valid confidence interval for α for any $\gamma \in G$.

We argue that the falsification exercises reported in Tables 3 and A7 represent our best estimates of a direct effect of the sex ratio on older adults' health. The point estimates reported in these tables are small and insignificant, but the standard errors indicate that we cannot exclude the possibility of larger direct effects that are comparable in magnitude to the reduced form effects reported in Table A6. We therefore construct confidence intervals for our causal effect that allow for direct effects of the instruments that as large as the standard errors reported in Table A7 for each outcome.

4 Results

4.1 The effect of grandparental childcare provision on health

Before estimating the causal effect of grandparents' childcare provision on health using the sex ratio instrument, we examine this relationship 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 childcare provision. Table 5 shows the results for our preferred model specification for all four health outcomes. Note that for all health indicators higher values represent worse health outcomes.

Table 5: OLS estimates

Dependent variable	ADL	IADL	Self-reported health	CESD Score
Grandparenting	-0.109***	-0.111***	-0.108***	-0.140***
	(0.007)	(0.006)	(0.010)	(0.018)
Mean of dependent variable	0.30	0.25	2.88	1.49
Number of clusters	25,045	25,044	25,052	24,072
Observations	119,222	119,222	119,234	111,449

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell shows OLS estimates from our preferred model specification with full controls for each dependent variable. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. 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, and the number of children of individuals. 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 grandchild care provision 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 a causal effect of childcare provision 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 grandchild care leads to worse health in the form of more functional limitations, more depressive symptoms, and worse self-reported health. Estimates for limitations in ADL and IADL as well as estimates for self-reported health are significant at the 5% level, whereas the effect on the CESD score is not significant. The Anderson-Rubin test (AR) reported in Table 6 is robust to potential problems caused by weak instruments and confirms the significance of these effects. Tables A8-A11 in the online appendix show that these results are overall robust to the inclusion or exclusion of covariates. The magnitude of these effects ranges from 0.78 standard deviations (SD) for ADLs, 0.86 SD for IADLs, to 1.05 SD for self-reported health. These are substantial effect sizes, which suggest that childcare provision can be a strenuous activity for grandparents. We report estimates for alternative health outcomes measuring limitations in mobility and physical activity as well as cognitive functioning in Appendix Table A12. All estimates are statistically significant and indicate that grandparental childcare provision is detrimental to grandparents' health.

4.2 IV validity and "plausibly exogenous" estimates

We address potential violations of the exclusion restriction by constructing confidence intervals for the estimates in Table 6 using the "union of confidence intervals" approach developed by Conley et al. (2012). Following equations (3)-(4), we plot these confidence intervals against a sensitivity parameter δ , which defines the support of direct effect of the instrument on the outcome as $[0, 2\delta]$. This means that for a given value of δ the shown confidence interval remains valid as long as the violation of the exclusion restriction (i.e.,

Table 6: 2SLS estimates

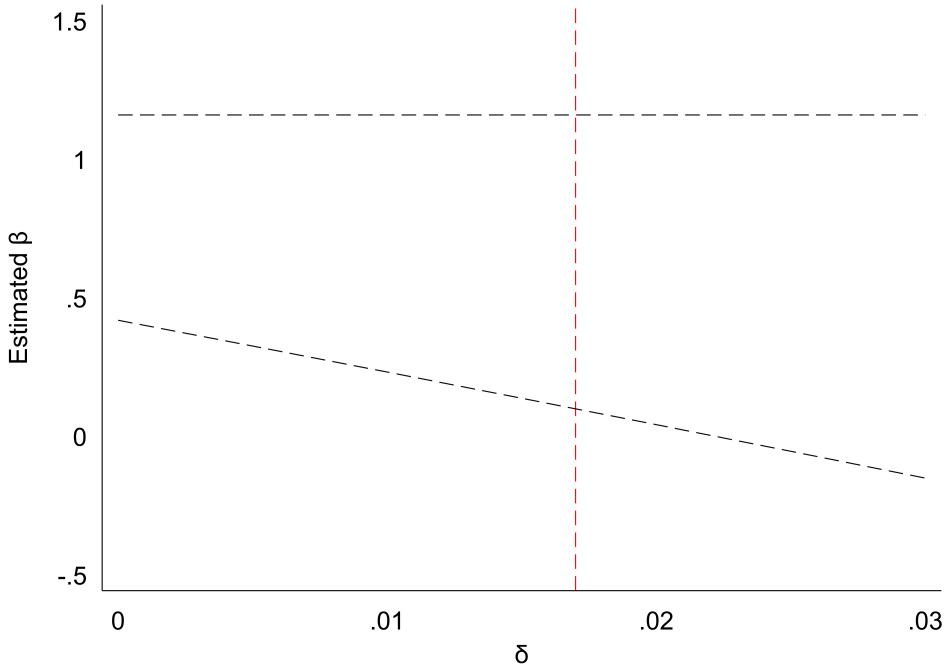
Dependent variable	ADL	IADL	Self-reported health	CESD Score
Grandparenting	0.683**	0.683**	1.178**	0.938
	(0.335)	(0.292)	(0.489)	(0.729)
Mean of dependent variable	0.30	0.25	2.88	1.49
Number of clusters	25,045	25,044	25,052	24,072
Observations	119,222	119,222	119,234	111,449
Kleibergen-Paap F-statistic	46.89	46.82	47.41	50.77
AR F statistic	4.80	6.60	7.00	1.75
AR p-value	0.029	0.010	0.008	0.186

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell shows 2SLS estimates from our preferred model specification with full controls in equation (2) for each dependent variable. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. 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, and the number of children of individuals. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

the direct effect of the instrument on the outcome) does not exceed 2δ .

Figure 6 shows the confidence interval for the estimated effect of grandchild care provision on ADL limitations. The upper grey line shows the upper limit of the 95% confidence interval, which remains constant as we only consider violations of the exclusion restriction that would bias our estimates towards zero. The lower grey line represents the lower limit of the 95% confidence interval as a function of the sensitivity parameter δ. For any given value of δ, the shown confidence interval will be robust to a violation of the exclusion restriction that is at most of magnitude 2δ. The red vertical line shows the value of δ at which the largest permissible violation of the exclusion restriction corresponds to the size of the standard error on the estimate for ADLs in Table A7. Even in the presence of violations of the exclusion restriction that are as large as or slightly larger than the

Figure 6: 95% interval estimates on ADL



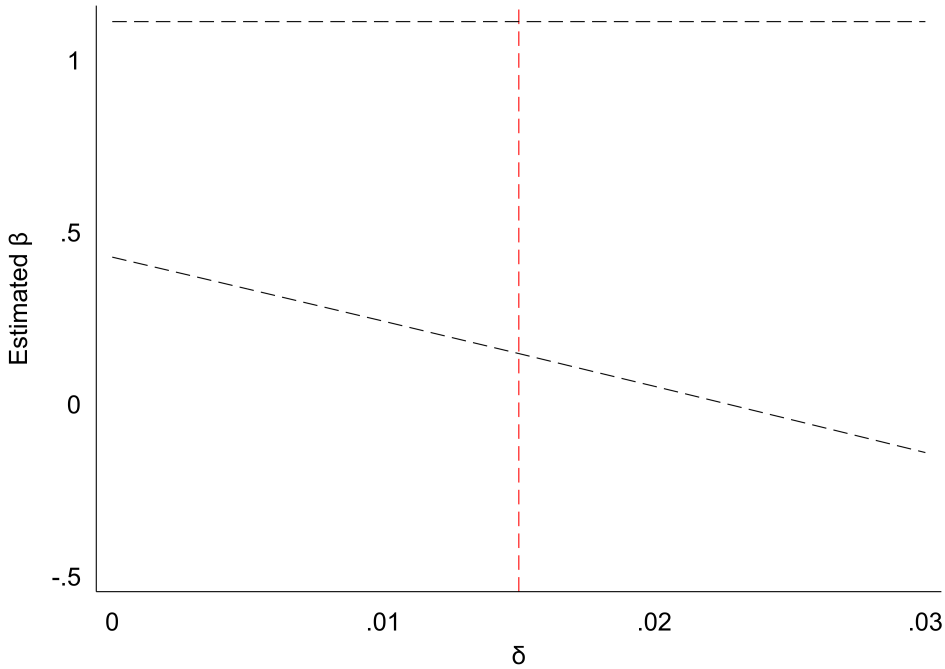
Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. This figure presents 95% confidence intervals for the effect of grandchild care provision on the ADL outcome for violations of the exclusion restriction that do not exceed 2δ. The intervals were constructed using the "union of confidence intervals" by Conley et al. (2012) from equations (3)-(4). The vertical line corresponds to a value of γ equivalent to the size of the standard errors in Appendix Table A7.

standard error for the reduced form estimate in the falsification exercise in Table A7, the 95% confidence interval for our estimate would still exclude zero. This includes potentially very large direct effects of the instrument. For example, a value of $\delta = 0.02$ is equivalent to a direct effect of the instrument (0.04) that is larger than our reduced form estimate for our working sample in Table A6 (0.033). The link between the sex ratio and marital status (-0.020) or informal care receipt (+0.019, see Table A5) is weaker than the first stage reported in Table 2 (+0.048), and therefore we would expect that any direct effects of the sex ratio on health that occur due to a higher risk of divorce or a higher likelihood of informal care receipt should plausibly be smaller than the reduced form effect observed in our working sample.[19] Figure 6 thus implies that our estimate is robust to larger violations

[19]The Wald IV estimator can be derived as the ratio of the reduced form effect to the first stage effect,

of the exclusion restriction than what we would consider plausible.

Figure 7: 95% interval estimates on IADL

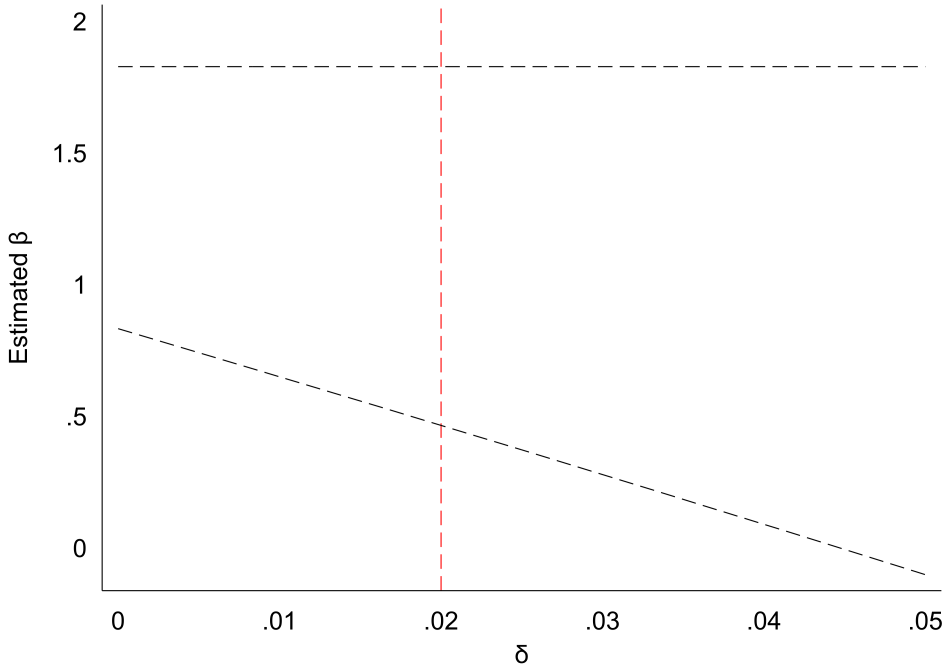


Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. This figure presents 95% confidence intervals for the effect of grandchild care provision on the IADL outcome for violations of the exclusion restriction that do not exceed 2δ. The intervals were constructed using the "union of confidence intervals" by Conley et al. (2012) from equations (3)-(4). The vertical line corresponds to a value of γ equivalent to the size of the standard errors in Appendix Table A7.

Similarly, Figures 7 and 8 show that even in the presence of considerable violations of the exclusion restriction (denoted by the red line) our estimates of the negative health effects of grandparental childcare provision remain statistically significant. While we cannot rule out that the exclusion restriction for the sex ratio instrument is violated, the estimated confidence intervals suggest that our conclusions remain robust even to moderately sized

which in turn implies that the reduced form effect is the product of the first stage effect and the treatment effect. To illustrate our argument, consider the extreme case where the entire reduced form effect documented in Table A6 can be attributed to a violation of the exclusion restriction, because the sex ratio affects, e.g., marital status. Since the link between the sex ratio and marital status is less than half of the first-stage estimate in Table 2, the treatment effect of marital status on health would need to be more than twice as large than the effect of grandchild care provision on health to result in the same reduced form effect. We consider this highly implausible, and therefore argue that any direct effect of the instrument on health operating through changes in marital status should be smaller than our reduced form effect.

Figure 8: 95% interval estimates on self-reported health



Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. This figure presents 95% confidence intervals for the effect of grandchild care provision on the self-reported health outcome for violations of the exclusion restriction that do not exceed 2δ. The intervals were constructed using the "union of confidence intervals" by Conley et al. (2012) from equations (3)-(4). The vertical line corresponds to a value of γ equivalent to the size of the standard errors in Appendix Table A7.

direct effects of the sex ratio on health.

4.3 Robustness

We assess the sensitivity of our results in a series of further 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 A13) 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 A13 are both quantitatively and qualitatively similar to our main results in Table 6.

Appendix Table A14 reports the estimates using several alternative definitions of grandchild care provision 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 panel A of Appendix Table A14 to facilitate comparisons. We consider two 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, and *(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.[20] The results in panels B-C of Appendix Table A14 confirm that our results are robust to these changes across different treatment definitions.

In addition, we check the sensitivity of our estimates to different outcome definitions. Appendix Table A15 reports the results for the health indicators defined with different cutoffs.[21] 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, grandchild care provision significantly increases the likelihood of reporting between one and three

[20] 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.

[21] All regressions are estimated using linear probability models. The results are robust to nonlinear IV estimation (results available on request).

limitations. For IADLs and subjective health, grandchild care is consistently harmful across different definitions. The magnitude of these effects generally decreases as the threshold increases, which indicates that most grandparents experience relatively mild limitations. 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. Finally, we also present first stage estimates using the first-born instrument in Appendix Table A16. As the estimates show, the sex of the first-born child is only weakly or not at all related to the provision of grandchild care, and we therefore do not show 2SLS estimates based on this instrumental variable.

4.4 Mechanisms and heterogeneity

Finally, we consider potential mechanisms that might explain why childcare provision has adverse effects on grandparents' health. We also examine effect heterogeneity across demographic subgroups, since differences in the magnitude of these effects may also provide insights towards such mechanisms. Caring for small children can be physically and mentally demanding, even more so for older adults who might experience declines in physical functioning. If grandchild care itself is a strenuous activity, we would expect that the effects of grandchild care provision increase with intensity. To test this, we re-estimate our IV regressions using indicators based on different threshold values to distinguish between low- and high-intensity childcare provision - any care provision, ¿50 hours, ¿200 hours, ¿500 hours, and ¿1000 hours of grandchild care over the past two years. Panels D-H in Table A14 show that effect sizes are generally larger for more intensive grandchild care provision - when we consider only older adults as treated if they provide at least 500 hours of grandchild care in the past two years, effect sizes are roughly twice as large as in our main specification using a cut-off of 100 hours.

We also consider specifications using the actual hours of grandchild care provision in

Appendix Tables A17 and A18. In line with the estimates in Table A14, estimates for grandchild care hours suggest that an increase in hours spent on grandchild care has negative effects on grandparents' health. The first stage estimates in Appendix Table A17 suggest that the instrument is only weakly related to hours, which is not surprising given the distribution of hours shown in Appendix Figure A2.

Second, it is possible that grandchild care provision crowds out time investment in other activities that are beneficial to grandparents' health. We test this hypothesis by estimating our preferred IV specification using engagement in a range of different activities as outcome variables. We note that the information on social participation is only available for a smaller subsample of the HRS data.[22] The results in Table A19 remain inconclusive - we find negative effects of grandchild care provision on the frequency of watching TV and writing, but no significant effects on, e.g., social participation (e.g., volunteering, charity work) or exercise.

Finally, we examine effect heterogeneity by splitting our sample by gender and race/ethnicity.[23] Table A20 suggests that the effects in our baseline specification are primarily driven by grandmothers. This is not surprising - grandmothers tend to provide more childcare than grandfathers, and consequently our instrument is only weakly related to grandchildcare provision for men. We find few differences between White and Black/Hispanic respondents.

4.5 Complier and external validity

The IV estimates in Table 6 represent a local average treatment effect (LATE), i.e., the effect of grandparental childcare provision on health for individuals whose decision to care for their grandchildren is determined by the sex ratio. Determining the size or characteristics of this complier population is not possible with our multi-valued instrument; however, based on the

[22]Social participation is covered in the Psychosocial and Lifestyle Questionnaire as a left-behind survey in HRS, which was introduced in 2006 and is given at every wave to a 50% subsample of core respondents. The consistent activity questions in this questionnaire are available from 2008.

[23]We choose to estimate regressions on separate samples rather than modelling interactions to allow the effects of covariates to also differ across subsamples.

discussion of the mechanisms connecting the sex ratio to grandchild care provision we can draw some tentative conclusions. As noted in section 3.1, complier with more daughters are more likely to provide grandchild care, because they transition to grandparenthood earlier and because they invest more in their daughters (e.g., due to stronger emotional bonds). This would suggest that complier providing childcare are on average younger and have a stronger bond with their children than "always taker" (i.e., grandparents providing childcare regardless of the values of the instrument).

If these younger grandparents are still active on the labor market, they might either have to reconcile their childcare provision with their working hours, creating a double burden; or they might choose to retire early (Rupert and Zanella 2018) with possible negative consequences for their health (Fitzpatrick and Moore 2018). It also seems plausible that younger grandparents (who have experienced less physical and cognitive decline) and those with a stronger bond to their children are more likely to continue providing grandchild care even if they perceive care provision to negatively affect their health. This suggests that treatment effects on the complier might be larger than effects in the general population.

We provide some tentative empirical evidence by re-estimating our IV regressions in a marginal treatment effect (MTE) framework (Heckman and Vytlacil 2007). The MTE is estimated under the same assumptions as standard IV models and measures the expected treatment effect as a function of an individual's unobserved resistance to treatment. Intuitively, individuals select into treatment based on their observable characteristics (incl. the instrument) and an unobserved resistance to treatment (e.g., based on their expected gains from treatment). In this setting, the instrumental variable is used to estimate the propensity score of treatment, which in turn is used to estimate the marginal treatment effect at different values of the unobserved resistance to treatment (Brinch et al. 2017).[24] One attractive feature of the MTE framework is that the estimated MTEs can be used to

[24] Under the assumption that individuals are treated if their propensity score is larger than their resistance to treatment.

derive different treatment effects, including the LATE or the average treatment effect.

Figures A3-A6 show the results for our four health outcomes. We note that for ADLs and subjective health, the marginal treatment effects are broadly stable across the distribution of the resistance to treatment, which suggests that the estimated effects from our IV regressions are likely to hold more broadly beyond the complier population. In line with the discussion above, we observe that for IADLs the marginal treatment effect is a downward-sloping function, i.e., individuals who are more likely to be treated are more heavily affected than individuals who are less likely to be treated. For the CESD score, we observe an upward-sloping pattern, but we note that the confidence bands only exclude zero for a very small region of the unobserved resistance to treatment (in line with our finding of a non-significant effect). We interpret these findings as suggestive evidence that our estimated effects might have some validity beyond the complier population.

5 Discussion

This study examines the effect of childcare provision on grandparents' health in the U.S.. We use the sex ratio as an instrument 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 suggest that the exclusion restriction for the sex ratio may not hold, because having a daughter is linked to other characteristics such as marital status and the receipt of informal care. We address these violations of the exclusion restriction by deriving 95% confidence intervals that remain valid in the presence of small or moderately sized direct effects of the sex ratio on health using the "plausibly exogenous" approach by Conley et al. (2012).

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 grandchild care provision on health are predominantly negative.

We find that grandparental childcare provision leads to an increase in ADLs by 0.79 standard deviations (SD), an increase in IADLs by 0.86 SD and worsens self-reported health by 1.05 SD. These are substantial negative effects. However, we argue that the magnitude of these effects should be interpreted with caution. First, as we discuss in section 3.3.4, it is possible that the exclusion restriction might be violated. In particular, a higher likelihood of divorce would likely exert a negative influence on health, thus biasing our results away from zero. We construct confidence intervals that are robust to moderate violations of the exclusion restriction, yet this means that our results are set- and not point-identified. In other words, the magnitude of the effects might be lower than our 2SLS point estimates suggest. Second, theoretical considerations about the complier population (see section 4.5) suggest that the LATE identified for the complier population might be larger than an average treatment effect. We provide some tentative evidence that for ADLs and self-reported health the marginal treatment effects are stable, however, for IADLs we indeed find that the marginal treatment effects follow a downward-sloping curve. Third, we note that all our health measures are discrete, and a 1-point change is the smallest possible change that an individual can experience. An estimated effect size of 1.2 for self-reported health (or 1.05 SD) does therefore not necessarily mean that some treated individuals experience a substantial health shock. Instead, it could also be that most treated individuals experience a small 1-point change in their health status. For these reasons, we focus in our interpretation primarily on the qualitative direction of these health effects rather than their magnitude.

We provide some tentative evidence that the negative effects of grandchild care provision on health are stronger for high-intensity childcare provision, whereas activity substitution

does not seem to play an important role. The effects are more pronounced for grandmothers than for grandfathers.

Our findings suggest that previously reported positive associations between grandchild care 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. In particular, exploring potential mechanisms in more detail would require a more reliable measure of actual care hours, e.g., based on time use diaries. Similarly, it seems plausible that the effects of grandchild care 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. We alternatively considered engagement in a range of activities as potential mechanisms. While we find no evidence for activity substitution, this should only be considered as suggestive, since the results are based on a much smaller sample than our main findings.

In summary, our results show that grandparental childcare provision does not improve the health of grandparents, rather it may be detrimental. 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 childcare provision should not be considered as "active ageing" – a socially desirable activity that preserves or improves older adults' health. Instead, childcare provision 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. Consequently, family policies that improve the availability

and affordability of formal childcare may generate positive externalities as grandparents may feel less obliged to help out their adult children to the detriment of their own health. Rather, they may enjoy the grandparents' privilege and pass on the baton when the going gets rough.

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.

Bratti, M., Frattini, T., and Scervini, F. (2018). Grandparental availability for child care and maternal labor force participation: Pension reform evidence from Italy. *Journal of Population Economics*, 31:1239–1277.

Brinch, C. N., Mogstad, M., and Wiswall, M. (2017). Beyond LATE with a discrete instrument. *Journal of Political Economy*, 125(4):985–1039.

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.

Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272.

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.

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.

Dimova, R. and Wolff, F.-C. (2011). Do downward private transfers enhance maternal labor supply? evidence from around Europe. *Journal of Population Economics*, 24:911–933.

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.

Fitzpatrick, M. D. and Moore, T. J. (2018). The mortality effects of retirement: Evidence from social security eligibility at age 62. *Journal of Public Economics*, 157:121–137.

Frimmel, W., Halla, M., Schmidpeter, B., and Winter-Ebmer, R. (2022). Grandmothers' labor supply. *Journal of Human Resources*, 57(5):1645–1689.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of Econometrics*, 6:4875–5143.

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.

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.

Salm, M., Siflinger, B. M., and Xie, M. (2021). The effect of retirement on mental health: Indirect treatment effects and causal mediation.

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.

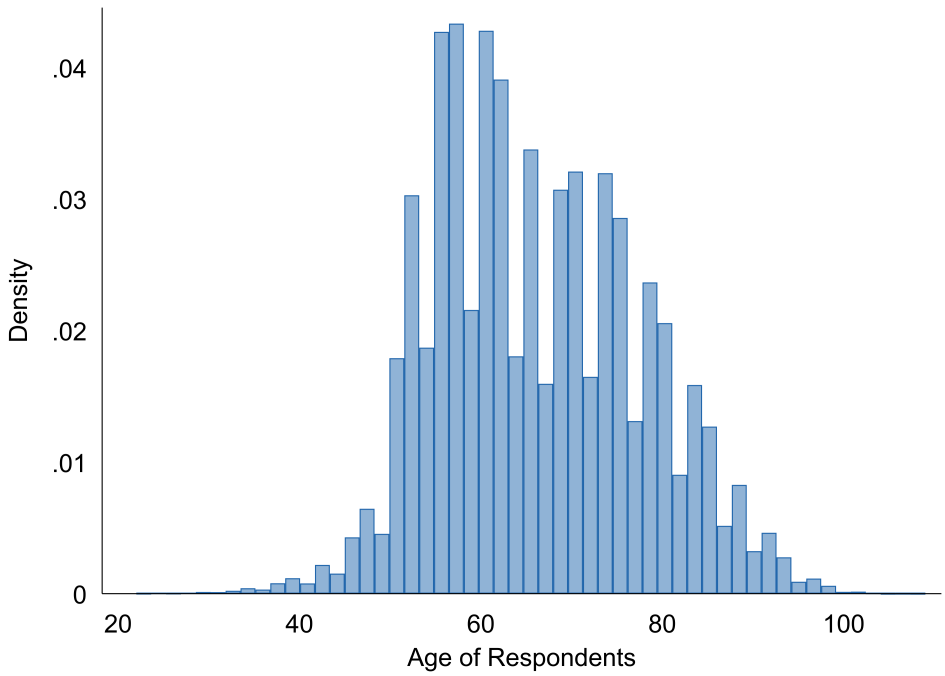
Van Houtven, C. H. and Norton, E. C. (2004). Informal care and health care use of older adults. *Journal of Health Economics*, 23(6):1159–1180.

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.

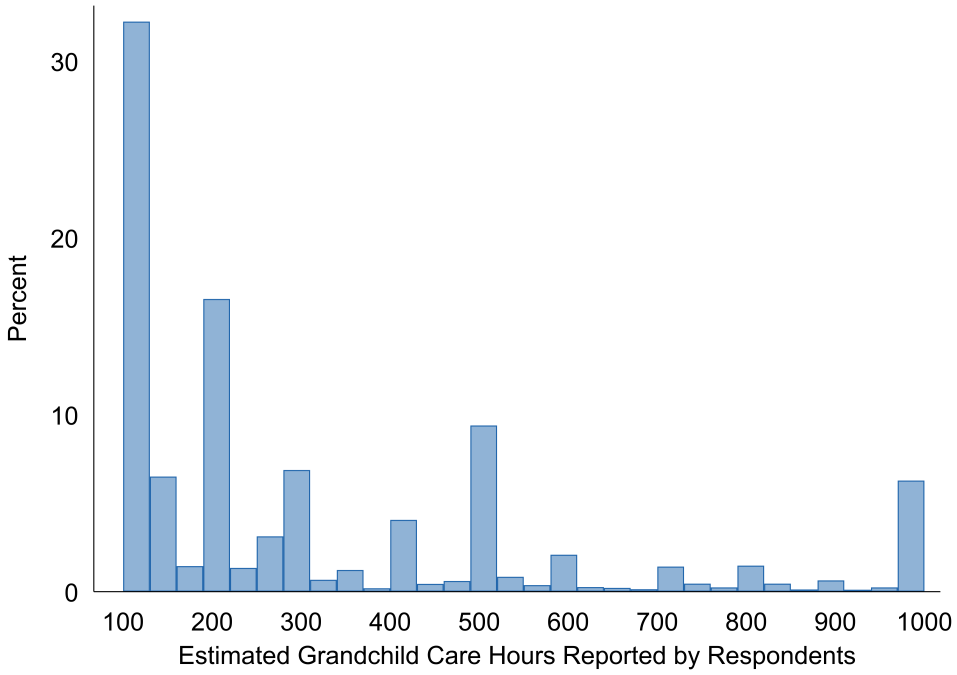
Zeng, Y., Chen, Y.-C., and Lum, T. Y. S. (2021). Longitudinal impacts of grandparent caregiving on cognitive, mental, and physical health in China. *Aging & Mental Health*, 25(11):2053–2060. Publisher: Routledge.

Figure A1: Distribution of Age of HRS Respondents



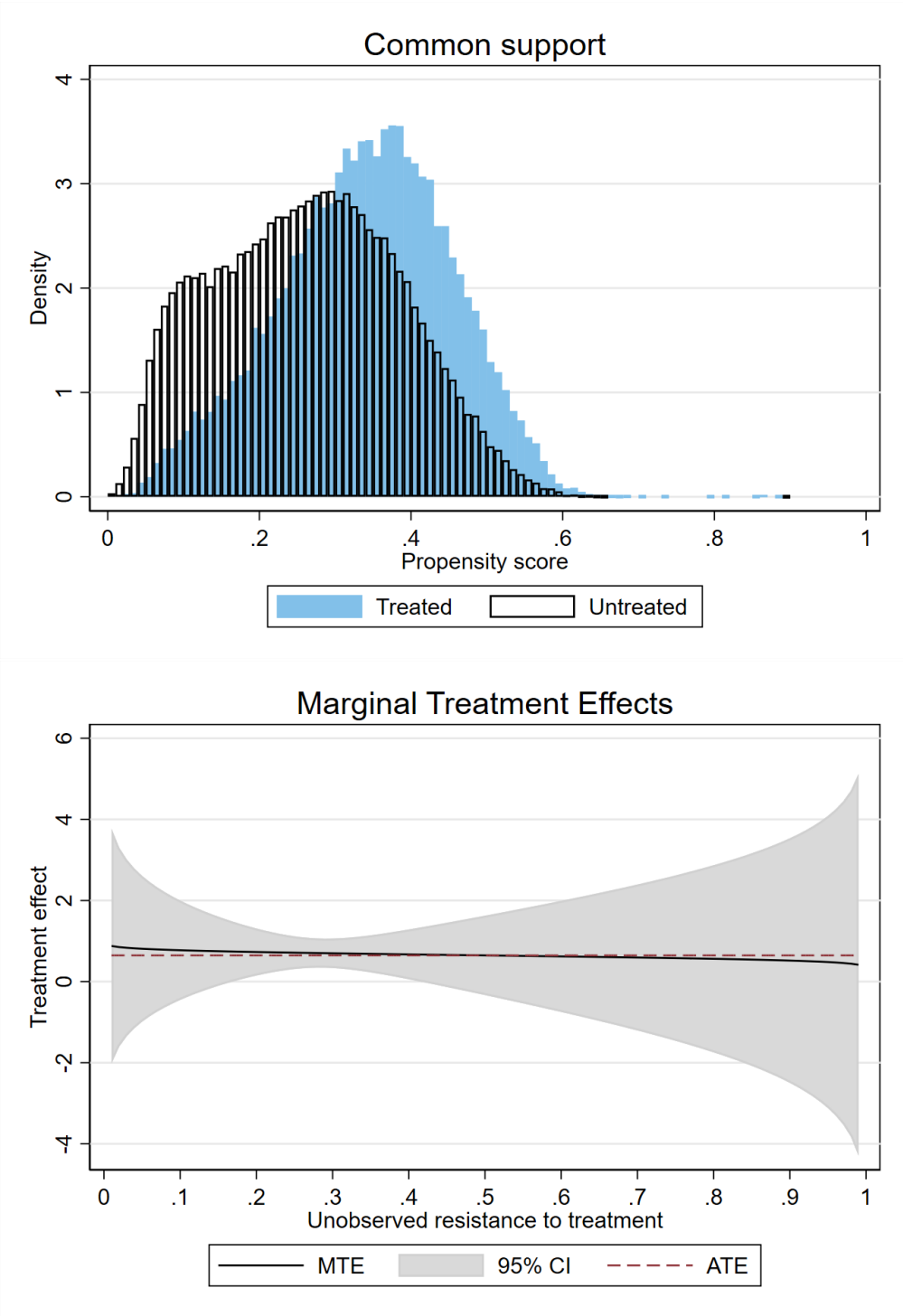
Notes: The data is the HRS from 1992 to 2014. This graph draws the distribution of age of HRS respondents. The vertical axis shows the density of age. The age eligibility of respondents is above 50 and the age of their spouses can be any age as shown in the plot.

Figure A2: Distribution of grandchild care hours



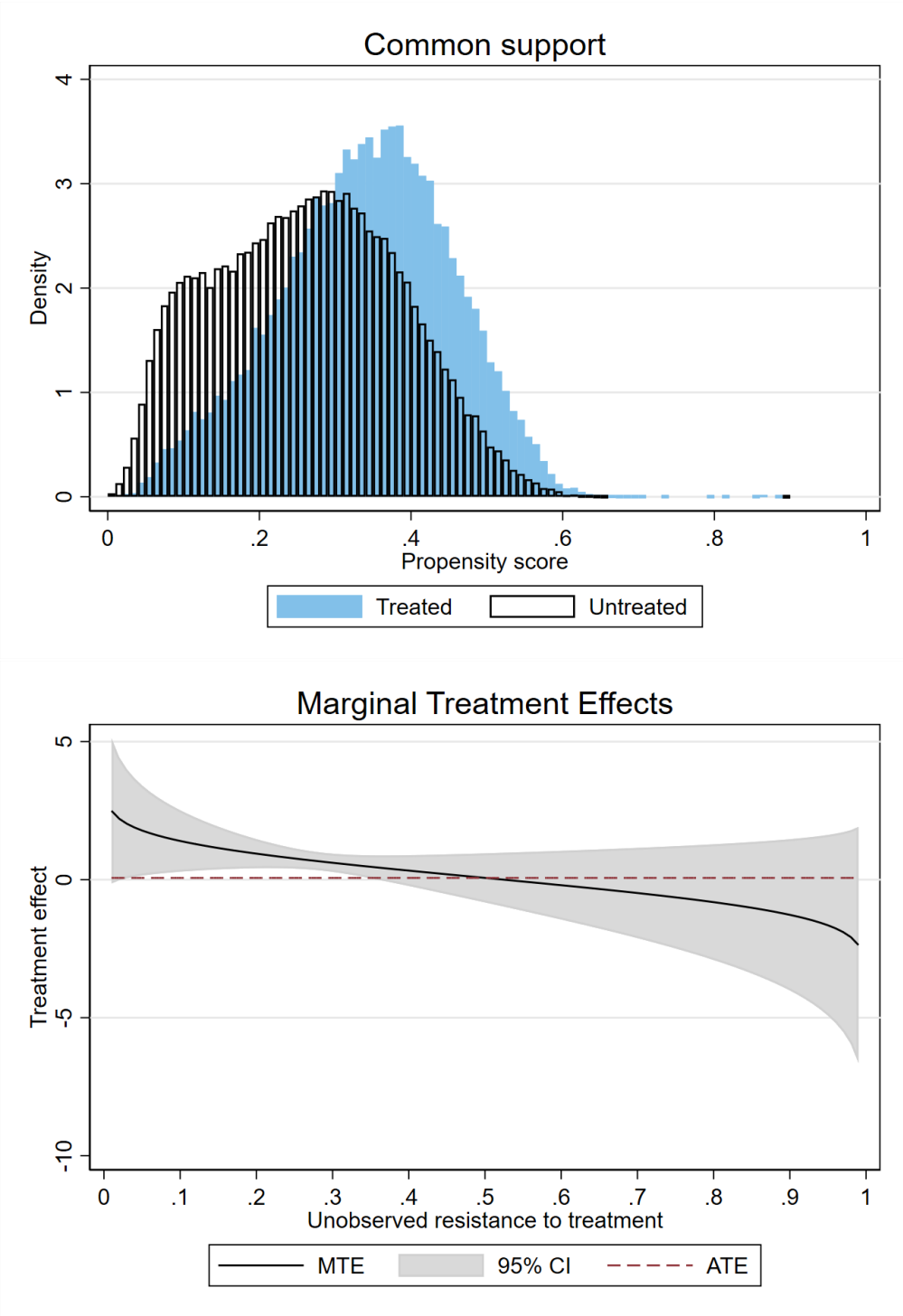
Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. This graph draws the distribution of grandchild care hours for respondents who are grandparenting and provide less than 1,000 hours over the last two years. Respondents providing less than 100 hours of grandchild care (defined as not grandparenting in our study) are omitted for clarity.

Figure A3: Generalization of LATE estimate on ADL outcome



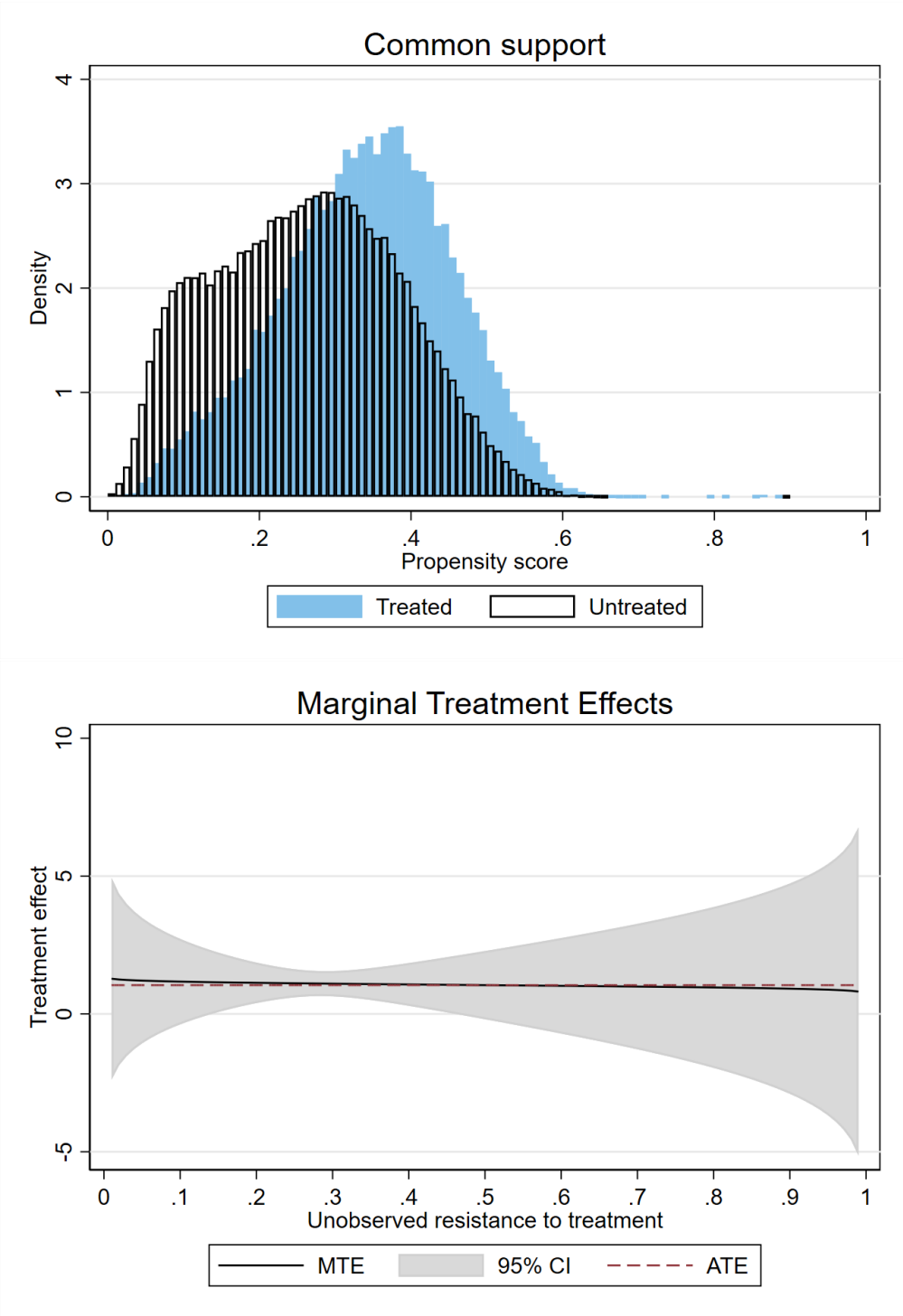
Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. The upper graph shows the overlapping propensity score for individuals who provide grandchild care provision and who do not. The lower graph plots the marginal treatment effect across the distribution of the unobserved resistance to treatment. The standard errors are bootstrapped 50 times.

Figure A4: Generalization of LATE estimate on IADL outcome



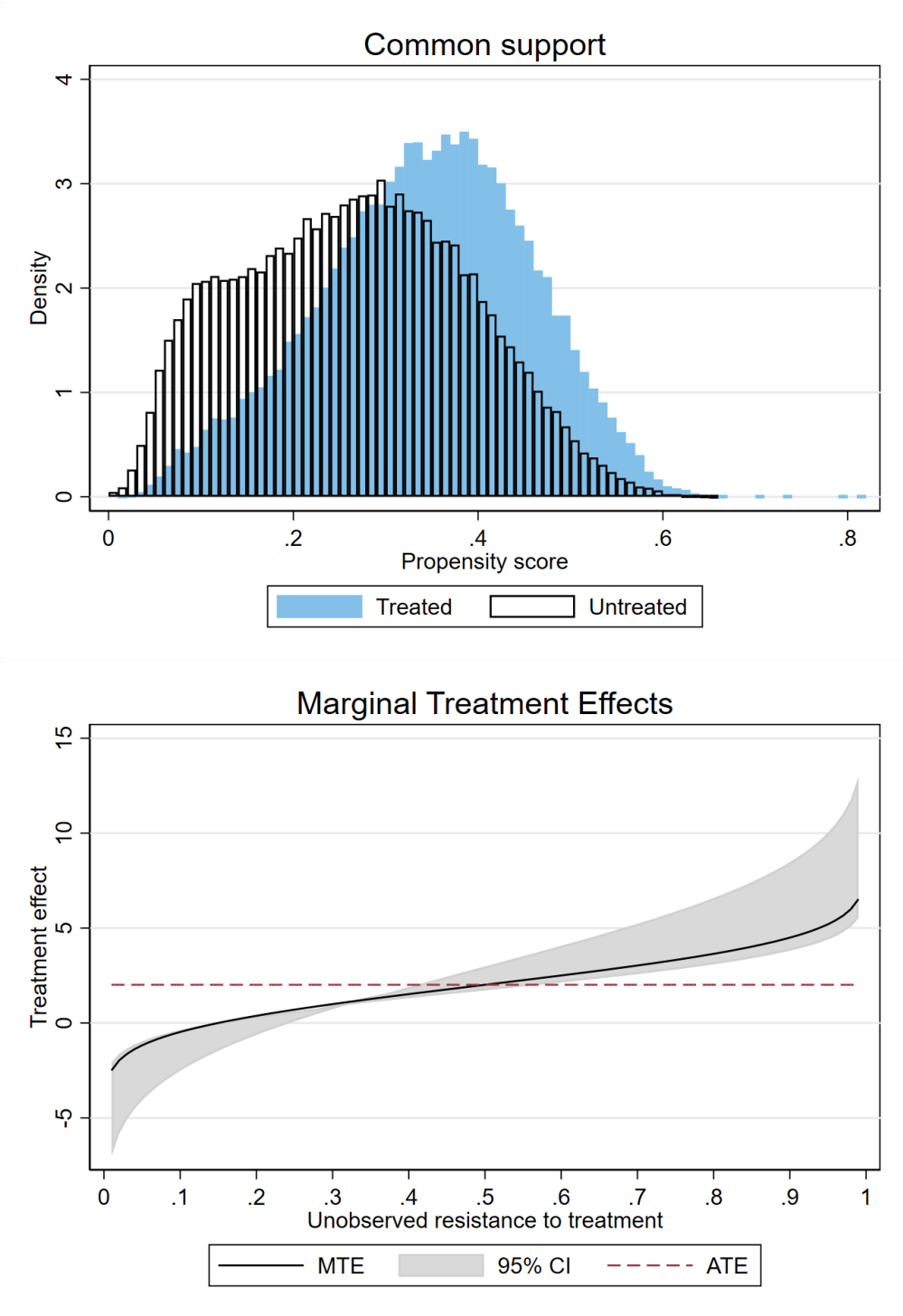
Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. The upper graph shows the overlapping propensity score for individuals who provide grandchild care provision and who do not. The lower graph plots the marginal treatment effect across the distribution of the unobserved resistance to treatment. The standard errors are bootstrapped 50 times.

Figure A5: Generalization of LATE estimate on self-reported health outcome



Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. The upper graph shows the overlapping propensity score for individuals who provide grandchild care provision and who do not. The lower graph plots the marginal treatment effect across the distribution of the unobserved resistance to treatment. The standard errors are bootstrapped 50 times.

Figure A6: Generalization of LATE estimate on CESD outcome



Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. The upper graph shows the overlapping propensity score for individuals who provide grandchild care provision and who do not. The lower graph plots the marginal treatment effect across the distribution of the unobserved resistance to treatment. The standard errors are bootstrapped 50 times.

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	
Self-reported health	Respondent's self-reported general health status, 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", and 5 for "poor".
ADL	Index of difficulties in Activities of Daily Living (ADL) with the range from 0 to 5, indicating respondents having any problem in bathing, eating, getting dressed, getting in/out of bed, and walking across a room
IADL	Index of difficulties in Instrumental Activities of Daily Living (IADL) with the range from 0 to 5, indicating respondents having any problem in using the phone, managing money, taking medications, shopping for groceries, and preparing hot meals
CESD Score	Center for Epidemiological Studies Depression (CESD) scale with the range from 0 to 8: 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: Robustness of 2SLS estimates using another definition of sex ratio

Dependent variable	ADL	IADL	Self-reported health	CESD Score
Grandparenting	0.918**	0.870***	1.746***	1.787**
	(0.362)	(0.317)	(0.546)	(0.781)
Mean of dependent variable	0.303	0.249	2.878	1.486
Number of clusters	24,968	24,967	24,975	23,988
Observations	117,593	117,593	117,608	109,902
Kleibergen-Paap F-stat	43.99	43.89	44.43	49.07
AR F statistic	8.063	9.896	14.4	6.103
AR p-value	0.005	0.002	0.000	0.014

Notes: Each cell shows 2SLS estimates from our preferred model specification with full controls in equation (2) for each sub-sample. This sample uses data from the HRS 1996 to 2014 who are 50 to 80. The sex ratio is defined by dividing the number of daughters on the number of living children as an alternative measure. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. All models use the main specification with full controls. See text for details. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A3: Robustness of 2SLS estimates with both FE on youngest and oldest child

Dependent variable	ADL	IADL	Self-reported health	CESD Score
Grandparenting	0.705**	0.707**	1.174**	0.974
	(0.338)	(0.296)	(0.492)	(0.735)
Mean of dependent variable	0.304	0.25	2.88	1.492
Number of clusters	25046	25045	25053	24073
Observations	119,223	119,223	119,235	111,450
Kleibergen-Paap F-stat	46.35	46.28	46.86	50.2
AR F statistic	5.071	7.011	6.884	1.863
AR p-value	0.024	0.008	0.009	0.172

Notes: Each cell shows 2SLS estimates from our preferred model specification with full controls in equation (2) for each sub-sample. This sample uses data from the HRS 1996 to 2014 of all individuals who are 50 to 80. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. All models use the specification with full controls and replaces the fixed effects for the year of birth of the first-born child of an individual with the fixed effects for the age of the oldest child of an individual. See text for details. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A4: Robustness of 2SLS estimates dropping individuals with time-varying sex ratio

Dependent variable	ADL	IADL	Self-reported health	CESD Score
Grandparenting	0.744**	0.671**	1.257**	0.961
	(0.346)	(0.299)	(0.510)	(0.767)
Mean of dependent variable	0.295	0.243	2.858	1.447
Number of clusters	21457	21456	21464	20548
Observations	97,024	97,023	97,030	90,644
Kleibergen-Paap F-stat	43.6	43.55	43.97	44.75
AR F statistic	5.497	6.081	7.513	1.658
AR p-value	0.019	0.014	0.006	0.198

Notes: Each cell shows 2SLS estimates from our preferred model specification with full controls in equation (2) for each sub-sample. This sample uses data from the HRS 1996 to 2014 who are 50 to 80 and drops individuals who have time-varying sex ratio values. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. All models use the main specification with full controls. See text for details. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A5: The first stage estimates on potential confounders

	Marital Status			Help from Children			Living Proximity or Contact w/ Children			
Dependent variable	Married	Divorced	Widowed	Any Help	ADL	IADL	Close	w/ 10 mi	w/ Children	In Contact
A. Estimates with controls										
Sex ratio	-0.020**	0.019**	0.000	0.019***	0.009***	0.014***	0.005	0.014	-0.033***	0.000
	(0.009)	(0.007)	(0.007)	(0.003)	(0.002)	(0.003)	(0.009)	(0.010)	(0.006)	(0.003)
Mean of DV	0.71	0.12	0.16	0.05	0.03	0.04	0.70	0.60	0.24	0.95
Observations	119,195	119,195	119,195	119,293	119,286	119,278	117,411	115,685	119,293	77,372
KP F-stat	4.75	6.35	0.00	32.27	19.20	24.04	0.33	2.00	26.19	0.01
B. Estimates without controls										
Sex ratio	-0.025**	0.020***	0.002	0.021***	0.011***	0.015***	0.011	0.014	-0.022***	-0.005
	(0.010)	(0.007)	(0.008)	(0.003)	(0.002)	(0.003)	(0.010)	(0.010)	(0.008)	(0.004)
Mean of DV	0.71	0.12	0.16	0.05	0.03	0.04	0.70	0.60	0.24	0.95
Observations	119,967	119,967	119,967	120,066	120,059	120,051	118,108	116,358	120,066	77,906
KP F-stat	6.49	7.33	0.07	38.95	24.57	29.46	1.25	1.91	8.19	1.93

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. The sex ratio is defined as the number of daughters divided by the total number of children of an individual. Section A reports estimates with controls 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, and the number of children of individuals. Section B shows estimates without any controls. Columns 1-3 report the estimates of sex ratio on marital status. Column 1 is an indicator of whether an individual is married or partnered; column 2, separated or divorced; and column 3 indicates whether an individual is widowed. Columns 4-6 focus on whether an individual's children provide help. Column 4 shows whether a child provides any help, help with basic activities of daily living (ADLs), such as bathing, dressing, eating, toileting, transferring, and walking in column 5, help with Instrumental Activities of Daily Living (IADL), like using the phone, managing money, taking medications, shopping for groceries, and preparing hot meals in column 6. Columns 7-10 report whether children are living within close proximity with the respondent. Column 7 displays whether an individuals children live in close proximity such as living together or living within 10 miles of individuals, column 8, living within 10 miles, column 9, living together, and column 10, whether an individual is in contact with their children. The mean of each dependent variable (DV) is shown in each column. "KP F-stat" denotes the cluster-robust Kleibergen-Paap (KP) F-statistic on testing weak instruments. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A6: Reduced form regression of the working sample

Dependent variable	ADL	IADL	Self-reported health	CESD Score
Sex ratio	0.033**	0.033**	0.057***	0.048
	(0.015)	(0.013)	(0.021)	(0.036)
Mean of dependent variable	0.30	0.25	2.88	1.49
Number of clusters	25,045	25,044	25,052	24,072
Observations	119,222	119,222	119,234	111,449

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell shows reduced form estimates from our preferred model specification with full controls. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. 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, and the number of children of individuals. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A7: Falsification exercise: Reduced form regression

Dependent variable	ADL	IADL	Self-reported health	CESD Score
Sex ratio	-0.003	0.007	0.027	0.008
	(0.034)	(0.030)	(0.041)	(0.081)
Mean of dependent variable	0.32	0.27	2.98	1.66
Number of clusters	7,990	7,988	7,993	7,320
Observations	8,638	8,636	8,640	7,915

Notes: The data used are from the HRS 1996 to 2014 of the subsample of respondents who are 50 to 80 and who do not have grandchildren now but have later. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. 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, and the number of children of individuals. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A8: Robustness of 2SLS estimates on ADL

Dependent variable: ADL							
Model	1	2	3	4	5	6	7
Grandparenting	0.789**	0.722**	0.681**	0.677**	0.683**	0.688**	0.530*
	(0.330)	(0.325)	(0.324)	(0.335)	(0.335)	(0.336)	(0.317)
Year FE + Birth year FE of first born		Y	Y	Y	Y	Y	Y
Cohort FE + Birth year FE of youngest born			Y	Y	Y	Y	Y
Demographics of R				Y	Y	Y	Y
Family size					Y	Y	Y
Age FE						Y	Y
Labor Force Participation and Living Proximity							Y
Mean of dependent variable	0.305	0.305	0.305	0.304	0.304	0.304	0.281
Number of clusters	25289	25217	25192	25045	25045	25045	24878
Observations	119,992	119,895	119,807	119,222	119,222	119,222	116,295
Kleibergen-Paap F-statistic	46.71	49.30	49.78	46.95	46.89	46.80	41.62
AR F statistic	6.936	5.754	5.081	4.714	4.800	4.865	3.115
AR p-value	0.008	0.017	0.024	0.030	0.029	0.027	0.078

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell reports estimates from a separate specification using equation (2) for the dependent variable. ADL is the number of limitations reported by an individual with the range from 0 to 5. The details of this health outcome can be referred to Appendix Table A1. 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 replaces the quadratic form of age of respondent with age fixed effects. Column 7 contains control for the labor force participation and the living proximity (within 10 miles with children). Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A9: Robustness of 2SLS estimates on IADL

Dependent variable: IADL							
Model	1	2	3	4	5	6	7
Grandparenting	0.768***	0.690**	0.652**	0.676**	0.683**	0.690**	0.631**
	(0.292)	(0.284)	(0.282)	(0.292)	(0.292)	(0.293)	(0.272)
Year FE + Birth year FE of first born		Y	Y	Y	Y	Y	Y
Cohort FE + Birth year FE of youngest born			Y	Y	Y	Y	Y
Demographics of R				Y	Y	Y	Y
Family size					Y	Y	Y
Age FE						Y	Y
Labor Force Participation and Living Proximity							Y
Mean of dependent variable	0.251	0.251	0.251	0.250	0.250	0.250	0.225
Number of clusters	25287	25215	25190	25044	25044	25044	24878
Observations	119,991	119,894	119,806	119,222	119,222	119,222	116,295
Kleibergen-Paap F-statistic	46.60	49.20	49.67	46.87	46.82	46.73	41.61
AR F statistic	8.902	7.160	6.390	6.476	6.604	6.725	6.576
AR p-value	0.003	0.007	0.012	0.011	0.010	0.010	0.010

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell reports estimates from a separate specification using equation (2) for the dependent variable. IADL is the number of limitations reported by an individual with the range from 0 to 5.The details of this health outcome can be referred to Appendix Table A1. 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 replaces the quadratic form of age of respondent with age fixed effects. Column 7 contains control for the labor force participation and the living proximity (within 10 miles with children). Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A10: Robustness of 2SLS estimates on self-reported health

Dependent variable: self-reported health							
Model	1	2	3	4	5	6	7
Grandparenting	1.328***	1.302***	1.260***	1.170**	1.178**	1.182**	1.124**
	(0.490)	(0.485)	(0.483)	(0.488)	(0.489)	(0.490)	(0.492)
Year FE + Birth year FE of first born		Y	Y	Y	Y	Y	Y
Cohort FE + Birth year FE of youngest born			Y	Y	Y	Y	Y
Demographics of R				Y	Y	Y	Y
Family size					Y	Y	Y
Age FE						Y	Y
Labor Force Participation and Living Proximity							Y
Mean of dependent variable	2.882	2.881	2.881	2.88	2.88	2.88	2.871
Number of clusters	25297	25224	25199	25052	25052	25052	24877
Observations	120,005	119,907	119,819	119,234	119,234	119,234	116,244
Kleibergen-Paap F-statistic	47.19	49.82	50.26	47.46	47.41	47.32	41.88
AR F statistic	9.386	8.985	8.380	6.918	7.004	7.044	6.281
AR p-value	0.002	0.003	0.004	0.009	0.008	0.008	0.012

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell reports estimates from a separate specification using equation (2) for the dependent variable. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". The details of this health outcome can be referred to Appendix Table A1. 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 replaces the quadratic form of age of respondent with age fixed effects. Column 7 contains control for the labor force participation and the living proximity (within 10 miles with children). Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A11: Robustness of 2SLS estimates on CESD Score

Dependent variable: CESD Score							
Model	1	2	3	4	5	6	7
Grandparenting	1.574**	1.295*	1.234*	0.935	0.938	0.951	0.914
	(0.741)	(0.726)	(0.727)	(0.729)	(0.729)	(0.730)	(0.758)
Year FE + Birth year FE of first born		Y	Y	Y	Y	Y	Y
Cohort FE + Birth year FE of youngest born			Y	Y	Y	Y	Y
Demographics of R				Y	Y	Y	Y
Family size					Y	Y	Y
Age FE						Y	Y
Labor Force Participation and Living Proximity							Y
Mean of dependent variable	1.495	1.494	1.494	1.492	1.492	1.492	1.486
Number of clusters	24299	24234	24207	24072	24072	24072	23959
Observations	112,116	112,028	111,942	111,449	111,449	111,449	109,212
Kleibergen-Paap F-statistic	50.19	53.17	53.1	50.79	50.77	50.66	44.63
AR F statistic	5.108	3.453	3.106	1.735	1.747	1.79	1.535
AR p-value	0.024	0.063	0.078	0.188	0.186	0.181	0.215

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell reports estimates from a separate specification using equation (2) for the dependent variable. CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of this health outcome can be referred to Appendix Table A1. 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 replaces the quadratic form of age of respondent with age fixed effects. Column 7 contains control for the labor force participation and the living proximity (within 10 miles with children). Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A12: 2SLS estimates on other health outcomes

Dependent variable	Gross motor skill	Mobility	Large muscle	Fine mobility	Cognition score
Grandparenting	0.841**	1.470**	1.362**	0.519**	-5.026*
	(0.428)	(0.624)	(0.577)	(0.203)	(2.569)
Mean of dependent variable	0.49	1.07	1.29	0.19	22.36
Number of clusters	25,046	25,042	25,042	25,045	21,389
Observations	119,226	119,187	119,204	119,225	72,590
Kleibergen-Paap F-statistic	46.97	46.87	46.98	46.90	30.79
AR F statistic	4.446	6.649	6.490	7.920	4.682
AR p-value	0.035	0.010	0.011	0.005	0.031

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell shows 2SLS estimates from our preferred model specification with full controls in equation (2) for each dependent variable. Gross motor skills limitation in column 1 is an index of gross motor skill difficulties ranging from 0 to 5, indicating respondents having any problem in walking one block, walking across a room, climbing one flight of stairs, getting in or out of bed, and bathing activities. Mobility difficulty in column 2 is an index of mobility difficulties ranging from 0 to 5, indicating respondents having any problem in walking one block, walking several blocks, walking across a room, climbing one flight of stairs, and climbing several flights of stairs. Large muscle limitation in column 3 indexes for difficulty items such as sitting for 2 hrs, getting up from a chair, stooping, kneeling or crouching, and pushing or pulling large objects activities with the range from 0 to 5. Fine motor skills limitation in column 4 indexes for any difficulty in picking up a dime, eating, and dressing activities with the range from 0 to 3. Cognition score in column 5 is the total cognition score which is the sum of the total word recall and mental status test scores ranging from zero to 35. The word recall index sums the immediate and delayed word recall test scores. The mental status index includes the scores for serial 7's, counting backwards from 20, naming objects, recalling dates, and naming the president/vice-president. The higher the cognitive score, the better the health. 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, and the number of children of individuals. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A13: Robustness of estimates to sample change

Dependent variable	ADL	IADL	Self-reported health	CESD Score
A. Sample without age restrictions				
Grandparenting	0.821**	0.853**	1.227**	0.437
	(0.398)	(0.379)	(0.516)	(0.742)
Mean of dependent variable	0.396	0.369	2.924	1.526
Number of clusters	27565	27566	27574	26227
Observations	143,381	143,381	143,386	131,643
Kleibergen-Paap F-statistic	48.00	47.89	48.52	53.49
AR F statistic	4.937	6.005	6.766	0.353
AR p-value	0.026	0.014	0.009	0.553
B. Sample restricted to grandparents				
Grandparenting	0.830**	0.809**	1.307**	1.241
	(0.383)	(0.335)	(0.557)	(0.828)
Mean of dependent variable	0.302	0.247	2.871	1.477
Number of clusters	24631	24630	24639	23650
Observations	110,028	110,030	110,039	103,044
Kleibergen-Paap F-statistic	36.80	36.76	37.20	40.04
AR F statistic	5.731	7.500	6.873	2.448
AR p-value	0.017	0.006	0.009	0.118

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 50 to 80 and excludes those who report no grandchildren. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. All models use the main specification with full controls. See text for details. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A14: Robustness of estimates to treatment definitions and treatment intensity

Dependent variable	ADL	IADL	Self-reported health	CESD Score
A. Main specification	0.683**	0.683**	1.178**	0.938
	(0.335)	(0.292)	(0.489)	(0.729)
Mean of dependent variable	0.30	0.25	2.88	1.49
Kleibergen-Paap F-statistic	46.89	46.82	47.41	50.77
B. Couples' hours	0.623**	0.622**	1.075**	0.880
	(0.305)	(0.265)	(0.443)	(0.684)
Mean of dependent variable	0.30	0.25	2.88	1.49
Kleibergen-Paap F-statistic	50.71	50.74	51.23	52.39
C. Excluding reported maximum hours	0.804**	0.813**	1.408**	1.072
	(0.382)	(0.336)	(0.557)	(0.807)
Mean of dependent variable	0.31	0.25	2.88	1.50
Kleibergen-Paap F-statistic	39.88	39.81	40.30	43.79
D. > 0 hours	0.643**	0.643**	1.110**	0.889
	(0.316)	(0.276)	(0.461)	(0.691)
Mean of dependent variable	0.30	0.25	2.88	1.49
Kleibergen-Paap F-statistic	47.48	47.42	47.91	50.92
E. > 50 hours	0.680**	0.679**	1.173**	0.929
	(0.335)	(0.293)	(0.490)	(0.723)
Mean of dependent variable	0.30	0.25	2.88	1.49
Kleibergen-Paap F-statistic	44.55	44.48	44.97	48.79
F. > 200 hours	0.915**	0.923**	1.573**	1.235
	(0.439)	(0.386)	(0.637)	(0.947)
Mean of dependent variable	0.31	0.25	2.88	1.49
Kleibergen-Paap F-statistic	36.82	36.82	37.07	38.57
G. > 500 hours	1.341**	1.328**	2.103**	1.557
	(0.614)	(0.543)	(0.870)	(1.285)
Mean of dependent variable	0.31	0.26	2.89	1.49
Kleibergen-Paap F-statistic	30.67	30.66	30.87	31.78
H. > 1000 hours	3.161**	3.163**	4.651**	3.573
	(1.455)	(1.309)	(2.022)	(2.971)
Mean of dependent variable	0.32	0.27	2.89	1.48
Kleibergen-Paap F-statistic	20.95	20.95	21.12	20.68

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell shows 2SLS estimates from our preferred model specification with full controls in equation (2) for each dependent variable. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. 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, and the number of children of individuals. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A15: Robustness of estimate to outcome definitions

A. ADL						
Definition	Baseline (Original scale)	≥ 1	≥ 2	≥ 3	≥ 4	≥ 5
Grandparenting	0.683**	0.296**	0.186*	0.111*	0.062	0.028
	(0.335)	(0.136)	(0.096)	(0.067)	(0.048)	(0.030)
Mean of dependent variable	0.30	0.15	0.08	0.04	0.02	0.01
Kleibergen-Paap F-statistic	46.89	46.89	46.89	46.89	46.89	46.89
AR F statistic	4.800	5.551	4.306	3.013	1.746	0.882
AR p-value	0.029	0.019	0.038	0.083	0.186	0.348
B. IADL						
Definition	Baseline (Original scale)	≥ 1	≥ 2	≥ 3	≥ 4	≥ 5
Grandparenting	0.683**	0.232*	0.117	0.136**	0.126***	0.072***
	(0.292)	(0.121)	(0.080)	(0.059)	(0.044)	(0.027)
Mean of dependent variable	0.25	0.13	0.06	0.03	0.02	0.01
Kleibergen-Paap F-statistic	46.82	46.82	46.82	46.82	46.82	46.82
AR F statistic	6.604	4.146	2.304	6.406	10.700	8.275
AR p-value	0.010	0.042	0.129	0.011	0.001	0.004
C. Self-reported health						
Definition	*Baseline (Original scale)*	"Poor" or "Fair"	*"Poor"*			
Grandparenting	1.178**	0.439**	0.241**			
	(0.489)	(0.184)	(0.103)			
Mean of dependent variable	2.88	0.29	0.09			
Kleibergen-Paap F-statistic	47.41	47.41	47.41			
AR F statistic	7.004	6.846	6.542			
AR p-value	0.008	0.009	0.011			
D. CESD Score						
Definition	Baseline (Original scale)	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6
Grandparenting	0.938	0.084	0.168	0.196	0.132	0.087
	(0.729)	(0.157)	(0.139)	(0.119)	(0.099)	(0.079)
Mean of dependent variable	1.49	0.33	0.22	0.15	0.11	0.07
Kleibergen-Paap F-statistic	50.77	50.77	50.77	50.77	50.77	50.77
AR F statistic	1.747	0.289	1.528	2.918	1.857	1.237
AR p-value	0.186	0.591	0.216	0.088	0.173	0.266

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. 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 original scale is the baseline scale measured in HRS questionnaire. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. The other cutoffs are used as alternative definitions to create the health indicators. For example, a dichotomous indicator equals one if an individual reports ADL (Panel A)/IADL (Panel B) limitations for at least 1 to 5 items or 2 to 6 items for CESD score (Panel D), and zero otherwise. In Panel C, an indicator for poor or fair self-reported health is one if self-reported health is "fair" or "poor", and zero otherwise. 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, and the number of children of individuals. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A16: First stage estimates of the gender of first born

Dependent variable: Grandparenting					
Model	1	2	3	4	5
Gender of first born	0.008*	0.007*	0.007*	0.006	0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Year FE + Birth year FE of first-born		Y	Y	Y	Y
Cohort FE + Birth year FE of youngest-born			Y	Y	Y
Demographics				Y	Y
Family size					Y
Mean of dependent variable	0.29	0.29	0.28	0.29	0.29
Number of clusters	25,313	25,238	25,198	25,051	25,051
Observations	120,189	120,087	119,865	119,278	119,278
1st stage Kleibergen-Paap F-statistic	3.79	2.98	2.83	2.34	2.30

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell reports estimates from a separate specification using equation (1). The gender of first born is equal to 1 if the first born of respondents is female, 0 for male. 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. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A17: First stage estimates of sex ratio on hours of grandchild care provision

Dependent variable: Grandparenting hours					
Model	1	2	3	4	5
Sex ratio	44.031***	43.351***	41.032***	37.939**	37.936**
	(14.947)	(14.783)	(14.801)	(14.800)	(14.801)
Year FE + Birth year FE of first-born		Y	Y	Y	Y
Cohort FE + Birth year FE of youngest-born			Y	Y	Y
Demographics				Y	Y
Family size					Y
Mean of dependent variable	258.0	258.2	258.3	258.3	258.3
Number of clusters	24,782	24,711	24,686	24,543	24,543
Observations	104,639	104,543	104,461	103,934	103,934
1st stage Kleibergen-Paap F-statistic	8.68	8.60	7.69	6.57	6.57

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. 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 hours is the estimated grandchild care hours reported by respondents over the last two years. 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. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A18: 2SLS estimates of sex ratio on hours of grandchild care provision

Dependent variable	ADL	IADL	Self-reported health	CESD Score
Grandparenting hours	0.00105*	0.00105*	0.00155*	0.00115
	(0.001)	(0.001)	(0.001)	(0.001)
Mean of dependent variable	0.32	0.27	2.89	1.48
Number of clusters	24,533	24,532	24,540	23,522
Observations	103,872	103,874	103,880	96,909
Kleibergen-Paap F-stat	6.54	6.54	6.63	7.03
AR F statistic	6.243	8.342	7.179	1.575
AR p-value	0.013	0.004	0.007	0.210

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Each cell shows 2SLS estimates from our preferred model specification with full controls in equation (2) for each dependent variable. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. 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, and the number of children of individuals. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A19: Mechanisms of grandparental childcare provision on health

Dependent variable	Care	Volunteer	Charity	Education	Club	Non-religious	Pray	Read	TV	Games	Play	Write	Computer Use	Maintenance	Bake	Sew/Knit	Hobby	Sport	Walk
Grandparenting	0.555	-0.679	1.598	0.333	1.323	0.425	-0.233	1.691	1.299*	-0.809	0.437	3.036**	2.900	0.904	1.437	-1.116	0.687	1.767	0.379
	(1.026)	(0.694)	(1.002)	(0.575)	(1.006)	(0.656)	(1.350)	(1.033)	(0.698)	(1.400)	(1.076)	(1.301)	(1.797)	(1.257)	(1.183)	(0.917)	(1.210)	(1.456)	(1.314)
Mean of DV	5.804	6.117	5.741	6.240	5.574	6.134	2.908	2.012	1.266	4.534	5.281	5.420	3.550	3.524	4.032	6.107	4.724	4.260	3.557
Number of clusters	11976	11997	11995	11966	11995	11992	12040	12090	10460	12041	11999	12007	12006	12060	12038	11929	11954	11980	12085
Observations	17,875	17,903	17,938	17,853	17,914	17,920	18,027	18,140	13,723	18,019	17,957	17,973	17,969	18,043	18,032	17,778	17,798	17,900	18,098
KP F-stat	15.76	15.65	15.72	15.92	14.93	14.51	15.66	15.90	12.49	15.25	16.06	15.63	15.42	14.69	15.06	14.41	15.09	14.70	14.69
AR F statistic	0.296	0.942	3.082	0.340	1.969	0.427	0.0294	3.338	4.950	0.335	0.167	8.246	3.241	0.543	1.669	1.628	0.329	1.682	0.0834
AR p-value	0.587	0.332	0.0792	0.560	0.161	0.513	0.864	0.0677	0.0261	0.563	0.683	0.00409	0.0719	0.461	0.196	0.202	0.566	0.195	0.773

Notes: The data used are from the HRS left behind sample in 2008 to 2014 of individuals who are 50 to 80. The HRS leave-behind questionnaires were administered to a randomly selected 50% subsample of HRS respondents who were interviewed in person or via telephone, by proxy, or residing in a nursing home. The questionnaires, presented in a dyadic manner to both spouses within a household, were self-administered. Response rates varied between 72% and 84%. The psychosocial leave-behind questionnaire comprehensively explores various dimensions, including well-being, lifestyle, social relationships, personality, work and self-related beliefs. Grandparental childcare provision is defined as an indicator that is 1 if the estimated grandchild care hours reported by respondents are at least 100 hours. The dependent variable in each column shows the frequency of doing activities on caregiving for a sick or disabled person (Care), do volunteer work with children or young people (Volunteer), participate in any other volunteer or charity work (Charity), attend an educational or training course (Education), go to a sport, social or other club (Club), attend meetings of non-religious organizations such as political, community or other interest groups (Non-religious), pray privately in places other than a church or synagogue (Pray), read books, magazines or newspapers (Read), watch television (TV), do word games such as crossword puzzles or Scrabble (Games), play cards or games such as chess (Play), do writing such as letters, stories, or journal entries (Write), use a computer for e-mail, internet or other tasks (Computer Use), do home or car maintenance or gardening (Maintenance), bake or cook something special (Bake), make clothes, knit, embroider, etc. (Sew/Knit), work on a hobby or project (Hobby), play sports or exercise (Sport), walk for 20 minutes or more (Walk). Each variable is assessed on a scale ranging from 1 to 7, indicating the frequency of the respective activity. 1 corresponds to "Daily," 2 to "Several times a week," 3 to "Once a week," 4 to "Several times a month," 5 to "At least once a month," 6 to "Not in the last month," and 7 to "Never/Not relevant". "KP F-stat" denotes the cluster-robust Kleibergen-Paap (KP) F-statistic on testing weak instruments. All models use the main specification with full controls. See text for details. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.

Table A20: The Heterogeneous Effects of Grandchild Care Provision

Dependent variable	ADL	IADL	Self-reported health	CESD Score
A. Female subsample				
Grandparenting	0.803**	0.668**	0.812*	0.516
	(0.358)	(0.299)	(0.478)	(0.782)
Observations	69,625	69,626	69,645	67,343
KP F-stat	46.12	46.06	46.42	46.50
B. Male subsample				
Grandparenting	0.462	0.856	2.913*	2.784
	(0.839)	(0.817)	(1.731)	(1.945)
Observations	49,709	49,708	49,701	44,213
KP F-stat	6.23	6.22	6.39	8.16
C. White subsample				
Grandparenting	0.579	0.631*	1.081*	1.076
	(0.408)	(0.358)	(0.637)	(0.940)
Observations	93,390	93,387	93,383	87,447
KP F-stat	26.94	26.94	27.18	29.20
D. Black/Hispanic subsample				
Grandparenting	0.787	0.691	1.072*	0.367
	(0.558)	(0.476)	(0.634)	(1.058)
Observations	25,832	25,835	25,851	24,002
KP F-stat	25.33	25.24	25.74	27.91

Notes: The data used are from the HRS 1996 to 2014 of individuals who are 50 to 80. Grandparental childcare provision is defined as an indicator that is 1 if the estimated grandchild care hours reported by respondents are at least 100 hours. The table shows the heterogeneous effects of childcare care provision. ADL and IADL are the number of limitations reported by an individual with the range from 0 to 5. Self-reported health is general health status reported by HRS individuals, with values of 1 for "excellent", 2 for "very good", 3 for "good", 4 for "fair", to 5 for "poor". CESD score is the number of depressive symptoms reported by an individual with the range from 0 to 8. The details of each health outcome can be referred to Appendix Table A1. 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, and the number of children of individuals. "KP F-stat" denotes the cluster-robust Kleibergen-Paap (KP) F-statistic on testing weak instruments. Standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.10.