

IZA DP No. 10022

Is Childcare Bad for the Mental Health of Grandparents? Evidence from SHARE

Giorgio Brunello
Lorenzo Rocco

June 2016

Is Childcare Bad for the Mental Health of Grandparents? Evidence from SHARE

Giorgio Brunello

*University of Padova,
Cesifo, IZA and ROA*

Lorenzo Rocco

University of Padova

Discussion Paper No. 10022
June 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Is Childcare Bad for the Mental Health of Grandparents? Evidence from SHARE*

We estimate the causal effects of regular and occasional grandchild care on the depression of grandmothers and grandfathers, using data from the Survey on Health, Ageing and Retirement in Europe and an instrumental variables strategy which exploits the variation in the timing of interviews across individuals and the fact that childcare declines with the age of grandchildren. We find that 10 additional hours of childcare per month, a 31 percent increase with respect to the sample average, increases the probability of developing depressive symptoms by 3.0 to 3.2 percentage points for grandmothers and by 5.4 to 5.9 percentage points for grandfathers. These results suggest that policies that substitute informal with formal childcare can improve the mental wellbeing of grandparents.

JEL Classification: J13, I12

Keywords: childcare, grandparents, depression, Europe

Corresponding author:

Lorenzo Rocco
Department of Economics and Management
University of Padova
via del Santo 33
35100 Padova
Italy
E-mail: lorenzo.rocco@unipd.it

* We are grateful to Marco Bertoni, Maria De Paola, Giovanni Mastrobuoni, Raffaele Miniaci, Enrico Rettore and the audiences at seminars in Brescia, Bolzano and Trento (IRVAPP) for comments and suggestions. This paper uses data from SHARELIFE release 1 and SHARE release 2.6.0. The SHARE data collection has been primarily funded by the European Commission through the 5th frame work programme (project QLK6-CT-2001- 00360 in the thematic programme Quality of Life), through the 6th framework programme (projects SHARE-I3, RII-CT- 2006 062193, COMPARE, CIT5-CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th frame work programme (SHARE-PREP, 211909 and SHARE-LEAP, 227822). The usual disclaimer applies.

Introduction

Increased life expectancy across the developed world means that it is now quite common for a child to grow up while grandparents and even great-grandparents are alive. By providing unpaid and flexible child care, grandparents can facilitate maternal labour force participation and benefit both their children and society at large.¹ But does child caring benefit grandparents as well?

In most countries, childcare is provided both formally and informally. Formal childcare typically includes registered child-minders, nurseries or playgroups, before-school care, after-school care and holiday clubs. Informal childcare is instead the one involving friends, relatives and grandparents, and other arrangements such as non-registered child-minders, baby-sitters and au-pairs. The relative importance of formal childcare varies across European countries. According to the EU Statistics on Income and Living Conditions survey (EU-SILC), the average total number of weekly hours spent in 26 European countries during 2013 looking after children aged 12 or less was 33.59, highest in Poland (50.84 hours) and lowest in Bulgaria (18.43). Most of these hours (87.6 percent) were spent for formal care.

Across EU Member States, grandparents are the most common providers of informal childcare. Using data from the European Social Survey, Jappens and Van Bavel, 2012, show that less than 15 percent of elderly parents in Northern and Western European countries co-reside with their children, in contrast to 48 percent or more in Southern and Eastern Europe. Co-residence clearly facilitates reciprocal support and care provision to and by older respondents. Therefore, mothers in Italy, Greece, Hungary and parts of Spain and Austria are the most likely to mainly rely on grandparents as childcare providers. In contrast, grandparents in Sweden, Denmark, Iceland and France are least likely to provide this type of care.

While much economic research has investigated the effects of the provision of formal and informal childcare on maternal labour supply, less is known on whether informal childcare has beneficial effects on the welfare of grandparents. On the positive side, care can be rewarding and source of higher satisfaction, and lead to more physical exercise and mental activity, which improve health (see Silverstein, Giarrusso and Bengtson, 2003; Coall & Hertwig, 2011; Powdthavee, 2011). On the negative side, caring means also responsibility and stress, which may harm health and lead to depression, and also reduces leisure and alternative uses of time (see Minkler, 1999). The balance of these positive and negative effects is not clear a priori.

¹ The effects of childcare provisions on female labour supply have been extensively studied, with controversial results. See for instance Posadas and Vidal – Fernandez, 2012, Del Boca, 2002, Rupert and Zanella, 2016, Zamarro, 2011. Blau and Currie, 2006, and Havnes and Mogstad, 2011, are recent reviews of this literature.

The early literature in this area has focused on grandparents who are the primary carers of their grandchildren. Research based on nationally representative U.S. data, such as the Health and Retirement Study, has found that grand-parenting is associated to poorer health, lack of privacy and leisure time, a heightened risk of isolation, and depression (Baker & Silverstein, 2008). Clearly, providing care on a supplementary rather than on a primary basis is likely to have different effects. In a study that examines both primary and supplementary grandparent childcare, Hughes et al, 2007, found that grandmothers raising children in skipped-generation households had higher rates of depression, whereas grandmothers providing supplementary care were in better health (see the discussion in Arpino and Bordone, 2014).

In this paper, we estimate the causal effects of regular and occasional grandchild care on the depression of grandmothers and grandfathers, using data from the Survey on Health, Ageing and Retirement in Europe (SHARE), a representative survey of European individuals aged 50+.² We are not the first to use these data to address a similar question. Di Gessa, Glaser and Tinker, 2015, for instance, use SHARE data to explore the association between the care of grandchildren and the health of grandparents, and find that caring grandparents experience some health benefits, independently of the intensity of care. This positive association, however, could simply reflect the fact that grandparents in better health are more likely to be engaged in childcare.

Since the allocation of grandparents to childcare is unlikely to be random, but depends instead both on individual characteristics and on perceived mental health, the identification of causal effects requires a source of exogenous variation in the provision of childcare. Recently, a few studies have tried to address the endogenous selection of grandparents into childcare using an instrumental variables strategy. Arpino and Bordone, 2014, for instance, look at the effect of childcare on grandparents' cognitive functioning and find that providing childcare has a positive effect only on one of the four cognitive tests considered, verbal fluency. Their selected instrument is the availability of grandchildren, under the assumption that children do not take the provision of grandchild care by their parents into account when making fertility choices. On the one hand, it is questionable whether this assumption is tenable. On the other hand, the availability of grandchildren could affect grandparents' welfare independently of whether they engage in childcare or not.

Ku et al, 2012, use Taiwanese data and show that caregivers have better self-rated health, fewer mobility limitations and fewer depressive symptoms compared to non-caregivers. Their analysis relies for identification on two instruments: the number of grandchildren and marital status of adult children. However, one may wonder whether these variables affect grandparents' health only by influencing their

² Both regular and occasional childcare are done without the presence of parents. In these data, we cannot distinguish between primary and supplementary care.

engagement in childcare activities. Finally, Reinkowski, 2013, estimate the effects of childcare on grandparents' health in the European countries participating to SHARE, using the gender of the firstborn child as instrument for the provision of childcare. She finds that childcare has no statistically significant effect on both physical and mental health. Since the gender of the oldest child may affect the quality of social interactions between grandparents and children independently of whether the former provide childcare, it seems difficult to defend the key exclusion restriction.

In this paper, we propose an alternative instrument, the distance (in months) of the interview from the onset of the survey (wave-specific) fieldwork, which exploits the variation in the timing of interviews across individuals and the fact – documented below - that childcare declines with the age of grandchildren. We argue that grandparents with the same observable characteristics and with grandchildren born in the same date are likely to be less involved in childcare if they are interviewed later, when their grandchildren are older. Even a few months can make a difference for the demand of informal childcare, because alternative services for very young children are not widespread in many European countries,³ and because younger children are often ill and unable to attend nurseries or kindergartens compared to older children.

Given the age of grandparents and grandchildren at the onset of the fieldwork, assignment to a later interview is equivalent to a positive exogenous shock to the age of grandchildren that can be used as instrument of childcare if two conditions are met, excludability and independence. Excludability holds if this shock affects depression only via childcare, and if there are no relevant macroeconomic shocks occurring during the length of the fieldwork. Independence holds if the distance from the start of interviews is uncorrelated with grandparent characteristics that influence their depression. If these assumptions do not hold, we should find that variations in measured distance affect the depression symptoms of all seniors, including those who are not involved in any childcare. Yet, when we run a falsification test both on the sub-sample of seniors without grandchildren, i.e. those who cannot provide childcare by definition, and on the sub-sample of grandparents with grandchildren older than 16, who are unlikely to require significant care, we find that depression is unaffected.

Using the exogenous variation provided by our selected instrument, we estimate that the effects of childcare on grandparents' depression are positive and sizeable, especially for grandfathers. Our results indicate that 10 additional hours of childcare per month, a 31 percent increase with respect to the sample average (32.3 hours), increase the probability of developing depressive symptoms by 3.0 to 3.2 percentage points for grandmothers (equivalent to 10 to 11 percent with respect to the sample average)

³In Europe, there is a sharp difference in the provision of formal childcare between children younger than three and older children. See Janta, 2014.

and by 5.4 to 5.9 percentage points for grandfathers (equivalent to 37 to 41 percent with respect to the sample average).

Our findings suggest that the negative effects of childcare on mental health prevail on the potential benefits, and that males suffer significantly more than females. To further investigate these effects, we decompose depression into two main factors, called in the literature affective suffering – which include depression, tearfulness, wishing death, lack of sleep, guilt, irritability and fatigue - and lack of motivation, which includes pessimism, lack of enjoyment, interest and concentration, and find that childcare affects both factors for grandfathers but only lack of motivation for grandmothers.

In the presence of heterogeneous responses to treatment, our estimates do not capture the average treatment effect, but the effects on the sub-group of the population – the compliers – who change their behaviour because of the exogenous variation in the age of their grandchildren. We characterize compliers by showing that they are more likely to be located in Southern and Eastern Europe (Italy, Poland and Spain), inactive, younger than average, have one child and at most two grandchildren, high household income, and have lower education and numeracy skills. We test whether our results can be generalised from compliers to the entire population but find that the necessary conditions for this to happen are not met in our data.

High income and the low number of children and grandchildren suggest that compliers are characterised by a relatively high alternative value of time, and that giving up this time in order to care for grandchildren may be particularly stressful. There is also evidence that a division of roles in the household that assigns childcare mainly to mothers and grandmothers is more widespread in Southern and Eastern Europe, suggesting that grandmothers in these countries are more used than grandfathers to care for other family members, and therefore less likely to be emotionally affected by grandparenting.

The development of formal childcare is at the forefront of European policy. At the Barcelona Summit in 2002, for instance, the European Council has set targets for providing formal childcare in EU Member States, specifying that at least 90 per cent of children aged between 3 years and the mandatory school age, and at least 33 per cent of children aged under 3 should have access to formal childcare provision. Increasing formal childcare is costly but also likely to reduce the need for informal childcare, which is often provided by grandparents. Our paper suggests that a potential benefit of policies raising formal childcare is to improve the mental wellbeing of senior European citizens by reducing their

engagement in informal grand-parenting, and by so doing to contribute to the reduction of the health costs of depression.⁴

The paper is organized as follows. We introduce the data in Section 1, describe our empirical strategy in Section 2, and our results and a few extensions in Section 3. Conclusions follow.

1. The Data

Our data are drawn from the Survey of Health, Ageing and Retirement in Europe (SHARE), a multidisciplinary and cross-national European data set containing current and retrospective information on labor market activity, retirement, health and socioeconomic status for more than 25,000 individuals aged 50 or older. Interviewed individuals who have grandchildren are asked whether, during the time since the last interview, they have regularly or occasionally looked after their grandchildren without the presence of parents. Those who answer positively are then asked about both the frequency of care (less than monthly, almost every month, and almost every week, almost daily) and the number of hours of care spent to look after all their grandchildren in a typical month during the past year.

Since the effects of childcare on mental health are likely to depend on its intensity, in this paper we focus on the latter measure, which is available only in Waves 1 and 2 of the survey. In our selected sample, 39.2 percent of grandparents do no childcare, and 28.2 percent do at least 30 hours of childcare a month. For those doing any childcare, the median and mean number of hours of childcare per month are 24 and 53.2 respectively. Table 1 shows the average number of hours of informal childcare provided by grandparents in the sample that we use, which includes 12 countries (Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Switzerland, Belgium, Czechia and Poland).⁵ The intensity of informal childcare is higher for grandmothers than for grandfathers, relatively high in Poland and Italy and low in Sweden and Denmark.

We measure depression using the Euro-D scale (Prince et al, 1999), which was originally developed in an effort to harmonize data on late-life depression in Europe as part of the EURODEP collaboration. The scale focuses on current depression and adds up the following twelve symptoms – taken from the Geriatric Mental State (GMS) – as they are reported by the interviewee: depressed mood, pessimism, wishing death, guilt, lack of sleep, lack of interest, irritability, lack of appetite, fatigue, lack of

⁴According to Sobocki et al, 2006, the total annual cost of depression in Europe is estimated at Euro 118 billion in 2004, which corresponds to a cost of Euro 253 per inhabitant. Direct costs alone comprise outpatient care (Euro 22 billion), drug costs (Euro 9 billion) and hospitalization (Euro 10 billion). Indirect costs due to morbidity and mortality are estimated at Euro 76 billion. The cost of depression corresponds to 1% of total European GDP.

⁵ We exclude Israel as a non – European country, Ireland because of the very few observations and Greece because the short fieldwork period in the first wave (due to the beginning of the 2004 Olympic Games) and the use of the telephone directory as the sampling frame cast doubts on the representativeness of the Greek sample. See Mazzonna and Peracchi, 2014. Table 1 includes also grandparents doing no childcare.

concentration, lack of enjoyment and tearfulness.⁶ Its validation has shown adequate internal consistency, and the optimal cut-off point of 3/4 (in the range 0-12) has been used to predict both GMS depression and short-care pervasive depression (see Castro-Costa et al, 2008 and Larraga et al, 2006).

We define the depression dummy D , that takes value one if the Euro-D scale is equal to four or higher, and zero otherwise. Following Castro-Costa et al, 2008, we also use principal component analysis to extract from the twelve symptoms two orthogonal factors, affective suffering (A) and lack of motivation (M), with depressed mood, tearfulness, wishing death, lack of sleep, guilt, irritability and fatigue loading mainly on the first factor, and pessimism, lack of interest, enjoyment and concentration loading mainly on the second.⁷ Table 2 reports the percent of individuals affected by depression ($D=1$), by gender and country. Females are much more likely to report depression than males, and the percent depressed is highest in Italy, Poland and Spain, the countries in our sample with the highest average number of monthly hours of informal childcare done by grandparents (40 in Spain, 55 in Italy and 58 in Poland, relative to 32.3 in the entire sample).

We match the information in the former two waves with the retrospective life histories contained in Wave 3 in order to control for early life conditions, including physical and mental health during childhood, which could persist into adult age. We retain only individuals aged 50 to 75 who have one to four children, because SHARE provides socio-economic information only for the first four children,⁸ including the year of birth of the youngest grandchild. To obtain our final sample, we further retain only seniors whose youngest grandchild is 16 or younger⁹ and exclude the few cases when the reported number of monthly hours of childcare is higher than 450. Our final sample consists of 13,091 observations, 7,397 for females and 5,694 for males. Table 3 reports the summary statistics for the variables used in the empirical analysis.

2. The Empirical Setup

We investigate the effects of childcare on the depression of grandparents by estimating the following empirical model on our final sample

⁶ Most of these symptoms refer to the last month or to the current situation.

⁷ These two factors are associated to the two eigenvalues higher than one and explain about 33 percent of total variance.

⁸ In SHARE, the information on children is reported by only one household member, the so called respondent. We impute this information to the non-respondent partner in all households with two eligible members (59 percent of the sample). About 85 percent of individuals aged 50 to 75 have one to four children. At least one child is required to qualify as potential grandparent.

⁹ Close to 94 percent of the youngest grandchildren in our sample are aged 0 to 16. We expect that childcare is not an issue for grandchildren older than 16.

$$D_{it} = \alpha HC_{it} + \beta X_{it} + \gamma W_{it} + \lambda_i + \varepsilon_{it} \quad (1)$$

where D is the depression dummy, HC is the number of monthly hours of childcare, X is a vector of controls from Waves 1 and 2 of the survey, which includes: gender, a quadratic in the age of the grandparent at the start of each wave of the survey, wave by country dummies – that capture country specific macroeconomic effects and institutional changes that may affect both childcare and depression - and a dummy indicating whether the individual is present in both waves, educational attainment, height, the presence of siblings, birth order, the number of children, marital status, dummies indicating whether parents are alive and their health if alive, the age of the oldest child and the age of the youngest grandchild at the beginning of the fieldwork; W is a vector of the following indicators of early life conditions from Wave 3 of the survey: relative (self-reported) performance in math at age 10, the number of books in the house at age 10, the main occupation of the breadwinner at age 10, poor health, whether parents were in the house, smoke, drank or had mental problems and the presence of hardship episodes during childhood, stress and mental problems before age 15;¹⁰ λ is an individual error component accounting for individual un-observables and ε is a IID error term.

Table 4 presents the ordinary least squares estimates of the association between monthly hours of childcare and the depression dummy D , showing that this association is very close to zero in the full sample, equal to zero in the sub-sample of females and small, positive and statistically significant in the sub-sample of males.¹¹ The OLS estimate of (1) is unlikely to identify the causal effect of childcare on depression for three reasons: a) depressed grandparents are less likely to be involved in childcare activities (reversed causation); b) hours of childcare might be measured with error; c) unobserved individual effects might influence both depression and childcare. An example is the individual opportunity value of time: seniors with low values are more likely to take up child-caring and to develop depression symptoms.

To identify the causal effect of childcare on depression, we use an instrumental variables strategy. We exploit the fact that – for each wave of the survey - the timing of interviews varies across households. In the first wave, interviews started in February 2004 and ended in February 2005, one year later; in the second wave, they started in September 2006 and ended in September 2007. Consider two grandparents who are equal in all observable characteristics and have a single grandchild born in the same date. The grandparent interviewed later has a marginally older grandchild at the time of the interview, which is likely to require less childcare, as suggested by Figure 1, where we plot a polynomial smooth of monthly

¹⁰ For each variable, we set missing values to zero and add to each regression a dummy equal to 1 if the value of the relevant variable is missing.

¹¹ We cluster standard errors at the household level in the full sample (males and females), and at the individual level in the samples of males and females.

hours of childcare as function of the age of the youngest grandchild, measured at the beginning of the (wave-specific) fieldwork.¹²

Define Z as the distance in months from the wave-specific start of the interviews. We construct Z as follows

$$\begin{aligned} Z &= (YI - 2004) * 12 + (MI - 2) && \text{for Wave 1} \\ Z &= (YI - 2006) * 12 + (MI - 9) && \text{for Wave 2} \end{aligned} \tag{2}$$

where YI and MI are the year and the month when interviews started. Figure 2 reports the distribution of Z in the selected sample. The modal distance is four months, the median is five months and the interquartile range is three months. Since we always condition in our estimates for country by wave dummies, Table 5 shows the variability of Z within country and wave: the difference between maximum and minimum distance ranges from four to eleven months, and the interquartile range varies between one and four months.

The distance from the first interview Z is a valid instrument if three conditions hold: relevance, independence and exclusion. Concerning relevance, Table 6 reports the estimated coefficient of the first stage regression of the endogenous variable (HC) on the instrument (Z). We estimate that one additional month of delay from the wave – specific start of the fieldwork reduces the monthly hours of childcare in the full sample by 1.51 hours, which corresponds to a 4.6 percent decline with respect to the sample average. The estimated effect is very similar for males and females. We also find that the F-test for the inclusion of Z in the first stage regression rejects the null at the 1 percent level of confidence and assumes values well above 10, the rule – of – thumb threshold for instrument weakness.

Stated in terms of the potential outcomes model with heterogeneous treatment effects, independence requires that the instrument is independent of both potential outcomes and potential treatment status. Roughly, this condition requires that the instrument is as good as randomly distributed across individuals belonging to the treatment and control groups, conditional on some observables. Although this assumption is untestable, it implies that the date of interview should be uncorrelated with the observable characteristics of the two groups, which in our case include more than forty variables in vectors X and W .

¹² In the Figure we allow the age of the youngest grandchild to vary between 1 and 16. Before age 1, maternity and paternity leaves – available to different extent in the countries in our sample – imply that parents do the bulk of child caring. In the raw data, hours are close to 33 per month when the youngest grandchild is less than 1, about 38 per month when she is aged 1 to 3, about 36 per month when she is aged 4 to 6, close to 27 per month for age 7 to 9 and about 17 per month when age is from 10 to 16.

We perform a balancing test and compare individuals interviewed near the beginning and the end of the fieldwork by regressing each interviewee’s characteristic on the dummy DZ – equal to 1 if Z is at or above its median value and to 0 otherwise. In these regressions, we always control for the characteristics that, being known to the interviewers, could have influenced how they organized their fieldwork (gender, respondent’s age at the start of interview and participation to both waves). We also add country by wave dummies to control for the slightly different procedures adopted by survey agencies. Table 7 reports the results for the full sample, showing that balancing fails (at the 5 percent level of confidence) only for one educational dummy and one of the five dummies reporting the number of books in the house at age 10.¹³

Our identification strategy requires that the date of interview is neither affected by interviewee’s characteristics nor results from a bargain between the interviewers and interviewee. Overall, the balancing tests support this assumption. We also perform a balancing test on lagged depression, focusing on the sub-sample of individuals interviewed both in the first and in the second wave, and find that past depression does not affect the date of interview in the second wave. This result dispels the concern that depressed individuals could manage to systematically anticipate or delay this date.¹⁴

The claim that the date of the interview is (conditionally) independent on interviewee’s characteristics is also supported by the analysis of the date of the first contact. By definition, this date is not the outcome of a negotiation between interviewer and interviewee and depends only on how interviewers organize their work. In all countries, excepting Sweden, the interim reports transmitted by the national survey agencies to the survey headquarters show that households were progressively contacted during the period of fieldwork and, with good approximation, both the date of the first contact and the date of the interview were uniformly distributed over the period of fieldwork. These facts suggest that in the large majority of cases the date of the interview followed closely the date of the first contact.¹⁵

The third condition for instrument validity, exclusion, requires that Z affects depression only by influencing the intensity of childcare. Clearly, this condition is not necessarily granted, because, conditional on a given age at the onset of wave – specific fieldwork, grandparents who are interviewed later are both marginally older (aging effect), and might have been affected by shocks that occurred

¹³ In the case of females, only marital status – whether widowed or not - turns out to be imperfectly balanced. In the case of males, this happens for two out of the four educational attainment dummies, for whether parents are alive and for relative performance in math at age 10. More detailed results by gender are available from the authors upon request. We have also performed a balancing test for variables that are not necessarily pre-determined, such as the age of youngest grandchild at the start of the interviews, grandparent’s activity and retirement status. In these tests, we condition for all available pre-determined variables in X and W. Reassuringly, we do not observe any unbalancing.

¹⁴ Since the dependent variable is only available for the individuals in wave 2 who were also present in wave 1, we correct for endogenous selection into the sample using the Heckman two step model. Our estimates for the full sample and by gender - reported in Table A1 in the Appendix – clearly reject the hypothesis that a correlation exists at the standard levels of statistical significance. The selection equation includes two variables excluded from the main equation: the probability of having received a gift or inheritance and the probability of having changed house between the two waves.

¹⁵ The Appendix provides further details on the organization of fieldwork in SHARE.

during the fieldwork (time effect). Not only grandparents but also their grandchildren are marginally older. These effects may have a direct impact on depression, thereby invalidating the selected instrument. However, if the distance from the first (wave-specific) interview Z affects depression independently of childcare, because of the joint impact of aging and time, we should find that it significantly affects depression also among seniors without grandchildren, who are precluded from childcare. In a similar vein, since grandchildren older than 16 are unlikely to require much childcare, we should find that in the sample of grandparents with grown up grandchildren distance Z affects depression.¹⁶

We run these falsification tests for the full sample and separately for males and females. As shown in Table 8, we find that the effect of Z on depression is always small and never statistically significant at conventional levels, both in the sample without grandchildren and in the sample with older grandchildren. In the last row of the table, we also report the results of similar regressions for the sample of grandparents with grandchildren younger than 17. In sharp contrast with the rest of the table, there is evidence of a statistically significant relationship between Z and depression (at the one percent level in the full sample and for males and at the ten percent level for females), supporting our claim that the effect of the selected instrument on depression operates exclusively by affecting the intensity of childcare.¹⁷

To further inquire into the existence of a direct effect of grandchild aging on depression, we consider the possibility that a few more months near the start of pre-adolescence (at age 9) could turn grandchildren rather abruptly into rebellious adolescents, with negative consequences for the mental health of grandparents. However, when we replicate our estimates in the sub-sample of grandparents whose youngest grandchildren are (strictly) younger than 9, our results are qualitatively unaffected.

Since equation (1) is defined in the sub-sample of grandparents with grandchildren, the exclusion restriction could fail if the distance Z affects the probability of having grandchildren. We verify whether this is the case by regressing the dummy GC (1 for the presence of grandchildren and 0 otherwise) on Z and the vector of controls X and W , using the sample of grandparents with and without grandchildren. Reassuringly, the results reported in Table A2 in the Appendix indicate that the coefficient associated to Z is very small and not statistically significant at conventional levels.

¹⁶ While in the former case we test for the presence of an effect due to the aging of grandparents, in the latter we test for the presence of the joint effect of the aging of grandparents and grandchildren.

¹⁷ These tests indicate that the modest age difference between grandparents interviewed at the beginning and towards the end of each wave-specific fieldwork does not affect depression. To evaluate whether the effect of a marginal increase in age varies by cohort, we have also run specifications including the interaction of Z with the age of grandparents at the beginning of the fieldwork. Our results confirm the absence of aging effects, conditional on grandparents' age at the beginning of the fieldwork.

3. Results

We present in Table 9 the 2SLS estimates of equation (1) both for the full sample and separately for males and females.¹⁸ We find that childcare increases depression symptoms in all samples, and that the estimated effect is statistically significant in the full sample and in the sub-sample of grandfathers. Since the dependent variable in (1) is a dummy, we also show in the table the marginal effects estimated using an IV probit model, which are very similar to the 2SLS effects and somewhat more precise.

Our estimates suggest that 10 more hours of childcare per month, which correspond to a 31 percent increase with respect to the sample average - increase the probability of developing depression symptoms by 3.0 to 3.2 percentage points for grandmothers (a 10 to 11 percent increase) and by 5.4 to 5.9 percentage points for grandfathers (a 37 to 41 percent increase), depending on whether we use a linear or a nonlinear specification. These findings indicate that in our data the negative effects of childcare prevail on the positive effects, that the overall effects are sizeable and that grandfathers are more affected than grandmothers.

Following psychiatric research (see Prince et al, 1999, and Castro-Costa et al, 2008), we decompose the Euro-D index into two (orthogonal) components using factor analysis and call these orthogonal components – each with zero mean and unit standard deviation - affective suffering and lack of motivation. We estimate model (1) using these two factors as dependent variables and present our results in Table 10. For grandfathers, we find that the positive effects of childcare on affective suffering and lack of motivation are very similar. On the other hand, childcare only affects positively and significantly (in a statistical sense) the motivation of grandmothers. These estimates suggest that adding 10 hours of monthly childcare reduces the motivation of both grandmothers and grandfathers by about 0.14 standard deviations and increases the affective suffering of grandfathers by 0.16 standard deviations, without any significant effect on grandmothers.

When the response of depression to the intensity of childcare is heterogeneous in the population, 2SLS estimates do not identify average treatment effects, but the effects on the sub-group of the population (compliers) who is prompted to change behaviour because of the exogenous variation induced by the selected instrument (LATE).¹⁹ The natural question to ask is whether our estimates have any external validity outside the sample of compliers. We address this question in two steps. First, following Angrist and Pischke, 2008, we describe the sub-group of compliers by using the result that the prevalence of a given characteristic among compliers - relative to the overall sample - can be obtained by computing

¹⁸ As for the OLS regressions, we include among the regressors the variables in vector W to account for the few cases of unbalancing and to increase precision.

¹⁹ To interpret 2SLS estimates as LATE, the additional assumption of monotonicity must hold, i.e. all grandparents change the intensity of childcare in the same direction as the age of their grandchildren is subject to an exogenous positive shock.

the ratio of the first stage for individuals with that characteristic and the overall first stage.²⁰ As reported in Table 11 for the full sample of males and females, compliers are much more likely than average to be located in Southern and Eastern Europe, be inactive, have a single child and high household income, and less likely to be older than average, with higher education or high numeracy skills and to have more than two grandchildren.²¹

Second, we implement a test recently suggested by Black et al, 2015. For the LATE effect to be generalizable to the whole population it should be that compliers (C), always takers (AT) and never takers (NT) are equivalent with respect to potential outcomes (Y_1, Y_0), possibly after conditioning on observables. The two necessary (but not sufficient) conditions for this to happen

$$\begin{aligned} E\{Y_1|AT\} &= E\{Y_1|C\} \\ E\{Y_0|NT\} &= E\{Y_0|C\} \end{aligned} \tag{3}$$

can be tested by regressing depression D on Z , X and W separately for those with $HC>0$ and $HC=0$. Conditions (3) hold if the coefficient associated to Z is not statistically different from zero in either regression. Unfortunately, this is not the case in our data when $HC=0$ and only partially the case when $HC>0$ (see Table 12). We conclude from this that our LATE estimates, while internally consistent, cannot be easily generalized outside the current sample.

If parents who leave their children to the care of grandparents are under economic or personal stress, the depression symptoms of grandparents could be due not to childcare but to the situation of parents. The survey SHARE includes information on the first four children of each grandparent. To verify whether it is the poor situation of children rather than child-caring that affects depression, we augment our regressions with indicators of whether at least one of the first four children is disabled, unemployed, not married or divorced. We find that these indicators influence positively depression, as expected. Their inclusion, however, changes only marginally the estimated effect of childcare on depression, which declines in the sample of males and females from 0.0043 in the baseline to 0.0041.

Because of attrition, the inclusion of early life condition variables from the third wave of the survey in our regressions reduces the final sample by about 30 percent, a considerable amount. We verify whether our results hold in the larger sample that excludes these variables and find that, reassuringly,

²⁰ To ease interpretation, we have transformed both HC and Z in dummy variables, DHC , that takes value 1 if $HC>0$ and 0 otherwise, and DZ , that takes value 1 if Z is at or above its median value and 0 otherwise.

²¹ The analysis by gender is likely to yield imprecise results because of the reduction in sample size induced by the focus on single characteristics. Even in the full sample, due to the small sample size, the estimated ratio is significantly different from one only for residence in Southern and Eastern Europe.

they do (see Table A3). There is also evidence that the falsification tests discussed in Section 2 produce qualitatively similar results in the extended sample.

We also consider alternative outcomes, including indicators of poor health, cognitive skills and memory (numeracy and recall of words), life satisfaction, control, autonomy, pleasure and self-realization. For grandmothers, we find evidence that childcare reduces both life satisfaction and the degree of self-realization – see Table A4. For grandfathers, hours of childcare negatively affect self-reported health. These results are broadly consistent with our findings that childcare affects mainly motivation for grandmothers, and both motivation and affective suffering for grandfathers.

Conclusions

Informal childcare is the result of a process involving grandparents, their offspring and the availability of alternative of formal childcare. On the one hand, demand depends on the number of grandchildren and their age distribution, that reflect the fertility decisions of both grandparents and parents. For parents, these decisions are likely to take into account whether grandparents can or are willing to look after their new children. On the other hand, grandparents supply childcare by considering their health conditions, their alternative value of time and the proximity of parents. The grandparents who eventually look after their grandchildren are likely to be different from the population of grandparents, for instance because they are in better physical and mental health. Because of endogenous sorting, the simple association between grandparents' health and childcare cannot inform us on the causal effect of childcare.

In this paper, we propose to identify the causal effect of childcare with an instrumental variables strategy, which relies on the variation in the time of the interview in the Survey on Health, Ageing and Retirement in Europe to provide exogenous variation in the age of grandchildren. We supply evidence that the time of the interview is independent of interviewee's characteristics and rather depends on how survey agencies and interviewers organize their fieldwork. Our falsification tests also suggest that the modest aging that a later interview causes to grandparents and grandchildren does not directly affect depression.

Our results indicate that childcare increases the likelihood of depression for grandmothers and especially for grandfathers. The estimated effect is rather large, as 10 more hours of childcare per month cause the probability of depression to increase by 3.0 to 3.2 percentage point among grandmothers and by 5.4 to 5.9 percentage points among grandfathers. A test recently proposed by Black et al., 2015, confirms that our LATE estimates cannot be generalized to the entire sample.

Rather, they represent the average effect for the subset of compliers, i.e. the grandparents who reduce the intensity of their care when their grandchildren are a few months older.

We have characterized compliers as a group of seniors who are richer and younger than the entire sample. Their choice of having a single child may indicate that they have a relatively high opportunity cost of time. They typically reside in Spain, Italy and Poland, where educational attainment and retirement age are historically lower – and co-residence higher - than in Central and Northern Europe.

According to the 2002 Wave of the International Social Survey Program, Spain and Poland have a higher than average percentage of individuals agreeing or strongly agreeing with the view that men should do more childcare and household chores. Data from the European Working Condition Survey also suggest that Italy, Spain and Poland have lower than average participation of men in caring activities (European Union, 2012). Finally, the 2004 wave of the European Social Survey shows that in Poland, Italy and Spain the burden of housework is largely on the shoulders of women, who devote four times more hours per week to housework than their male partners. The higher prevalence of a traditional division of labour by gender suggests that males in these countries are less used than females to deal with children and grandchildren, and therefore more exposed to the negative consequences of child-caring on depression symptoms.

Additional formal childcare is costly but also likely to reduce the need for informal childcare, which is often provided by grandparents. Our paper suggests that a potential benefit of policies raising formal childcare is to improve the mental wellbeing of senior European citizens by reducing their engagement in informal grand-parenting, and by so doing to contribute to the reduction of the health costs of depression.

References

- Arpino, B and Bordone, I, 2014, "Does Grandparenting Pay Off? The Effect of Child Care on Grandparents' Cognitive Functioning", *Journal of Marriage and Family*, 76 (April 2014): 337–351
- Baker, L. A., & Silverstein, M. (2008). "Preventive Health Behaviors among Grandmothers Raising Grandchildren". *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 63, S304–S311
- Black, Dan A., Joo, J., LaLonde, R., Smith, J. and Taylor, E., 2015. "Simple Tests for Selection Bias: Learning More from Instrumental Variables," IZA Discussion Papers 9346, Institute for the Study of Labor (IZA).
- Blau, D.M., Currie, J., 2006. Pre-school, Day Care, and After-school Care: Who's Minding the Kids? In: Hanushek, E.A., Welch, F. (Eds.), *Handbook of the Economics of Education*, vol.2. Elsevier, pp. 1163–1278.
- Börsch-Supan, A. and Krieger, U. (2013) "Investigating Response Behavior" in Malter, F. and Börsch-Supan, A. (eds) *SHARE Wave 4 Innovations & Methodology*, Munich: Munich Center for the Economics of Aging (MEA)
- Castro-Costa E, Dewey M, Stewart R, Banerjee S, Huppert F, Mendonca-Lima C, Bula C, Reisches F, Wancata J, Ritchie K, Tsolaki M, Mateos R, Prince M., 2008, "Ascertaining Late-life Depressive Symptoms in Europe: an Evaluation of the Survey Version of the Euro-D scale in 10 Nations. The SHARE Project", *International Journal of Methods in Psychiatric Research*, 17, 1, 12-29.
- Coall, D. A., & Hertwig, R. (2011). "Grandparental Investment: A Relic of the Past or a Resource for the Future?" *Psychological Science*, 20, 93–98
- De Luca, G. and Lipps, O. (2005) "Fieldwork and Survey Management in SHARE" in Börsch-Supan, A. and Jürges, H. (eds.) *The Survey of Health, Aging, and Retirement in Europe – Methodology*, Mannheim: Mannheim Research Institute for the Economics of Aging
- Del Boca, Daniela. 2002. "The Effect of Child Care and Part Time Opportunities on Participation and Fertility Decisions in Italy." *Journal of Population Economics*. Vol.15., pp.549-573
- Di Gessa, G., Glase, K. And Tinker, A., 2015, "The Health Impact of Intensive and Non-intensive Grandchild Care in Europe: New Evidence from SHARE", *Journal of Gerontology: Social Sciences*, 1-13.
- European Commission, 2012, "*The Role of Men in Gender Equality: European Strategies and Insights*", Brussels.
- Havnes T. and Mogstad M., 2011. "Money for Nothing? Universal Child Care and Maternal Employment", *Journal of Public Economics*, 95, 1455-65.
- Hughes, M. E., Waite, L. J., LaPierre, T.A., Luo, Y., 2007. "All in the Family: The Impact of Caring for grandchildren on Grandparents' Health". *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 62, S108–S119.
- Ku L, 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
- Larraga L, Saz P, Dewey M, Marcos G and Antonio Lobo, 2006, "Validation of the Spanish Version of the EURO-D Scale: an Instrument for Detecting Depression in Older People", *International Journal of Geriatric Psychiatry* 2006; 21: 1199–1205.
- Janta, B, 2014, "*Caring for Children in Europe*", RAND Europe.
- Jappens, M., & J. van Bavel. 2012. "Regional Family Norms and Child Care by Grandparents in Europe." *Demographic Research* 27 (4): 85–114.
- Mazzonna F and Peracchi F, 2014. "Unhealthy Retirement?," EIEF Working Papers Series 1409, Einaudi Institute for Economics and Finance (EIEF).
- Minkler, M. (1999). "Intergenerational Households Headed by Grandparents: Contexts, Realities, and Implications for Policy". *Journal of Aging Studies*, 13, 199–218
- Posadas, J and Vidal-Fernandez, M., 2012, "Grandparents' Childcare and Female Labor Force Participation", IZA Discussion Paper n.6398.
- Powdthavee, N. (2011). "Life Satisfaction and Grandparenthood: Evidence from a Nationwide Survey". IZA Discussion Paper No. 5869, Institute for the Study of Labor (IZA)

- Rupert, P and Zanella, G, 2016, "Grandchildren and their Grandparents' Labour Supply", mimeo
- Reinkowski, J, 2013, "Should we Care that they Care? Grandchild Care and its Impacts on Grandparent Health", Ifo working paper 165, Munich.
- Silverstein, M., Giarrusso, R., & Bengtson, V. L., 2003, "Grandparents and Grandchildren in Family Systems: A Social-developmental Perspective" In V. L. Bengtson & A. Lowenstein (Eds.), *Global Aging and Challenges to Families* (pp. 75–103). New York: Aldine de Gruyter.
- Sobocki P, Jönsson B, Angst J, Rehnberg C., 2006, "Costs of Depression in Europe", *Journal of Mental Health Policy Economics*, 9(2):87-98.
- Zamarro, Gema. 2011. "Family Labor Participation and Child Care Decisions. The Role of Grannies" *RAND Working Paper WR-833*

Appendix - The fieldwork in SHARE

Fieldwork management and monitoring are crucial to guarantee that the total survey error is minimized and data quality is high. Accordingly, the SHARE coordination team has continuously improved the organization of fieldwork. An attentive monitoring activity was set up from the beginning of the study to be able to intervene and fix problems in real time. Bi-weekly reports were prepared during the fieldwork with evidence on the ongoing survey, including the contact rate (i.e. the proportion of sample units contacted since the beginning of the survey) and the collaboration rate (i.e. the proportion of sample units with at least one interview out of all contacted sample units).

In this appendix we discuss a number of aspects of the SHARE fieldwork that are relevant to explain variation in the date of interview across households.

a) The fieldwork period

The SHARE coordinator team sought to harmonize as much as possible the fieldwork period across all participating countries. Fieldwork was expected to be start in May 2004 for wave 1, and in October 2006 for wave 2. In practice, because of several reasons generally due to funding delays, interviewers did not start working simultaneously in all countries. Depending on when each country started its operations, the intensity of data collection changed. Typically, countries starting later tried to catch up by hiring more interviewers. In wave 1, France started later and completed operations in two months, Belgium - because of delay in financing - started operations when all other countries were about finishing their fieldwork and Sweden prolonged the fieldwork because of its larger number of interviews than in the other countries. In wave 2, Netherlands and Flanders started three and two months later while the Czech Republic was the first to conclude (May 2007) and Sweden and Spain the last (November 2007).

Actually, according to SHARE microdata, some countries started their operations earlier than expected. Specifically, the first interviews of wave 1 were carried out in Germany already in February 2004 (three months earlier), followed by a few other interviews in Denmark and Sweden in March and 90 interviews distributed across several countries in April. Because of this, we select February 2004 as the starting point of wave 1 operations. Similarly for wave 2, the first interviews were in September 2006 (in Germany, one month in advance). Therefore, we set September 2006 as the starting point of the wave 2 campaign.

b) The contacts

Interviewers were instructed to attempt contacting sample units at least 5 times in wave 1 and 8 times in later waves. Contacts could be by telephone or in person, although a certain number of contacts in person were compulsory, and had to occur at different times of the day and in different days of the week. Contact attempts result in (successful) contacts if the subject opens the door or answer the telephone. After a successful contact, there may be either an appointment for an interview or a refusal. In case of refusal, the subject is not immediately dropped from the sample but further attempts can be made, at least if the refusal is a soft one.

In general not all sample units in SHARE are immediately contacted. The reason is that a relatively small number of interviewers are active at the beginning of the fieldwork. Interviewers, in planning their work, progressively

contact new households as long as they complete previous interviews. With the exception of Sweden, that made an initial effort to contact most sampled households, households in other countries were progressively contacted during the fieldwork period so that the date of first contact is approximately uniformly distributed, as documented by figures 7.1 and 7.2 in De Luca and Lipps, 2005.

c) *The sample management system.*

An important device to organise fieldwork is the sample management system SMS, a software installed on the servers of the national survey agencies and the laptops of the interviewers to assign sample units, record contacts and their outcomes, collect and transmit data. A specific SMS was developed for SHARE and was constantly updated across waves. Starting from Wave 2, all national agencies and interviewers adopted the same SMS.

d) *Interviewers*

The SMS assigned all sampled households to interviewers, leaving to the latter the task of deciding how to organize their work over the established fieldwork period (each interviewer was assigned 43 households on average). According to Börsch-Supan and Krieger, 2013, "...driving distances between households is [interviewers'] foremost consideration when planning their work. While interviewers get some compensation for travel cost, their main income results from finalized interviews. Thus, they try to optimize driving distances between addresses... (p.53)". Most countries implemented incentive schemes for interviewers in order to increase their motivation. For households with more than one eligible person, interviewers received more money for the first respondent. Such payment system compensates for the higher effort which is normally needed to make the first interview. In Austria, Italy and Switzerland, special premia were also paid to stimulate response rates. Each survey agency fixed its own threshold response rate. Once the target rate was reached, interviewers received higher payments for additional interviews. Finally, an extension to the contract with the survey agencies contained incentives for the survey agency to reach the target response rate. While interviews in excess of the target response rates were paid better, not reaching the target response rate triggered a contractual penalty. In spite of the fact that interviewers were left to organize their own fieldwork, its progress was rather smooth (see Figure 7.3 in De Luca and Lipps, 2005).

e) *Supporting participation*

To increase households' propensity to collaborate, SHARE emailed or mailed an advance letter to each household to inform of upcoming calls or visits by an interviewer, to communicate the nature and motivation of the survey and explain the importance of participating. After the interview, a thank-you letter was mailed out to each respondent in order to increase the propensity to participate in future waves of the survey. In most SHARE countries, incentives for respondents were distributed in order to gain their cooperation. In several countries, individuals received a small gift before completing the interview (e.g. a lottery ticket in Sweden, a box with a set of ball-pens in Germany, a sweet in Austria, a voucher for a department store in Spain). In other countries, incentives were given at the end of the interview (15 Euro per completed household in the Netherlands). Denmark was the only country in which incentives were considered inappropriate.

Tables and Figures

Figure 1. Smooth polynomial of monthly hours of childcare as function of the age of the youngest grandchild

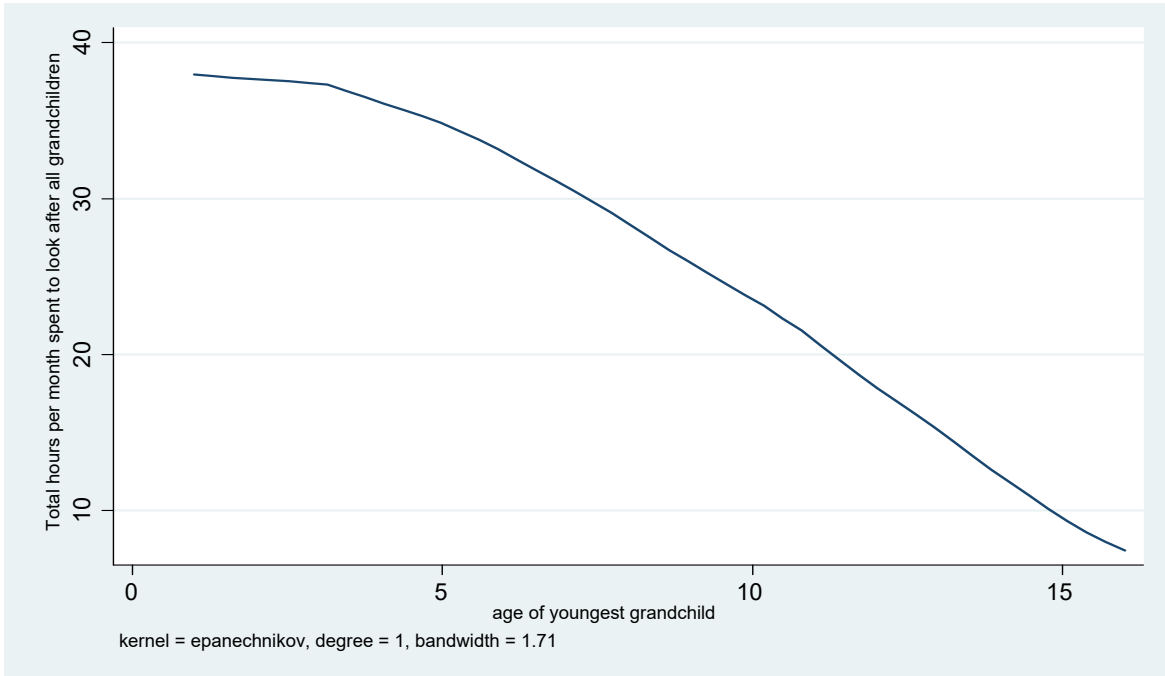


Figure 2. Empirical distribution of Z, the number of months of distance since the first interview

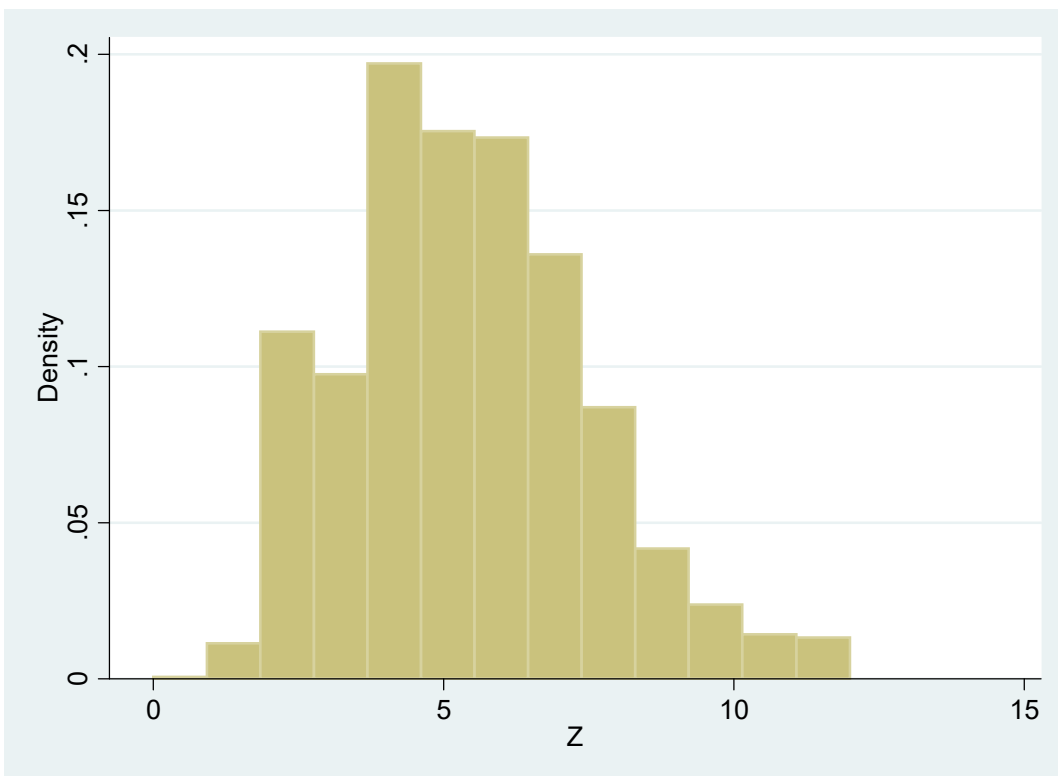


Table 1. Average monthly hours of childcare, by country. SHARE Waves 1 and 2

	Males all	Females all	Males - with positive hours	Females with positive hours
Austria	35.80	32.50	58.48	59.59
Germany	25.59	31.20	56.90	63.75
Sweden	12.25	18.52	58.13	69.16
Netherlands	19.62	25.40	68.89	75.69
Spain	31.20	46.68	37.41	45.40
Italy	38.77	67.31	46.10	63.45
France	28.12	35.97	57.37	61.34
Denmark	16.33	19.03	63.76	69.43
Switzerland	17.09	32.03	55.91	67.66
Belgium	34.60	38.82	65.95	70.34
Czechia	18.94	33.46	38.71	56.19
Poland	38.02	72.72	49.01	61.92

Table 2. Percent depressed (Euro-D equal to or higher than 4), by gender and country.

	Males	Females
Austria	8.04	20.93
Germany	10.26	22.42
Sweden	8.73	19.33
Netherlands	12.05	20.93
Spain	16.96	42.59
Italy	19.87	39.70
France	19.32	38.96
Denmark	9.58	17.84
Switzerland	13.64	19.14
Belgium	14.55	30.66
Czechia	9.95	22.92
Poland	32.29	53.56

Table 3: Summary statistics, by gender.

	Males - means	Males - standard deviations	Females - means	Females - standard deviations
Depressed	0.144	0.351	0.289	0.454
Hours of childcare per month	25.897	54.505	37.297	66.535
Z	5.385	2.293	5.353	2.250
Age at the beginning of the survey	63.653	6.158	62.236	6.237
Age of youngest grandchild at the beginning of survey	4.266	4.040	4.772	4.287
Percent college	0.203	0.402	0.151	0.358
Was vaccinated in early life	0.940	0.237	0.947	0.223
Poor health in early life	0.077	0.266	0.095	0.293
Parents smoke in early life	0.651	0.477	0.624	0.485
Parent drank in early life	0.084	0.278	0.087	0.281
Parents with mental problems in early life	0.020	0.141	0.030	0.172
Lived with both parents at age 10	0.849	0.358	0.852	0.355
Had nervous problems by age 15	0.008	0.088	0.015	0.123
Any hardship episodes by age 15	0.005	0.067	0.004	0.062
Periods of stress by age 15	0.008	0.088	0.003	0.052
Mother in poor health if alive	0.092	0.289	0.110	0.313
Father in poor health if alive	0.025	0.158	0.033	0.178
No parent alive	0.180	0.384	0.206	0.404
Relative math score at 10	0.379	0.485	0.307	0.461
0-10 books at home at age 10	0.427	0.495	0.416	0.493
11-25 books at home at age 10	0.228	0.419	0.229	0.420
26-100 books at home at age 10	0.214	0.410	0.215	0.411
101-200 books at home at age 10	0.061	0.239	0.063	0.243
More than 200 books at home at age 10	0.058	0.234	0.062	0.242
Main breadwinner at 10 was a professional	0.117	0.321	0.115	0.319
Main breadwinner at 10 was a clerk	0.141	0.348	0.141	0.348
Main breadwinner at 10 was a blue collar	0.522	0.500	0.515	0.500
Main breadwinner at 10 was in elementary occupations	0.192	0.394	0.203	0.402
Number of children	2.428	0.862	2.418	0.855
Age of oldest children	37.595	5.771	38.768	6.037
Widowed	0.041	0.198	0.142	0.349
Height in cm	174.673	10.445	162.702	12.717
Number of siblings	0.910	0.286	0.909	0.288
Oldest sibling	0.314	0.464	0.303	0.460
Number of observations	5,694		7,397	

Table 4. OLS estimates. Dependent variable: D. Full sample and by gender.

	Males and females	Males	Females
Hours of childcare per month	0.0001 (0.000)	0.0002** (0.000)	0.0000 (0.000)
Observations	13,091	5,694	7,397
R-squared	0.102	0.062	0.093

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X and W. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.

Table 5. Variability of Z within country and wave

	Wave 1 – minimum	Wave 1 – maximum	Wave 1 – first q.le	Wave 1 – third q.le	Wave 2 – minimum	Wave 2 – maximum	Wave 2 – first q.le	Wave 2 – third q.le
Austria	3	8	4	7	2	11	4	7
Germany	2	8	4	6	0	10	3	6
Sweden	1	10	5	9	1	10	2	5
Netherlands	3	10	4	7	4	12	5	8
Spain	3	10	5	8	1	12	4	7
Italy	2	10	4.5	7	1	12	4	6
France	3	9	7	8	1	4	2	3
Denmark	1	8	4	7	2	11	3	6
Switzerland	2	7	4	5	1	12	4	6
Belgium	9	12	11	12	1	12	4	6
Czechia					1	9	3	6
Poland					2	11	4	6

Table 6. First stage regressions of hours of childcare on the instrument Z. Full sample and by gender.

Variables	Males and females	Males	Females
Z	-1.507*** (0.330)	-1.356*** (0.376)	-1.594*** (0.404)
F-test	20.82	13.04	15.54
Observations	13,091	5,694	7,397

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X and W. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.

Table 7. Balancing tests. Full sample.

Variable	Coefficient and standard error	Variable	Coefficient and standard error
mother in poor health if alive	0.001 (0.006)	was vaccinated in early life	-0.003 (0.005)
father in poor health if alive	-0.001 (0.003)	poor health in early life	0.002 (0.006)
no parents alive	0.012* (0.007)	parents smoke in early life	-0.000 (0.010)
relative math score at 10	0.009 (0.010)	parent drank in early life	0.008 (0.006)
0-10 books at home at age 10	-0.011 (0.010)	parents with mental problems in early life	0.003 (0.003)
11-25 books at home at age 10	0.001 (0.009)	lived with both parents at age 10	-0.006 (0.007)
26-100 books at home at age 10	0.019** (0.008)	had nervous problems by age 15	0.000 (0.002)
101-200 books at home at age 10	-0.005 (0.005)	any hardship episodes by age 15	0.000 (0.001)
more than 200 books at home at age 10	-0.001 (0.005)	periods of stress by age 15	-0.001 (0.001)
main breadwinner at 10 was a professional	0.001 (0.007)	ISCED1	-0.020** (0.008)
main breadwinner at 10 was a clerk	-0.003 (0.007)	ISCED 2	-0.004 (0.008)
main breadwinner at 10 was a blue collar	0.007 (0.010)	ISCED 3-4	0.013 (0.009)
main breadwinner at 10 was in elementary occupations	-0.005 (0.008)	ISCED 5-6	0.007 (0.008)

oldest sibling	0.009 (0.009)	height in cm	-0.278 (0.240)
number of siblings	-0.001 (0.006)		
widowed	-0.009 (0.006)	age of youngest grandchild at start of survey	-0.036 (0.086)
number of children	-0.002 (0.021)	active	-0.002 (0.007)
age of oldest children	0.038 (0.085)	retired	0.005 (0.008)

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence. Additional controls are: gender, age at the start of the interview, country by wave dummies and an indicator of whether the individual is present in both waves. For the age of youngest grandchild at the start of the survey and the dummies “grandparent active” and “grandparent retired” we also control for the pre-determined variables in vectors X and W.

Table 8. Falsification tests. Dependent variable: whether depressed.

	Males and females	Males	Females
No grandchildren	-0.0009 (0.003) [7,504]	-0.0045 (0.003) [3,715]	0.0026 (0.004) [3,789]
Youngest grandchild older than 16	-0.0048 (0.011) [597]	-0.0144 (0.018) [200]	-0.0087 (0.015) [397]
Youngest grandchild aged 16 or younger	-0.0065*** (0.002) [13,091]	-0.0080*** (0.003) [5,694]	-0.0051* (0.003) [7,397]

Note: number of observations within brackets, standard errors within parentheses. Standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X and W. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.

Table 9. 2SLS and IV Probit estimates. Dependent variable: depression D. Full sample and by gender.

	2SLS		IV Probit		IV	
	Males and females	2SLS Males	2SLS Females	Males and Females	Probit Males	IV Probit Females
Monthly hours of childcare	0.0043*** (0.002)	0.0059** (0.002)	0.0032 (0.002)	0.0040*** (0.001)	0.0054*** (0.001)	0.0030** (0.001)
Observations	13,091	5,694	7,397	13,091	5,694	7,397

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X and W. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.

Table 10. 2SLS estimates. Dependent variable: affective suffering (A) and lack of motivation (M). Full sample and by gender.

	Males and females	Males	Females
Affective Suffering	0.0076** (0.003)	0.0160** (0.007)	0.0029 (0.004)
Lack of motivation	0.0148*** (0.004)	0.0145** (0.007)	0.0138*** (0.005)
Observations	13,091	5,694	7,397

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X and W. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.

Table 11. Complier characteristics. Full sample. Ratio of the first stage for characteristic x to the overall first stage.

	Males and females
age at time survey at or above third quartile	0.937
high school or college education	0.615
inactive	1.432
family income at or above third quartile	1.289
numeracy skills at or above third quartile	0.661
one child	1.274
have more than 2 grandchildren	0.292
widowed	0.474
South or Eastern Europe	2.299

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X and W. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.

Table 12. Tests of external validity. OLS regressions of the dummy D on Z and the vector X when HC>0 and HC=0. Full sample and by gender.

	Males and females	Males	Females
hours of child care per month >0	-0.0038 (0.003)	-0.0072** (0.003)	-0.0012 (0.004)
hours of child care per month =0	-0.0112*** (0.003)	-0.0097** (0.004)	-0.0130*** (0.005)

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X and W. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.

Appendix – Additional Tables

Table A1. The effect of Z on lagged depression.

	Males and females	Males	Females
Z	-0.005 (0.003)	-0.002 (0.004)	-0.005 (0.005)
Observations	3,869	1,696	2,173

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X and W. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.

Table A2. The effect of Z on the probability of having grandchildren.

	Males and females	Males	Females
Z	0.0001 (0.002)	-0.0011 (0.002)	0.0012 (0.002)
Observations	21,380	9,667	11,713

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X and W. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.

Table A3. 2SLS estimates. Dependent variable: depression D. Sample without controls from Wave 3. Both genders and by gender.

	Males and females	Males	Females
Hours of childcare per month	0.0060*** (0.002)	0.0070** (0.003)	0.0050** (0.002)
First stage coefficient of Z	-1.1748*** (0.277)	-1.1333*** (0.327)	-1.1888*** (0.334)
Observations	19,421	8,516	10,905
F test first stage	18.03	11.99	12.65

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.

Table A4. The effect of hours of childcare per month on indicators of individual health, cognition and life satisfaction. 2SLS estimates. Full sample and by gender.

	Males and females	Males	Females
Poor health	0.0050*** (0.002)	0.0088*** (0.003)	0.0025 (0.002)
High numeracy	-0.0003 (0.001)	0.0003 (0.002)	-0.0007 (0.002)
High recall of words, first trial	-0.0014 (0.002)	-0.0014 (0.002)	-0.0008 (0.002)
High recall of words, delayed	0.0018 (0.002)	0.0007 (0.003)	0.0027 (0.002)
Satisfied with life	-0.0060** (0.003)	-0.0073 (0.006)	-0.0054** (0.003)
Control over own life	0.0001 (0.002)	0.0025 (0.005)	-0.0015 (0.003)
Autonomy in own life	-0.0022 (0.002)	-0.0024 (0.005)	-0.0025 (0.002)
Pleasure in life	-0.0014 (0.001)	-0.0032 (0.003)	-0.0004 (0.002)
Self-realization	-0.0065** (0.003)	-0.0095 (0.006)	-0.0053** (0.003)

Note: standard errors clustered at the household level in the full sample and at the individual level in the sample by gender. Each regression includes the variables in vectors X and W. One, two and three stars for statistical significance at the 10, 5 and 1 percent level of confidence.