The Effect of Retirement on Mental Health: Indirect Treatment Effects and Causal Mediation

Salm, Martin; Siflinger, Bettina; Xie, Mingjia

Link to publication in Tilburg University Research Portal

THE EFFECT OF RETIREMENT ON MENTAL HEALTH: INDIRECT TREATMENT EFFECTS AND CAUSAL MEDIATION

By

Martin Salm, Bettina Siflinger, Mingjia Xie

12 April 2021

Discussion Paper

CentER

TILBURG ◆ UNIVERSITY

The Effect of Retirement on Mental Health: Indirect Treatment Effects and Causal Mediation

Martin Salm[*]   Bettina Siflinger[†]   Mingjia Xie[‡]

April 10, 2021

Abstract

People experience multiple changes in their lives after retirement which can affect their mental health. In this paper, we examine the mediating impact of grandparental childcare in the effect of retirement on mental health among elderly women in Europe. We apply a semi-parametric estimation strategy to disentangle the total effect of retirement on mental health into a direct effect, and an indirect effect mediated through grandparental childcare. We find that retirement directly leads to a significant increase in mental health problems. However, this effect is completely offset by a significant reduction in mental health problems generated by a mediating effect of grandparental childcare. As a result, the total effect of retirement on mental health is close to zero. We then examine country-specific heterogeneity in the provision of public childcare and find that the mediating effect unfolds its full compensating strength in countries in which grandparental childcare is supplemental to public childcare. Our results have important implications for designing old-age social policies.

JEL Classification: I10, J13, J26, C14, C36

Key words: retirement; mental health; grandparental childcare; causal mediation analysis; semi-parametric estimation

---

[*]Tilburg University, Department of Econometrics & OR, m.salm@uvt.nl

[†]Tilburg University, Department of Econometrics & OR, b.m.siflinger@uvt.nl

[‡]Tilburg University, Department of Econometrics & OR, m.xie_1@uvt.nl

1 Introduction

Retirement is an important transition in life and it has substantial consequences on economic outcomes, such as income, consumption, and health. In this study, we investigate the effects of retirement on mental health. As pointed out by Piccio and van Ours (2019), mental health may be shaped immediately by retirement, whereas physical health is affected more gradually after retirement. While the empirical literature has extensively studied this relationship, evidence on the magnitude and direction of this effect is mixed, strongly depending on the institutional context and country under study and the chosen identification strategy. Some studies find a positive impact of retirement on mental health, well-being, and related outcomes (for instance Charles, 2004; Johnston and Lee, 2009; Eibich, 2015; Kolodziej and García-Gómez, 2019). Other studies conclude there is a negative effect (e.g. Dave et al., 2008; Rohwedder and Willis, 2010; Bonsang et al., 2012; De Grip et al., 2012; Heller-Sahlgren, 2017; Mazzonna and Peracchi, 2017; Atalay et al., 2019) or no effect (Coe and Zamarro, 2011; Behncke, 2012; Fe and Hollingsworth, 2016).[1]

While this literature has closely studied the effect of retirement on mental health, much less is known about *why* retirement has an impact on mental health (Gelman and Imbens, 2013; Imai et al., 2011). Only a few studies analyze the underlying mechanisms in more detail (Eibich, 2015; Belloni et al., 2016; Atalay et al., 2019), typically regressing the mediator of interest on retirement. This approach allows estimating the causal effect of retirement on the mediator but it cannot determine to what extent the effect of retirement on mental health is caused by a mediator. In other words, even though retirement might affect the mediator, this does not necessarily imply that the mediator is a causal mechanism for the effect of retirement on mental health.

In this paper, we study the impact of retirement on mental health among elderly women in Europe, applying a causal mediation analysis. The idea of such an approach is to split the total effect of retirement on mental health into two distinct effects – an indirect effect

---

[1]For an excellent literature overview on mental health and retirement, see Piccio and van Ours (2019).

and a direct effect. The indirect effect refers to the effect of retirement on mental health that operates through a particular mediator of interest. The direct effect is the effect of retirement on mental health that operates through all channels of retirement other than the mediator. The mediator of interest is grandparental childcare i.e. childcare provided by grandmothers mediates the impact of retirement on mental health. As discussed in Eibich (2015) the time grandparents spent on childcare is positively associated with retirement and with mental health.

To estimate direct and indirect effects, we follow a non-parametric instrumental variable approach developed by Frölich and Huber (2017). This approach allows decomposing the local average treatment effect (LATE) into a direct treatment effect and an indirect treatment effect. The treatment and the mediator are allowed to be endogenously related to unobserved confounders and identification is achieved using separate instrumental variables. As the non-parametric identification does not impose the assumption of constant direct and indirect treatment effects, it allows for arbitrary interactions between treatment, mediator, individual characteristics, and unobservables. Compared to a parametric model, this approach thus does not impose rigid assumptions on the interaction between these quantities. To keep the analysis tractable, we estimate the model semi-parametrically.

The analysis is based on cross-sectional data of women aged 50-80 years from ten European countries who participated in the Survey of Health, Ageing, and Retirement in Europe (SHARE). We measure mental health problems using the EURO-D depression scale. The data contain detailed information on retirement, the time spent taking care of grandchildren, and a large number of sociodemographic characteristics. As an instrumental variable for retirement, we use country-specific statutory retirement ages. Grandparental childcare is instrumented with the sex ratio of a woman's children. We discuss the assumptions of the model and of our instruments in detail, and we show in a wide range of tests that our empirical strategy is valid.

We estimate a total effect of retirement on mental health that is close to zero among com-

pliers. This total effect consists of opposing direct and indirect treatment effects. The estimated direct effect is negative and statistically significant. Accordingly, retirement leads to an increase in depression by about 0.58 points or 0.26 standard deviations. The estimated indirect effect has the opposite direction, is somewhat bigger in magnitude, and also significantly different from zero. The estimate suggests that a retirement-induced increase in grandparental care reduces mental health problems by about 0.68 points or 0.30 standard deviations. It implies that the mediating effect of grandparental childcare is large enough to fully compensate for the negative effect of retirement on mental health operating through other potential mechanisms. This result highlights the importance of taking account of potential effect heterogeneity when analyzing direct and indirect treatment effects.

We next investigate whether the importance of the mediator effect differs across countries. In some European countries, such as Italy and Spain, care provided by grandparents is a major source of childcare, while in Northern European countries formal childcare is dominant. We thus analyze two groups of countries that differ in the extent to which children are enrolled in formal childcare. For countries with high enrollment rates in public childcare, retirement significantly decreases mental health problems by about 0.59 points or 0.26 standard deviations among compliers. By contrast, in countries with low enrollment rates, the total effect of retirement is negative and implies an increase in mental health problems by about 0.70 points (0.32 standard deviations). In these countries, the mediating effect of grandparental childcare is not strong enough to offset the negative direct effect.

The results of this study contribute to several strands of the literature. Many studies estimate the LATE of retirement on mental health and related outcomes using age eligibility of retirement as exogenous variation (e.g. Coe and Zamarro, 2011; Bonsang et al., 2012; Eibich, 2015; Mazzonna and Peracchi, 2017; Atalay et al., 2019). To shed light on potential mechanisms, the LATE is often estimated for sub-groups, by regressing retirement on the potential mediator, or by adding it as a covariate in the main specification. A study that uses such an approach to shed light on potential mechanisms linking retirement and health is

4

Eibich (2015). Potential mediators are higher sleep duration, less work stress, more physical activity, and increased childcare, albeit to a lower extent. While these results point toward potential mediators, the analysis is investigating associations rather than causal mediator effects. Our study goes beyond this and other correlational analyses of mechanisms carried out in the literature. We implement a causal mediation approach that allows estimating a causal mediator effect of retirement on mental health.

Our analysis moreover highlights the importance of grandparental childcare as a mediator for the effect of retirement on mental health, relative to all other mechanisms of retirement. By showing that the zero total effect is a result of offsetting direct and indirect effects, our results add to the understanding of *why* retirement affects mental health. Besides, our country-specific analysis sheds light on why results in this literature are sensitive to the countries analyzed (Mazzonna and Peracchi, 2017).

We also add to a handful of studies evaluating the impact of grandparental care on grandparent's health. For instance, Brunello and Rocco (2019) find that more childcare considerably increases depression among European grandparents. For Taiwan, Ku et al. (2012) show that grandparental childcare significantly reduces functional limitations but does not increase mental health problems. While these studies directly estimate a LATE of grandparental childcare on mental health, our mediation analysis suggests that it can alleviate the negative consequences of retirement for the mental health of grandmothers.

From the perspective of grandmothers' mental health, grandparental childcare should be a supplement to formal childcare rather than a substitute. Only in countries in which grandparents are typically not primary caregivers, the negative effects of retirement on mental health through other factors may be mitigated. That grandmothers are in good mental health also is an important factor for maternal labor supply. As discussed by Del Boca (2015) parents tend to rely on grandparents, in particular in countries where formal childcare is rationed or otherwise not available. Without an extended family, many women may not be able to participate in the labor market (e.g. Zanella, 2017). Since healthy grandparents are a major

source of informal childcare (Del Boca et al., 2005), universal access to public childcare could promote both, maternal labor supply and the mental health of grandmothers.

The rest of the paper is organized as follows. Section 2 introduces the econometric model, outlines our identification strategy and discusses the estimation approach. Section 3 describes the data. Section 4 provides a detailed discussion of our instruments and further assumptions necessary for identification. Section 5 presents the main results and compares them to a number of alternative specifications. Section 6 concludes.

2 Empirical strategy: a semi-parametric instrumental variable approach

The goal of our study is to disentangle the total effect of retirement on mental health into an indirect and a direct effect. The indirect effect corresponds to the effect of retirement on mental health that is mediated by a change in grandparental childcare. The direct effect is the effect of retirement on mental health that operates through all potential channels of retirement other than grandparental childcare.

In the causal mediation literature, there is an ongoing discussion about the identification of treatment and mediator effects. Parametric causal mediation models are often based on the assumption of no-interaction between treatment and mediator, or they assume exogeneity of the treatment as well as the mediator conditional on the treatment (sequential ignorability, for an overview, see Huber (2020)). A few studies relax sequential ignorability on the treatment and/or the mediator, and use parametric IV models to estimate direct and indirect treatment effects (see for instance Powdthavee et al., 2015; Chen et al., 2019, 2020). This however comes at the cost of either ignoring interactions between the treatment and the mediator or imposing assumptions such as additivity between observables and unobservables. As pointed out by Imai et al. (2011) the identification of a causal mechanism by definition requires inference about changes in the mediator as a response to treatment. Ignoring effect

6

heterogeneity leads to inconsistent estimates of the mediator effect.

To address these concerns in our application, we adopt a nonparametric causal mediation framework by Frölich and Huber (2017). The method makes use of two distinct instruments to overcome endogeneity problems with the treatment and the mediator. With these instruments, all direct and indirect effects for all treatment compliers are identified. Moreover, Frölich and Huber (2017) generalize the nonparametric identification results to situations with a continuous mediator and continuous instruments. Unlike a linear model with interactions between treatment and mediator, the nonparametric model allows for arbitrary effect heterogeneity through flexible interactions between the treatment, the mediator, and all individual characteristics.

For our application, such a flexible nonparametric instrumental variables approach is preferred for several reasons. First, the direct effect captures all possible channels other than grandparental childcare through which retirement affects mental health and thus may interact with the mediating effect of childcare in complex, nonlinear ways. Second, our treatment is the retirement decisions of women and our mediator is a woman's decision to provide childcare to her grandchildren. Consequently, the treatment and the mediator are both endogenous. Third, the amount of childcare provided by a grandmother is a continuous measure, and the corresponding instrument is continuous as well. For estimation we follow the suggestion of Frölich and Huber (2017), and we estimate the model semi-parametrically. This mitigates problems often accompanying nonparametric methods but retains flexibility and still allows for a great amount of effect heterogeneity.

2.1 The econometric model

The following section shows how the total effect of retirement on mental health can be decomposed into a direct effect and an indirect mediator effect operating through grandparental care. Building on Frölich and Huber (2017), we consider the following system of

non-separable nonparametric equations.

$$Y = \varphi(D,\ M,\ X,\ U), \tag{1}$$

$$M = \zeta(D,\ Z_2,\ X,\ V), \tag{2}$$

$$D = 1\{\chi(Z_1,\ X,\ W) \geq 0\} \tag{3}$$

where φ, ζ, χ are unknown functions. Y is mental health which is a function of the retirement status D, the amount of childcare provided to grandchildren M, and covariates X. U, V, W are unobservables which are allowed to be correlated. This correlation raises two endogeneity issues. First, since U and W can be correlated, retirement D is endogenous, even after conditioning on X. The problem is resolved by using a binary instrument for retirement, Z_1. Our first instrument is whether a woman is older than the statutory retirement age or not. Second, grandparental childcare M is confounded by V which in turn can be correlated with U and W. Thus, a second instrument Z_2 is required, and it has to induce variation in M but is independent of variation in D. We use the sex ratio of a woman's children as the second instrument.[2]

Equations (1)–(3) show that retirement D is linked to mental health Y in two ways. On the one hand, retirement may have an indirect effect on mental health which operates through grandparental childcare. On the other hand, retirement may have a direct effect on mental health which comprises any mechanisms other than grandparental childcare, e.g. changes in income or social interactions. The treatment effect of retirement on mental health explained by the 'grandparental childcare mechanism' is denoted as *indirect effect*. The retirement effect explained by 'all other mechanisms' is denoted as *direct effect*. Adding up those two effects provides us with the *total* average treatment effect among compliers (LATE).

To define the direct and indirect effect as well as the total effect, we closely follow Frölich and Huber (2017) using a potential outcome framework. According to Equations (1) and (2),

---

[2]Section 4 will discuss the validity of our instruments in detail.

three potential outcomes can be considered: the potential mental health outcome Y^d; the potential mediator outcome under treatment, M^d; and the potential mental health outcome, $Y^{d,M^{d'}}$, as a function of the retirement status, $D = d$, $d \in \{0,1\}$, and the potential mediator, $M^{d'}$, under treatment, $D = d'$, $d' \in \{0,1\}$.[3]

The total effect among compliers corresponds to the LATE. It is defined as the mean difference in the potential mental health outcome for retired women, $D = 1$ and nonretired women, $D = 0$.

$$\Delta = \mathrm{E}[Y^1 - Y^0 | T = co] = \mathrm{E}[Y^{1,M^1} - Y^{0,M^0} | T = co], \tag{4}$$

where $T = co$ denotes the complier subpopulation. Following from Equation (3), compliers are defined by $D(Z_1 = 1, X, W) = 1$ and $D(Z_1 = 0, X, W) = 0$. Thus, conditional on X, the type T is a function of the unobservable W only.

The right-hand side of Equation (4) emphasizes that the LATE comprises both, the direct treatment effect and the indirect treatment effect via M. These two effects can be separated by keeping either grandparental care fixed at its potential value for $D = d$, or the retirement itself fixed at $D = d$.

The *natural* indirect treatment effect (IATE) among compliers is obtained by taking the mean differences in mental health when retirement is fixed at $D = d$, while grandparental care is shifted to its potential value for being retired, M^1 and not being retired, M^0.

$$\delta(d) = \mathrm{E}[Y^{d,M^1} - Y^{d,M^0} | T = co], \quad d \in \{0,1\} \tag{5}$$

The *natural* direct treatment effect (DATE) among compliers is the effect of retirement D on mental health Y when keeping grandparental care M fixed at its potential value for $D = d$.

---

[3]The potential mental health outcome is a function of the retirement status, $D = d$, $d \in \{0,1\}$ and the potential mediator, M, under treatment, $D = d'$, $d' \in \{0,1\}$ is $Y^{d,M^{d'}} = \mathrm{E}[\varphi(D = d, M = \zeta(D = d'))]$.

In this way, we shut down the 'grand-parenting mechanism'.

$$\theta(d) = \mathrm{E}[Y^{1,M^d} - Y^{0,M^d}|T = co], \quad d \in \{0,1\} \tag{6}$$

Table 1 gives an overview of the different complier treatment effects from observed and counterfactual outcomes.

Table 1: Definition of DATE, IATE, and LATE

	$M = M^0$	$M = M^1$	IATE
$D = 0$	$\mathrm{E}[Y^{0,M^0}|T = co]$	$\mathrm{E}[Y^{0,M^1}|T = co]$	$\delta(0)$
$D = 1$	$\mathrm{E}[Y^{1,M^0}|T = co]$	$\mathrm{E}[Y^{1,M^1}|T = co]$	$\delta(1)$
DATE	$\theta(0)$	$\theta(1)$	

For any observation, only Y^{0,M^0} or Y^{1,M^1} can be observed, and neither Y^{0,M^1} nor Y^{1,M^0} can be observed. All potential outcomes and effects are conditional on compliers.

The LATE is the diagonal difference between $\mathrm{E}[Y^{1,M^1}|T = co]$ and $\mathrm{E}[Y^{0,M^0}|T = co]$. The mixed terms in the off diagonal, $\mathrm{E}[Y^{1,M^0}|T = co]$ and $\mathrm{E}[Y^{0,M^1}|T = co]$, are counterfactual outcomes and represent the interactions between retirement and the mediator. $\mathrm{E}[Y^{1,M^0}|T = co]$ is the expected outcome in the mental health of a woman who is retired and provides childcare as if she was not retired, while $\mathrm{E}[Y^{0,M^1}|T = co]$ is the expected mental health of a woman who is not retired but provides childcare as if she was retired. We will estimate both counterfactual outcomes with our data. From Table 1 it becomes clear that the interactions between the treatment and the mediator states generate two direct effects and two indirect effects. We will discuss these now step by step.

- $\theta(0)$ represents the direct effect of retirement on mental health if grandparental care is fixed at the pre-retirement level. The effect captures all potential channels other than grandparental care that link retirement and mental health.

10

- $\theta(1)$ represents the direct effect of retirement on mental health if grandparental child-care is fixed at the post-retirement level. This effect represents all channels other than grandparental childcare.

- $\delta(0)$ represents the indirect effect of retirement on mental health. This effect is generated through a retirement-induced change in grandparental childcare while other channels of retirement are fixed at the pre-retirement level.

- $\delta(1)$ represents the indirect effect of retirement on mental health generated through retirement-induced changes in grandparental childcare. Here, other channels of retirement are fixed at the post-retirement level.

By summing over the complements of one direct and one indirect effect one can obtain the LATE, $LATE = \delta(0) + \theta(1) = \delta(1) + \theta(0)$. Note that $\theta(0)$ and $\theta(1)$, as well as $\delta(0)$ and $\delta(1)$, do not need to be necessarily numerically identical because the effects are evaluated at different states of the treatment and of the mediator. For instance, $\theta(0)$ is positive if all mechanisms other than grandparental care interact positively with the pre-retirement level of grandparental childcare. An example could be that grandparental childcare is less stressful before retirement. Keeping grandparental care at this level would reinforce other positive aspects of retirement, such as increased time for leisure or exercising, while negative mechanisms are dampened. By contrast, $\theta(1)$ can be negative if the post-retirement grandparental care level would incur a large amount of stress, such that other negative mechanisms linking retirement and mental health are not offset but even amplified. For $\delta(0)$ and $\delta(1)$ the rationale is very similar. Strong differences in the pairs $(\delta(0), \delta(1))$ and $(\theta(0), \theta(1))$ indicate strong heterogeneity in the interactions between treatment and mediator, thus stressing the need for a highly flexible empirical specification.

2.2 Identification and estimation

Identification of the direct and indirect treatment effects of retirement on mental health is based on a set of assumptions common to identification approaches using instruments. Besides, additional assumptions are required to secure identification of the counterfactual outcomes on compliers $T = co$, $\mathrm{E}[Y^{0,M^1}]$ and $\mathrm{E}[Y^{1,M^0}]$, when the mediator M and the respective instrument Z_2 are both continuous. The key idea of identification is as follows: the treatment D is changed through variations in Z_1, while the mediator M is fixed through a variation in Z_2, offsetting the effect Z_1 has on M. The identification results in this section are based on Frölich and Huber (2017).

Assumption 1 states that the instruments Z_1 and Z_2 are independent of unobservables in Equations (1)–(3), conditional on covariates X.

Assumption 1

$$(Z_1, Z_2) \perp (U, V, W)|X$$

In particular, this assumption requires that the instrument Z_2 is conditionally independent of W. Since the type T (complier, never takers, always takers) is a function of W conditional on X, it implies that the probability of being complier is independent of Z_2. Since the model we apply identifies the fraction of compliers, this assumption can be tested.[4]

A second assumption needed for identification is that Z_1 and Z_2 are conditionally independent of each other.

Assumption 2

$$Z_1 \perp Z_2|X$$

Intuitively, this assumption ensures that once we vary D through a variation in Z_1, it is ruled out that variation in Z_2 affects D. If Z_1 and Z_2 were correlated, we could not make

---

[4]Frölich and Huber (2017) show that instead of assumption 1 two weaker assumptions are sufficient for identification: $(Z_1, Z_2) \perp (U, V)|T, X$ and $Z_1 \perp (U, V, T)|Z_2, X$. In contrast to Assumption 1, this set of assumptions allows Z_2 and W to be dependent, conditional on X. In Section 4.1, we show that the correlation between Z_2 and the probability of being complier is close to zero which supports the stronger Assumption 1.

sure whether they are shifting M or D or both.

To rule out defiers and to ensure the existence of compliers, we need monotonicity of the treatment D in Z_1.

Assumption 3

$$Pr(T = de) = 0 \quad and \quad Pr(T = co) > 0.$$

Under these assumptions, the complier subpopulation can be identified as

$$Pr(T = co) = E\left[\frac{D}{Pr(Z_1 = 1|Z_2, X)}\frac{Z_1 - \bar{\bar{\Pi}}}{1 - \bar{\bar{\Pi}}}\right],$$

where $\bar{\bar{\Pi}} = Pr(Z_1 = 1|Z_2, X) = \Pi = Pr(Z_1|X)$ under Assumption 2.

In our application, the mediator M is the amount of childcare provided by grandmothers. The instrument Z_2 for M is the sex ratio of a grandmother's children. Frölich and Huber (2017) suggest a control function approach in which the mediator M is assumed to be monotone in unobservables V (see Equation (2)). This is achieved by restricting V to be a continuous, random variable with a strictly increasing CDF in the support of V, conditional on covariates among compliers. Moreover, the function in Equation (2) is strictly increasing in its last argument, V. These two conditions are summarized by the following assumption.

Assumption 4

(a) V is a continuously distributed random variable with a CDF $F_{V|X=x,T=co}(v)$
    that is strictly increasing in the support of V, for almost all values of x.

(b) $\zeta(d, z_2, x, v)$ is strictly increasing in v for almost all d, z_2 and x.

Intuitively, Assumption 4 implies that one can obtain conditional distributions of V from conditional distributions of M, by defining a control function C.[5] By replacing V with C

---

[5]The control function is defined as $C_i = C(M_i, D_i, Z_{1i}, X_i)$,

$$C(m, d, z_2, x) = \frac{E\left[(d + D - 1)\{Z_1 - \bar{\pi}(z_2, x)\}|M \leq m, Z_2 = z_2, X = x\right]}{E\left[D\{Z_1 - \bar{\pi}(z_2, x)\}|Z_2 = z_2, X = x\right]}F_{M|Z_2, X^{(m, z_2, x)}},$$

and invoking Assumptions 1-4, it can be shown that the potential outcome is identified by reweighting the complier subpopulation. This weight is a ratio of conditional densities of M under $Z_1 = 0$ versus $Z_1 = 1$, and is compensating the effect of Z_1 on M. To ensure that this weight is well defined (neither being zero nor infinity), we require common support in M.

Assumption 5

$$0 < Pr(Z_1 = 1 | M, V, X, T = co) < 1, \text{ almost surely.}$$

Under Assumptions 1–5, the potential outcome among compliers is identified.

$$\mathrm{E}(Y^{1,M^0} | T = co) = \mathrm{E}\left[Y D \Omega \frac{Z_1 \Pi}{\Pi(1 - \Pi)} \right] \frac{1}{Pr(T = co)}, \tag{7}$$

with

$$\Omega = \omega(M, C, X) = 1 - \frac{\mathrm{E}(Z_1 | M, C, X) - \pi(X)}{\mathrm{E}(D Z_1 | M, C, X) - \mathrm{E}(D | M, C, X) \pi(X)} \tag{8}$$

where Ω corresponds to the weight required to undo the effect of Z_1 on M and $\Pi = Pr(Z_1 | X)$. Equations (7) and (8) only depend on quantities that can be estimated from the data. We estimate the conditional expectations of C using OLS, and obtain estimates of conditional distributions using non-parametric kernel density estimators. Estimates of conditional probabilities in Equation (8), $\mathrm{E}(Z_1 | M, C, X)$, $\mathrm{E}(D Z_1 | M, C, X)$, and $\mathrm{E}(D | M, C, X)$, are obtained from logistic regressions on $(1, M, C, X)$. The fraction of compliers in Equation (7) is obtained by estimating $\Pi = Pr(Z_1 = 1 | X)$ using a logistic specification. Since under Assumption 2 $\Pi = \bar{\Pi} = \pi(X)$, estimates of $\pi(X)$ are used for Π in Equation (7). By plugging estimates in Equations (7) and (8), we obtain semi-parametric estimates of potential outcomes required to compute direct and indirect effects of retirement on mental health (for details, see Appendix B.1).

---

and identifies V_i, see Frölich and Huber (2017).

In practice, two difficulties associated with estimating Equations (7) and (8) can occur. First, for Assumption 5 to hold requires that the weights $\omega(.)$ are neither zero nor infinity. In our application, we remove observations if their estimated weights are close to zero or extremely large to guarantee common support. Second, the ratio of the estimated Ω and the estimated propensity score Π in Equation (8) can become arbitrarily large for some observations. As a consequence, the estimated potential outcome for these observations becomes extremely influential (Huber et al., 2013; Frölich and Huber, 2014). To prevent our estimates to be driven by such observations, we apply the following trimming procedure: an observation is discarded if the ratio of Ω and Π is outside the 0.25% quantile on both sides of its estimated distribution for a given mean potential outcome.[6]

In our setting, Z_1 is a binary variable indicating whether a woman's age is above statutory retirement age in a given country and year. Therefore, variation in Z_1 is jointly determined by age, country of residence, and calendar year, and conditioning on this information gives a propensity score $\Pi = Pr(Z_1 = 1|X)$ that is either zero or one. As a consequence, common support of M (Assumption 5) cannot be maintained.[7]

To address this issue, we partial out the variation of age, as one variable jointly determining variation in Z_1. To this end, we assume that mental health is a function of covariates that is separable and linear in age and age squared.

$$Y_i = \phi(D, \ M, \ X, \ U) + \beta_1 Age + \beta_2 Age^2 \tag{9}$$

Note that Equation (9) is the same as Equation (1) except for the additional assumption

---

[6]We show that our results are robust to different quantile levels.

[7]Controlling for age, country, and year in Equations (1)-(3), implies that we do not have any compliers in the population, thus Ω approaches ∞. However, by not controlling for this information we may miss potentially important determinants of mental health.

about age. The conditional expectation of Equation (9) is

$$\mathrm{E}[Y|D, M, X] = \mathrm{E}[\phi(D,\ M,\ X,\ U)|D, M, X] +$$

$$\beta_1\,\mathrm{E}[Age|D, M, X] + \beta_2\,\mathrm{E}[Age^2|D, M, X]. \tag{10}$$

Subtracting Equation (10) from (9) yields

$$Y - \mathrm{E}[Y|D, M, X] = \beta_1(Age - \mathrm{E}[Age|D, M, X]) + \beta_2(Age^2 - \mathrm{E}[Age^2|D, M, X]) +$$

$$\phi(D,\ M,\ X,\ U) - \mathrm{E}[\phi(D,\ M,\ X,\ U)|D, M, X] \tag{11}$$

Partialling out age results in Z_1 being no longer perfectly determined by covariates, and thus compliers exist. We estimate all conditional expectations and parameters in Equations (10) and (11) by OLS.[8] Our main outcome is now a transformed version of the original mental health. In Appendix B.2 we show that the interpretation of direct and indirect treatment effects is not affected by the transformation.

In addition to semi-parametric estimates of direct and indirect treatment effects, we also estimate parametric specifications for comparison. We moreover compare our estimated LATE with standard IV and OLS estimates. We will discuss the results obtained from these different empirical strategies in Section 5.

3   Data

We use data from the Survey of Health, Ageing, and Retirement in Europe (SHARE), a biannual cross-national panel data set that consists of individual-level data on health, socio-

---

[8]We first estimate $\mathrm{E}[Y|D, M, X]$, $\mathrm{E}[Age|D, M, X]$, and $\mathrm{E}[Age^2|D, M, X]$ running separate linear OLS regressions of Y, Age and Age^2 on $(1, D, M, X)$. We then obtain $\hat{\beta}_1$ and $\hat{\beta}_2$ from an OLS regression of $Y - \hat{Y}$ on $Age - \widehat{Age}$ and $Age^2 - \widehat{Age^2}$. Finally, we compute the transformed mental health outcome as $Y' = Y - (\hat{\beta}_1 Age + \hat{\beta}_2 Age^2)$. We use Y' throughout this study.

economic status, and family networks of individuals aged 50 or older.[9] SHARE follows over 140,000 individuals from 27 European countries and Israel. It consists of seven waves, covering the years 2004 to 2019. SHARE administers two types of surveys, a regular panel survey in waves 1,2, and 4-6, and a retrospective survey in waves 3 and 7. In the regular survey, respondents are asked about their mental health, work status, education, and other sociodemographic characteristics. Moreover, respondents report information about their adult children, such as age, gender, residence, and the number of their children. Also, SHARE asks respondents about how much time they devote to their grandchildren. In waves 3 and 7, SHARE conducted a retrospective survey (SHARELIFE), collecting information on all important areas of respondents' lives, ranging from partners and children over housing and work history to detailed questions on health and health care.

For the empirical analysis, we construct a pooled cross-section by using information from the first record of a respondent. We further restrict our sample to individuals who (1) are female, (2) are aged between 50 and 80, and (3) have at least one grandchild who is not older than 16 years. Finally, we keep only respondents from countries that participated in all waves, namely Austria, Germany, Belgium, Denmark, France, Italy, Netherlands, Spain, Sweden, Switzerland. Our analytic sample consists of 11,386 individuals.

3.1 Mental health

Our main outcome of interest is mental health which is measured by the EURO-D depression scale. The EURO-D scale was developed to assess and compare the prevalence of depression between European countries.[10] This scale consists of 12 items, asking respondents whether or not they experienced depression, pessimism, suicidality, guilt, trouble with sleep, lack of interest, irritability, changes in appetite, fatigue, problems with concentration, lack of enjoyment, and tearfulness in the last month. The items are coded as 0-1 dummy variables,

---

[9]This paper uses data from SHARE Waves 3, 4, 5, 6, and 7. (DOI: 10.6103/SHARE.w3.700, 10.6103/SHARE.w4.700, 10.6103/SHARE.w5.700, 10.6103/SHARE.w6.700, 10.6103/SHARE.w7.700), see Börsch-Supan et al. (2013) for methodological details

[10]An evaluation of the EURO-D depression scale can be found in Castro-Costa et al. (2008).

where 1 refers to the presence of a symptom, and 0 otherwise. The total score is the sum of all 12 items. It ranges from 0 to 12, and a higher score indicates more depressive symptoms.

3.2   Retirement status

To measure an individual's decision of being retired, we make use of a question in SHARE asking respondents which situation best describes their current work status: retired, employed or self-employed, unemployed, permanently sick or disabled, homemaker, or other. We compute a binary measure of retirement which takes the value 1 if an individual considers herself as retired, and 0 otherwise.[11]

A key problem in the analysis of retirement and health is that retirement decisions are endogenous (see e.g. Mazzonna and Peracchi, 2012; Insler, 2014; Mazzonna and Peracchi, 2017). Endogeneity arises from reverse causality, or unobserved confounders, such as cognitive functioning or health limitations. To overcome these issues, one strategy is to use statutory retirement ages in each country's social security scheme as an instrument (e.g. Rohwedder and Willis, 2010; Coe and Zamarro, 2011; Mazzonna and Peracchi, 2012, 2017). In our application, the instrument is a binary indicator that takes the value 1 if a woman is at or above the statutory retirement ages given country and birth cohort, and is 0 otherwise (a detailed discussion on this instrument is provided in Section 4.2). Column (1) of Table 2 displays the correlation between whether a woman is above the statutory retirement age and her retirement decision, controlling for covariates. Being at or above the statutory retirement age increases the probability of being retired by almost 40 percentage points. The F-statistic is 279.7, indicating that the instrument is sufficiently strong.

---

[11]There are several definitions of retirement status. Insler (2014) discusses two common definitions of being retired: self-reported retirement status, or not being in paid labor. Both have been used in the literature.

3.3 Grandparental childcare

We are interested in understanding which role the provision of childcare by grandmothers plays in the effect of retirement on mental health. To this end, we utilize a question in SHARE that asks respondents whether they have taken care of any of their grandchildren without the presence of the parents, and if yes, how often they took care of them in the past 12 months.[12] We combine these two questions into a single measure on the intensity of childcare provided by a woman for grandchildren below the age of 16.[13] The fraction of time on childcare on a given day, g_j, for children of child j is defined as

$$
g_j = \begin{cases}
1 & \text{if woman provides care to child } j\text{'s children daily} \\
1/7 & \text{if grandmother provides care to child } j\text{'s children weekly} \\
1/30 & \text{if grandmother provides care to child } j\text{'s children monthly} \\
0 & \text{if grandmother provides care to child } j\text{'s children less often.}
\end{cases}
$$

Some women report that they provide daily care to more than one child's children. This implies that the maximum amount of childcare is greater than 1. To deal with this, we impose the restriction that a woman cannot spend more than 100% of her time for childcare across all children's $j = 1, ...k$ children.

$$
CG \equiv \min \left(\sum_{j=1}^{k} g_j, \ 1 \right),
$$

where CG denotes an upper bound on the actual time spent on childcare. Consider a woman who has several grandchildren from her three children $j = 1, 2, 3$. If the woman takes care of all grandchildren every week, she spends $CG = \min(\sum_{j=1}^{3} \frac{1}{7}, 1) = \frac{3}{7}$ of her time on childcare. If the woman provides daily care to children of one child, but weekly care to

---

[12]The wording of this question is "How often (daily, weekly, monthly, less often) do you look after child(ren) of your i^{th} child?", where i denotes to the birth order of a child.

[13]We could also have computed a similar measure on a weekly or monthly basis which would just be a rescaling of the daily measure.

her other children's children, the maximum time she spends on childcare is constrained to $CG = \min(1 + \frac{2}{7}, 1) = 1$.

To deal with endogeneity in the intensity of grandparental childcare, we instrument our measure with the number of her daughters as a ratio of all her children. A similar instrument has been used e.g. by Rupert and Zanella (2018), and we will discuss the instrument and related assumptions in detail in Section 4.3. Column (2) in Table 2 shows that our instrument is highly relevant for a grandmother's time spent on childcare. Women increase childcare by 0.066 or about 1 day every two weeks if all children are daughters compared to all children being sons. The corresponding F-statistic is 47.8, suggesting that the instrument is sufficiently relevant.

Table 2: First stage regression results for retirement status and intensity of grandparental childcare

	retired: yes	intensity childcare
	(1)	(2)
age $\geq$ statutory retirement age	0.400	
	(0.024)	
ratio: nr daughters/all children		0.066
		(0.010)
F-statistic	279.737	47.869
number of observations	11,386	

Standard errors clustered on country-birth cohort level in parentheses; OLS regressions of treatment and mediator on instruments using 11,386 observations. Control variables: age, age quadratic, marital status, education level, number of children, number of grandchildren, SES in childhood and binary missing value indicator for childhood SES, wave FE, and country FE.

3.4 Summary statistics

Table 3 shows the summary statistics for all variables used in the empirical analysis. The average Euro-D depression score in our sample is about 2.60 and rather low given its range from 0 to 12. It indicates that the women in our sample on average have good mental health. As described in Section 2.2, we transform the original EURO-D score to maintain the common support assumption for the mediator. The transformed mental health variable takes values from 7.48 to 19.98 and has a mean of 10.49. The standard deviation of the transformed variable is 2.25 and is almost identical to the standard deviation of the original EURO-D score of 2.26.

In our sample, women are on average about 65 years old and almost 50% report being retired. Self-reported retirement is only slightly lower than the fraction of women who have passed the official statutory retirement age in their home country at the time of the interview (53.7%). The average fraction of time grandmothers spend on childcare is about 0.16, which refers to a bit more than one day per week. As expected, the sex ratio in a woman's children is well balanced. About 50% of her children are girls and boys, respectively.

On average women have 2.6 children and 3.6 grandchildren, and 70% of them are married. About 22% are highly educated, while about 36% have achieved a medium level of education, and 42% are low educated.[14] To take additionally account of differential SES background of women, we control for childhood SES using a country-specific, standardized index.[15]

4 Discussion of instruments & assumptions

In this section, we discuss the assumptions necessary to hold for identification of the causal mediator model. We first investigate whether Z_2 is related to W using the share of compliers.

---

[14] Levels of education are based on ISCED-1997 classification. Categories 1-2: basic/lower secondary education; categories 3-4: (upper) secondary/post-secondary education; categories 5-6: tertiary education.

[15] We compute the index following Kesternich et al. (2014). It unifies four measures of SES at age 10 obtained from SHARELIFE: logged number of books in a household, logged number of rooms and persons in a household, features in a household, occupation of the main breadwinner. The index is obtained from principal component analyses for each country.

Table 3: Summary statistics, analytic cross-country sample

variables	mean	sd	min	max
outcome				
mental health, Euro-D scale	2.604	2.256	0	12
mental health, transformed	10.486	2.251	7.484	19.981
treatment and mediator				
woman is retired	0.480	0.500	0	1
intensity childcare	0.158	0.299	0	1
instruments				
woman's age $\geq$ statutory retirement age	0.537	0.499	0	1
ratio: nr daughters/all children	0.501	0.339	0	1
control variables				
age	64.59	7.27	50	80
age quadratic	4225	946	2500	6400
number children	2.586	1.209	1	17
number grandchildren	3.570	2.677	1	20
married	0.700	0.458	0	1
education: low	0.421			
education: medium	0.356			
education: high	0.222			
socioeconomic status during childhood ($<$ age 10)	0.000	0.746	-3.132	3.144
number of observations			11,386	

We moreover carry out a test to show that the instruments Z_1 and Z_2 are conditionally independent, and we illustrate that the common support assumption of M holds. We then

examine the validity of the instruments for the treatment, Z_1, and the mediator, Z_2. In particular, we discuss the corresponding exclusion restrictions and conduct empirical tests on instrumental variable (IV) validity.

4.1 Conditional independence and common support

One requirement of Assumption 1 is that Z_2 and W are independent conditional on X. As outlined in Section 2.2, the subpopulation or type T is a function of W and X. This implies that Z_2 and the probability of complying should be unrelated if Assumption 1 holds. Since the complier subpopulation, $T = co$, is identified in our model, we can test this hypothesis by regressing the probability being complier on Z_2 and X. We find small and insignificant association of 0.007 (p-value = 0.722) between the complier probability and Z_2. Since the complier type T is a function of W and X, the conditional correlation between Z_2 and D should also be close to zero. An OLS regression of D on Z_2, conditioning on Z_1 and X, estimates a coefficient of -0.002 and a p-value of 0.830.[16] Together, these findings support the validity of Assumption 1 in our application.

Assumption 2 postulates that the instruments Z_1 and Z_2 are independent given covariates X. This rules out that the instruments are shifting endogenous variables other than the intended ones. To assess potential violations of Assumption 2, we regress the instrument for the treatment, Z_1, on the instrument for the mediator, Z_2, and covariates X using OLS (Frölich and Huber, 2017). The correlation between the instruments Z_1 and Z_2 is literally zero (0.000, p-value = 0.999). Thus, we do not reject the conditional independence of the instruments, as stated by Assumption 2.

Assumption 5 is the common support of M. It rules out that grandparental childcare, covariates, and unobservables jointly determine the age thresholds of statutory retirement age. It is equivalent to assuming that for all M, X, and V, there always exist some observations on both sides of the age threshold of statutory retirement age. We investigate the common

---

[16]p-values in both regressions are based on standard errors clustered at the country-birth cohort level.

support of M by plotting women's age against the amount of grandparental childcare and all covariates. Figure 1 illustrates the joint distribution of age and grandparental care using the example of Belgium.[17] The solid horizontal line represents the statutory retirement age during the observation period. The figure shows that for all values of grandparental care, there are observations above and below the statutory retirement age, supporting the common support of M.

Figure 1: Joint distribution of grandparental care and age around statutory retirement age, Belgian subsample
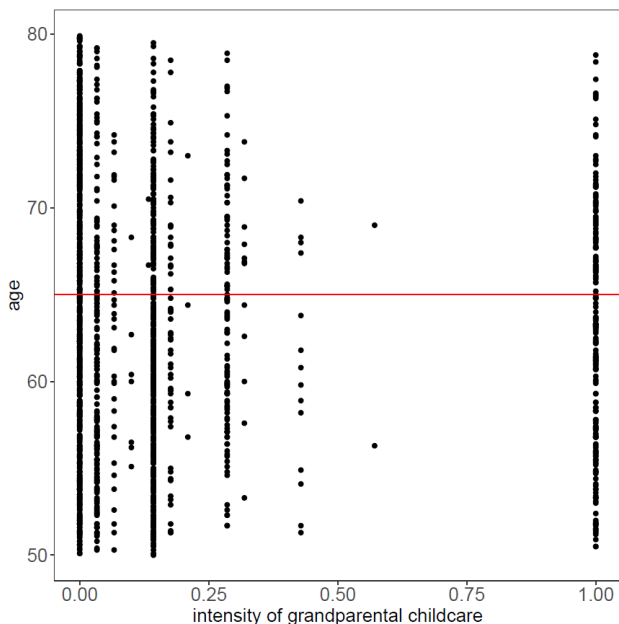
Horizontal red lines represent the age cut-off for statutory retirement age (65) in Belgium for the year 2011-2017.

4.2 Validity of statutory retirement age (Z_1) as instrument for retirement (D)

To address concerns with endogeneity in retirement, we instrument a woman's retirement decision with the statutory retirement age by country, age, and year. As discussed in Gruber and Wise (2009), retirement behavior responds very strongly to incentives set by social

---

[17]Figures for all countries and all covariates can be found in Appendix A.

security pension systems. Since such policies are defined on the national level and outside of individual control, they can provide exogenous variation to individual retirement decisions. It is therefore unlikely that mental health shows discontinuities around retirement eligibility ages that can be attributed to reasons other than retirement decisions.[18]

To examine the validity of Z_1 more formally, we conduct two IV validity tests suggested by Huber and Mellace (2014) and Kitagawa (2015). These methods jointly test whether (a) Z_1 is independent of unobservables U and V given covariates X (implied by Assumptions 1 and 2), and (b) whether D is monotonic in Z_1 (Assumption 3). Both tests are based on a set of inequalities that must hold under the null hypothesis of IV validity. Huber and Mellace (2014) define a set of testable inequalities on mean potential outcomes; Kitagawa (2015) derives distributional inequalities, allowing for more testable alternatives.[19] Given the null hypothesis of the tests they cannot verify but only reject IV validity.

Panels A and B in Table 4 present the results of the IV validity tests for Z_1 with respect to the outcome Equation (1) and the mediator Equation (2). Columns (1)-(3) are p-values obtained from the test of Huber and Mellace (2014). Column (4) shows the p-values from the test by Kitagawa (2015) using a trimming constant of $\xi = 0.075$.[20] According to the p-values shown in Table 4, there is no evidence against the validity of Z_1 at conventional levels of significance, at least not for violations in assumptions the tests cover.

---

[18]One concern with the instrument could be that health insurance benefits are correlated with retirement schemes. Since European health insurance benefits are not contingent on age, this is not an issue here.

[19]The idea of Huber and Mellace (2014) is that the mean potential outcome of always takers under treatment are point identified under LATE assumptions. For this observed conditional mean, one can derive upper and lower bounds which correspond to the relative share of always takers and compliers in mixed populations. If the instrument is valid, all point identified potential outcomes must lie within these bounds. The authors derive a set of inequalities that must hold jointly under IV validity and can be tested using data. The idea of Kitagawa (2015) is that the conditional density for the treatment outcome under $Z = 1$ must nest the respective density under $Z = 0$ (same applies for control outcome). This allows deriving a set of inequalities that have to hold under IV validity. Since the conditional densities are observed, one can test the inequalities using a variance-weighted Kolmogorov-Smirnov test statistic.

[20]The user-specific trimming constant, ξ, determines the weight given to the variance in the test statistic. Kitagawa (2015) recommends to specify ξ to be 0.05-0.1. We additionally perform the test for $\xi = 0.05$ and $\xi = 0.1$. The corresponding p-values for Panel A and B are $(0.501, 0.304)$ for both values of ξ. The p-value in Panel C is 0.595 for both values of the trimming constant.

Table 4: Results from IV validity tests

	Huber and Mellace (2014)			Kitagawa (2015)
	(1)	(2)	(3)	(4)
A. validity of instrument Z_1 in outcome equation, Y				
p-value	1.000	1.000	0.500	0.501
B. validity of instrument Z_1 in mediator equation, M				
p-value	1.000	1.000	0.500	0.304
C. validity of instrument $\tilde{Z}_2$ in outcome equation, Y				
p-value	0.626	1.000	1.000	0.595

Tests are based on a sample of 11,386 observations. Columns (1)-(3): distribution of test statistic obtained from bootstrapping with 1000 replications. Column (1): minimum p-value (partial recentering), Bennett (2009); Column (2): minimum p-value (full recentering), Bennett (2009); Column (3): p-value using the smoothed indicator-based method by Chen and Szroeter (2014). Column (4): user-specific trimming constant, $\xi = 0.075$. Panel C: M is replaced by $\tilde{M}$; Z_1 is replaced by $\tilde{Z}_2$.

4.3 Validity of gender ratio in children (Z_2) as instrument for grandparental childcare (M)

To instrument a woman's decision on how much time to spend on childcare, we use the ratio of the number of her daughters to the total number of her children, Z_2. In the literature, the gender of the firstborn child is commonly used as an instrument for grandparental childcare (Rupert and Zanella, 2018). It is based on the assumption that a child's gender is as good as random, and that parents of firstborn girls become grandparents earlier than parents of firstborn boys. Thus, the gender of the firstborn child provides exogenous variation to the timing of becoming grandparents. Our instrument is slightly different but follows a similar rationale: a higher number of daughters on average increases the probability of becoming a grandmother at an earlier age. Thus, more childcare is provided to daughters' children relative to sons' children. We investigate this mechanism, by regressing the age of the

youngest grandchild on the instrument using OLS.[21] As shown by Column (1) in Table A.2 (Appendix A), the youngest grandchild is almost nine months older if all children are daughters compared to all children being sons. For an equal share of daughters and sons, a grandchild is on average a bit more than four months older.

There could be several reasons why the exclusion restriction for our instrument is violated, i.e. that a higher number of daughters relative to all children leads to better mental health. First, if women may have stronger preferences for sons, then more daughters could reduce a woman's mental health. However, there is little evidence for gender preferences, at least in industrialized, non-traditional societies (see Lundberg, 2005). Another violation of the exclusion restriction occurs if more daughters increase divorce rates and marital dissatisfaction, leading to worse mental health (Katzev et al., 1994; Dahl and Moretti, 2008). To investigate this channel, we regress the probability of having ever been divorced on the instrument. The estimated association is close to zero and not statistically significant (see Column (2) of Table A.2, Appendix A).

The exclusion restriction could also be violated for reasons underlying residential patterns between women and their children. As documented by Compton and Pollak (2015) for the US, daughters live closer to their mothers than sons. If gender-specific proximity correlates with mental health, then the exclusion restriction may be violated. Column (5) in Table A.2 in Appendix A, shows that the correlation between a woman's average living distance to her children and the number of daughters relative to all children is close to zero (0.007) and not statistically significant.

To additionally rule out that the instrument has affected the mental health of women through these (and other) channels in the past, we moreover regress mental health on the instrument for a subsample of 1,852 women who were observed in SHARE before they became grandmothers. As indicated by Column (4) of Table A.2 (Appendix A), the number of daughters relative to all children does not significantly affect mental health before women became

---

[21]In all checks we perform in this Section, we control for a full set of covariates.

grandmothers. In sum, these exercises make us confident that the instrument exogenously shifts the intensity of childcare provided by grandmothers.

We finally apply the IV validity tests of Huber and Mellace (2014) and Kitagawa (2015) to the instrument Z_2. By doing so, we test implications of Assumptions 1, 2, and 4(b). Since both tests are available only for binary instruments and endogenous variables, we dichotomize M using the median as threshold (Frölich and Huber, 2014). Moreover, we define $\tilde{Z}_2$ as a binary indicator for Z_2 which takes the value 1 if a woman has at least as many daughters as sons, and is 0 otherwise. Panel C in Table 4 show the p-values from the two IV validity tests. Again, there is no evidence against the validity of Z_2 at conventional levels of significance.

5   Results

In this section, we discuss our main estimation results. We first present the semi-parametrically estimated mean potential outcomes as well as the direct and indirect effects. We then compare our results to different parametric specifications. Finally, we present the estimates obtained when stratifying the sample by countries with high and low enrollment rates in formal center-based childcare.

5.1   Estimates of counterfactuals and treatment effects

Table 5 presents the estimated potential outcomes and the corresponding treatment effects among compliers (see Table 1 for definitions).[22] For given childcare intensity, women who are not retired have an average depression score of 10.36 ($\mathrm{E}[Y^{0,M^0}|T = co]$). Women who are retired have a somewhat lower score of 10.26 ($\mathrm{E}[Y^{1,M^1}|T = co]$).

The estimated counterfactual outcomes are $\mathrm{E}[Y^{1,M^0}|T = co]$ and $\mathrm{E}[Y^{0,M^1}|T = co]$. The former denotes the average level of mental health for women who provide grandparental

---

[22]The estimated quantities are obtained by discarding observations which have a relative weight below the 0.25% or above the 99.75% quantile of the estimated distribution of relative weights. Table A.3 in Appendix A shows the distribution of discarded observations when a threshold of 0.25% is applied. When estimating each of the potential outcomes, 58 observations are discarded.

care as if they were not retired, keeping childcare fixed at its pre-retirement level. In this counterfactual, the average mental health score of 10.43. This score is higher than for both observed outcomes, implying that mental health in retirement is worst if changes in grandparental care are netted out.

The second counterfactual we consider is $E[Y^{0,M^1}|T = co]$, i.e. women are not retired but provide care to their grandchildren as if they were retired. In this counterfactual, women have an average mental health score of 9.68 – the lowest score among all potential outcomes. Our interpretation of this finding is that in this scenario women get mental health benefits from retirement-induced changes in grandparental childcare but without potential costs associated with retirement, e.g. a loss in income or social contacts.

The bottom line in Table 5 shows the direct effect (DATE) of retirement on mental health. Depending on the counterfactual scenario, the direct effect is either the effect of retirement on mental health when grandparental childcare is fixed at its pre-retirement level ($\hat{\theta}(0)$) or the effect with childcare fixed at its post-retirement level ($\hat{\theta}(1)$). The very right column in Table 5 shows the indirect treatment effect (IATE). Again, the estimated indirect effect depends on the chosen counterfactual, either being $\hat{\delta}(0)$ or $\hat{\delta}(1)$. The sum of $\hat{\delta}(0) + \hat{\theta}(1)$ or $\hat{\delta}(1) + \hat{\theta}(0)$ denotes the LATE estimate, It captures both retirement-induced changes in grandparental care (from M^0 to M^1) as well as changes in other potential channels (through a change from $D = 0$ to $D = 1$). The estimated LATE is -0.098 points and not statistically significant at the 5% level. This suggests that the overall impact of retirement on mental health is negative but close to zero. This result is similar to e.g. Coe and Zamarro (2011) who estimate a decrease in mental health of -0.069 points due to retirement.

To understand whether the LATE estimate is driven by the direct effect or by the indirect effect of retirement on mental health and for which counterfactual, we first consider the pair $(\hat{\theta}(1), \hat{\delta}(0))$. We estimate a large and significant direct effect. Retirement increases mental health problems by 0.58 points or 0.26 standard deviations if the grandparental care is fixed at its post-retirement level. This large negative impact is mainly driven by fewer mental

health problems in the counterfactual $E[Y^{0,M^1}|T = co]$. With actual retirement, channels other than grandparental care are activated. The interaction of such channels with post-retirement grandparental childcare amplifies their negative influence on mental health and causes in total a significant, negative direct effect.

Table 5: Semi-parametric estimation of mean potential outcomes, direct and indirect treatment effects and LATE

	$M = M^0$	$M = M^1$	IATE
$D = 0$	$E[Y^{0,M^0}|T = co]$	$E[Y^{0,M^1}|T = co]$	$\hat{\delta}(0)$
	10.362	9.682	-0.680
	[9.730, 10.628]	[6.972, 9.886]	[-2.858, -0.514]
$D = 1$	$E[Y^{1,M^0}|T = co]$	$E[Y^{1,M^1}|T = co]$	$\hat{\delta}(1)$
	10.430	10.264	-0.166
	[10.090, 11.505]	[10.151, 10.379]	[-1.218, 0.168]
DATE	$\hat{\theta}(0)$	$\hat{\theta}(1)$	LATE
	0.068	0.582	-0.098
	[-0.302, 1.710]	[0.384, 3.356]	[-0.358, 0.568]

95% asymmetric, equal-tailed confidence intervals in brackets, obtained from 4000 bootstrap replications clustered on country-birth cohort level. Estimates are based on a sample of 11,250 observations after trimming 0.25% of observations on both sides of the distribution Ω/Π in each potential outcomes. Counterfactual outcomes are estimated semi-parametrically. Control variables: age, age quadratic, marital status, education level, number of children, number of grandchildren, SES in childhood and binary missing value indicator for childhood SES, wave FE, and country FE.

The corresponding indirect effect, $\hat{\delta}(0)$, is negative and statistically significant on the 5% level. By keeping other factors of retirement fixed at their pre-retirement level, we estimate an indirect reduction in mental health problems of 0.68 points or 0.30 standard deviations. It implies that retirement may have large benefits for women's mental health. This positive role of grandparental childcare was also found in previous studies (e.g. Arpino et al., 2018). Grandparental childcare can provide a range of positive experiences, such as emo-

tional closeness and strengthened family ties, which alleviate feelings of isolation or loneliness. The estimated indirect effect $\hat{\delta}(0)$ is even larger than the corresponding direct effect, $\hat{\theta}(1)$. This suggests that mental health returns to retirement through grandparental care are large enough to compensate for the large negative impacts of retirement through all other channels.

Our estimation approach provides us with a second pair of direct and indirect effects, $(\hat{\theta}(0), \hat{\delta}(1))$. They are obtained for an alternative counterfactual in which the grandparental childcare is fixed on its level before retirement. $\hat{\theta}(0)$ indicates that retirement directly increases mental health problems by 0.068 points or 0.03 standard deviations at the sample mean. This estimated coefficient is only about 12% of the magnitude of $\hat{\theta}(1)$ and not statistically significant. It suggests that the interaction between pre-retirement levels of grandparental care and other retirement-induced channels is less detrimental for mental health than the interaction with post-retirement childcare levels. While we cannot determine which channels are responsible for the differences in effect sizes of $\hat{\theta}(0)$ and $\hat{\theta}(1)$, interactions with childcare at pre-retirement levels seems to reinforce positive aspects of the direct retirement effect and/or dampen negative aspects.

The indirect effect of retirement of mental health, $\hat{\delta}(1)$ is negative. A change in grandparental care due to retirement reduces mental health problems by 0.166 points or 0.07 standard deviations. This effect is not statistically significant on conventional levels of significance. Compared to $\hat{\theta}(0)$, the indirect effect, $\hat{\delta}(1)$, is more than twice as large and thus more than compensates a negative direct impact of retirement on mental health.

We investigate the robustness of our estimates by changing the trimming cut-off for influential observations to 0.2% and 0.3% respectively. As shown in Table A.4 the point estimates for the indirect effects are insensitive to changes in the trimming levels, while the magnitude of both direct effects changes somewhat. However, all estimated direct and indirect effects lie well within the respective estimated confidence intervals in Table 5. By comparing the estimated mean potential outcomes under different trimming, we find that differences in

point estimates are small.[23]

Our results allow us to draw the following conclusion. The pure benefits from grandparental care can be achieved if other channels that are (negatively) associated with retirement can be mitigated ($\hat{\delta}(0)$). If these other channels kick in though, they produce negative interactions in a way that the full benefits of retirement-induced changes in grandparental care are almost offset ($\hat{\delta}(1)$). Moreover, the negative direct effect of retirement on mental health fully kicks in when interactions with post-retirement grandparental childcare are incorporated ($\hat{\theta}(1)$). Such a negative effect can occur because either post-retirement grandparental care amplifies effects from negative channels, or it attenuates or substitutes effects from positive channels. By contrast, the corresponding direct retirement effect ($\hat{\theta}(0)$) is close to zero. This indicates that positive and negative channels almost offset each other when not triggered by retirement-related grandparental childcare.

5.2   Comparison to parametric estimation procedures

One main advantage of estimating a semi-parametric specification is that it allows for arbitrary effect heterogeneity through flexible interactions between all observables and unobservables. In the following, we compare our main results with two parametric specifications that do not allow for this amount of flexibility.

We first estimate a parametric IV model that addresses endogeneity in the treatment and in the mediator but does not allow for interactions between treatment and mediator (this model is similar to Powdthavee et al., 2015). The model is estimated in several steps. First, we regress D on $(1, Z_1, X)$ using a logistic specification to predict the treatment $\hat{D}$. In the second step we run an OLS regression of M on $(1, Z_2, \hat{D})$ and X to obtain the predicted values $\hat{M}$. Third, we regress the outcome Y on $(1, \hat{D}, \hat{M}, X)$ using OLS. The direct effect corresponds to the estimated coefficient on $\hat{D}$ in this regression. In a last step, we run an OLS

---

[23]We also assess a higher quantile for trimming (0.5%) and an alternative trimming rule suggested by Huber et al. (2013). In neither case, the point estimates of direct and indirect effects exceed the confidence intervals obtained for the main specification. The results are available upon request.

regression of M on $(1, \hat{D}, X)$. The estimated coefficient on $\hat{D}$ in this regression multiplied by the estimated coefficient on $\hat{M}$ in the regression of Y from the third step corresponds to the indirect treatment effect. The estimated coefficients obtained from this multistep estimation procedure are shown in Panel B in Table 6. Compared to the semi-parametric estimate (see Panel A in Table 6), the estimated direct effect of retirement on mental health is negative and not statistically significant. The estimated indirect effect is negative but small when compared to our semi-parametric results and not statistically significant. These findings suggest that ignoring effect heterogeneity may bias estimated direct and indirect treatment effects.[24]

The second parametric specification we are applying is a linear IV model that adds an interaction between the treatment and the mediator. To implement this model and to compute direct and indirect effects, we follow Chen et al. (2019). More specifically, we estimate a linear IV model, $Y = \beta_0 + \beta_1 D + \beta_2 M + \beta_3 D \times M + \gamma \boldsymbol{X} + U$, where $D \times M$ denotes the interaction between the treatment and the mediator. The three endogenous variables D, M, and $D \times M$ are instrumented by Z_1, Z_2 and $Z_1 \times Z_2$. The average direct effect $\theta(1)$ is computed as $\hat{\beta}_1 + \hat{\beta}_3 \, \mathrm{E}(M|D = 1)$. The average indirect effect $\delta(0)$ is obtained from $\hat{\beta}_2 [\mathrm{E}(M|D = 1) - \mathrm{E}(M|D = 0)]$.[25] The identification of the average direct and indirect effect is based on two properties: (a) the distribution of unobservables U is independent of the treatment D given M; and (b) the conditional means of M given treatment D are the same for observed and counterfactual mediator outcomes.

We estimate this model and compute the average direct and indirect effects at the sample mean of M conditional on treatment D. As can be seen in Table 6 the corresponding estimates are small and not statistically significant. Moreover, the direct and the indirect effect offset each other such that the estimated total average effect is literally zero. The magnitude of these estimates are considerably smaller than those we obtained from the

---

[24]As pointed out by Frölich and Huber (2017) the bias from multistep IV model is almost as large as for OLS when permitting effect heterogeneity.

[25]The estimated coefficients $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ can be found in Table A.5 of Appendix A

Table 6: Estimated direct, indirect and total effects of retirement on mental health comparing different empirical approaches

model specification	DATE θ		IATE δ		total effect
	$\hat{\theta}(1)$	$\hat{\theta}(0)$	$\hat{\delta}(1)$	$\hat{\delta}(0)$	
A. estimates from the semi-parametric model (Table 5)					
semi-parametric	0.582	0.068	-0.166	-0.680	-0.098
	[0.384, 3.356]	[-0.302, 1.710]	[-1.218, 0.168]	[-2.858, -0.514]	[-0.358, 0.568]
B. estimates from parametric models					
multistep IV, no interaction	-0.109		-0.076		-0.186
	[-0.448, 0.229]		[-0.200, 0.047]		[-0.498, 0.127]
linear IV with interaction $D \times M$	0.036		-0.036		0.000
	[-0.417, 0.488]		[-0.136, 0.064]		[-0.442, 0.442]
IV					-0.049
					[-0.398, 0.300]
OLS					-0.152
					[-0.248, -0.057]

95% confidence intervals in brackets; Semi-parametric, multistep IV, linear IV with interactions: confidence intervals obtained from 4000 bootstrap replications clustered on country-birth cohort level. LATE, OLS: analytic confidence intervals, standard errors clustered on country-birth cohort level. Panel A: point estimates obtained from sample of 11,250 observations after on both sides of the distribution Ω/Π in each potential outcome. Panel B: estimates obtained from a sample of 11,386 observations. IV model with interaction $D \times M$: conditional F-statistic for D, M, and $D \times M$ are (15.79, 15.89, 13.30). Control variables in all specifications: age, age quadratic, marital status, education level, number of children, number of grandchildren, SES in childhood and binary missing value indicator for childhood SES, wave FE, and country FE.

semi-parametric estimator. The main reason for these differences is that the two properties are unlikely to hold in our application. Essentially, these properties state that the treatment D is random. As discussed in Section 3, retirement decisions are endogenous, and retirement cannot be considered as random. Therefore, the linear IV model with interactions is not suitable for our application.[26]

---

[26]In Chen et al. (2019) the treatment is the gender of a child which arguably is random.

The final column in Table 6 shows the estimated LATE for the different specifications as well as the IV and OLS estimate of the total effect. Except for OLS, none of estimated total effects is statistically significant on conventional levels. The lowest estimated total effect is obtained from the linear IV model with an interaction, while the largest estimated coefficient is obtained when interactions between treatment and mediator are not permitted (multistep IV). Across all specifications, the estimated total effect from the semi-parametric specification is closest to the standard LATE estimate. Overall, the comparison of different empirical specifications highlights the importance to allow for arbitrary effect heterogeneity when one is interested in understanding why a treatment affects an outcome in a specific way and how a mediator contributes to this effect.

5.3   Childcare systems across countries

Our results suggest that grandparental childcare is a significant mechanism through which retirement affects the mental health of elderly women in Europe. Yet, our estimates may mask a considerable amount of heterogeneity in the importance of grandparental childcare as a mediator.

In some European countries, grandparents are often primary caregivers, while in other countries formal childcare is the major source of childcare. In northern Europe, such as Sweden or Denmark, public childcare is universally available. In southern European countries a combination of private and public childcare is most common, and grandparents are an important contributor. Del Boca (2015) shows that in Southern Europe twice as many grandparents take care of grandchildren than in Nordic countries. Also, about 30% of grandparents in Italy and Spain provide care daily compared to only about 2% in Denmark or Sweden. It implies that whether the negative effect of retirement on mental health can be compensated by grandparental care may depend on the institutional setting.

To examine the heterogeneity in the importance of grandparental childcare as a mediator, we define two groups of countries: countries in which formal public care is the most common

care arrangement; and countries with less utilization of public childcare. We define these two groups using country-specific OECD data on enrollment rates in public care for children under the age of two (Table A.6, Appendix A). Countries with high use of public childcare are the Netherlands, France, Belgium, Denmark, and Sweden. Countries with lower use of public care are Switzerland, Germany, Spain, Italy, and Austria.

Table 7: Estimated direct, indirect and total effects of retirement on mental health by groups of countries

DATE θ		IATE δ		total effect
$\hat{\theta}(1)$	$\hat{\theta}(0)$	$\hat{\delta}(1)$	$\hat{\delta}(0)$	
A. All countries (Table 5)				
0.582	0.068	-0.166	-0.680	-0.098
[0.384, 3.356]	[-0.302, 1.710]	[-1.218, 0.168]	[-2.858, -0.514]	[-0.358, 0.568]
B. Countries with developed formal childcare system				
0.013	-0.686	0.094	-0.606$^+$	-0.592
[-0.896, 0.672]	[-1.794, 0.142]	[-0.694, 0.999]	[-1.149, 0.082]	[-1.123, -0.188]
C. Countries with less developed formal childcare system				
0.881	0.592$^+$	0.105	-0.184$^+$	0.697
[0.547, 2.648]	[-0.049, 1.507]	[-0.457, 0.670]	[-1.712, 0.063]	[0.393, 1.316]

95% asymmetric, equal-tailed confidence interval in brackets obtained from 4,000 bootstrap replications clustered on country-birth cohort level. $^+$: statistical significance based on 90% confidence interval. Panel A: point estimates are based on 11,250 observations after trimming 0.25% on both sides of the distribution Ω/Π in each potential outcome. Panel B: estimates are based on 6,113 observations after trimming 0.25% on both sides of the distribution Ω/Π in each potential outcome; EURO-D score: sample mean=2.549, sd=2.193. Panel C: estimates are based on 5,133 observations after trimming in each potential outcome; EURO-D score: sample mean=2.670, sd=2.328. Control variables in all subsamples: age, age quadratic, marital status, education level, number of children, number of grandchildren, SES in childhood and binary missing value indicator for childhood SES, wave FE, and country FE.

The estimated direct and indirect effects are presented in Table 7. Panel B shows the results for countries with high use of public childcare. We find a negative and statistically significant

total effect. Retirement decreases mental health problems by 0.592 points or 0.27 standard deviations. Depending on the decomposition, this total effect is either driven by the direct effect, $\hat{\theta}(0)$, or the indirect effect, $\hat{\delta}(0)$, of retirement on mental health. In states of the world in which either grandparental childcare is provided as if not being retired, or all other channels are fixed at their pre-retirement level, women benefit in their mental health from retirement. This result is in line with finding by Arpino et al. (2018): grandparenthood has a stronger positive association with subjective well-being in countries where intensive grandparental childcare is not common and less socially expected. $\hat{\theta}(1)$ and $\hat{\delta}(1)$ are the estimated direct and indirect effects of retirement on mental health if channels are fixed at their post-retirement level. Both estimated effects are close to zero, suggesting that the complex interactions between post-retirement levels of grandparental childcare and all other channels net out each other.

Panel C presents the estimated effects of retirement on mental health for countries with lower rates of public childcare utilization. In these countries, retirement leads to an increase in mental health problems by 0.69 points or 0.3 standard deviations at the sample mean. This negative total effect seems to be entirely driven by large, negative direct effects. By contrast, both estimated indirect effects are small. Consider, for instance, the estimated pair $(\hat{\theta}(1), \hat{\delta}(0))$. Retirement directly leads to an increase in mental health problems by 0.881 points or 0.38 standard deviations, while the corresponding indirect effect is small, albeit negative and statistically significant on the 10% level. It suggests that the benefits from retirement generated through a change in grandparental care cannot compensate for the large losses in mental health through all other channels induced by retirement.

According to Table 7, the mediating effect of grandparental care strongly depends on the utilization of formal childcare. In countries with low rates of formal care utilization grandmothers may have the role of primary caregivers before or after retirement while parents work. In countries in which formal childcare is most common, grandparental care is not

essential for parent's labor supply (for a discussion, see Del Boca, 2015).[27] An implication could be that grandparental childcare must not be too demanding upon retirement to offset the negative effect of retirement on mental health among women.

6 Conclusion

People experience multiple changes in their lives after retirement which in turn can affect their mental health. In this paper, we examine how grandparental childcare contributes to the effect of retirement on mental health. We use a nonparametric identification approach and semi-parametric estimation strategy to disentangle the total effect of retirement on mental health into a mediating, indirect effect of grandparental care and a direct retirement effect operating through other mechanisms among women in Europe. We find that retirement directly increases mental health problems but this negative impact is offset by a significant reduction in mental health problems generated through retirement-induced changes in grandparental childcare. As a result, the overall effect of retirement on mental health is small and insignificant – a finding that is in line with other studies estimating local effects of retirement on mental health. To understand the importance of the mediating effect of grandparental care in more detail, we then examine the direct, indirect, and total effect by country groups. The results imply that whether grandparental care is a supplement or a requirement for parent's labor supply crucially depends on whether the mediating effect unfolds its full compensating strength or not.

The results of our study have important implications for understanding how retirement affects mental health. By disentangling the causal mechanism of grandparental childcare from other mechanisms of retirement, we can show under which conditions retirement has positive or negative effects on mental health. Also, we illustrate how important the grandparental care channel is relative to the sum of other channels. The relative magnitude of such effects

---

[27]In our sample women from low utilization countries provide about 60 % more care to their grandchildren than women from countries with high utilization rates.

can be used to efficiently design social security policies for retirement.

Our findings also suggest that informal childcare could be more beneficial if it is a supplement to center-based formal childcare rather than replacing formal childcare. Only in countries where grandmothers are not too engaged in childcare, negative effects of retirement on mental health can be mitigated. This also has implications for policymakers when designing childcare policies. In countries in which public childcare is restricted parents heavily rely on grandparents to participate in the labor market. However, this may come at the cost of their mental health around retirement. Thus, policymakers should take into account the costs associated with mental health problems among grandparents when imposing restrictions on public childcare. Providing more formal childcare could allow for both an increase in maternal labor supply and an enhancement in mental health of grandmothers.

References

ARPINO, B., V. BORDONE, AND N. BALBO (2018): "Grandparenting, education and subjective well-being of older Europeans," European Journal of Ageing, 15, 251–263.

ATALAY, K., G. F. BARRETT, AND A. STANEVA (2019): "The effect of retirement on elderly cognitive functioning," Journal of Health Economics, 66, 37–53.

BEHNCKE, S. (2012): "Does retirement trigger ill health?" Health Economics, 21, 282–300.

BELLONI, M., E. MESCHI, AND G. PASINI (2016): "The effect on mental health of retiring during the economic crisis," Health Economics, 25, 126–140.

BENNETT, C. J. (2009): "Consistent and asymptotically unbiased minP tests of multiple inequality moment restrictions," Vanderbilt University Department of Economics Working Papers, 908.

BONSANG, E., S. ADAM, AND S. PERELMAN (2012): "Does retirement affect cognitive functioning?" Journal of Health Economics, 31, 490–501.

BÖRSCH-SUPAN, A., M. BRANDT, C. HUNKLER, T. KNEIP, J. KORBMACHER, F. MALTER, B. SCHAAN, S. STUCK, AND S. ZUBER (2013): "Data resource profile: the Survey of Health, Ageing and Retirement in Europe (SHARE)," International Journal of Epidemiology, 42, 992–1001.

BRUNELLO, G. AND L. ROCCO (2019): "Grandparents in the blues. The effect of childcare on grandparents' depression," Review of Economics of the Household, 17, 587–613.

CASTRO-COSTA, E., M. DEWEY, R. STEWART, S. BANERJEE, F. HUPPERT, C. MENDONCA-LIMA, C. BULA, F. REISCHES, J. WANCATA, K. RITCHIE, ET AL. (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, 12–29.

CHARLES, K. K. (2004): "Is retirement depressing?: Labor force inactivity and psychological well-being in later life," Research in Labor Economics, 23, 269–299.

CHEN, L.-Y. AND J. SZROETER (2014): "Testing multiple inequality hypotheses: A smoothed indicator approach," Journal of Econometrics, 178, 678–693.

CHEN, S. H., Y.-C. CHEN, AND J.-T. LIU (2019): "The impact of family composition on educational achievement," Journal of Human Resources, 54, 122–170.

CHEN, Y.-T., Y.-C. HSU, AND H.-J. WANG (2020): "A stochastic frontier model with endogenous treatment status and mediator," Journal of Business & Economic Statistics, 38, 243–256.

COE, N. B. AND G. ZAMARRO (2011): "Retirement effects on health in Europe," Journal of Health Economics, 30, 77–86.

COMPTON, J. AND R. A. POLLAK (2015): "Proximity and co-residence of adult children and their parents in the United States: Descriptions and correlates," Annals of Economics and Statistics, 117/118, 91–114.

DAHL, G. B. AND E. MORETTI (2008): "The demand for sons," The Review of Economic Studies, 75, 1085–1120.

DAVE, D., I. RASHAD, AND J. SPASOJEVIC (2008): "The effects of retirement on physical and mental health outcomes," Southern Economic Journal, 75, 497–523.

DE GRIP, A., M. LINDEBOOM, AND R. MONTIZAAN (2012): "Shattered dreams: The effects of changing the pension system late in the game," Economic Journal, 122, 1–25.

DEL BOCA, D. (2015): "Child care arrangements and labor supply," Tech. rep., IDB Working Paper Series.

DEL BOCA, D., M. LOCATELLI, AND D. VURI (2005): "Child-care choices by working mothers: The case of Italy," Review of Economics of the Household, 3, 453–477.

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

FE, E. AND B. HOLLINGSWORTH (2016): "Short- and long-run estimates of the local effects of retirement on health," Journal of the Royal Statistical Society - Series A, 179, 1051–1067.

FRÖLICH, M. AND M. HUBER (2014): "Treatment Evaluation With Multiple Outcome Periods Under Endogeneity and Attrition," Journal of the American Statistical Association, 109, 1697–1711.

FRÖLICH, M. AND M. HUBER (2014): "Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables," IZA Discussion Paper, 8280.

——— (2017): "Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables," Journal of the Royal Statistical Society. Series B: Statistical Methodology, 79, 1645–1666.

GELMAN, A. AND G. IMBENS (2013): "Why ask why? Forward causal inference and reverse causal questions," NBER working paper, 19614.

GRUBER, J. AND D. A. WISE (2009): Social security programs and retirement around the world: Micro-estimation, University of Chicago Press.

HELLER-SAHLGREN, G. (2017): "Retirement blues," Journal of Health Economics, 54, 66–78.

HUBER, M. (2020): "Mediation analysis," in Handbook of Labor, Human Resources and Population Economics, ed. by K. F. Zimmermann, Springer.

HUBER, M., M. LECHNER, AND C. WUNSCH (2013): "The performance of estimators based on the propensity score," Journal of Econometrics, 175, 1–21.

HUBER, M. AND G. MELLACE (2014): "Testing instrument validity for LATE identification based on inequality moment constraints," Review of Economics and Statistics, 97, 398–411.

IMAI, K., L. KEELE, D. TINGLEY, AND T. YAMAMOTO (2011): "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies," American Political Science Review, 105, 765–789.

INSLER, M. (2014): "The health consequences of retirement," Journal of Human Resources, 49, 195–233.

JOHNSTON, D. W. AND W. LEE (2009): "Retiring to the good life? The short-term effects of retirement on health," Economics Letters, 103, 8–11.

KATZEV, A. R., R. L. WARNER, AND A. C. ACOCK (1994): "Girls or boys? Relationship of child gender to marital instability," Journal of Marriage and the Family, 89–100.

KESTERNICH, I., B. SIFLINGER, J. P. SMITH, AND J. K. WINTER (2014): "The effects of World War II on economic and health outcomes across Europe," Review of Economics and Statistics, 96, 103–118.

KITAGAWA, T. (2015): "A Test for instrument validity," Econometrica, 83, 2043–2063.

KOLODZIEJ, I. W. AND P. GARCÍA-GÓMEZ (2019): "Saved by retirement: Beyond the mean effect on mental health," Social Science & Medicine, 27, 85–97.

KU, L. J. E., S. C. STEARNS, C. H. VAN HOUTVEN, AND G. M. HOLMES (2012): "The health effects of caregiving by grandparents in Taiwan: An instrumental variable estimation," Review of Economics of the Household, 10, 521–540.

LUNDBERG, S. (2005): "Sons, daughters, and parental behaviour," Oxford Review of Economic Policy, 21, 340–356.

MAZZONNA, F. AND F. PERACCHI (2012): "Ageing, cognitive abilities and retirement," European Economic Review, 56, 691–710.

——— (2017): "Unhealthy retirement?" Journal of Human Resources, 52, 128–151.

PICCIO, M. AND J. C. VAN OURS (2019): "Mental health effects of retirement," De Economist, 168, 419–452.

POWDTHAVEE, N., W. LEKFUANGFU, AND M. WOODEN (2015): "What's the good of education on our overall quality of life? A simultaneous equation model of education and life satisfaction for Australia," Journal of Behavioral and Experimental Economics, 54, 10–21.

ROHWEDDER, S. AND R. J. WILLIS (2010): "Mental retirement," Journal of Economic Perspectives, 24, 119–138.

RUPERT, P. AND G. ZANELLA (2018): "Grandchildren and their grandparents' labor supply," Journal of Public Economics, 159, 89–103.

ZANELLA, G. (2017): "How does grandparent childcare affect labor supply?" IZA World of Labor.

Online Appendix

A   Additional Tables and Figures

Tables

Table A.1: Statutory retirement ages for women across countries and years

	year							
country	2010	2011	2012	2013	2014	2015	2016	2017
Austria	60	60	60	60	60	60	60	60
Belgium	65	65	65	65	65	65	65	65
Denmark	65	65	65	65	65	65	65	65
France	60	60.3	60.6	60.9	61.2	61.4	61.6	62
Germany	65	65	65.1	65.2	65.3	65.3	65.4	65.5
Italy	60	60	62	62.3	62.3	63.8	65.6	65.6
Netherlands	65	65	65	65	65	65.3	65.5	65.8
Spain	65	65	65	65.1	65.2	65.3	65.3	65.4
Sweden	65	65	65	65	65	65	65	65
Switzerland	64	64	64	64	64	64	64	64

Data sources: `https://tradingeconomics.com` and `https://www.ssa.gov/policy/docs/progdesc/ssptw/`. Sweden has a flexible retirement age scheme, age 65 is the guaranteed pension age.

Table A.2: Tests on exclusion restriction for the mediator instrument, Z_2

	age youngest grandchild	ever been divorced	living distance > 25 km	mental health
ratio: nr daughters/all children	0.728***	0.003	0.007	-0.127
	(0.108)	(0.008)	(0.012)	(0.152)

*** p<0.01, ** p<0.05, * p<0.1; standard errors clustered at country-birth cohort level in parentheses; Column (1): OLS regression of age of youngest grandchild on number of daughters relative to all children. Column (2): OLS regression of ever been divorced on number of daughters relative to all children. Estimates in Column (1)-(2) are based on a sample of 11,386 observations. Column (3): OLS regression of woman's living distance being > 25 km from nearest child on number of daughters relative to all children for a sample of 11,386 observations. Column (4): OLS regression of original Euro-D score on number of daughters relative to all children based on a subsample of 1,852 women; mental health is measured before women became grandmothers. Control variables in all specifications: age, age quadratic, marital status, wave dummy, education level, country dummy, number of children, number of grandchildren, SES in childhood, a binary missing value indicator for childhood SES, wave FE, and country FE.

Table A.3: Distribution of estimated relative weights for observations discarded in the estimation of treatment effects based on quantile-specific trimming level

trimming level: 0.25% on both sides of the estimated distribution of the ratio Ω/Π

	$E[Y^{0,M^0}]$	$E[Y^{1,M^0}]$	$E[Y^{0,M^1}]$	$E[Y^{1,M^1}]$
minimum value	-8.52	-147.29	-18.05	-156.61
trimming cutoff, 0.25%	-4.25	-8.06	-6.05	-5.97
trimming cutoff, 99.75%	15.36	11.02	18.01	5.66
maximum value	391.64	13485.80	5548.66	16.15
# obs. trimmed	58	58	58	58

Sample size $N = 11,386$; extreme weights are identified by exceeding the level a on both sides of the estimated distribution of the ratio obtained from Equation (7).

Table A.4: Estimated direct, indirect and total effects of retirement on mental health trimming at quantiles 0.2% and 0.3%

DATE θ		IATE δ		total effect
$\hat{\theta}(1)$	$\hat{\theta}(0)$	$\hat{\delta}(1)$	$\hat{\delta}(0)$	
A. trimming level: 0.2% of estimated distribution of the ratio Ω/Π				
0.433	-0.025	-0.195	-0.653	-0.22
[0.183, 3.275]	[-0.429, 1.656]	[-1.268, 0.172]	[-2.902, -0.436]	[-0.490, 0.489]
B. trimming level: 0.3% of estimated distribution of the ratio Ω/Π				
0.681	0.128	-0.121	-0.674	0.007
[0.505, 3.397]	[-0.249, 1.725]	[-1.143, 0.224]	[-2.793, -0.504]	[-0.223, 0.660]

95% confidence intervals clustered on country-birth cohort level in parentheses. Compared to results in Table 5, more observations are discarded for 0.3%; less observations are trimmed for 0.2%.

Table A.5: Estimated coefficients from linear IV model with additive interaction between treatment and mediator

	D	M	$M \times D$	constant
estimated coefficients	1.080	2.203	-6.983*	7.355**
	(0.702)	(2.131)	(4.239)	(2.941)

*** p<0.01, ** p<0.05, * p<0.1; standard errors clustered on country-birth cohort level in parentheses; estimated coefficients from a linear IV model with additive interactions with mental health as outcome: $Y = \beta_0 + \beta_1 D + \beta_2 M + \beta_3(D \times M) + \gamma \boldsymbol{X} + U$; estimation based on a sample of 11,386 observations. Control variables: age, age quadratic, marital status, wave dummy, education level, country dummy, number of children, number of grandchildren, SES in childhood, a binary missing value indicator for childhood SES, wave FE and country FE.

Table A.6: Enrollment rates in early childhood education and care services, children below the age of two, 2017

country	NL	FR	BE	DK	SE	CH	DE	ES	IT	AT
enrollment rate (%)	59.3	56.3	56.1	55.4	46.6	38.0	37.2	36.4	29.7	21.0

source: `https://www.oecd.org/els/soc/PF3_2_Enrolment_childcare_preschool.pdf`, OECD Family Database (last update: 05-11-2019). Enrollment rates in public childcare arrangements for children aged 0–2 years.

Figures

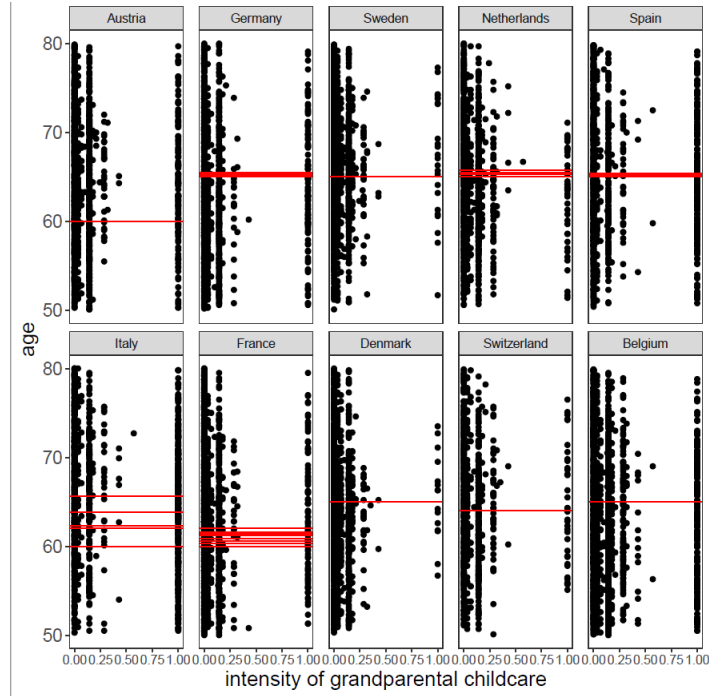Figure A.1: Joint distribution amount grandparental care and age, around statutory retirement ages

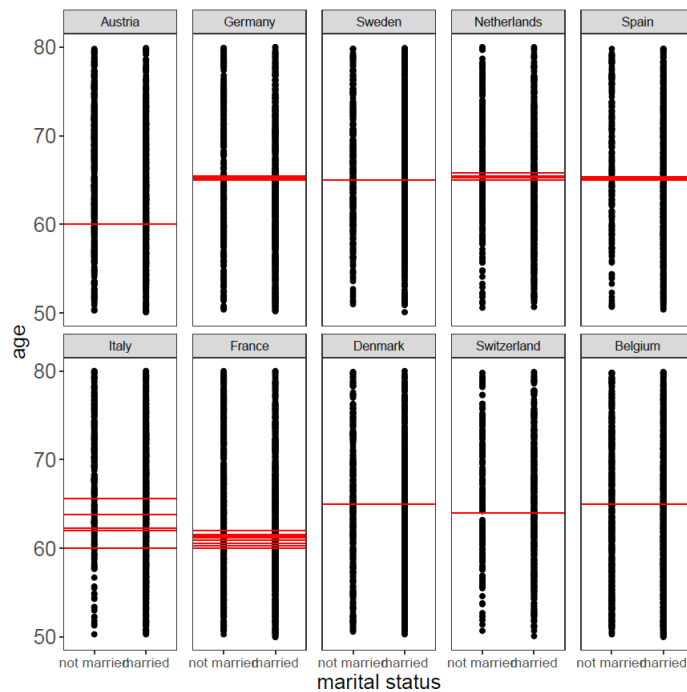Figure A.4: Joint distribution marital status and age, around statutory retirement ages

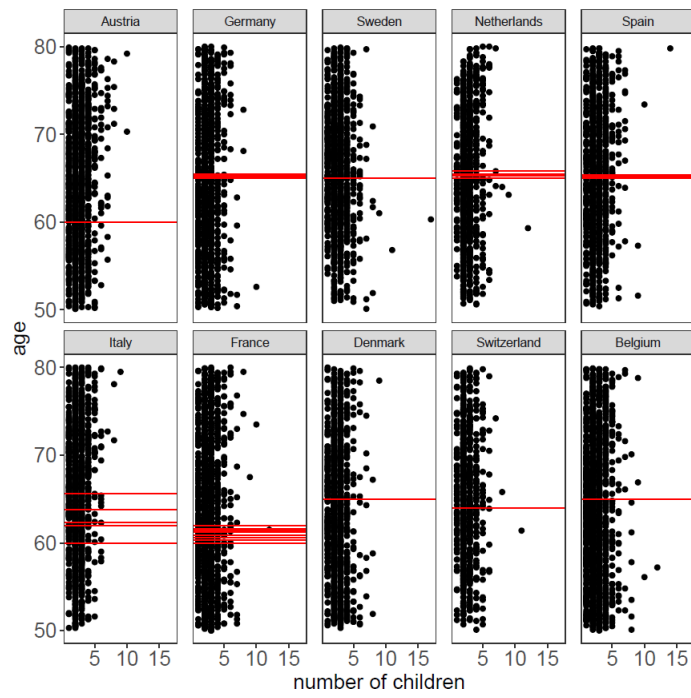Figure A.2: Joint distribution number of children and age, around statutory retirement ages

Figure A.5: Joint distribution level of edcuation and age, around statutory retirement ages
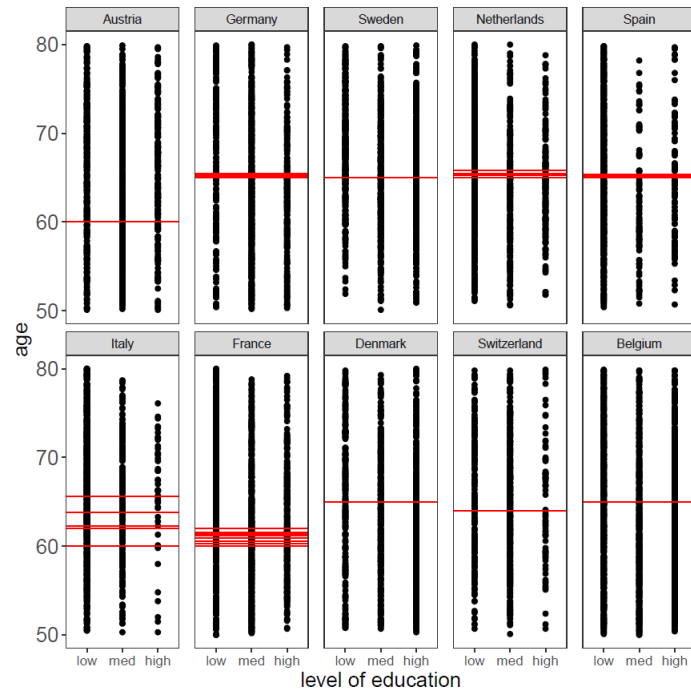
Figure A.3: Joint distribution number of grandchildren and age, around statutory retirement ages
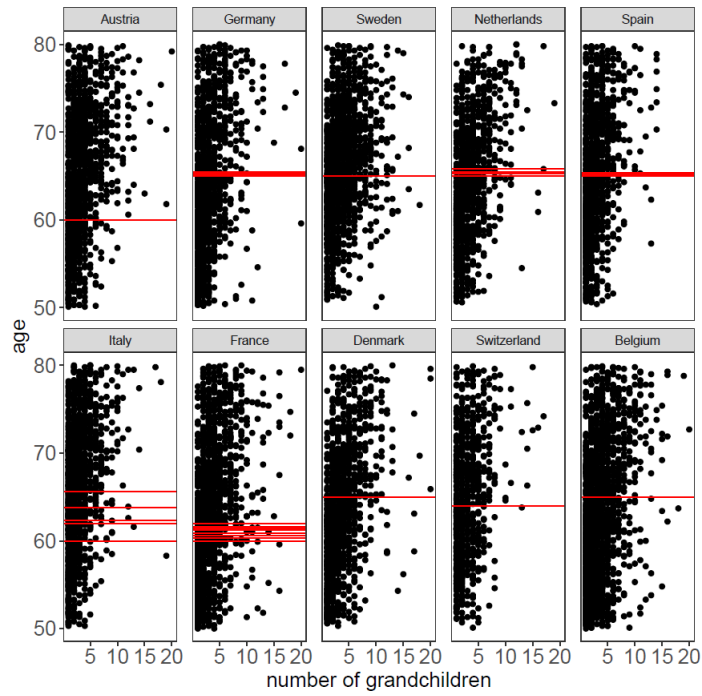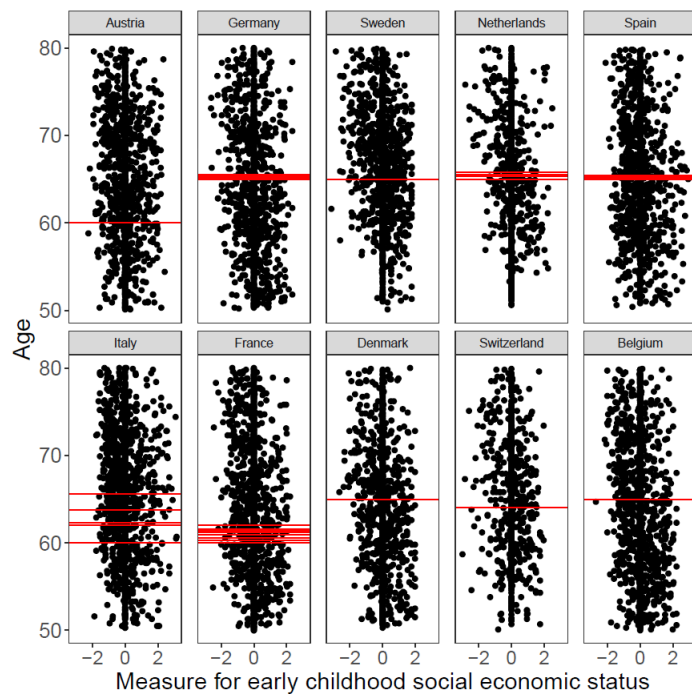

Figure A.6: Joint distribution childhood SES and age, around statutory retirement ages

B Estimation procedure and proofs

B.1 Estimation of propensity score Π, compliers subpopulation, and weight Ω

The following subsection provides an overview over single estimation steps.

1. Estimate the propensity score Π_i Given $\Pi = Pr(Z_1 = 1|X)$, we use the fitted value $\bar{\Pi}$ from a logistic regression of Z_1 on X to estimate Π.

2. Obtain the ratio of compliers in the population $Pr(\text{compliers})$

$$Pr(\text{compliers}) = \mathrm{E}[\frac{D(Z_1 - \bar{\Pi})}{\bar{\Pi}(1 - \bar{\Pi})}] \tag{B.1}$$

   Since we know D_i, Z_{1_i}, and $\bar{\Pi}_i$, we can use the sample average of the right hand side of Equation (B.1) to estimate $Pr(\text{compliers})$. The intuition behind Equation (B.1) is that $D = 1$ can occur when the respondent is either complier with $Z_1 = 1$ or always-taker. In the case of being complier with $Z_{1_i} = 1$, the right hand side of Equation (B.1) equals 1. In the case of being always-taker, the right hand side of Equation (B.1) equals 0. Based on this, the right hand side of the equation is $Pr(\text{compliers})$, and the equation holds.

3. The weighting function, Ω, is given by

$$\Omega = 1 - \frac{\mathrm{E}[Z_1|M, C, X] - \bar{\Pi}}{\mathrm{E}[DZ_1|M, C, X] - \mathrm{E}[D|M, C, X]\bar{\Pi}} \tag{B.2}$$

   where $C = C(M, D, Z_2, X)$ defines the control function

$$C(m, d, z_2, x) = \tag{B.3}$$

$$\frac{\mathrm{E}[(d + D - 1)[Z_1 - \bar{\Pi}]|M \le m, Z_2 = z_2, X = x]}{\mathrm{E}[D[Z_1 - \bar{\Pi}]|Z_2 = z_2, X = x]} F_{M|Z_2,X}(m, z_2, x)$$

that identifies V (see Frölich and Huber (2017) for the proof). We obtain conditional expectations of binary variables, Z_1, DZ_1, and D, in Equation (B.2), by using fitted values from logit models. Conditional expectations of $(d + D - 1)[Z_1 - \bar{\Pi}]$ and $D[Z_1 - \bar{\Pi}]$ in Equation (B.3), come from OLS fitted values. The conditional CDF, $F_{M|Z_2,X}(m, z_2, x)$, is obtained from a nonparametric kernel regression with Gaussian kernel.

4. Trimming rules

Estimated relative weights can be extremely large if the estimated propensity score Π or the estimated Ω is close to 0 or 1. To trim too influential observations, we specify the following trimming rule based on quantiles of the estimated distribution of relative weights. We discard observation i if the relative weight, $\hat{w}_i$, is either $> \hat{w}_{1-p}$ or $< \hat{w}_p$ where $\hat{w}_a$ is the a_{th} quantile of the estimated distribution of $\hat{w}$. We choose $a = 0.0025$ as the trimming level for our main specification and $a = 0.002$ as well as $a = 0.003$ to examine robustness.

B.2 Equivalence of coefficients of transformed and original outcome

To maintain assumption 5, we use a transformed version of the Euro-D score, $Y' \equiv Y - \kappa(\text{Age})$, which captures age-adjusted variation in mental health. Using this transformed version does not changes the magnitude of the effects compared to using the original Euro-D scale, since it can be shown that the coefficients remain unchanged. For example, for the

direct treatment effect $\theta(1)$ we obtain

$$\mathrm{E}[Y'^{D=1,M^{D=1}} - Y'^{D=0,M^{D=1}}]$$

$$= \mathrm{E}[\phi(D=1,\ M=M^{D=1},\ X,\ U)] - \mathrm{E}[\phi(D=0,\ M=M^{D=1},\ X,\ U)]$$

$$= \mathrm{E}[\phi(D=1,\ M=M^{D=1},\ X,\ U)] + \mathrm{E}[\kappa(\mathrm{Age})]$$

$$\quad - (\mathrm{E}[\phi(D=0,\ M=M^{D=1},\ X,\ U)] + \mathrm{E}[\kappa(\mathrm{Age})])$$

$$= \mathrm{E}[Y^{D=1,M^{D=1}} - Y^{D=0,M^{D=1}}]$$

Thus, using the transformed outcome gives the same coefficients as using the original outcome.