

Will you still need me, will you still feed me when I'm 64?

The health impact of caregiving

P.L. de Zwart, P. Bakx, and E.K.A. van Doorslaer

Will you still need me, will you still feed me when I'm 64?

The health impact of caregiving

P.L. de Zwart¹, P. Bakx², E.K.A. van Doorslaer^{2,3,4}

Abstract

Informal care may substitute for formal long-term care that is often publicly funded or subsidized. The costs of informal caregiving are borne by the caregiver and may consist of worse health outcomes and, if the caregiver has not retired, worse labor market outcomes. We estimate the impact of providing informal care to one's partner on the caregiver's health using data from the Survey of Health, Ageing and Retirement in Europe (SHARE). We exploit the panel structure of the data and use statistical matching to deal with selection bias and endogeneity. We find that in the short run caregiving has a substantial negative effect on the health of caregivers. These negative effects should be taken into account when comparing the costs and benefits of formal and informal care provision. These negative effects are potentially short-lived, however: we do not find any evidence that the health effects persist after 4 or 7 years.

JEL codes: C21, I11, J14

Keywords: long-term care, informal care, SHARE, propensity score matching

¹University Medical Center Groningen, the Netherlands. ²Institute of Health Policy and Management, Erasmus University Rotterdam (EUR), the Netherlands. ³Erasmus School of Economics, EUR. ⁴Tinbergen Institute, Rotterdam, the Netherlands. Correspondence to Pieter Bakx (bakx@bmg.eur.nl). This research is part of the Netspar theme Optimal saving and insurance for old age: The role of public long-term care insurance. The authors thank Judit Vall-Castellò, Eduard Suari-Andreu, seminar participants at Universitat Pompeu Fabra and participants at the 5th Health Econometrics Workshop in Bari and the Netspar Pension Day for comments. We use data from SHARE. The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

1. Introduction

Informal caregiving to frail elderly by family or friends is often regarded as an important way to sustainably meet the rising demand for long-term care (LTC), in particular because its direct costs appear much lower than the costs of formal care provided by professionals: informal caregivers generally do not receive monetary compensation (OECD 2011). Even though informal care is usually unpaid, its provision may entail substantial direct and indirect opportunity costs, both for the caregiver and for the public purse: it may negatively affect the mental and physical health of the caregiver and these health problems may, combined with time constraints, in turn lead to increased health spending, absenteeism, or even reduced labor force participation.

We study the impact of providing informal care on the health and health care use of caregiving spouses using data from the Survey on Health, Aging and Retirement in Europe (SHARE). Spouses provide about a third of all informal care (OECD 2011), while they are likely to have retired and to be in frail health themselves. Hence, for this subgroup the caregiving tasks may be more burdensome and the health effects more severe than for younger caregivers. Appropriate justification of LTC policy requires a thorough understanding of all relevant costs and benefits associated with the provision of formal and informal care. As the health effects of caregiving may be substantial and difficult to compensate monetarily, we need evidence on the significance and relevance of these health effects.

The decision to provide informal care

In a simple version of a Roy model (Heckman and Sedlacek 1985; Borjas 1987; Heckman and Honoré 1990), individuals choose between a paid job or home production, which includes

caregiving tasks¹. The additional payoff of having a paid job is D , which is a random variable reflecting differences in ability. But a paid job also involves a cost C of hiring a formal caregiver, which is influenced by the government through subsidization² but is a given (i.e. exogenous) to individuals. As a result, the utility maximizing individual favors home production as long as $D < C$ and favors taking up a paid job otherwise. As ability and health status are correlated, we need to account for selection when studying the effect of caregiving on health.

If the government subsidizes formal care more (thus lowering C), we expect the share of caregivers in the population and the number of hours provided by these caregivers to be lower. Hence, if the health effects H of an additional hour of caregiving are individual-specific but do not depend on the total number of hours of caregiving, we expect that the average health effect of being a caregiver is larger when i) C is higher and ii) when the demand for informal care is higher, i.e. when the health problems of the spouse are more severe.

In addition to differences in C and D , the decision between formal and informal care may also be affected by factors influencing the non-monetary cost of the two alternatives. The latter group of factors may include the expected health costs of providing informal care and cultural differences – including gender patterns – changing the perception of potential caregivers of what is expected from them.³ These cultural differences may thus lead to differences in the propensity to provide care and the amount of care provided and thus to differences between countries and between men and women.

Most individuals providing care to their spouse are retired. However, the retired face a related choice: to enjoy leisure time and hire a formal caregiver, or to give up leisure time to provide

¹ We assume that the amount of informal care provided by other family members is exogenous and insufficient.

² The governments of virtually all European countries pay for at least some formal long-term care (OECD 2011).

³ Additionally, gender patterns in caregiving may partially persist when there is a gender pay gap on the labor market.

informal care to their spouse. Hence, with diminishing marginal benefits from leisure and increasing marginal disutility of informal care provision and in the presence of individual-level differences in (i) the marginal decrease in utility resulting from providing care (e.g. because of functional limitations and bad health⁴) and (ii) the value of an additional unit of leisure (e.g. because of differences in income and wealth), the decision to provide informal care will be affected by these characteristics. As these characteristics also influence the health of the caregiver, this self-selection into informal caregiving needs to be accounted for when studying the effect of caregiving on health, as in the simplified Roy model.

Previous findings

Two recent studies have addressed the endogeneity caused by selection bias using an instrumental variable approach to estimate a causal effect between caregiving and caregiver's health (Coe and Van Houtven 2009, Do et al. 2013)⁵. Using data from the US Health and Retirement Survey on the characteristics of siblings and the death of the mother as instrumental variables, Coe and Van Houtven (2009) find significant negative effects on mental health and self-reported health (married respondents), and heart conditions (single men), both immediately and a few years later (depressive symptoms in married women only). Using South Korean data, Do et al. (2015) instrument daughter-in-law's informal caregiving status by their parents-in-law's health endowments, and conclude that caregiving increases the probability of reporting pain affecting daily activities, a fair or poor self-rated health and the use of prescription drugs and outpatient care by the caregiver. Part of the differences in the results may be the result of differences in the specific subpopulations studied – Coe and Van Houtven (2009) include all adults; Do et al. (2015) only study married women – and the fact

⁴ Note that health, which is correlated with ability, may have opposite effects before and after retirement.

⁵ A related strand of the literature, which deals with a similar endogeneity issue, studies the impact of caregiving on labor market outcomes. Recent contributions include Crespo and Mira (2014), Skira (2015), and Van Houtven et al. (2013).

that they use different instruments and thus estimate local average treatment effects for different groups of “compliers”: respondents whose caregiving status is influenced by the instrument. The instruments used in both of these studies only apply to the case in which the children provide care to parents. Children are an important source of caregiving but still make up only about a third of all caregivers (OECD 2011).

Several other studies have addressed the endogeneity of caregiving using fixed effects regressions (Leigh et al. 2010; Schmitz and Stroka 2013; Van den Berg et al. 2014) or statistical matching (Brenna and Di Novi 2015 and Schmitz and Westphal 2015)⁶. The fixed-effects regressions yield mixed results: while Leigh et al. (2010) find no effect on life satisfaction, Van den Berg et al. (2014) report a negative impact. Schmitz and Stroka (2013) find an effect on the use of antidepressants and tranquilizers for those who provide care while working fulltime but not on drugs for physical health problems, suggesting that caregiving affects mental health only. Schmitz and Westphal (2015) arrive at the same conclusion using the mental and physical scales underlying the SF-12 composite health measure that is collected as part of the German Socio-Economic Panel; Brenna and Di Novi (2013) analyze one cross-section of SHARE data and confirm that caregivers are more likely to show depressive symptoms, but only find a significant effect for caregivers in southern European countries.

We contribute to the existing literature in three respects. First, we study the health effects of informal caregiving on *spouses* rather than parents. These effects may be larger than the effects of caregiving to parents because a spouse caregiver is generally older and in worse health than a child caregiver to a parent and hence spouse caregiver health may be more frail. Second, we use statistical matching exploiting the panel nature of the SHARE data⁷ to overcome identification problems. Matching also provides an opportunity to obtain more

⁶ Lechner (2009) provides a discussion of the advantages of statistical matching over fixed-effects regressions.

⁷ Brenna and Di Novi (2013) also apply statistical matching to the SHARE data, but rely on one wave only.

information about the degree of selective attrition, which is a common problem when using panel data from a survey, but one that has not gained much attention in the previous studies. We are the first to describe how attrition affects the estimates of the long-term effects of caregiving when panel survey data are used – and may have affected results in prior studies. Third, we study not only health but also health care use. A health loss in itself leads to welfare losses that may be expressed in monetary terms but this monetization requires many assumptions. Health losses also lead to higher health care spending through increased health care use. As such, this helps to quantify at least some of the health-related costs of informal caregiving. By zooming in on one specific subgroup and making two incremental – yet potentially important – methodological improvements compared to previous studies we are able to make better inferences on the private costs of caregiving in terms of health.

2. Data

Sample

We use waves 1-2 and 4-5 from the Survey of Health, Ageing and Retirement in Europe (SHARE).^{8,9} SHARE respondents are people aged 50 years and older and their spouses. The first two waves contain 47,100 individuals (68,263 observations) from 12 countries. However, we exclude all respondents who either (i) are not present in both waves 1 and 2, (ii) who live without a spouse (including those living separate from their spouse), (iii) for whom some information is missing, (iv) or who provided care to their spouse in wave 1.¹⁰ After removal of these individuals, the final dataset that is used for the matching contains 10,472 individuals,

⁸ This study is based on version 2.6.1 of the 2004 and 2006 waves, version 1.1.1 of the 2010 wave and version 1.0.0 of the 2013 wave. See Börsch-Supan et al. (2013) for methodological details.

⁹ While information on life histories from the retrospective third wave (SHARELIFE) could potentially be valuable, we chose not to add it because of attrition and because the life history information appeared to be only distantly related.

¹⁰ The vast majority of the observations is removed for reasons i) and ii) meaning that the results are likely to be still representative for all spousal caregivers (see appendix).

with roughly equal numbers of men (5,185) and women (5,108). Outcome data for waves 4 and 5, collected approximately 4 and 7 years after wave 2, are used to estimate the longer-term effects of caregiving. Because of attrition, the sample size falls in wave 4 (to 5,992) and wave 5 (to 5,229).

Model and variables

Respondents are classified as a caregiver when they provided any help to their partner daily or almost daily during at least three months within the past 12 months, where help is defined as personal care, e.g. washing, getting out of bed or dressing.

We study the impact of caregiving on four health measures: i) whether the respondent uses prescription drugs, ii) the number of doctor visits in the past 12 months, iii) depression as measured by the EURO-D scale (Prince et al. 1999) ranging from 0 (not depressed at all) to 12 (severely depressed)¹¹, iv) self-perceived health (SPH) on a five-point scale.¹²

To ensure high quality matching, it is important to include information about all characteristics affecting both one's health and whether one is a caregiver. From the introduction, it follows that there are a number of factors affecting the decision to provide informal care. Schmitz and Westphal (2015) summarize these as follows: (i) the need to provide care, (ii) the willingness to provide care, and (iii) the ability to provide care.

The need to provide care to the partner is related to characteristics of the partner, as well as to the availability of alternative sources of care (both formal and informal). Hence we include information on the age and the health of the partner. The availability of formal care is proxied

¹¹ The Euro-D scale is a 12-item questionnaire specifically designed to measure depression late in life. All 12 items have weight 1, meaning the score ranges from 0 to 12 (Prince et al. 1999).

¹² SPH is expected to pick up both health and well-being effects. Caregiving may affect the well-being of a spousal caregiver in two ways. That is, in addition to the direct effect of caring *for* one's spouse, there is a "family effect": a caregiver may care *about* the care recipient (Bobinac et al. 2010). In this article, however, the latter effect is partly mitigated by matching on the partner's health status at t-2.

by indicators for the region of Europe in which the respondent lives¹³, by household income and wealth (which may also proxy for willingness to provide care); the availability of informal care by the presence of children living at home, the total number of children and the number of siblings.

How respondents respond to demand for care (i.e. by providing informal care or hiring a professional caregiver) depends on i) their ability to provide informal care and ii) their willingness to give up leisure when retired or paid labor otherwise. The respondent's ability may be proxied by age, gender, whether he provided informal care at t-2 and his health status; the willingness to provide informal care by employment status, education level, income, wealth and proxies for personality.

3. Methods

The decision to provide informal care depends, among other things, on the ability of the potential caregiver to provide care and on his or her opportunity cost of time (section 1). To address any selection bias, we use statistical matching. That is, every individual providing informal care is matched to a set of individuals not providing informal care with similar observable characteristics. These two matched groups together then form a reduced sample in which being a caregiver is uncorrelated with the other observed characteristics (and therefore the propensity to become a caregiver is equalized). To estimate the average treatment effect of caregiving on health in this sample we perform a regression analysis to control for any residual differences between the treatment group and the controls (Stuart 2010).

¹³ Using i) propensity scores estimated using country dummies and ii) exact matches on country dummies yields poorer matches but final results that are not much different from the preferred estimates (available upon request).

Identifying assumptions

Matching yields an unbiased estimate of the treatment effect if two assumptions hold (Rosenbaum and Rubin 1983). The first assumption is the Stable Unit Treatment Value Assumption, which requires that there is a unique health outcome h_{iT} for individual i and caregiving activity T and that this outcome does not depend on treatment assignment (i.e. caregiving activity) of another individual j . In our case, the health outcome of one partner is certainly dependent on the caregiving activity of the other partner. Therefore, this assumption only holds if partners are separated during the analysis, which is achieved by sample stratification with respect to gender.^{14,15}

The second assumption is that of a Strongly Ignorable Treatment Assignment (SITA). This assumption has two components. First, no combination of covariates should be fully predictive of caregiving or non-caregiving to one's partner. For this reason, individuals living without a partner are removed from the dataset before matching.

Second, there should be no relevant remaining unobserved differences between the matched groups of caregivers and non-caregivers, conditional on the observed covariates. To fully satisfy this assumption, all information related to both health and caregiving should be included. Typically, information on some characteristics is not available, however, yet this may not be a problem because these unobserved characteristics are often correlated with observed characteristics. Therefore, matching on observed covariates implies at least some degree of matching on unobserved covariates that are correlated with the observed ones and hence this assumption is reasonable when the set of observed characteristics is fairly complete.

¹⁴ This strategy does not separate same-sex couples. However, their number is very small (40 households in 2004) and thus has a negligible impact on the results.

¹⁵ This separation is also useful because it yields results that are easily comparable to the existing literature, as virtually all other studies either study health effects on men and women separately (e.g. Coe and van Houtven (2009)) or on women only (e.g. Do et al. (2015); Schmitz and Westphal (2015); Brenna and Di Novi (2015)).

To further increase the likelihood that this assumption holds, we exploit the fact that respondents are interviewed four times and follow the same approach as Lechner (2009) and Schmitz and Westphal (2015): we use the first wave to define the covariates and the second wave to define the treatment status. We then match respondents providing care in the second wave to respondents not providing care in the second wave. The main advantage of this strategy is that the treatment status cannot influence the covariates.

Another advantage of this strategy is that we may stratify by previous caregiving activity. Previous caregiving activity is likely to capture much of the unobserved heterogeneity affecting caregiving activity in later years (and thus treatment assignment as well). Moreover, this stratification thus mitigates endogeneity caused by persistence in caregiving (Lechner 2009). Therefore, the SITA assumption is much more likely to hold if matched individuals have the same previous caregiving status. We focus on the stratum not providing care in wave 1 in the remainder of this article, because the stratum providing care in wave 1 is too small to perform a similar analysis.

The second wave is also used to assess the immediate impact of caregiving, while the last two waves are only used to estimate the longer-term impact.¹⁶ To maximize the sample size in each of the analyses, the matching is performed separately to determine the immediate, medium-term (4 years) and long-term effects (7 years).

Matching procedure

Including more information improves the quality of the matching and means that it is more likely that the SITA assumption holds. However, it also complicates the matching: exact

¹⁶ Respondents in the treatment group and in the control group may or may not be caregivers in subsequent waves. This means that the analysis presented here thus resembles an intention to treat analysis as the health status in the later waves may be affected by differences in caregiving activities in subsequent waves. The results presented here are nonetheless meaningful because individuals face a choice to become a caregiver at $t=0$ without being able to foresee how long they will be providing care.

matching on all variables is not feasible when the number of variables on which respondents are matched is large. Instead, we match on the propensity score, i.e. the probability of being a caregiver conditional on the relevant covariates, which is estimated using a probit regression. Propensity scores have two key properties. First, matching on the propensity score ensures that the conditional distribution of the observed covariates given the propensity score is the same for caregivers as for non-caregivers. Second, if the SITA assumption holds given the covariate vector X , then it also holds given the propensity score (Rosenbaum and Rubin 1985). In sum, the propensity score summarizes all relevant covariate information in a single value, while not compromising on the necessary assumptions (Stuart 2010).

We match observations using kernel weighting matching: each treated observation i is matched to multiple non-treated individuals, with weights determined by the absolute difference in propensity scores with individual i and the particular kernel function that is used. The main advantage of this method is that few individuals need to be excluded from the analysis, and thus little information is lost. A disadvantage is that it requires that nontrivial decisions are made regarding the matching process.¹⁷

Following Rubin (2001), we test the similarity of the covariate distributions using three statistics: i) the absolute standardized difference of the means of the propensity score between the treated and matched control group (Rubin's B), ii) the ratio of variance of the propensity score of the treated and the matched control group (Rubin's R) and iii) the ratio of the variances of the residuals from regressions of each of covariates on the propensity score for the treated and the control group.¹⁸

¹⁷Choices have to be made about the value of a bandwidth parameter and the type of kernel function. The bandwidth measures how similar the propensity scores of two individuals should be for them to be regarded as a match. The choice of bandwidth quantifies the trade-off between bias and variance of treatment effect estimates. The bandwidth is set at 0.03 because this bandwidth is the smallest bandwidth for which many treated individuals can be matched. The kernel used is a standard Epanechnikov kernel.

¹⁸ The samples are usually sufficiently balanced when the first statistic is smaller than 0.25 and the second statistic is between 0.5 and 2 (Rubin 2001) The third statistic is between 0.8 and 1.25 for variables for which the matching was successful.

To calculate the average treatment effect on the treated (ATT)¹⁹, the outcome is regressed on the treatment indicator and all covariates for the matched sample²⁰. Through this regression we correct for small residual variation in covariate distributions between the matched groups when determining the ATT (Rubin 1973; Lechner 2009). All these regressions are estimated with cluster-robust standard errors at the household level.

As described above, roughly 40% of the respondents that were interviewed in wave 1 and wave 2 were not interviewed in wave 4. Hence, the estimates of the longer-term effects of caregiving may be influenced by selective attrition. If the effect of caregiving on health is not of equal size for all subgroups of respondents and if some of the subgroups for which the effect is expected to be particularly large are more likely to drop out, attrition may lead to an underestimation of the effect of caregiving on health. To examine whether selective attrition is a threat to the identification of long-term effects, we analyze whether there is an association between the propensity to be a caregiver and the probability of dropping out of the sample after wave 2.

4. Results

Descriptive statistics and the propensity to provide care

Compared to non-caregivers, caregivers in wave 2 are older, have an older spouse and are in worse health in wave 1 (table 1). Furthermore, caregivers earn lower incomes and are less likely to be highly educated than non-caregivers. These differences highlight that selection bias may affect the results and therefore needs to be addressed.

4.3% of the women and 3.6% of the men in the study sample are caregivers at $t = 0$. The probit regression of caregiver status on the covariates, which is used to estimate the

¹⁹ The analysis is restricted to the treated individuals for whom suitable matches were found and hence the ATT is only applicable to these subgroups.

²⁰ The weights from the kernel matching are used as probability weights.

propensity score²¹, confirms that having a spouse who is in bad health and of high age is strongly *positively* associated with being a caregiver, as are being retired or unemployed (table 2).

Some of the treated individuals are removed because their estimated propensity score is too high to be reliably matched to untreated individuals. However, the share of individuals that is removed is less than 1% (table 3).

The matching procedure (full details available upon request) mitigates virtually all differences in the means of observed characteristics between the treated and control group: the absolute standardized differences in the means decrease to zero for the propensity score and covariates. Finally, the ratio of the variance of regression residuals between the treated and control group indicates that many covariates were classified as 'of concern' or 'bad' before the matching on the basis of this ratio, while only very few are still of concern after the matching. For instance, in the case of the analysis of the immediate impact on female caregivers, the variance ratio for 18 covariates is 'of concern' or 'bad' before, of which only one remains of concern after matching. Based on these results, we conclude that the matching sufficiently reduces the differences between the treatment and control groups.

Short-term health effects

The regressions on the matched sample show a substantial and significant immediate effect of caregiving on the health of caregivers. For females, it leads to more symptoms of depression; for this subgroup, caregiving causes an immediate increase of 0.582 points of the depression score on a scale of 0 to 12. Considering that the full sample mean is 1.948 and the difference between a person with no symptoms of depression and a person that is likely to be clinically

²¹ As explained in the methods section, the propensity score is estimated separately for each of the samples. To save space, table 2 contains the estimates for the sample used to estimate the immediate impact of caregiving on health for females who did not provide care in wave 1 only; the estimates for the other samples, which are very similar, are available upon request.

identified as depressed is only 4 points (Prince et al. 1999) on this scale, providing care to the partner can contribute substantially.

Furthermore, caregiving is also found to lead to worse self-reported health, a higher probability of using prescription drugs (6.2 percentage) and 1.4 more doctor visits per 12 months for female caregivers; for males, effects tend to be smaller and only the effects on depressive symptoms and self-reported health are significant.

Longer-term effects and selective attrition

With one exception (reduced doctor visits for males after 7 years), we find no health effects at all after 4 and 7 years. This is not because of lack of power as the estimates are almost as precisely estimated as the immediate effects. Rather, it appears the result of selective attrition: individuals with higher propensity scores are indeed somewhat more likely to drop out after wave 2 (table 5). Ill health and use of care of the partner are the most important determinants of the higher propensity scores (table 2) and disproportionate dropout of individuals with partners with greater care burden lowers the average care burden in the later waves. If care burden is associated with negative health effects (as it is hypothesized to be), it is likely that the negative intermediate-term and long-term effects of providing care to the partner on the caregiver's health are underestimated in this study, especially for men.

A comparison of the original estimates of the immediate effect with estimates of the immediate effect of caregiving on health for the subsample that does not drop out after wave 2 shows that this is indeed the case: the point estimates of the health effects of caregiving in table 5 are always smaller than the estimates in table 3, which together with larger standard errors caused by a smaller sample size leads to fewer statistically significant results. So while the findings in table 3 indicate that the health effects of caregiving disappear over time, it may be that selective attrition has played a role in this phenomenon.

5. Conclusion and discussion

Informal caregivers provide much of the help with daily activities that frail elderly receive and the great majority of this care is provided by family members. Caregiving is time-consuming and may be burdensome, and any such negative effects need to be accounted for when comparing the costs and benefits of formal care and informal care. The impact of caregiving on the caregiver is also relevant as both economic theory and empirical evidence suggest that because of self-selection into caregiving any adverse health effects of caregiving will make socioeconomic gradients in health and income steeper. The health effects of caregiving by *spouses* are of particular interest because they are likely to have retired and to be older and be in a relatively worse health when starting to providing care. Hence, caregiving tasks can be expected to be more burdensome for this subgroup of caregivers, and the health effects more severe. To estimate the health effects of caregiving by spouses, we exploit the availability of exceptionally good information on the health, well-being and activities of both spouses in the SHARE data. We use this information to overcome endogeneity bias through statistical matching.

The main findings are as follows. First, for women and men alike, caregiving leads to an increase in depressive symptoms and to a reduction in self-assessed health. These effects are larger for women than for men, e.g. women face an increase in depressive symptoms of about 0.58, while the score for men increases by 0.45, which is large compared to the full sample mean of 1.948 (on a 12-point scale). Second, for women caregiving significantly raises their use of medical care: on average, it increases the probability for a woman to use prescription drugs by 6.2 percentage points and leads to 1.4 more doctor visits per year.

Estimates of these health effects facilitate a more informed comparison of the costs and benefits of formal and informal care which are relevant inputs in policy decisions that influence the mix of informal and formal care through subsidies and other incentives (OECD

2011). These health effects may play a particularly important role in such a cost-benefit analysis because they may worsen existing inequalities in health and because health effects are more difficult to compensate than reduced labor market opportunities, as monetary transfers are less likely to suffice as compensation for health losses than for a reduction in labor supply.

Third, we find that the short term negative health effects do, however, disappear over time: four and seven years after the treatment assignment, the health of the respondents who were caregivers initially was not significantly different from the health of the matched control group. The finding that the negative health effects disappear over time is in line with the findings of Schmitz and Westphal (2015) for German caregivers. There may be three reasons for this. First, it may be that the caregiving activities are temporary and that their effects do not last. Second, caregivers may find ways to cope with the burden of caregiving. Third, selective attrition may have biased the results. Selective attrition is a major issue with any panel survey data, and may thus also have affected the results from previous studies that aim to identify long-run effects of caregiving using similar data. As the identification of longer-term effects will bring us a step closer to understanding all the costs associated with informal care provision, future research should focus on finding ways to avoid the problem of selective attrition in panel surveys. The growing availability of administrative population register data with information about long-run health outcomes is one in this context possible avenue to explore.

Bibliography

- Bobinac A, van Exel N, Rutten F, Brouwer W. 2010. Caring for and caring about: disentangling the caregiver effect and the family effect. *Journal of Health Economics* 29(4): 549-556.
- Borjas G. 1987. Self-selection and the earnings of immigrants. *American Economic Review* 77: 531-553
- Borsch-Supan A, Brandt M, Hunkler C, Kneip T, Korbmacher J, Malter F, Schaan B, Stuck S, Zuber S. 2013. Data resource profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*: 1-10.
- Coe N, Van Houtven C. 2009. Caring for mom and neglecting yourself? The health effects of caring for an elderly parent. *Health Economics* 18(9): 991-1010.
- Crespo L, Mira P. 2014. Caregiving to elderly parents and employment status of European mature women. *Review of Economics and Statistics*, 96(4): 693–709.
- Di Novi C, Brenna E. 2015. Is caring for elderly parents detrimental for women's mental health? The influence of the European North-South gradient. *Review of Economics of the Household*: 1-34.
- Do YK, Norton E, Stearns S, Van Houtven C. 2015. Informal care and caregiver's health. *Health Economics*: 224–237.
- Heckman J, Honoré B. 1990. The empirical content of the Roy model. *Econometrica* 58: 1121-1149
- Heckman J, Sedlacek G. 1985. Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market. *Journal of Political Economy* 93: 1077-1125
- Lechner M. 2009. Long-run labour market and health effects of individual sports activities. *Journal of Health Economics* 28(4): 839-854.

- Leigh A. 2010. Informal care and labor market participation. *Labour Economics* 17(1): 140-149.
- OECD. 2011. *Help Wanted? Providing and paying for long-term care*. OECD, Paris.
- Prince M, Reischies F, Beekman A, Fuhrer R, Jonker C, Kivela S, Lawlor B, Lobo A, Magnusson H, Fichter M, et al. 1999. Development of the EURO-D scale: a European Union initiative to compare symptoms of depression in 14 European centres. *The British Journal of Psychiatry* 174(4): 330-338.
- Rosenbaum P, Rubin D. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1): 41-55.
- Rosenbaum P, Rubin D. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39(1): 33-38.
- Rubin D. 1973. The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*: 185-203.
- Rubin D. 2001. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology* 2(3-4): 169-188.
- Schmitz H, Stroka M. 2013. Health and the double burden of full-time work and informal care provision: evidence from administrative data. *Labour Economics* 24: 305-322.
- Schmitz H, Westphal M. 2015. Short-and medium-term effects of informal care provision on health. *Journal of Health Economics* 42: 174-185.
- Skira M. 2015. Dynamic wage and employment effects of elder parent care. *International Economic Review* 56 (1): 63-93.
- Stuart E. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science* 25(1): 1.

Van den Berg B, Fiebig D, Hall J. 2014. Well-being losses due to care-giving. *Journal of Health Economics* 35: 123-131.

Van Houtven C, Coe N, Skira M. 2013. The effect of informal care on work and wages. *Journal of Health Economics* 32 (1): 240-252.

Table 1: Descriptive statistics

Variable	Full sample	Non-caregivers	Caregivers
<i>Treatment indicator</i>			
Caregiver to spouse in wave 2	0.040	0.000	1.000***
<i>Health outcomes in wave 2</i>			
Depressive symptoms	1.948	1.904	3.025***
% Using prescription drugs	0.701	0.693	0.872***
Number of doctor visits	5.769	5.641	8.808***
Self-perceived health	2.035	2.061	1.428***
<i>Health outcomes in wave 4</i>			
Depressive symptoms	2.227	2.199	2.919***
% Using prescription drugs	0.783	0.779	0.893***
Number of doctor visits	6.343	6.253	8.576***
Self-perceived health	1.949	1.966	1.551***
<i>Health outcomes in wave 5</i>			
Depressive symptoms	2.266	2.233	3.141***
% Using prescription drugs	0.798	0.794	0.906***
Number of doctor visits	6.575	6.522	7.955***
Self-perceived health	2.066	2.050	2.487***
<i>Covariates (measured in wave 1)</i>			
Depressive symptoms	1.990	1.965	2.591***
Physical health	0.636	0.644	0.452***
Doctor visits	5.268	5.170	7.589***
Prescription drug use	0.661	0.654	0.823***
Self-perceived health	2.203	2.224	1.704***
Age	62.05	61.80	67.998***
Number of children	2.317	2.319	2.260
Living with a child	0.308	0.314	0.168***
Number of siblings	4.839	4.847	4.648*
Education: low	0.479	0.474	0.608***
Education: medium	0.315	0.317	0.270**
Education: high	0.205	0.209	0.123***
Fraction of household income earned by respondent	0.201	0.206	0.095***
Log of standardized household income	9.602	9.612	9.361***
Assets (in 10.000 PPP-adjusted euros)	36.382	36.712	28.542**
Employed	0.332	0.341	0.106***
Retired	0.454	0.445	0.669***
Homemaker	0.160	0.161	0.151
Unemployed	0.054	0.053	0.073
Motivation for charity or voluntary work: contribute to something useful	0.219	0.220	0.199
Motivation for charity or voluntary work: I am needed	0.284	0.287	0.217***
Caregiver to a parent	0.089	0.089	0.069
Living in northern Europe	0.282	0.285	0.213***
Living in central Europe	0.443	0.442	0.473
Living in southern Europe	0.275	0.274	0.314*
Living in an urban environment	0.481	0.483	0.426**
Age of the spouse	61.97	61.68	68.716***
Depressive symptoms of the spouse	2.117	2.075	3.123***
Spouse activity limitations	0.651	0.642	0.884***
Spouse prescription drug use	5.116	4.970	8.570***
Spouse doctor visits	0.645	0.659	0.319***
Spouse self-perceived health	2.208	2.239	1.478***
Number of observations ^a	10,472	10,049	423

Notes: *, **, and *** indicate that differences with the sample of non-caregivers are significant at $p < 0.10$, 0.05 and 0.01, respectively. ^a Number of observations is smaller for health outcomes in wave 4 and wave 5 (see main text).

Table 2: Propensity score estimation

	Coefficient
Depressive symptoms	0.008
Activity limitations	-0.007
Doctor visits	0.005
Prescription drug use	0.003
Self-perceived health	-0.082*
Age	-0.153***
Age ²	0.001***
Number of children	-0.039
Living with a child	-0.057
Number of siblings	0.003
Education: medium	0.047
Education: high	0.007
Fraction of household income earned by respondent	0.048
Log of standardized household income	-0.031
Assets (in 10.000 PPP-adjusted euros)	0.000
Retired	0.401**
Homemaker	0.575
Unemployed	0.368*
Motivation for charity or voluntary work: contribute to something useful	0.115
Motivation for charity or voluntary work: I am needed	0.044
Caregiver to a parent	0.105
Living in northern Europe: Denmark, Sweden or the Netherlands	0.060
Living in southern Europe: Greece, Italy or Spain	0.087
Living in an urban environment	0.030
Age of the spouse	0.026***
Depressive symptoms of the spouse	0.031**
Spouse prescription drug use	0.033
Spouse doctor visits	0.012**
Spouse activity limitations	-0.330***
Spouse self-perceived health	-0.108**
Intercept	1.690
Number of observations	5,185

Notes: Results from the propensity score estimation for the subgroup of male respondents for the estimates of the immediate effects. The estimates of the propensity scores for the other subgroups are available upon request. *, **, and *** indicate that differences with the sample of non-caregivers are significant at $p < 0.10$, 0.05 and 0.01, respectively.

Table 3: Matched sample

		Total observations	Treated observations	
			Matched	Not matched
Females	Immediate	5,108	219	1
	4 years	3,025	133	0
	7 years	2,653	114	0
Males	Immediate	5,185	185	0
	4 years	2,957	96	2
	7 years	2,576	76	1

Table 4: Estimation results

		Depressive symptoms	Self-reported health	Prescription drug use	Doctor visits	n
Females	Immediate	0.582*** (0.158)	-0.207*** (0.061)	0.062** (0.019)	1.427** (0.471)	5,103
	4 years	-0.104 (0.179)	0.011 (0.072)	0.012 (0.024)	0.152 (0.512)	3,024
	7 years	-0.117 (0.196)	0.005 (0.075)	0.044* (0.020)	-1.349* (0.579)	2,763
Males	Immediate	0.454** (0.162)	-0.159* (0.069)	-0.001 (0.023)	0.629 (0.499)	5,181
	4 years	-0.208 (0.177)	0.071 (0.088)	-0.019 (0.033)	0.955 (0.642)	2,950
	7 years	0.184 (0.229)	0.016 (0.097)	-0.019 (0.035)	1.150 (0.767)	2,575

Notes: *, **, and *** indicate that the results are significant at $p < 0.10$, 0.05 and 0.01 , respectively.

Table 5: Sensitivity test for selective attrition

		Full sample	Respondents present in waves 4 and 5
Females	Mean propensity in 2006	0.043	0.040
	Effect of caregiving on depressive symptoms	0.582*** (0.158)	0.516* (0.234)
	Effect of caregiving on self-reported health	-0.207*** (0.061)	-0.141 (0.079)
	Effect of caregiving on prescription drug use	0.062** (0.019)	0.053* (0.026)
	Effect of caregiving on doctor visits	1.427** (0.471)	0.546 (0.636)
	Number of observations	5,103	2,292
Males	Mean propensity in 2006	0.036	0.028
	Effect of caregiving on depressive symptoms	0.454** (0.162)	-0.050 (0.190)
	Effect of caregiving on self-reported health	-0.159* (0.069)	-0.163 (0.106)
	Effect of caregiving on prescription drug use	-0.001 (0.023)	-0.007 (0.036)
	Effect of caregiving on doctor visits	0.629 (0.499)	-0.062 (0.668)
	Number of observations	5,181	2,463

Notes: *, **, and *** indicate that the results are significant at $p < 0.10$, 0.05 and 0.01 , respectively.

Appendix

Table A1: Descriptive statistics sample selection

	Number of observations
Wave 2 sample	37,447 (100%)
Interviewed in wave 1 too	21,163 (56.5%)
Living with a spouse	13,798 (36.8%)
No information is missing	10,847 (29.0%)
Did not provide care to spouse in wave 1	10,472 (28.0%)

% indicates the percentage of the original wave 2 sample that is left after applying the sample selection criterion