RUHR

ECONOMIC PAPERS

Hendrik Schmitz
Matthias Westphal

Short- and Medium-term Effects of Informal Care Provision on Health

#426

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, email: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Christoph M. Schmidt
RWI, Phone: +49 (0) 201/81 49-227, e-mail: christoph.schmidt@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #426

Responsible Editor: Volker Clausen

All rights reserved. Bochum, Dortmund, Duisburg, Essen, Germany, 2013

ISSN 1864-4872 (online) – ISBN 978-3-86788-482-2

The working papers published in the Series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Ruhr Economic Papers #426

Hendrik Schmitz and Matthias Westphal

Short- and Medium-term Effects of Informal Care Provision on Health

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788482
ISSN 1864-4872 (online)
ISBN 978-3-86788-482-2

Hendrik Schmitz and Matthias Westphal[1]

Short- and Medium-term Effects of Informal Care Provision on Health

Abstract

This paper estimates the effect of informal care provision on female caregiver's health. We use data from the German Socio-economic Panel and assess effects up to seven years after care provision. A simulation-based sensitivity analysis scrutinizes the sensitivity of the results with respect to potential deviations from the conditional independence assumption in the regression adjusted matching approach. The results suggest that there is a considerable short-term effect of informal care provision on mental health which fades out over time. Five years after care provision there are no significant effects left. Both short- and medium-term effects on physical health are much smaller and insignificant throughout.

JEL Classification: I10, I18, C21, J14

Keywords: Informal care; regression adjusted matching; propensity score matching; mental health; physical health

July 2013

1 Both UDE and CINCH – We thank Martin Fischer and Stefan Pichler for valuable suggestions. Moreover, we are grateful for comments at the Annual Meeting of the dggö (Essen), the CINCH health economics seminar in Essen, the CINCH academy, The economics of disease conference in Darmstadt, and seminars in Bayreuth and Paderborn. Financial support by the Fritz Thyssen Stiftung is gratefully acknowledged. – All correspondence to Hendrik Schmitz, Universität Duisburg-Essen, Fachbereich Wirtschaftswissenschaften, Schützenbahn 70, 45117 Essen, Germany, E-Mail: hendrik.schmitz@uni-due.de.

1 Introduction

Responding to the ongoing demographic change appropriately is a serious policy challenge. Besides adjusting the pension insurance, managing the demand for long-term care in particular is an important issue in aging societies. In Germany, for example, there have been about 2.5 million people in need of care in 2011 (Statistisches Bundesamt, 2011) and this number is estimated to grow steadily, reaching at least 3 million care recipients according to an optimistic-case scenario in 2030 (Statistisches Bundesamt, 2010). Traditionally, provision of long-term care in Germany has been predominantly the task of the family (Schulz, 2010). Even the introduction of the German social long-term care insurance system in 1995 stressed the family as the main provider of care, as it is thought to provide care cheaper[1], more agreeable and more efficiently. This inherent principle is also mirrored in numbers: 1.76 million care recipients (70% of all) are cared at home by professional caregivers and/or relatives. More than 1 million are exclusively cared by family members rendering informal care the most important part of the German long-term care system.

However, provision of informal care is both mentally and physically challenging. We, therefore, analyze the question whether there are some hidden costs – or costs, not discussed so far in the public debate – that make informal care provision not as economic as often thought. This could be the case if informal care provision goes along with health impairments of the caregivers. Other costs (not considered here) are forgone income for those who leave the labor force to provide care.

The economic literature on health effects of caregiving is fairly scarce.[2] In particular, to the best of our knowledge, there are only two studies on the effect of care provision on health in a narrow sense. Coe and van Houtven (2009) estimate health effects of informal caregiving in the US using seven waves of the Health and Retirement Survey (HRS). They use sibling characteristics and the death of the mother as instrumental variables that control for selection into and out of caregiving in order to identify causal effects. They find that continued caregiving leads to a significant increase in depressive symptoms for both sexes while physical health does not seem to be affected. Do et al. (2013) use data from South Korea where informal care is quite common among females caring

[1]For instance, in 2012, the German social long-term care insurance pays 700€ per month for care recipients of the highest care level who are cared by family members and 1,550€ per month to the same recipient cared by professional caregivers.

[2]In the medical literature, there is a large amount of studies on the relationship of health and care provision. They mainly stem from the US (see e.g. Schulz et al. (1995), Stephen et al. (2001), Gallicchio et al. (2002), Tennstedt et al. (1992), Beach et al. (2000), Ho et al. (2009), Shaw et al. (1999), Lee et al. (2003), Dunkin and Anderson-Hanley (1998), or Colvez et al. (2002)). In general these studies use non-representative samples and widely neglect endogeneity problems. Furthermore, they often concentrate on more specific definitions of care, such as caring for people with dementia, etc.

for their parents-in-law. The data facilitates identifying a health effect for daughters-in-law where selection of health into care is alleviated by instrumenting the informal care decision with parents-in-law's health endowment. Their findings suggest that there is an increased probability of worse physical health by providing informal care.

Two further papers evaluate the relationship of care provision and drug utilization (which, however, can also be seen as a health measure). Van Houtven et al. (2005) assess the impact of caring on the intake of drugs using a rich database on caregivers for US veterans. One finding is that the intensive care margin is an important factor determining drug intake. Stroka and Schmitz (2013) exploit data of a large German sickness fund that enables to consider prescriptions of anti-depressants and drugs to restore physical health. Although their data exhibit detailed measures for care and health together with a large amount of observations, this comes at the cost of not observing a large set of socio-economic characteristics including good candidates for instruments. Thus, they need to rely on fixed effects methods. Their results also support van Houtven et al. (2005), providing some evidence that caregiving increases the intake of anti-depressants in particular if coupled with having a job.

Other studies look at broader welfare consequences of caring and use life satisfaction as a proxy. Bobinac et al. (2010) assert that the caregiver and the family effect (the effect due to seeing a family member decline) are closely connected. They suggest using the health of the care recipient to isolate the family effect and the care intensity of the caregiver to disentangle the caregiving effect. Using a Dutch cross-sectional sample of caregivers, they find that the caregiving effect persists even after having controlled for the family effect. Van den Berg and Ferrer-i Carbonell (2007) regress life satisfaction on monthly hours of care provided by an informal caregiver and household income. This set up facilitates estimating the effect of health in order to value this effect monetarily. In contrast to the studies above, Leigh (2010) does not find significant effects of care provision on life satisfaction when fixed-effects are employed. One issue with these latter three studies is that they do not address reverse causality and selection problems based on time-varying unobserved heterogeneity.

We use representative household data from the German Socio-economic panel to estimate the effects of informal care provision on female caregivers' health. The outcome variables are mental and physical summary scale measures (called MCS and PCS) for the years 2002 to 2010 that capture the multidimensional aspect of health. Our contributions to the literature on health and informal care are twofold: First, we use a different approach to address selection into and out of care provision. Previous studies that deal with these problems all use instrumental variables approaches. However, good (i.e. valid and strong) instruments are not always available and most employed instruments are far from being undisputed among researchers. Since their validity cannot be tested they need to be justified on a theoretical basis. Our approach is to fully exploit the time dimension and richness of panel data in order to justify the conditional independence assumption (CIA) that allows for a causal interpretation of the results. Moreover, we use a regression adjusted matching approach. Although we argue below that given our data at hand we can justify the CIA, we allow in a sensitivity analysis that follows Ichino et al. (2008) for certain deviations from this assumption.

Second, to the best of our knowledge, this is the first study that does not only look at contemporary, or short-term effects of informal care provision on health, but also on medium-term effects of up to seven years after care provision. This adds on work by Coe and van Houtven (2009) who also aim at showing persistence of health effects but need to stick to a two year period. Medium-term consequences could be more severe than instantaneous short-term health impacts restricted to the period of providing care.

Our results suggest that there are considerable short-term effects of informal care provision on mental health which, however, fade out over time. Five years after care provision there are no significant effects left. Both short- and medium term effects on physical health are much smaller and insignificant throughout. The sensitivity analysis suggests that sensible deviations from the CIA do not change these results.

The paper is organized as follows. Section 2 discusses the empirical approach, Section 3 presents the data. The results are reported in Section 4 while Section 5 assesses the sensitivity of the results with respect to certain deviations from the CIA. Section 6 concludes.

2 Empirical Strategy

Basic set-up

We aim at estimating the effect of informal care provision on health in later years. Because it is not possible to set up an experiment where care is randomly assigned to individuals who are at the margin of indifference between caring and not caring, this paper applies methods that replicate this ideal to retain causal estimates. Certainly, the decision to provide care is not random *per se*. Given that someone close becomes care dependent, some individuals choose to provide care while others do not. The willingness to provide care depends on factors such as the financial and temporal affordability, the quality of the extra and intra familial social environment, own health endowment as well as innate tendencies such as personality traits.

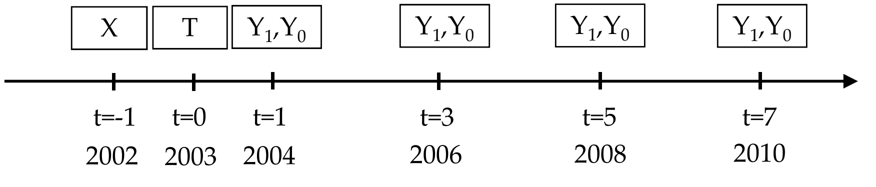
To deal with this problem we apply the model of Rubin (1974). Following his notation we observe $Y = T \cdot Y_1 + (1 - T) \cdot Y_0$, where T indicates whether an individual is assigned to treatment (care provision) or control group, Y is the outcome (health), and the index

$\{0,1\}$ indicates the potential health outcome if being a caregiver or not. If we simply compare the realized outcomes, i.e. $E(Y_1|T=1) - E(Y_0|T=0)$, selection bias will most likely arise due to the non-randomness of care provision.

However, the average treatment effect on the treated (ATT) can be identified if all the determinants that simultaneously influence the health outcome and the selection into treatment are observed. Then, the Conditional Independence Assumption (CIA) holds and assignment to treatment is random (conditional on controls): $Y_1, Y_0 \perp T|X$. $ATT = E(Y_1 - Y_0|T=1, X)$ is the causal ceteris paribus impact of informal care on health that can be estimated either with ordinary linear regression models or with propensity score methods, where the latter are becoming increasingly popular. The reason for this is that propensity score matching methods make less parametric assumptions and are shown to yield more robust estimates (see Dehejia and Waba, 2002). We, thus, use these "more sophisticated methods for adjusting for the difference in covariates" (Imbens and Wooldridge (2009), p.24) as these models might be more robust to non-linearities in covariates.

We estimate short- and medium-term effects of care provision on health. In doing this, we use the time structure as presented in Figure 1. Assignment to treatment T occurs in $t = 0$ or year 2003.[3] We condition on a large set of covariates in $t = -1$, thus reducing the potential problem that covariates are affected by the treatment status.[4] We, then, compute the the treatment effect four times: 1 year after treatment, 3 years after treatment, 5 years after treatment, and 7 years after treatment. Note that conditioning variables and treatment group assignment are always the same and determined in $t = 0$ and $t = -1$.

Figure 1: Basic time structure



Regression Adjusted Matching Approach

In this paper both, matching and regression methods, are applied jointly which is called the regression adjusted matching approach (see, for example, Rubin, 1979). The advantage is that it yields consistent estimates if either one of each method fails to remove the selection bias. This is called the double robustness property by Bang and Robins (2005).

[3]The data and their availability are presented in Section 3.

[4]In the regression framework this is referred to as the problem of 'bad controls' (Angrist and Pischke, 2009).

The basic estimation strategy is a two-step process, originally proposed by Bang and Robins (2005) and employed for example by Marcus (2012). As a first step, the probability of being carer conditional on relevant covariates is estimated with a probit model. Subsequently, treatment and control group are matched. We use an Epanechnikov kernel with a bandwidth of 0.03 in the basic specification. To further increase precision, the sample is restricted to the common support of the propensity scores of the treatment and control group. By kernel matching, the trade-off between bias and variance might be transferred to the choice of the bandwidth in the kernel algorithm. The extent to which degree the differences in the covariates are sensitive to this parameter will be addressed in the section on matching quality.

As a second step, the health outcome is regressed on informal care and, again, all control variables where the observations are weighted by the kernel weights estimated by the matching algorithm: $\hat{\beta} = (X'WX)^{-1}X'y$. Standard errors are computed according to the suggestion of Marcus (2012) who employs robust standard errors of the regression above since they are slightly more conservative but easier to estimate than bootstrapped standard errors that in addition are not formally justified.[5]

Selection issues

The decision to provide informal care is complex, affecting not only one's own utility but also the utility of the care recipient. Hence, the personal assessment of the pros and cons drives the selection into care. This decision can be structurally modeled where the informal care indicator is clearly endogenous. Another alternative is to tackle the problem the other way around by sampling the control group in principle such that treatment is random conditional on controls. Even though we condition on a large set of covariates that are supposed to capture the process of the decision to provide care, there are probably some threats to the CIA assumption. First, there might be health driven selection into treatment. Individuals who are confronted with the question to provide care but are themselves in poor health might not be able to do so. As informal care provision is both physically and mentally challenging, this possible selection holds for both dimensions of health. If this is indeed the case and informal care provision has negative health effects, ignoring this reverse causality problem would lead to an underestimation of the true effects (in absolute values). We follow, e.g. Lechner (2009) and García-Gómez (2011) and match individuals on pre-treatment outcomes (here, health status in $t = -1$), thus only comparing individuals of the same baseline health status before treatment. This rules out

[5]We can confirm this finding in our data. Bootstrapped standard errors yield slightly less conservative standard errors.

that individuals in the control group are in worse health due to a selection of healthy individuals into care provision.[6]

A second issue is unobserved heterogeneity, confounders that both affect treatment and outcome, but are not observable for the researcher. As Lechner (2009) suggests, this problem can be mitigated by stratifying the sample according to care provision in $t = -1$. Comparing only individuals with the same care status in $t = -1$ accounts for a lot of unobserved heterogeneity that affects treatment participation. Hence, the CIA is much more likely to hold within the strata. Moreover, stratifying the sample at least mitigates the problem in the care starter stratum that control variables, though dated back to $t = -1$, could be determined by care provision in $t = 0$ through confounders that both affect past control variables and current treatment status.

Table 1: Stratified sample

Stratum	t=-1	t=0
1	care	care
		no care
2	no care	care
		no care

Hence, we generate two samples based on information in $t = -1$ and estimate the treatment effects independently for each sample as depicted in Table 1. Both estimated treatment effects and their variances for each stratum are merged as weighted means.[7]

Note that treatment is only defined as care provision in $t = 0$ while we leave future care status unrestricted as shown in Table 2. This greatly simplifies the analysis. An important advantage of this is that reverse causality (selection out of care provision due to bad health) is no problem in identifying medium-term effects of care provision because future health status – potentially affected by care and leading to selection out of care provision in later years – does not affect the treatment group assignment at all. While this definition of a treatment does not affect the identification of short-term effects, it implies that we actually identify an *intention-to-treat* (ITT) parameter for the medium-run effects. This is because not all individuals stop caring after $t = 0$. In a sense, some do not comply and continue caregiving.

All in all, we argue that there is certainly selection into care provision but that the data source presented below is informative enough to control for it in the way that was just described.

[6] Note that Heckman et al. (1997) call matching along with controlling on pre-treatment outcome "Difference-in-Difference matching".

[7] $\widehat{ATT} = \frac{1}{n}\sum_{i\in 1,2} n_i \cdot \widehat{ATT}_i, \quad \widehat{se} = \sqrt{\frac{1}{n^2}\sum_{i\in 1,2} n_i^2 \cdot se_i^2}$.

Table 2: Group assignment rules

	2002 $t=-1$	2003 $t=0$	2004 $t=1$	2006 $t=3$	2008 $t=5$	2010 $t=7$
Treatment group	X	1	X	X	X	X
Control group	X	0	0	0	0	0
Excluded	X	0	X	X	X	X

Note: 1 = providing care; 0 = not providing care; X = care status not specified (= either 1 or 0). As explained in more detail in Section 3, the control group consists of individuals who never provided care over the entire observation period. We exclude individuals who did not care in $t=0$ but in any of the later periods.

3 Data

We use data from the German Socio-economic Panel (SOEP) which is a yearly repeated representative longitudinal survey of households and persons living in Germany that started in 1984. The SOEP covers a wide range of questions on the socio-economic status like family status and habits, work, education, and further topics like health and life satisfaction (see Wagner et al., 2007 for details). Currently, some 22.000 individuals from more than 10.000 households are interviewed each year.

We identify caregivers depending on how individuals respond to the following question: "What is a typical day like for you? How many hours do you spend on care and support for persons in need of care on a typical weekday?" which was included into the SOEP questionnaire in 2001. If an individual states caring a positive amount of hours we consider her as a caregiver, i.e. we transform this information into a binary variable. We argue that individuals who burden themselves even with one hour of informal care each weekday to someone close should be considered as serious caregivers. We, however, also try stricter cut-offs of at least three hours. The question does not allow for a link between caregiver and care recipient. Hence, we have no information on the care recipient and we are not able to stratify our analysis with respect to her (e.g., in order to evaluate differences between caring for spouses or parents). This is a common shortcoming in this literature.

Table 3 shows the number of observations in the sample. We restrict the sample to women that have complete information on treatment status in 2003 and control variables in 2002. Moreover, we drop professional caregivers from the sample, as they might mix up professional and personal affairs.[8] Finally, in order to have a clean control group, we drop all individuals who do not provide care in 2003 but in any of the following years. Thus,

[8]To be exact we drop 313 females (1169 person-year observations) who work either as social work associate professionals or as institution-based personal care workers.

the control group exclusively consists of individuals who never provided care between 2003 and 2010. In total, we end up with 8,785 women in 2003, 550 of whom provide care according to the definition above. We track these individuals over the subsequent years, since we are primarily interested in evaluating a lagged impact on informal caregiver's health. Due to panel attrition, the sample declines somewhat over time. However, we use unbalanced data in order to exploit as much information as possible. In total, attrition amounts to roughly about one third from $t = 0$ to $t = 7$. There is no significant different attrition pattern detectable between carers and non-carers which in turn reduces the need to balance the panel.

Caregiving among men is much less common, we observe some 4% of all men in 2003 as caregivers. In case of men, it was very difficult to properly model the treatment participation (the propensity scores yielded only very low values). Therefore, we restrict the analysis to women.

Table 3: Sample Size

		t=0	t=1	t=3	t=5	t=7
All observations		8,785	8,388	7,841	6,850	5,776
$T = 1$ (Caregiver in 2003)	min 1 hour	550	524	489	431	369
	min 3 hours	182	170	150	124	112

Source: SOEP, own calculations. *min 1 hour* means care provision for at least one hours per day. *min 3 hours* means care provision for at least three hours per day.

The intensity and the duration of care among this defined treatment group are not equally distributed across all members. The first dimension where caregivers could differ from each other is the daily amount of hours cared per weekday. Table 4 presents the distribution. As a matter of fact, the majority (i.e. 43%) of caregivers states caring one hour per day, whereas only 15 percent in the sample care more than 4 hours per day on average. The other dimension is the duration of care across waves. Table 5 reveals the length of the care spells in our sample, e.g. the total number of years cared between 2003 and 2011. Here, again, the relative majority of caregivers (20%) only cared once in 2003 and then ceased caring subsequently. But, nonetheless, the remainder is burdened with care in multiple periods.

These are the two parameters we can use to alter the treatment definition later in the section on robustness checks in order to reveal whether the magnitudes of potential effects are sensitive to different treatment definitions.

Table 4: Care Intensity

Stated hours of care in 2003	1hr	2hrs	3hrs	4hrs	>4hrs
$T = 1$ (Caregiver in 2003)	239	129	57	45	80
Share	43%	23%	11%	8%	15%

Source: SOEP, own calculations. Note that here, *1hr* means exactly one hour, *2hrs* means exactly two hours, and so on.

Table 5: Care Duration

Years of care in 2003-2010	1	2	3	4	5	6	7	8	9
$T = 1$ (Caregiver in 2003)	112	100	87	60	48	40	43	17	43
Share	20%	18%	16%	11%	9%	7%	8%	3%	8%

Source: SOEP, own calculations. Note: Care does not have to be provided in consecutive years.

Outcome variables

The two outcome measures are a mental and a physical health score that are based on information from the so called SF-12v2 questionnaire in the SOEP which includes twelve questions on mental and physical health. Answers to these questions are collapsed into the Mental Component Summary Scale (MCS) and the Physical Component Summary Scale (PCS) by explorative factor analysis (see, Andersen et al., 2007). Thus, both variables capture the multidimensional aspect of health. The scales range from 0 to 100, normalized to mean values of 50 and standard deviations of 10 in the 2004 reference sample. Higher values mean a better health status. MCS loads information on perceived melancholy, time pressure, mental balance and emotional problems into one summary scale.[9] This resulting variable is informative about any mental strain the corresponding person is suffering from and thus well-suited for detecting an impact caused by the burden of informal care. MCS is widely used in the previous literature, see, e.g. Reichert and Tauchmann (2011), Schmitz (2011), or Marcus (2013). MCS and PCS were first introduced in the SOEP in 2002 and subsequently sampled every other year. This is why we restrict our observation period to the years 2002–2010.

Control variables

Taking on the burden of care could theoretically be modeled as a three-stage process. People provide care if (*i*) they need to. Given that they need to provide care, they (*ii*) must be willing to do so. Finally, (*iii*), they need to be able to provide care.

[9]The physical component comprises: - Physical fitness (2 Questions), general health, bodily pain, role physical (2); the metal component comprises: - Mental health (2), role emotional (2), social functioning, vitality. See the questionnaire in Table A2 in the Appendix.

At the first stage, the hazard that someone close becomes care dependent is a prerequisite of the need to provide informal care. This first stage in general depends on the age and the quantity of (intra-familial) social environment. We model the social environment by using indicators whether parents are alive, their age as well as the number of siblings. The latter can reduce the need to provide care for frail parents as siblings could step in. Variables on this stage are often employed as instruments for care provision in other papers.

At the second stage, given that someone close is in need of care, the willingness to provide care can be modeled as a function of socio-economic characteristics and personality traits. Socio-economic characteristics grouped in here are, e.g., own age, marital status, employment status, and level of education. Note, however, that family background variables might also belong to the first stage. For instance, single individuals do not need to care for a spouse or parents-in-law.[10] Furthermore, we use character traits measured in the Big Five Inventory (BFI), well-known in psychology for being a proxy of human personality (see McRae and John, 1992 or Dehne and Schupp, 2007) as well as positive and negative reciprocity. Although the SOEP captures each item of the BFI with relatively few questions in the 2005 and 2009 questionnaires, surveys revealed sufficient validity and reliability (see Dehne and Schupp, 2007). The items of the BFI are: neuroticism, the tendency of experience negative emotions; extraversion, the tendency to be sociable; openness, the tendency of being imaginable and creative; agreeableness, the dimension of interpersonal relations and conscientiousness the dimension of being moral and organized (see Budria and Ferrer-i Carbonell, 2012). There are three questions for each of these items which are gathered on a 7-item scale. Furthermore, there is positive reciprocity, the tendency of being cooperative and negative reciprocity, the tendency of being retaliatory. For each personality measure, the score is generated by averaging over the outcome of the corresponding questions per individual. Although these questions are only prompted twice in the SOEP and in years after the treatment assignment,[11] they are useful controls because these measures are supposed to be stable over the life cycle. The individual average of each measure is taken over all years as a proxy for time invariant personality.[12]

Finally, on the third stage, the own health status determines the ability to provide care. This was discussed in Section 2 as a potential problem of selection into care provision. As mentioned, we control for pre-treatment health measured as MCS and PCS to account for this. Moreover, we control for health satisfaction and life satisfaction.

[10] Note that we do not explicitly model this three-stage process but that we just have it in mind. Which variable belongs to which stage is then just a matter of interpretation.

[11] The BIG5 are included in the surveys in 2005 and 2009, whereas questions on negative and positive reciprocity are asked in 2005 and 2010.

[12] We do not think that these parameters are influenced by informal care. A conceivably high impact due to a trauma resulting from care is unlikely and not supported by our results below.

According to these stages that make people select into care, the control variables are grouped into three categories and listed in Table 6. Variables that might theoretically belong into the model but were not significant in the propensity score regression are left out. This holds, for instance, for income, the age of the father, or the number of brothers.

4 Results

4.1 Matching Quality

The selection process is modeled implicitly by a probit model where the informal care indicator is regressed on the mentioned observable covariates that affect the propensity to provide care.

Table 6 reports descriptive statistics of all these covariates for different subgroups. It clearly reveals that the mean as well as the standard deviation of the covariates are significantly different in the unweighted baseline sample. Column 4 gives the standardized difference between both means. Without matching almost all confounders are different at the 5% significance level between the carer and non-carer sample. Because these variables have been realized in $t = -1$, they might be causes that drive the selection into care. In particular age, the age of the mother, and marital status exhibit large differences but also personality traits seem to be quite strong predictors of care provision. The kernel matching algorithm equalizes both samples by assigning different weights to each member of the control group. In order to compute these weights we employ an Epanechnikov kernel with a bandwidth of 0.03 and 0.06. Taking a bandwidth of 0.06 means, on the one hand, exploiting information of more individuals but on the other hand these individuals are not as similar in terms of the propensity score as individuals that are confined by a narrower bandwidth. This principle is also mirrored in the standardized difference presented in the last two columns of Table 6: whereas a bandwidth of 0.06 does not accomplish to equalize all covariates, a bandwidth of half the size balances every single control variable such that no test at the conventional significance level of 5% would indicate any difference.

As regards the propensity score, the regions of common support are roughly $[0.04, 0.14]$ for the stratum of women who did not provide care in $t = -1$ and $[0.23, 0.87]$ for those who did provide care. The overlap within each stratum is good as we do not lose treatment observations by restricting the sample to the common support.[13] The low probabilities in the first stratum are simply due to the small amount of caregivers. This indicates

[13]Of course, this also means that the required overlap condition stating that some randomness is needed is ensured in our model (see Heckman et al., 1998).

Table 6: Descriptive statistics of according to treatment and matching status

	Treated		**Controls**		**Matched controls**		**Standardized bias**		
	mean	**sd**	**mean**	**sd**	**mean**	**sd**	**unmatched sample**	**matched sample** (0.06)	(0.03)
Stage i): care obligations									
Age mother	44.94	34.76	37.92	30.65	44.81	33.3	21.43	4.37	0.40
Mother alive	0.44	0.5	0.49	0.5	0.44	0.5	-9.13	-1.58	-0.35
Father alive	0.22	0.41	0.34	0.47	0.23	0.42	-28.5	-6.16	-2.28
Number of sisters	0.89	1.29	1.03	1.33	0.89	1.25	-10.51	-2.13	-0.48
Stage ii): willlingness to provide care									
NEURO	4.53	0.71	4.37	0.74	4.53	0.73	21.98	4.21	0.49
CONSC	5.97	0.8	5.97	0.81	5.96	0.82	0.52	0.87	1.13
AGREE	5.61	0.86	5.58	0.85	5.6	0.84	2.88	0.60	0.43
OPENN	4.43	1.16	4.49	1.14	4.44	1.15	-5.21	-1.63	-0.75
EXTRA	5.00	0.94	5.03	0.97	5.00	0.96	-3.56	-0.39	0.19
Positive reciprocity	5.68	0.97	5.54	1.03	5.69	.98	14.39	2.11	-0.36
Negative reciprocity	2.77	1.22	2.89	1.27	2.78	1.26	-9.8	-2.39	-1.16
Acceptance of private funding	3.29	0.84	3.3	0.80	3.29	0.81	-1.26	-0.89	-0.86
Age	52.59	13.47	46.89	16.68	51.99	13.92	37.56	8.80	3.91
Age squared	2946.22	1393.79	2477.08	1655.84	2896.86	1428.82	30.65	7.26	3.23
Married	0.77	0.42	0.62	0.49	0.77	0.42	34.03	6.53	1.82
Divorced	0.06	0.23	0.08	0.27	0.06	0.23	-8.30	-1.92	-0.61
Single	0.10	0.30	0.20	0.40	0.10	0.30	-28.03	-4.84	-1.22
Children in hh	0.22	0.41	.34	0.47	0.23	0.42	-27.77	-6.82	-2.5
Educ general	0.16	0.37	.19	0.39	0.16	0.37	-7.72	-1.96	-0.78
Educ middle	0.48	0.50	.49	0.5	0.49	0.5	-0.66	-1.14	-1.37
Foreign	0.04	0.20	.07	0.26	0.05	0.21	-14.22	4.13	-2.36
West	0.70	0.46	0.75	0.43	0.69	0.46	-12.41	1.70	0.56
Full time	0.19	0.39	0.27	0.45	0.19	0.39	-20.23	-3.82	-0.92
Stage iii): ability to provide care									
MCS	47.11	10.39	49.18	10.16	47.05	10.86	-20.09	-3.03	0.62
PCS	47.69	9.86	49.65	10.04	47.86	10.34	-19.73	-4.30	-1.71
Satisfaction health	6.44	2.04	6.75	2.17	6.48	2.2	-14.35	-3.58	-1.53
Satisfaction life	6.88	1.8	7.1	1.69	6.89	1.79	-12.6	-2.55	-0.42
N		524		7,864		7,864			

The standardized difference is calculated according to: $Diff = 100 \cdot \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{\frac{1}{2}(\sigma_1^2 + \sigma_0^2)}}$ where 0.06 and 0.03 refer to the employed Kernel bandwidth.

that there is a large unobserved component determining caregiving. But we argue that this unobserved heterogeneity is not a big concern given the estimation strategy outlined in Section 2. Yet, there is one advantage of this fuzziness: It brings about a sufficiently large amount of observations in the control group having a similar value of the estimated propensity score. This provides a hint that the results are not sensitive to a different choice of the matching methods. Moreover, as a consequence of this, this decent overlap is a further justification why a smaller bandwidth works well.

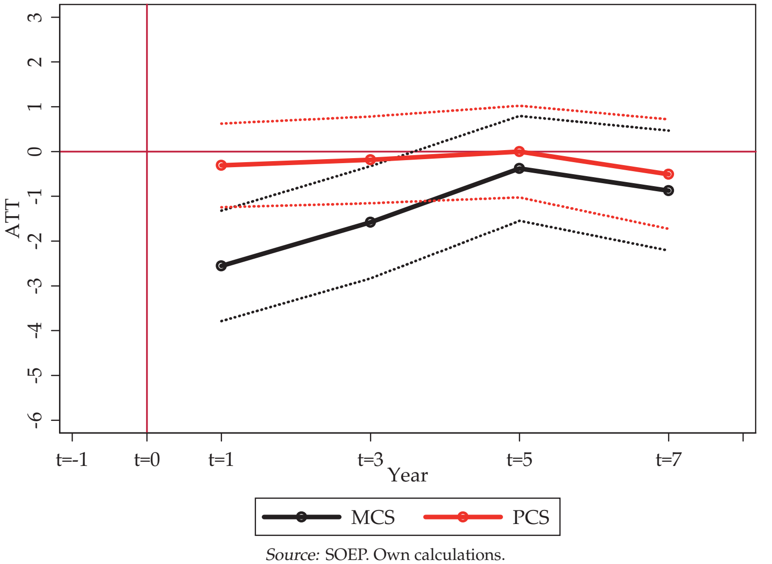
4.2 Estimation Results

The baseline estimation results are reported in Figure 2 for both outcome variables MCS (black points) and PCS (red points). For convenience, we restrict this section to a graphical presentation of the results. Table A1 in the Appendix gives an overview of all results shown in this section. We refer to baseline results as to the effect of the broadest possible treatment definition, i.e. at least one hour of care in $t = 0$. The dotted lines denote 95% confidence bands for the corresponding treatment effect. For PCS, there is no significant effect throughout the periods with a magnitude close to zero, whereas, regarding MCS, considerable short term effects (in $t = 1$) of care provision are detectable. If a woman cares at least one hour per day, her mental health score decreases by 2.55 units (or 25.5 percent of a standard deviation, sd) in the first year, all other things equal. Three years after informal care is observed, this effect reduces to 15.8 percent of a sd before becoming negligible five and seven years afterwards. The confidence bands indicate significant results at the 5 percent level one and three years after assignment to treatment.

These results provide some evidence for significant short-term mental health effects and negligible effects for physical health. Concerning the significant mental health effects, they are in line with findings from previous studies (e.g. Coe and van Houtven, 2009).

Given the absence of physical health effects in this study, we restrict our analysis to mental health in the following parts. We alter the treatment definitions in order to disentangle potential reasons for the observed picture. First of all, we test to what extent treatment intensity, measured as the amount of daily care provision, affects the results. Figure 3 presents the results for two alternative daily intensities. The left panel compares the baseline definition (in black) with a treatment that defines care as at least three hours per day. In this specification, we drop all individuals who care between one and two hours in $t = 0$ and compare three and more hours of care with no care (the red line). The results suggest a much stronger short term effect for individuals who provide more care. One year after caring, the impact is 57 per cent of a sd which is fairly large compared

Figure 2: Results MCS and PCS



Source: SOEP. Own calculations.

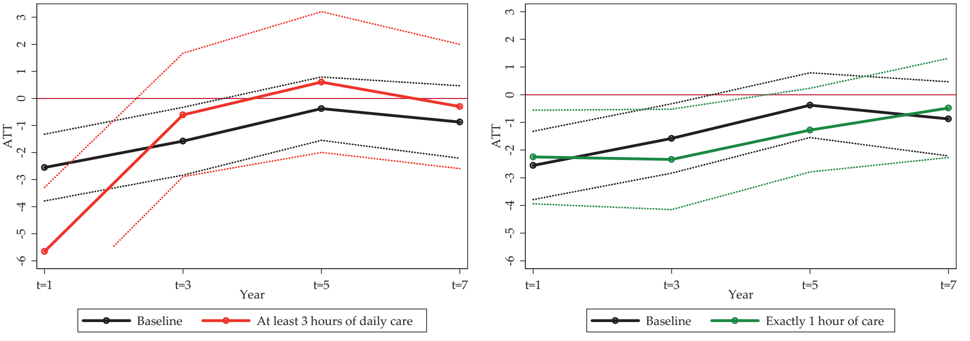
to other studies.[14] Subsequently, however, the effect does not remain on this high level. It immediately drops back to insignificant regions with a coefficient value around zero. Most notably, the qualitative result of a considerable short-term effect and absence of a medium-term effect remains unchanged regardless of the care intensity.

Given that the short-term effect of at least three hours of care per day is much higher than the one of at least one hour, it may be asked whether the baseline effect is fully driven by the higher care intensities. To test this, we define the treatment as providing exactly (instead of at least) one hour of care in $t = 0$. The green line in the right panel of Figure 3 shows the results of this specification. They suggest that also exactly one hour of daily care goes along with mental strain that is comparable in magnitude to the baseline effect in the first period. Three periods after treatment, the effect is even somewhat higher (not significantly different from the baseline effect, however), before fading out in a similar way.

We conclude that one-hour-per-day carers also have to be considered as serious caregivers and stick to the treamtent definition of at least one hour care provision. Besides yielding qualitatively the same results as the definition with higher intensity, the major advantage of this one is the higher number of observations than in the group with higher intensities.

[14]If an individual with a median mental health endowment cares 3 hours per day, his mental health declines from the median to the 29 percent quantile in the next period all other things equal.

Figure 3: Impact on MCS for various definitions of informal care



Source: SOEP. Own calculations. Numbers of observations refer to baseline specification.

One issue in interpreting the results is that the treatment definition only considers care provision in $t = 0$ and leaves future care status unrestricted. Note again that this definition rules out problems of selection out of care due to bad health in measuring medium-term effects. However, individuals who care in $t = 0$ are also likely to care in $t = 1$ and even subsequent years. Thus, it might not be fully convincing to speak of a medium-term effect if individuals provided care throughout the observation period.[15]

Given that some individuals in the treatment group provide care in subsequent periods, we actually identify an *intention-to-treat* (ITT) parameter of the medium-term effect. The difference between the ITT and the ATT is the mistake we are making by falsely assigning people to treatment (of caring only in $t = 0$) who continued to care. The left panel of Figure 4 alters the treatment definition and considers only those individuals who cared at least one hour in $t = 0$ and did not care afterwards. They are compared to the control group of individuals who never cared. Interestingly, the results are almost identical to those in the baseline definition, only being slightly smaller in $t = 3$ and $t = 5$.[16] While this treatment definition is not the preferred one due to potential selection problems out of care provision, the results suggests that the ITT and the ATT are probably close to each other.[17]

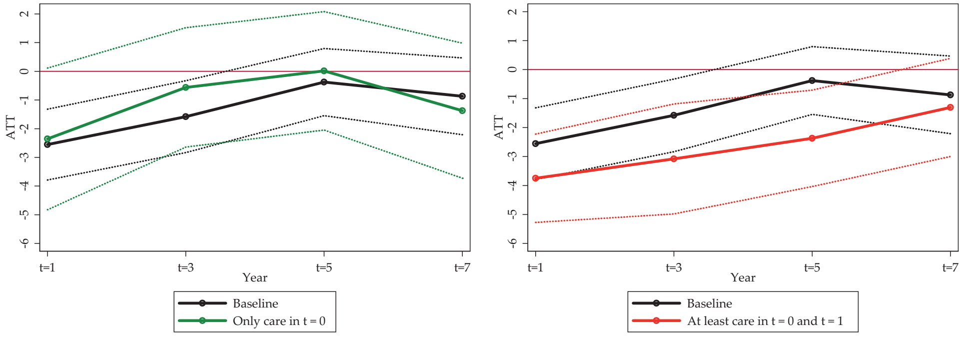
Finally, confining treatment to those having cared in $t = 0$ and $t = 1$ and leaving further periods unrestricted increases the effects in magnitude as the red line shows in the right panel of Figure 4. This shows that more years of care provision have a stronger effect. However, again, these effects vanish after some years.

[15] Although, in principle, a medium-term effect could be defined as an effect some years after care provision (as we have in mind) or some years after the start of a care episode.

[16] The significance bands are broader throughout since the treatment group consists of less individuals in this case.

[17] Note again that this discussion is only relevant for the medium-term effects as, in this application, ATT and ITT always fall together in the short-run.

Figure 4: Impact on MCS for various definitions of informal care



Source: SOEP. Own calculations. Numbers of observations refer to baseline specification.

Interpretation of the results

The results suggest a significant short-term effect of informal care-provision on health while there is no medium-term effect. In order to interpret this finding, we briefly want to outline potential pathways why informal care provision might affect health. We have four different effects in mind: (i) the obligation effect, (ii) the scarring effect, (iii) the family effect, and (iv) the adaptation effect.

We call the "obligation effect" a negative health effect of contemporaneous care provision. As care provision is a challenging task, it might affect caregiver's health. This effect is intrinsically short-term. During the care provision episode, caregiver's health status might worsen. In the medium-run, after the care episode ceased, former health impairments might persist. That is, former care provision might scar individuals, making the implicit costs of care more significant. The family effect is both in the short an medium-run, implying negative consequences of seeing a close relative decline. Finally, an adaptation effect might come to play for individuals who provide care for a longer time period. Getting used to care provision might reduce the stress and improve mental health after an initial drop.

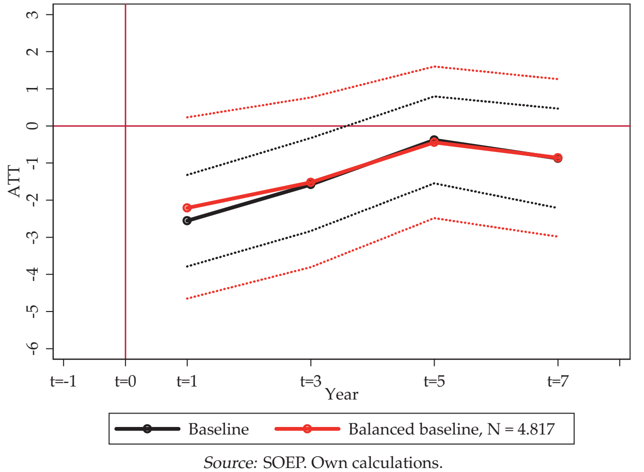
Basically, the estimated effects are weighted averages of these four effects. In the following, we will discuss which one probably dominates (i.e. which one has the highest weight). The observed short-term effect in $t = 1$ is necessarily due to the obligation effect and/or the family effect, as both other ones become active in the medium run. As most of the previous literature, we cannot disentangle the family effect from the active caregiving effect. As results of Bobinac et al. (2010) suggest, the overall effect is a mixture of both but a caregiving effect remains after controlling for the family effect.

Given that the effect in $t = 7$ is negligible, there is obviously no scarring effect. That is, informal care does not leave scars or traumata. Another interpretation is that there might be an adaptation effect driven by those individuals who continued caregiving after $t = 0$. Both results could be interpreted as good news. While there is a considerable negative short-term effect of contemporaneous caregiving, there is no scarring effect and/or continued caregivers get used to care provision resulting in no medium-term effects.

4.3 Robustness Checks

One issue that remains to be clarified is whether panel attrition affects the results. To examine this, Figure 5 plots the results for a balanced sample where each individual is observed consecutively until $t = 7$ against the unbalanced baseline results. The balanced sample consists of 4.817 individuals compared to 8.388 in the baseline subset. Although there is this fairly large attrition rate, the results are almost identical except lost precision that is indicated by broader significance bands.

Figure 5: Differences in the baseline results for the balanced vs. the unbalanced sample



Source: SOEP. Own calculations.

5 Sensitivity Analysis

Basic idea

So far, we argued that our estimation strategy allows us to interpret the results in a causal way since by fully exploiting the panel information in the SOEP the CIA can be justified. However, this inherently not testable assumption might nevertheless fail. In the context of care it might be particularly challenging to properly control for intrinsic willingness to provide care that determines, among others, the mental and physical capability of providing active care. The threat that might come from such a variable is unobserved heterogeneity which would bias our results. Yet, the CIA is not necessarily an "all or nothing" assumption and there might be different degrees of its violation. To examine to what extent the magnitude and the significance of our results depend on the exclusion of such a variable and hence, on the CIA, we follow an approach by Ichino et al. (2008) who refined the suggestions for sensitivity analyses by Rosenbaum and Rubin (1983) and Imbens (2003) and implemented them in a more practical and easy to interpret fashion.

Specifically, we assume that the CIA does not hold

$$Y_0 \not\perp\!\!\!\perp T|X$$

but that the failure is due to an unobserved variable U. Could we condition on it, we had

$$Y_0 \perp\!\!\!\perp T|(X,U).$$

Hence, all the unobserved heterogeneity that leads to endogeneity problems is captured by U. To keep things as simple as possible, Ichino et al. (2008) follow Rosenbaum and Rubin (1983) who proposed U to be binary. This is appealing, since the distribution of a binary variable is fully determined by its mean. To describe how U affects both treatment and outcome, we define four probabilities p_{ij}, $i \in \{0,1\}$; $j \in \{0,1\}$ as

$$\begin{aligned}
p_{01} &= Pr(U=1|T=0,\hat{Y}=1) \\
p_{00} &= Pr(U=1|T=0,\hat{Y}=0) \\
p_{11} &= Pr(U=1|T=1,\hat{Y}=1) \\
p_{10} &= Pr(U=1|T=1,\hat{Y}=0)
\end{aligned} \tag{1}$$

where $\hat{Y} = \begin{cases} 1, & \text{if } Y > \bar{Y} \\ 0, & \text{else} \end{cases}$

Treatment status T and outcome category $\hat{Y}$ are observed in the data and, hence, individuals can be assigned one out of the four probabilities p_{ij} where i denotes treatment status and j indicates whether the outcome exceeds the sample mean. The four above equations define the distribution of the hypothetical confounding variable U completely. Depending on how these probabilities are set, the degree of correlation between Y and T varies.

Given "reasonable" values for p_{ij} we simulate U by drawing 200 times from Bernoulli distributions with the respective parameters for each individual and estimate the ATT 200 times, conditioning on X as before, but also on U. Taking the average over all results provides us with robust point estimates as well as standard errors of the average treatment effect where the CIA is extended.[18]

Specification of p_{ij}

So far, we have not specified these probabilities exactly such that we learn most about the sensitivity of our above stated results. We follow one of the two approaches suggested by Ichino et al. (2008) and set p_{ij} such that we control the "outcome effect" and the "selection effect" of U. As an illustration, think of U again as general intrinsic willingness to provide care. $U = 1$ indicates generally willing, $U = 0$ meaning not willing. This unobserved variable certainly has a positive and strong selection effect such that willing people are more likely to provide care. It may also have a positive outcome effect if the general willingness is positively correlated with health endowment independent of treatment. Vice versa, there could also be a negative outcome effect if frail people are generally more willing to care.

Ichino et al. (2008) define the parameter $s = p_{1.} - p_{0.}$ as the selection effect where

$$p_{i.} = Pr(U = 1|T = i) = p_{i0} \cdot P(\hat{Y} = 0|T = i) + p_{i1} \cdot P(\hat{Y} = 1|T = i) \quad i \in \{0, 1\}.$$

The larger this effect, the larger is the effect of U on selection into treatment keeping the outcome fixed. The outcome effect, defined as $d = p_{01} - p_{00}$ reflects the influence of U on the untreated counterfactual outcome. As an example an outcome effect of $d = 0.1 > 0$ means that the unobserved U positively affects the outcome variables. In the group of non-carers (with p_{0j}), those who are in good health have a higher likelihood of $U = 1$ than those who are in bad health. The higher d the stronger is this correlation. Likewise, a selection effect of $s = 0.1 > 0$ implies that among caregivers the likelihood of $U = 1$ is higher than among non-caregivers. Once we set values for d and s we can derive the four p_{ij} by solving an equation system (as shown in the Appendix) and simulate U.

In principle, d and s could be arbitrarily chosen. One way to find reliable values is to go back to the equation system and – starting the other way around – use observed binary variables in the data set, substitute them for the unobserved U and calculate the selection and the outcome effect of these variables. Thus, we get a feeling how selection and outcome effect of important and observed variables are distributed in the data. Next, one could argue that the unobserved variable U should have a similar selection and outcome effect as important observed variables. We follow this approach and compute these effects for all variables in the sample. To discretize continuous variables we define dummy variables that indicate values above the respective means. Results are reported in Table

[18] We use a modified version of the Stata command sensatt that is written by Nannicini (2007).

Table 7: Distribution of p_{ij} across control variables in the sample

	p01	p00	d	p1.	p0.	s	Effect
Stage i): care obligations							
Age mother	0.66	0.64	0.03	0.65	0.61	0.04	(+)
Mother alive	0.45	0.51	-0.05	0.43	0.48	-0.05	(+)
Father alive	0.23	0.37	-0.14	0.23	0.35	-0.12	(+)
Number of sisters	0.63	0.62	0.01	0.61	0.63	-0.03	(-)
Stage ii): willingness to provide care							
NEURO	0.75	0.67	0.07	0.69	0.62	0.07	(+)
CONSC	0.56	0.56	0.00	.63	0.64	-0.01	(∓)
AGREE	0.58	0.57	0.02	.64	0.64	0.00	(∓)
OPENN	0.5	0.53	-0.02	0.55	0.57	-0.02	(+)
EXTRA	0.5	0.52	-0.02	0.59	0.62	-0.03	(+)
positive reciprocity	0.6	0.51	0.08	0.6	0.55	0.05	(+)
Negative reciprocity	0.46	0.48	-0.03	0.43	0.47	-0.04	(+)
Acceptance of private funding	0.51	0.49	0.01	0.51	0.5	0.01	(+)
Age	0.61	0.4	0.21	0.63	0.44	0.19	(+)
Age squared	0.54	0.35	0.18	0.55	0.39	0.16	(+)
Married	0.76	0.57	0.19	0.76	0.6	0.16	(+)
Divorced	0.06	0.09	-0.03	0.06	0.08	-0.02	(+)
Single	0.13	0.23	-0.1	0.11	0.21	-0.1	(+)
Children in hh	0.2	0.35	-0.14	0.22	0.33	-0.11	(+)
Educ gen	0.19	0.22	-0.03	0.19	0.21	-0.03	(+)
Educ middle	0.49	0.48	0.01	0.49	0.49	0.00	(∓)
Foreign	0.04	0.08	-0.04	0.04	0.08	-0.04	(+)
West	0.69	0.73	-0.05	0.71	0.77	-0.06	(+)
Full time	0.23	0.27	-0.04	0.19	0.26	-0.07	(+)
Stage iii): ability to provide care							
MCS	0.29	0.33	-0.03	0.42	0.52	-0.09	(+)
PCS	0.42	0.54	-0.12	0.47	0.59	-0.12	(+)
Satisfaction health	0.45	0.51	-0.06	0.52	0.6	-0.09	(+)
Satisfaction life	0.56	0.57	-0.01	0.64	0.7	-0.05	(+)

All variables are transformed into binary indicators where the threshold is the sample average. Note: (+) means an amplifying effect, whereas (-) means that the effect attenuates. ∓ indicates no clear effect.

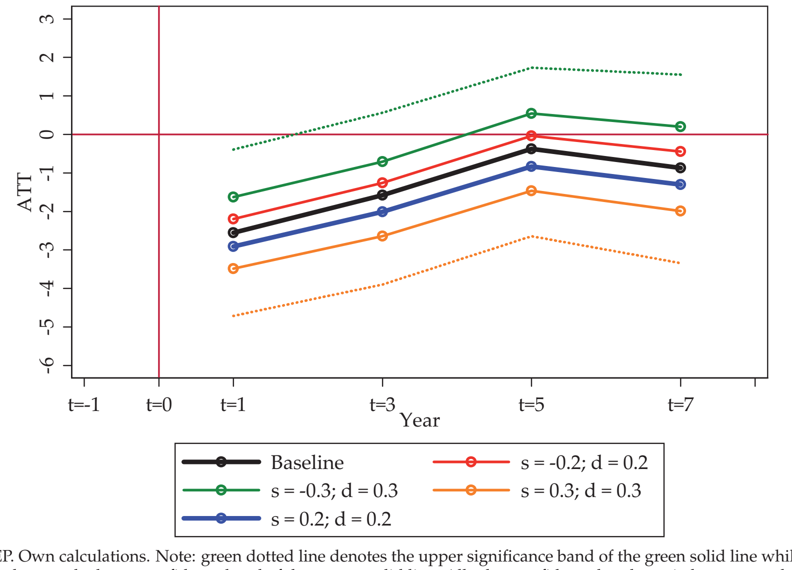
7. We see that most of the variables have selection and outcome effects of at most 0.1 in absolute values. In line with Table 6, the control variables with the strongest impacts are again age and being married exhibiting selection and outcome effects of up to 0.2.

Hence, we argue that given the unobserved variable has an effect on treatment and outcome in the same magnitude as important control variables with high effects, parameterizations of $s = 0.2$ and $d = 0.2$ or $s = -0.2$ and $d = 0.2$ are reasonable values. With these calibrations, no other confounder features such a high effect on mental health and no other makes people select into treatment like the simulated binary confounder U.

The first pair of d, s reflects a positive selection into treatment, i.e., more people with high values of U will take the treatment. Together with a positive outcome effect, meaning that the counterfactuals that we are trying to identify $Y_0|T = 1$ also exhibit higher values of U, we should underestimate effects of care provision on health. This is the case because if U is omitted, the impact is attenuated as more people select into treatment who have a high untreated outcome. But this selection is actually due to U and falsely attributed to the ATT parameter. It can be removed by matching on it.

The second pair reflects a negative selection into treatment leading to overestimation if this was neglected by the analysis so far. Table 7 lists the effect of the confounding variables on the ATT parameter. Except the variable for number of sisters, all confounders have an amplifying effect on the impact of informal care on health. This is supporting the conjecture that people with relatively good health select into care.

Figure 6: Results sensitivity analysis (MCS)



Source: SOEP. Own calculations. Note: green dotted line denotes the upper significance band of the green solid line while the orange dotted line denotes the lower confidence band of the orange solid line. All other confidence bands are in between and not shown.

Figure 6 presents the results of both specifications and the baseline specification, again only for MCS. Moreover, in order to make this result even more credible, we also calibrate confounders with parameters of $s = 0.3, d = 0.3$ and $s = -0.3, d = 0.3$. That is, we assume unobservables exhibiting a much stronger correlation with both treatment and outcomes than all the observed ones. As supposed including a confounder U with characteristics that lead to a positive selection into treatment ($s = 0.2$ and $d = 0.2$, the red dots) leads to larger effects of care provision than in the baseline case while we find weaker effects when including a confounder that induces as negative selection ($s = -0.2$ and $d = 0.2$, the blue line). The most important result, however, is that the overall picture does not change. The lines are parallel-shifted by the confounder, i.e. the results are equally affected by the confounders. In all three cases do we find a significant (both statistically and economically) short-term effect of care provision on mental health while there is no effect left after 5-7 years. This is even true for $s = -0.3$ and $d = 0.3$ in the first period. If there are further confounders that point in the same direction as most of the variables in our sample, our result will actually define a lower bound.

To sum up, in this sensitivity analysis we assume that, although we control for a lot of observed heterogeneity including pre-treatment outcomes and pre-treatment treatment status – which, in turn, captures more unobserved heterogeneity – there are unobserved effects such that the CIA is violated. As long as these unobserved effects do not have a drastically higher impact than observed control variables, we find that the average treatment effects we received in the main analysis are robust. Hence, even if the CIA does not hold, something we can neither demonstrate nor reject, a certain failure of the CIA still leads to the same effects as found in the main analysis.

6 Conclusion

This paper examines whether providers of informal care suffer from a higher mental or physical strain than those who do not care. We use the German Socio-economic panel that identifies informal caregivers by the daily time spent caring. In this paper, caregivers are individuals who care at least one hour per weekday (but stricter definitions of at least three hours lead to a similar picture). We evaluate the impact of caregiving on health by help of a regression adjusted matching technique. This approach is encouraged by theory as well as descriptive evidence indicating that selection into informal care takes place. The problems of unobserved heterogeneity and reverse causality are tackled by exploiting the panel structure of the data set. We use Lechner's (2009) approach and control for pre-treatment outcome and stratify by pre-treatment treatment status to avoid problems induced by selection of healthy individuals into care.

While we do not find effects of informal care on physical health neither in the short- nor in the medium-run, our results suggest that there are considerable short-term effects of informal care provision on mental health which, however, fade out over time. Five years after care provision there are no significant effects left. This leads to a discussion on the potential driving pattern behind these effects. Thus, contemporaneously, care provision is a mental burden but there is no scarring effect. The sensitivity analysis according to

Ichino et al. (2008) suggests that even sensible deviations from the CIA assumption do not change these results: the effects are still similar in magnitude even if we falsely have not incorporated a confounder that is stronger than every one else that we have controlled for before.

We contribute to the current debate on how to realign the care system in Germany and countries with similar demographic developments. The German government recently acknowledged that a realignment of the care system is necessary. As the Minister of Health, Daniel Bahr, puts it the long-term care system does require "radical changes to make it capable of meeting the challenges we will be facing in the near future."[19] Our results suggest that there are considerable short-term health effects. However, it seems to be good news that the effects are not long-lasting. Hence, in total, this paper does not provide evidence for a strong pressure to change the system in order to assist informal care providers.

However, the measured effect is only a mixed effect over different groups of care providers. Stroka and Schmitz (2013), for instance, focus on individuals who not only provide informal care but also work full-time. This double burden might well also have health effects in the long-run. This question is left for future research.

There are some limitations in our data. We do not observe any characteristics of the care recipient. Hence, we cannot distinguish between the family effect that occurs just because a close relative is in need of care and the caregiving effect. Moreover, the quality and intensity of care provision (apart from the hours spent caring) cannot be observed. Thus, we estimate a broad average effect over several different forms of informal care provision.

References

Andersen, H. H., Mühlbacher, A., Nübling, M., Schupp, J., and Wagner, G. G. (2007). Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP Version of SF-12v2. *Schmollers Jahrbuch*, 127:171–182.

Angrist, J. and Pischke, J. (2009). *Mostly Harmless Econometrics*. Princeton University Press.

Bang, H. and Robins, J. M. (2005). Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*, 61(4):962–973.

[19]Quotation taken from the Website of the Federal Ministry of Health: http://www.bmg.bund.de/ministerium/english-version/long-term-care.html.

Beach, S. R., Schulz, R., Yee, J. L., and Jackson, S. (2000). Negative and positive health effects of caring for a disabled spouse: Longitudinal findings from the caregiver health effects study. *Psychology and Aging*, 15(2):259–271.

Bobinac, A., van Exel, N. J. A., Rutten, F. F., and Brouwer, W. B. (2010). Caring for and caring about: Disentangling the caregiver effect and the family effect. *Journal of Health Economics*, 29(4):549–556.

Budria, S. and Ferrer-i Carbonell, A. (2012). Income comparisons and non-cognitive skills. *SOEPpapers No. 441*, pages 1–29.

Coe, N. B. and van Houtven, C. H. (2009). Caring for mom and neglecting yourself? The health effects of caring for an elderly parent. *Health Economics*, 18(9):991–1010.

Colvez, A., Joel, M.-E., Ponton-Sanchez, A., and Royer, A.-C. (2002). Health status and work burden of Alzheimer patients' informal caregivers: Comparisons of five different care programs in the European Union. *Health Policy*, 60(3):219,233.

Dehejia, R. and Waba, S. (2002). Propensity-Score Matching-Methods for Nonexperimental causal Studies. *Review of Economics and Statistics*, 84:151–161.

Dehne, M. and Schupp, J. (2007). *Persoenlichkeitsmerkmale im Sozio-ökonomischen Panel (SOEP) - Konzept, Umsetzung und empirische Eigenschaften*. PhD thesis, DIW Berlin.

Do, Y. K., Norton, E. C., Stearns, S., and Houtven, C. H. V. (2013). Informal care and caregiver's health. Working Paper 19142, National Bureau of Economic Research.

Dunkin, J. J. and Anderson-Hanley, C. (1998). Dementia caregiver burden - A review of the literature and guidelines for assessment and intervention. *Neurology*, 51(1):53–60.

Gallicchio, L., Siddiqi, N., Langenberg, P., and Baumgarten, M. (2002). Gender differences in burden and depression among informal caregivers of demented elders in the community. *International Journal of Geriatric Psychiatry*, 17(2):154–163.

García-Gómez, P. (2011). Institutions, health shocks and labour market outcomes across Europe. *Journal of Health Economics*, 30(1):200–213.

Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica*, 66(5):p 1017–1098.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, 64(4):p 605–654.

Ho, S. C., Chan, A., Woo, J., Chong, P., and Sham, A. (2009). Impact of Caregiving on Health and Quality of Life: A Comparative Population-Based Study of Caregivers for Elderly Persons and Noncaregivers. *The journals of gerontology. Series A, biological sciences and medical sciences*, 64(8):873–879.

Ichino, A., Mealli, F., and Nannicini, T. (2008). From Temporary Help Jobs To Permanent Employment: What Can We Learn From Matching Estimators and Their Sensitivity. *Journal of Applied Econometrics*, 23:305–327.

Imbens, G. W. (2003). Sensitivity to Exogeneity Assumption in Program Evaluation. *American Economic Review*, 93(2):126–132.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86.

Lechner, M. (2009). Long-run Labour Market and Health Effects if Individual Sport Activities. *Journal of Health Economics*, 28:839–854.

Lee, S., Colditz, G. A., Berkman, L. F., and Kawachi, I. (2003). Caregiving and risk of coronary heart disease in U.S. women: A prospective study. *American Journal of Preventive Medicine*, 24(2):113–119.

Leigh, A. (2010). Informal care and labor market participation. *Labour Economics*, 17(1):140–149.

Marcus, J. (2012). Does Job Loss Make You Smoke and Gain Weight? *DIW Discussion Paper 432*, (432).

Marcus, J. (2013). The effect of unemployment on the mental health of spouses – Evidence from plant closures in Germany. *Journal of Health Economics*, 32:546–558.

McRae, R. R. and John, O. P. (1992). An Introduction to the Five Factor Model and Its Applications. *Journal of Personality and Social Psychology*, 60(2):175–215.

Nannicini, T. (2007). Simulation-based Sensitivity Analysis for Matching Estimators. *Stata Journal*, 7(3):334–350.

Reichert, A. and Tauchmann, H. (2011). The causal impact of fear of unemployment on psychological health. *Ruhr Economic Papers 266*, (266).

Rosenbaum, P. R. and Rubin, D. B. (1983). Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(2):212–218.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 56(5):688–701.

Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74(366):pp. 318–328.

Schmitz, H. (2011). Why are the unemployed in worse health? The causal effect of unemployment on health. *Labour Economics*, 18(1):71–78.

Schulz, E. (June 2010). The Long-Term Care System for the Elderly in Germany.

Schulz, R., O'Brien, A., Bookwala, J., and Fleissner, K. (1995). Psychiatric and physical morbidity effects of dementia caregiving: prevalence, correlates, and causes. *Gerontologist*, 35(6):771–791.

Shaw, W. S., Patterson, T. L., Ziegler, M. G., Dimsdale, J. E., Semple, S. J., and Grant, I. (1999). Accelerated risk of hypertensive blood pressure recordings among alzheimer caregivers. *Journal of Psychosomatic Research*, 46(3):215–227.

Statistisches Bundesamt (2010). Demografischer Wandel in Deutschland - Heft 2: Auswirkungen auf Krankenhausbehandlungen und Pflegebedürftige.

Statistisches Bundesamt (2011). Pflegestatistik 2009 - Pflege im Rahmen der Pflegeversicherung Deutschlandergebnisse.

Stephen, M. A., Townsend, A. L., Martire, L. M., and Druley, J. A. (2001). Balancing parent care with other roles: interrole conflict of adult daughter caregivers. *The journals of gerontology. Series B, psychological sciences and social sciences*, 56(1):P24–P31.

Stroka, M. A. and Schmitz, H. (2013). Health and the Double Burden of Full-Time Work and Informal Care Provision: Evidence From Administrative Data. *Forthcoming*.

Tennstedt, S., Cafferata, G. L., and Sullivan, L. (1992). Depression among caregivers of impaired elders. *Journal of ageing and health*, 4(1):58–76.

van den Berg, B. and Ferrer-i Carbonell, A. (2007). Monetary valuation of informal care: the well-being valuation method. *Health Economics*, 16(11):1227–1244.

van Houtven, C., Wilson, M., and Clipp, E. (2005). Informal Care Intensity and Caregiver Drug Utilization. *Review of Economics of the Household*, 3(4):415–433.

Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP): scope, evolution, and enhancements. *Journal of Applied Social Science Studies (Schmollers Jahrbuch: Zeitschrift für Wirtschafts- und Sozialwissenschaften)*, 127(1):139–169.

Appendix

Table A1: Table of results

	t=1	t=3	t=5	t=7
Baseline	-2.55^{***}	-1.58^{***}	-0.38	-0.87
	(0.63)	(0.64)	(0.60)	(0.68)
Care 3 hours per day	-5.65^{***}	-0.61	0.61	-0.3
	(1.21)	(1.16)	(1.33)	(1.17)
Care exclusively 1 hour per day	-2.25^{***}	-2.34^{***}	-1.28^{*}	-0.48
	(0.86)	(0.93)	(0.77)	(0.91)
Care at least in $t = 0$ and $t = 1$	-3.75^{***}	-3.08^{***}	-2.37^{***}	-1.3
	(0.78)	(0.97)	(0.85)	(0.86)
Care exclusively in $t = 0$	-2.36^{*}	-0.56	0.02	-1.37
	(1.26)	(1.06)	(1.05)	(1.20)

Source: SOEP, own calculations. Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$ indicate the corresponding significance level. Standard errors are in parantheses.

Table A2: SF-12v2 questionnaire in the SOEP

	Very Good	Good	Satis-factory	Poor	Bad
How would you describe your current health?					
	Greatly	Slightly	Not at all	–	–
When you ascend stairs, i.e. go up several floors on foot: Does your state of health affect you greatly, slightly or not at all?					
And what about having to cope with other tiring everyday tasks, i.e. where one has to lift something heavy or where one requires agility: Does your state of health affect you greatly, slightly or not at all?					
Please think about the last four weeks. How often did it occur within this period of time, . . .	Always	Often	Some-times	Almost never	Never
◇ that you felt rushed or pressed for time? ◇ that you felt run-down and melancholy? ◇ that you felt relaxed and well-balanced? ◇ that you used up a lot of energy? ◇ that you had strong physical pains? ◇ that due to physical health problems . . . you achieved less than you wanted to at work or in everyday tasks? . . . you were limited in some form at work or in everyday tasks? ◇ that due to mental health or emotional problems . . . you achieved less than you wanted to at work or in everyday tasks? . . . you carried out your work or everyday tasks less thoroughly than usual? ◇ that due to physical or mental health problems you were limited socially, i.e. in contact with friends, acquaintances or relatives?					

Note. Source: SOEP Individual question form. Available at http://panel.gsoep.de/soepinfo2008/.

Calculation of the p_{ij}

Given values for d and s, the four parameters p_{ij} can be derived by solving an equation system of four equations. Assume that $d = 0.1$, $s = 0.1$ and $P(U = 1) = 0.5$. Then we have

$$\begin{aligned}
&P(U=1) = 0.5 \qquad \text{(A.1)}\\
&= p_{11} * P(\hat{Y}=1|T=1) * P(T=1) + p_{10} * P(\hat{Y}=0|T=1) + P(T=1)\\
&+p_{01} + P(\hat{Y}=1|T=0) * P(T=0) + p_{00} * P(\hat{Y}=0|T=0) * P(T=0) + p_{00} * P(T=0)
\end{aligned}$$

$$p_{11} - p_{10} = 0 \tag{A.2}$$

$$d = p_{01} - p_{00} = 0.1 \tag{A.3}$$

$$\begin{aligned}
&s = p_{1.} - p_{0.} = 0.1 \qquad \text{(A.4)}\\
&= p_{10} * P(Y=0|T=1) + p_{11} * P(Y=1|T=1)\\
&-p_{00} * P(Y=0|T=0) + p_{01} * P(Y=1|T=0)
\end{aligned}$$