Informal Care and Long-term Labor Market Outcomes*

Hendrik Schmitz†
University of Paderborn, RWI & CINCH

Matthias Westphal‡
RGS Econ, University of Duisburg-Essen & CINCH

March 11, 2016

Abstract

In this paper we estimate long-run effects of informal care provision on female caregivers' labor market outcomes up to eight years after care provision. We compare a static version, where average effects of care provision in a certain year on later labor market outcomes are estimated, to a partly dynamic version where the effects of up to three consecutive years of care provision are analyzed. Our results suggest that there are significant initial negative effects of informal care provision on the probability to work full-time. The reduction in the probability to work full-time by 4 percentage points (or 2.4 to 5.0 if we move from point to partial identification) is persistent over time. Effects on the probability of being in the labor force are quite small, however, high care intensity strongly reduces the probability to be in the labor force eight years after the start of the episode. Short-run effects on hourly wages are zero but we find considerable long-run wage penalties.

JEL Classification: I10, I18, C21, J14, J22

Keywords: Informal care, labor supply, inverse probability weighting, dynamic sequential models

*Financial support by the Fritz Thyssen foundation (Project: 10.12.2.096) is gratefully acknowledged. We are grateful for comments by Daniel Kamhöfer, Thorben Korfhage, Nicolas R. Ziebarth as well as participants of the health economics seminar at the University of Duisburg-Essen.

†University of Paderborn, Warburger Strasse 100, 33098 Paderborn, Germany, Tel.: +49 5251 603213, E-mail: hendrik.schmitz@uni-paderborn.de

‡RGS Econ: Ruhr Graduate School in Economics, University of Duisburg-Essen, Weststadttürme Berliner Platz 6-8, 45127 Essen, Germany, Tel.: +49 201 1832196, E-mail: matthias.westphal@uni-due.de

1 Introduction

The demographic transition is a major challenge for all European societies and puts pressure on their labor markets. In addition to shrinking working age populations an increasingly important issue are care obligations for the oldest old. Most European societies have a preference for informal long-term care, carried out by close relatives, as opposed to (sole) professional care. Yet, informal care is a challenging task for the caregiver and has potential spillover effects to their labor market participation, in particular of middle-aged women who carry the largest burden among working age individuals. Taking up the burden of care tightens the caregiver's time constraint, but whether labor supply is reduced depends on individual preferences and, ultimately, is an empirical question.

The effects of informal care provision on caregiver's labor force outcomes have been subject to a large literature in the previous two decades. Labor supply reactions (of females, mostly) have been studied by, e.g., Carmichael and Charles (1998), Heitmueller (2007), Casado-Marín et al. (2011), Bolin et al. (2008), Ettner (1995, 1996), Crespo and Mira (2014), Heger (2014), Meng (2012a,b), where the effects range from small to very large (up to 30 percentage points) reductions in the probability to work for pay.[1] The effect on working hours, as studied by, e.g., Wolf and Soldo (1994), Casado-Marín et al. (2011), Bolin et al. (2008), Ettner (1996), Johnson and Sasso (2000), and Van Houtven et al. (2013) are quite mixed, while wage penalties are more consistently found (Van Houtven et al., 2013, Carmichael and Charles, 2003, Heitmueller and Inglis, 2007).

All of these studies have in common that they look at the contemporaneous effect of caregiving on labor market outcomes. As many societies aim at increasing female labor force participation, one often reported policy implication is to set up more flexible work-arrangements to facilitate informal caregiving while keeping the job (Heitmueller, 2007). However, one might argue that negative short-term effects do not pose severe problems – both for the caregivers and societies as a whole – if they are not persistent. Caregiving spells typically last only a couple of years and as soon as caregivers who put their labor force participation on hiatus return to the labor market after cessation of their caregiving spell, the life-time opportunity costs of caregiving might not be too large. However, caregivers are often in the age of 50+ and might have problems to return into the labor force once they left it – either because they voluntarily decide to stay absent or because they cannot return due to labor market frictions.[2] This would imply negative consequences that potentially add up over many years after their caregiving period. Thus, to draw conclusions about the holistic costs of care, it is necessary to turn to a longer-run perspective

[1] Relatedly, Geyer and Korfhage (2015b,a) study incentive effects of the long-term care insurance on care provision and labor supply.

[2] The same holds for switching from full-time to part-time work.

since the cost of caring might be more complex than forgone income for the time spent caring.

This study looks at longer term labor market effects up to eight years after care provision. Evaluating the persistence of effects is the main contribution of this paper. As far as we are aware, only two papers explicitly move away from the contemporaneous perspective. Fevang et al. (2012) use Norwegian data to study labor market outcomes up to around 10 years before and 5 years after the death of a lone parent and do find notable effects on labor market participation (for women, not for men) around the death which, however, are not persistent. On the other hand, reliance on social assistance increases persistently for men. Although the authors do not observe actual care provision these effects can largely be ascribed to informal care obligations. Skira (2015) explicitly takes into account the dynamic effects on labor supply as one of the first papers in this literature. She estimates a dynamic discrete choice model that is underpinned with a theoretical framework. Her results highlight existing labor market frictions for caregivers as their reduced labor supply in the US due to caregiving persists over time.[3]

We use a representative German data set to assess short- and longer-term effects of care provision on labor market outcomes such as the probability to work full-time, to be in the labor force, the number of weekly working hours (conditional on working) and hourly wages. In Germany, the largest European economy, there were 2.6 million people in need of care in 2013 (Statistisches Bundesamt, 2015) and this number is estimated to increase steadily to outnumber 3.4 million people demanding care services by 2030 (Augurzky et al., 2013). Even according to these official – and probably underestimating – numbers, 1.9 million received care in their private home and 1.3 exclusively received informal care (typically by close relatives) making this the most important pillar of the German long-term care system. Thus, not only due to its size as the largest European labor market, Germany is an interesting country to study: it is rapidly aging and already now has a large informal care sector which is even going to increase in the future.

Apart form the longer-term perspective we, as another contribution to the literature, also take the dynamic nature of caregiving spells into account and use sequential inverse probability weighting (IPW) estimators as suggested by Lechner (2009) and Lechner and Miquel (2010) to estimate effects of up to three consecutive years of care provision. A further, if minor, contribution comes from the methodological side where we offer an identification strategy that relies on less functional form assumptions than the previous literature on short-run effects but also than Skira (2015).[4] Our strategy rests on (sequential) conditional independence assumptions (CIA) which we justify by exploiting cross-

[3] Michaud et al. (2010) also estimate a structural model in order to dynamically model the link between care and employment. Yet, effects beyond one period after caregiving are not reported.

[4] Certainly, this is not to say that we make less or weaker assumptions than the previous literature in general, merely that we make different ones, thereby complementing the picture.

sectional but also longitudinal information from our rich household survey, the German Socio-Economic Panel (SOEP). In auxiliary analyses, we relax the CIA to identify effect bounds under weaker assumptions. Other sensitivity tests such as placebo regressions imply that remaining time-invariant unobservables are unlikely to lead to an upward bias of our estimates.

Our main finding is that female caregivers reduce the probability to work full-time by 4 percentage points (at a baseline probability of 35 per cent). The effect is persistent over a period of eight years and seems to be mainly driven by switches to part-time work. High care intensities and longer episodes, however, also increase the long-run probability to leave the labor force. When we move away from point identification to effect bounds, the reduction in full-time work changes to an interval of 2.4 to 5.0 ppts. where we argue that, if any, 5.0 ppts. is more likely than 2.4 ppts. As another finding, wages seem to be unaffected contemporaneously but are significantly lower 8 years after the start of a care episode.

The paper proceeds as follows. Section 2 gives a brief introduction into the German long-term care system. Section 3 presents the data and how we exploit the panel structure. Section 4 lays out the estimation strategy and reports results of the baseline (static) model. Section 5 scrutinizes the identifying assumptions and allows for deviations. Results of the dynamic model are reported in Section 6, while some alternative specifications are carried out in Section 7. Section 8 concludes.

2 Institutional Background

The German social long-term care insurance system was introduced in 1995 as a pay-as-you-go system.[5] It is financed by a mandatory pay payroll tax deduction of currently 2.35 per cent of gross labour income (2.6 per cent for employees without children). In order to qualify for benefits, individuals need to be officially defined as care recipients and be classified into one of now four care levels. In care level one individuals need support in physical activities for at least 90 minutes per day and household help for several times a week. Individuals in need of more care are classified into care levels two or three, where the benefits increase in care levels. In addition, to acknowledge the care needs of people with dementia, care level 0 has been added in 2013, if they suffer from limited activities of daily living (but do not qualify for one of the other care levels).

Benefits also depend on the type of care, where monthly payments for informal care range from 123€ (level zero) to 244€ (level one) and to 728€ (level three), for professional ambulatory care from 468€ to 1,612€ and for professional nursing home care from 1,064€ to

[5]This section is taken almost unchanged from Schmitz and Westphal (2015).

1,550€ . The latter, in particular, does not fully cover the expenses for nursing home visits and copayments of up to 50 per cent are standard. Copayments for professional ambulatory care are smaller and amount to an average of 247€ or about 20 per cent (Schmidt and Schneekloth, 2011). Social welfare may step in if individuals are not able to bear the copayment. Thus, the decision for formal or informal ambulatory care is usually not driven by financial aspects as each care recipient who is assigned a care level is entitled to benefits for all kinds of care.

The introduction of the insurance system in 1995 stressed the family as the main provider of care, as it is thought to provide care cheaper, more agreeable, and more efficiently. From the care recipient's perspective, the decision to receive informal care typically expresses a preference for being cared by familiar relatives or friends. In some cases, informal care recipients are additionally supported by professional carers. These are, on average older recipients with a higher care level and, thus, a higher care burden (Schulz, 2010). Apart from the care burden, a reason for professional care can be the absence of appropriate informal caregivers, either because they chose to only participate in the labour market or because their own physical or mental health conditions prohibits the full amount of necessary care provision.

From the caregiver's perspective, affection and sense of responsibility towards a loved parent or spouse mainly drive the decision to provide care. Although the insurance benefits for informal care are often passed on to the care provider this comparably small amount cannot be regarded a financial incentive to provide care, as it is also needed to cover other expenses for care provision (see Schmidt and Schneekloth, 2011 for all points). Even if the the caregiver took the benefit fully as a remuneration, the hourly rate would amount to app. the 10% quantile of the female wage distribution. However, the insurance funds do pay pension contributions for informal carers who provide care at least 14 hours a week (Schulz, 2010). In 2002, people cared on average 14 hours per week for care recipients whose assessment of needs is at least classified as the lowest official category (Schneekloth and Leven, 2003).

Between 2001 and 2011 there were only minor adjustments to the German long-term care system. They were minor because benefits were increased but only to keep pace with the inflation (Rothgang, 2010) and, thus, did not change the incentives to provide care. As of 2008, employed individuals are allowed to take a 10 day (not repeatable) unpaid leave to organize or provide care in case of an incidence of care dependency in the family. However, only very few caregivers make use of this.[6] Thus, the tasks of informal caregivers, the composition of caregivers and care recipients as well as financial incentives remained fairly similar over time.

[6] Schmidt and Schneekloth (2011) report that only 9,000 out of possibly 150,000 made use of this until 2011. The most frequent reason for not making use in their survey was that individuals were not aware of the possibility.

3 Data

3.1 Sample selection

We use data from the German Socio-Economic Panel (SOEP) which is an annually repeated representative panel survey on households and persons living in Germany (Wagner et al., 2007). Since 1984 it covers many questions on different life domains such as work, health, time use and education. On average, the survey contains about 22,000 individuals. We use data from the waves 2001 – 2013 as these include information on informal care provision.

Informal care is defined by the answer 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 for care on a typical weekday?". Around 40% of those in the sample who state a positive number report to care for one hour per day. 25% care for two hours and the remaining 35% for three or more hours. Given that this is self-reported information from the time use questionnaire, we collapse this information into a binary variable which should considerably reduce measurement error – individuals are probably much more likely to recall *any* care provision than the exact number of hours. Our treatment variable D is defined as the indicator for providing care at least one hour per day. Specifications with two or three hours as relevant thresholds are also presented below.

Figure 1: Time structure of the data

	t=0	t=1	t=2	t=3		t=7	t=8
Starting wave 1	2001	2002	2003	2004	...	2008	2009
Starting wave 2	2002	2003	2004	2005	...	2009	2010
⋮			⋮				
Starting wave 6	2006	2007	2008	2009	...	2013	
⋮			⋮				
Starting wave 11	2011	2012	2013				
Starting wave 12	2012	2013					

Source: Own illustration. This figure shows the time structure of the data. Individuals who are observed in the wave of 2001 are called to be from "Starting wave 1". These individuals can, in principle, be observed in all following SOEP waves until the year 2013. Necessary information from future waves are merged to the information from the starting wave. Information of year 2002 is defined to be of year $t = 1$, information of year 2003 is defined to be of year $t = 2$ and so on. Individuals who are observed in the wave of 2002 are called to be from "Starting wave 2". Again, future information is merged, and so on. The same individuals, but at different points in time, can appear in different starting waves.

Figure 1 displays the time structure of our data set and shows that we pool observations from different waves. Multiple observations from the same individual are taken into account by using clustered standard errors on individual level in the estimation models.

Individuals from the first wave we use (the wave of 2001) can, in principle, be followed for 13 years until 2013. Individuals from wave 11, for instance, can be followed for three years until 2013. We standardize all calendar years across waves to years $t = 0$ to $t = 8$, where $t = 0$ is merely used to define a relevant sample of individuals who did not provide care in this starting period. We drop the years larger than $t = 8$ (relevant only for the first four starting waves) in order to also have a sufficient number of observations to estimate the long-run effects. We restrict the sample to women between 25 and 64 years, since they are in prime working age. Moreover, we only use individuals with full information in all conditioning variables.

3.2 Informal care paths

Table 1 reports numbers of observations used in this study. 63,372 person-year observations (from 13,393 different women) meet the sample restrictions (most importantly no care provision in $t = 0$ and age between 25 and 64) of which 2,171 start a care episode in $t = 1$ while 61,201 do not provide care.[7] Among these observations, 16,701 are still in the sample after 8 years (577 of them provided informal care in $t = 1$). This strong reduction has two main reasons: first, recall from Figure 1 that individuals who enter the estimation sample from late waves cannot be followed over many years and, second, individuals drop out of the sample if they reach the age of 65. Thus, the long-run effects will be estimated less precisely than the short-run effects.[8]

Turning to a dynamic perspective, 771 of all women with information in year 2 provide care in both years 1 and 2. More individuals, 1,045, only provided care in $t = 1$ but not in $t = 2$. This reflects the result that most care episodes only last for one year (see below). 222 women who care both in year $t = 1$ and $t = 2$ can be followed until year 8. The two bottom lines of Table 1 also report numbers of observations for those who cared or did not care in three consecutive years. 417 women are observed to care at least three consecutive years.

Figure 2 illustrates the distribution of care durations in our sample. It shows that 60% of all care spells in the sample that start in $t = 1$ last for one period, 18% for two periods, and 7% for three periods. The median care provision duration is one year, while the average is 1.8. Note, however, that these numbers only include spells of consecutive care provision. Interrupted spells like, e.g., care in $t = 1$, no care in $t = 2$, care in $t = 3$, count as duration of one period in Figure 2. A potential reason for this interruption could be

[7]These numbers hold for the outcome variables full-time work and labor force participation and are lower for wages (where we have 37,668 person-year observations and hours worked (38,357) as these outcomes are conditional on being employed.

[8]Robustness checks with an age cut-off of 57 that close the second channel for attrition yield the same results, see below.

Table 1: Numbers of observation

Care path	Numbers of observations in year 1	2	3	4	5	6	7	8
Static model:								
Care in $t = 1$	2,171	1,885	1,604	1,379	1,156	944	718	577
No care in $t = 1$	61,201	52,727	45,801	39,036	32,581	26,532	21,033	16,124
Dynamic model:								
Care in $t = 1$ and $t = 2$		771	650	555	462	365	285	222
Care in 1, No care in 2		1,045	879	757	640	538	403	328
No care in 1, Care in 2		1,481	1,281	1,080	910	743	601	442
No care in 1 and 2		48,715	41,800	35,728	29,901	24,362	19,327	14,830
Care in $t = 1, 2$ and 3			417	347	288	229	175	136
No care in $t = 1, 2$ and 3			40,669	34,347	28,797	23,454	18,607	14,247

Source: SOEP, own calculations. Note that, for example, the sum of individuals in all four paths in year 2 does not equal the sum of individuals from the static perspective in year 2. This is because these figures are based on the estimation samples and due to missing control variables in year 1, an issue that is irrelevant for the static case (explained in Section 4) but relevant for the dynamic one (explained in Section 6).

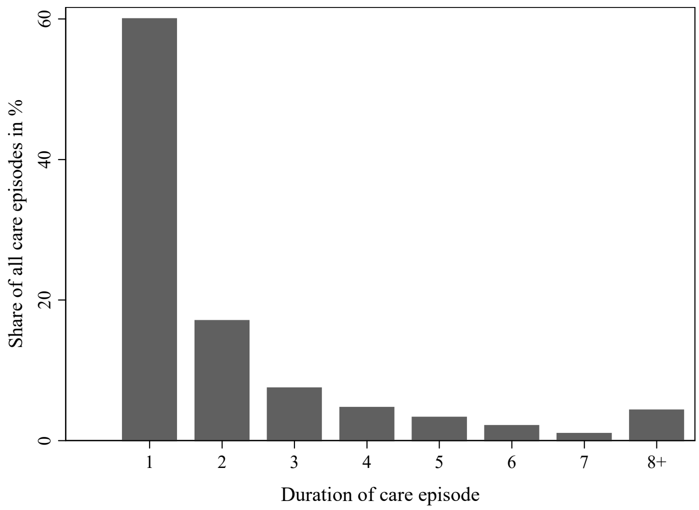
measurement error in self-reported care provision status, or, less likely, interruptions due to longer hospital or nursing home stays of the care recipient. Most likely, of course, it could also reflect the end of a care episode for a certain care recipient and a later start of a new one for another recipient. As we do not have complete information on the care recipient, this cannot be verified. Figure S1 in the supplementary materials shows the same figure for a case where we impute interrupted spells to consecutive spells.[9] This comprehensive change towards longer care spells still results in 75% of all spells lasting for up to three years (mean duration 2.3 years). The averages are in line with those reported in Müller et al. (2010) who use administrative data to estimate that care recipients receive informal care for 2.1 years, on average, in Germany and, thus, show that our care indicator seems to be a useful measure.

3.3 Outcome variables

We use four different outcome variables: an indicator of full-time work, an indicator of being employed, weekly hours worked (conditional on positive hours) and gross hourly wages (conditional on positive hours). Full-time work and employment are taken from the subfile "generated variables" in the SOEP and are based on a question on current employment status. Being employed means either working full-time, part-time, vocational training, or marginal and irregular part-time employment. Non-employment includes non-working individuals, those in military/community service, maternity leave, and employed persons in a phased retirement scheme whose current actual working hours are

[9] In a robustness check we also perform the main analysis using this sample with imputed care spells and the results hardly change.

Figure 2: The distribution of care spells



Source: SOEP, own calculations. This graph shows the distribution of episodes of consecutive care of at least one hour per day. The data are restricted to starting waves 1 to 5 as defined in Figure 1 to ensure that every spell can last for at least 8 years. Note that the vast majority of individuals does not provide care at all (about 90% of all individuals).

zero (SOEP Group, 2014). Working hours are current actual average working hours (including overtime) as reported by the individuals and not contracted working hours. Implausible answers are replaced to missing values by the SOEP group (SOEP Group, 2014). Gross hourly wages are defined as (deflated) gross monthly labor income divided by the product of 4.3 and weekly hours worked.

Table 2: Sample means of outcome variables by care status

	Caregivers in $t = 1$	Non-carers in $t = 1$
Full-time, $t = 1$	0.27 (0.44)	0.36 (0.48)
Full-time, $t = 0$	0.31 (0.46)	0.35 (0.48)
Employed, $t = 1$	0.73 (0.44)	0.77 (0.42)
Employed, $t = 0$	0.74 (0.44)	0.77 (0.42)
Hours, $t = 1$ if > 0	31.22 (13.94)	32.27 (13.18)
Hours, $t = 0$ if > 0	32.12 (13.81)	32.39 (13.23)
Hourly wage, $t = 1$	13.98 (8.18)	14.22 (8.74)
Hourly wage, $t = 0$	13.86 (7.99)	14.18 (8.78)

Source: SOEP, own calculations.

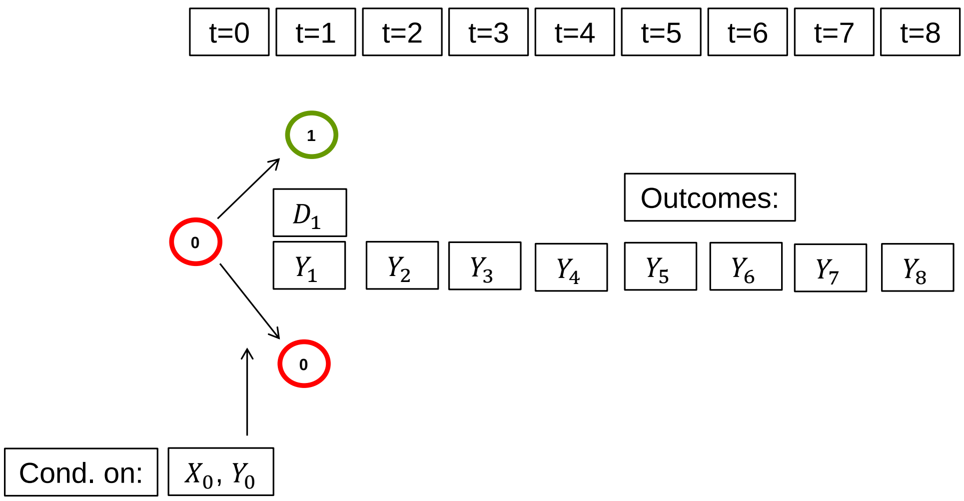
Table 2 shows sample means of the outcome variables in years 0 and 1 stratified by caregiver status in year 1. Non-carers have higher labor force participation and wages than caregivers. For instance, while the likelihood to work full-time is 36% for non-carers, it is 27% for carers. Somewhat less pronounced, yet significant differences can be found for the other variables. It remains to be seen whether these short-run differences are due to care provision or just reflect different compositions in both groups – and whether they are persistent if they are, at least in part, due to care provision. Obviously the groups of caregivers and non-carers do differ significantly with respect to their labor market attachment even without care provision. The table also shows average pre-treatment outcomes of year 0 – when both groups do not provide care – and only slightly less pronounced differences between both groups can be observed. Thus, it seems central to control for previous outcomes.

4 Baseline analysis

4.1 Empirical strategy I – A static design

We are interested in the effect of caregiving on labor market outcomes, both contemporaneously and up to eight years later. Figure 3 describes the basic design. In period 1, individuals receive the binary treatment D_1 (for all random variables to come, subscripts denote time in years), which could either be care provision ($D_1 = 1$ and a green circle in Figure 3) or no care provision ($D_1 = 0$ and a red circle).[10] We restrict the analysis to the subsample of individuals with $D_0 = 0$, that is, those who did not provide care in $t = 0$. We then observe outcomes Y_1 to Y_8. Y_t stands for the four different outcome variables.

Figure 3: Static design



Own illustration.

Each individual has two potential outcomes per period, Y_t^1 and Y_t^0, where the superscripts denote potential outcomes with or without care provision in period 1. The causal effect

[10] Note that, for simplicity, we drop subscripts i denoting individuals, such as D_{i1}, throughout the paper.

of providing care in period 1 on labor market outcomes in period t is $Y_t^1 - Y_t^0$. This individual treatment effect is a well-defined parameter but impossible to determine for the researcher as only the factual but not the counterfactual outcome is observed – the observational rule is $Y_t = D_1 Y_t^1 + (1 - D_1) Y_t^0$ if we only define it in terms of D_1 irrespective of any dynamics in D_t over time.

Ideally, we would like to randomly assign individuals to informal care in $t = 1$, follow them over the years and evaluate how they perform on the labor market compared to those who are not assigned to caregiving. This experiment would allow us to assess the average causal response of starting care in $t = 1$ irrespective of how long the person actually provides care. However, we have observational data and care is most naturally a voluntary decision. People individually (potentially altruistically) weight costs and benefits to make a choice that is roughly based on the opportunity, the willingness, and the ability to provide informal care. While it is not our aim to fully model the decision to provide care, we hope to control for all variables that affect both, treatment and outcomes, leaving the decision to care a random event (conditional on controls). In other words, in order to identify causal effects, we make a conditional independence assumption:

$$Y_t^1, Y_t^0 \perp\!\!\!\perp D_1 | X_0, Y_0 \qquad \forall t > 0$$

Below we go into detail about which variables we account for in X_0 and justify this assumption. We exploit detailed information on individuals' socio-economic background, potential caregiving obligations, health and, most importantly, pre-treatment outcomes Y_0 which capture a lot of time-invariant unobserved heterogeneity such as general attitudes towards labor market participation and other hard-to-measure factors such as intrinsic motivation, time preferences, or personality traits. The identifying assumption here is that, conditional on all covariates and past outcomes, the observed treatment is random.[11] We will relax this assumption later in Section 5.2.

The parameter we want to estimate is the sample average treatment effect on the treated (ATT), that is, the causal effect of caring for all caregivers in the sample. The identifying assumption enables us to use $E(Y_t^0 | D_1 = 0, X_0, Y_0)$ as a surrogate for the counterfactual $E(Y_t^0 | D_1 = 1, X_0, Y_0)$, and, hence, we overcome the identification problem and can calculate $ATT_t = E(Y_t^1 - Y_t^0 | D_1 = 1, X_0, Y_0)$. We use propensity score kernel matching and inverse probability weighting to achieve this.[12]

[11] The set of assumptions is completed by a common support assumption, the stable unit treatment value assumption and the assumption that no control variable is a direct product of the treatment. The first one is naturally fulfilled by restricting the sample to those individuals in the treatment and control group that share a common support of the propensity score. The third one is most likley fulfilled by including variables only that are measured one year before the treatment.

[12] As the results hardly differ at all between IPW and Matching, we will only report Matching results for the static version below. A comparison between IPW and Matching is shown in the Appendix in Figure A2.

Finally, note that the treatment status is only defined in $t = 1$ and not affected by later care provision as this might be endogenously affected by future realizations of the outcome or control variables. This is partly relaxed in a dynamic specification in Section 6. In a robustness check of the static version we restrict the control group to individuals that never provide care throughout the full observation period. This cannot be a preferred specification as the definition of the control group depends on future caregiver status. Yet, as the results do not differ compared to the baseline specification (see Figure A3 in the Appendix), it makes a strong case that it is no big issue that the control group according to the baseline definition above also includes women who will provide care after $t = 1$.

4.2 Control variables

The selection of the control variables to estimate the propensity score is crucial in order to make the conditional independence assumptions credible. Here we can make use of the major strength of high quality survey data: the abundance of individual level variables that potentially affect both treatment (paths) and potential outcomes as well as their changes over time. While a drawback compared to many administrative data sets is the comparably small sample size, the advantage is the widespread information on topics such as socio-economic background, health, or preferences that are usually not available in administrative data sets. We consider this crucial for the identifying assumptions.

In deciding for informal care provision one might have three basic blocks of prerequisites in mind. Individuals decide to provide care if (*i*) they need to, if (*ii*) they are willing to, and (*iii*), they are able to provide care. As of (*i*), individuals are only in the position to decide for care provision if someone close becomes care dependent. We model the intra-social environment by using indicators whether parents are alive, parents' age as well as the number of the potential caregiver's siblings. The latter can reduce the need to provide care for frail parents as siblings could step in.

As of (*ii*), we select socio-economic characteristics as covariates that also control for the willingness to provide care. This set contains age bin dummies, binary variables on marital status (married, divorced, widowed), whether children live in the household, as well as whether the individual is foreign born. Furthermore, we use character traits measured in the Big Five Inventory (Big5), 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. The items of the Big5 are: neuroticism, extraversion, openness, agreeableness, conscientiousness.[13] For each personality measure, the score is generated

[13]More specifically: 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

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,[14] they are useful controls because these measures are supposed to be stable over a shorter period of time. The individual average of each measure is taken over all years as a proxy for time invariant personality.

As of (*iii*), the own health status determines the ability to provide care. Here, we control for self-rated health, the number of doctor visits in the previous three months and the number of hospital visits in the previous year. Finally, we include pre-treatment outcome variables, regional dummies for the 16 federal states as well as a full set of year dummies. We interact all control variables with each other and include squared terms (as long they are not dummy variables) in order to capture potential non-linearities in the determinants to provide care. This, however, results in an extremely large number of control variables that is unfeasible to manage. Thus, we strongly reduce the dimension of the vectors of controls by using Lasso.[15] Specifically, we follow the "double selection"-procedure suggested by Belloni et al. (2014):

1. Select variables (using Lasso) from the full set of (X_0, Y_0) including interactions and squared terms that are relevant to predict D_1.
2. Select variables (using Lasso) from the full set of (X_0, Y_0) including interactions and squared terms that are relevant to predict Y_t.
3. Use the union of variables from steps 1 and 2 as controls.

This procedure works if the "approximate sparsity assumption" (Belloni et al., 2014) holds which, stated verbally, implies the following: the chosen subset using the procedure described above leads to an approximation of the true relationship between outcome and controls, where the approximation error is sufficiently small. Thus, if the CIA holds given the full set of controls and their interactions, it approximately also holds for the chosen subset of controls. All control variables and their sample means are reported in Table A1 in the Appendix. An example for the finally chosen ones by the double selection procedure is given in the supplementary materials.

on a 7-item scale. Although the SOEP captures each item of the Big5 with relatively few questions in the 2005 and 2009 questionnaires, surveys revealed sufficient validity and reliability (see Dehne and Schupp, 2007). Furthermore, there is positive reciprocity, the tendency of being cooperative and negative reciprocity, the tendency of being retaliatory.

[14] The Big5 are included in the surveys in 2005 and 2009, whereas questions on negative and positive reciprocity are asked in 2005 and 2010.

[15] We use the Stata ado lassoShooting provided by Christian Hansen on his website. Of course, any errors are our own responsibility.

4.3 Potential failure of the CIA

We can build on a large number of controls to condition on. These include observed preferences as well as predictors of general labor market prospects like age and education. Pre-treatment outcomes should capture time-invariant unobserved factors that might affect both care provision and labor market outcomes. These might, again, be general preferences but also baseline health, such as general unobserved frailty.

Yet, as we only condition on variables in $t = 0$ there might be shocks between $t = 0$ and $t = 1$ that both affect care provision and labor market participation which we are unable to capture. One such shock could be a health shock of the potential caregiver. Most likely this would lead to both a reduction in care provision and in labor market participation as an individual might not be able to perform any of the two tasks. If this is indeed the case and we do not account for a health shock, this would lead to an underestimation of the true effects (in absolute values). This is because, as those who provide care are healthier (conditional on other controls such as age), part of the positive effect of good health on labor force participation would falsely be ascribed to care provision.[16] A similar reasoning holds for the death of a parent or spouse. This should, on average, reduce the likelihood to provide care, if the parent was a care recipient. Moreover, if there is an own effect of a parent's death on labor force participation, it should be negative.

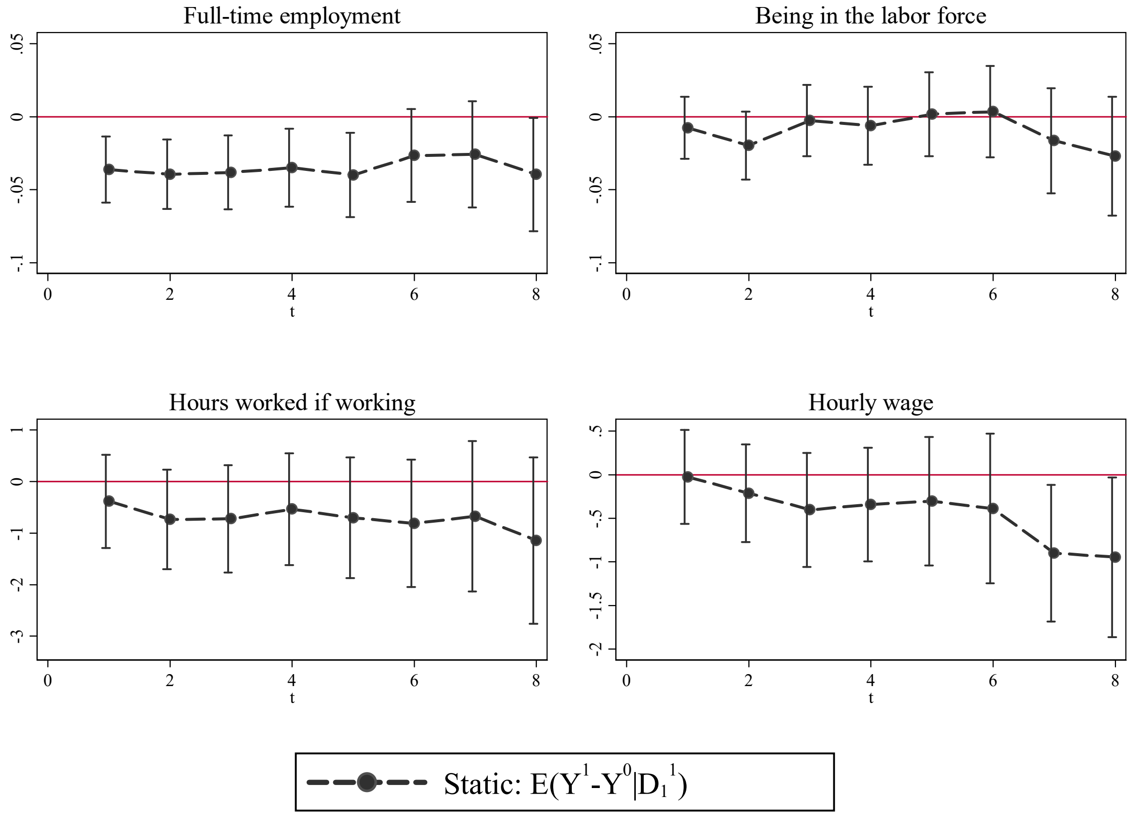
In Section 5 we present four types of analyses to defend our identification strategy. We, as a simple measure, drop individuals who experienced a health shock or who lost a parent between wave 0 and 1 and delete the two probably most important time-varying confounders. Thereby, we see that the health effect on care provision is not very strong and certainly not able to drive the results. More sophisticated, we openly allow for additional confounders by simulating them and taking them into account. By this we determine bounds that most likely include the true effects. Moreover, we use placebo estimates and check for common trends between carers and non-carers before they provide care. Finally, we scrutinize the results with respect to potential systematic measurement error in the treatment variable.

[16] This can also be seen in the omitted variable bias formula for the simple regression case. If the "true" model is $work = \beta care + \gamma health\ shock + \varepsilon$ and *health shock* is omitted in the estimation, the expectation of $\widehat{\beta}$ equals $\beta + \gamma\delta$ where δ is the coefficient of *health shock* in a regression of *care* on *health shock*. As both β, γ, and δ are probably negative, the expected value of $\widehat{\beta}$ should be smaller in absolute size than the true β. While this formula, in theory, does not hold for the multivariate model as the correlation structure between all control variables is also relevant, in practice it gives a first idea of the potential direction.

4.4 Estimation Results – Static model

Figure 4 reports the effects of caring in year 1 on all outcome variables across time until year 8. These are matching results.[17] Table A2 reports the exact numbers, standard errors, and bandwidths used. Figure A2 in the Appendix compares matching results to IPW. The differences are negligible. Therefore, we stick to matching as the method that is more standard in applied econometrics. The upper left panel shows the findings for full-time work. The probability to work full-time is reduced by around 4 percentage points (ppts.) when women start to provide care. This is a considerable effect given an average probability to work full-time of around 35 per cent. Moreover, it persists over the entire observation period, although the confidence bands widen over time due to fewer observations. As only a small fraction of caregivers in $t = 1$ still (or also) provides care in year 6, 7, or 8, this can be interpreted as evidence that some women leave full-time employment due to care obligations and, later, when the care spell ceased, do not return to full-time employment.

Figure 4: Labor market effects of informal caregiving for females – Static version



Source: SOEP. Own calculations. Note: The graph shows the point estimates and the 95% confidence intervals.

The estimates on being in the labor force (upper right panel) are very close to zero and thus insignificant throughout. Conditional working hours (lower left panel) are reduced

[17] Figure A1 in the Appendix reports matching quality for full-time work as an outcome variable and shows that covariate balance is achieved by matching. The results are comparable for the other outcomes.

by a little less than one hour, on average. Over time, the effect is persistent and slightly increasing but never significant and always rather small in magnitude. A quick back-on-the-envelope calculation shows how these findings fit together. Given that 77% of women in the sample are employed and 31% work full-time (42.2 hours on average, actual working time), 46% work part-time (23.8 hours on average). Assuming that no women leaves the labor force due to care provision but 4%-points switch to part-time work reduces the average conditional working hours from 32.1 to 31.2. This would imply a reduction by 0.9 hours, close to the observed effect.

Another potential labor market effect are wage penalties. This is explored in the lower right panel of Figure 4 where we assess the effect on hourly wages for all employed females. This graph reveals zero contemporaneous wage-effects. It seems to be, however, that fairly small negative effects add-up over time and become sizeable and significant after a couple of years. One reason for this could be forgone promotions due to caregiving with lagged effects that only materialize some time after care provision. Eight years after care provision, working women have a by 1 Euro per hour lower gross hourly wage which is around 7% in relative terms.

5 Scrutinizing and relaxing the CIA

So far we interpreted our estimates as causal conditional on the validity of our identifying assumption (the CIA) and found, in particular, considerable and persistent negative effects on full-time employment. We justified the identifying assumption by fully exploiting the panel information in the SOEP and using many observable characteristics to match on. As we also match on pre-treatment outcomes, time-constant unobserved heterogeneity that probably explains a lot of the willingness and ability to provide care is also taken into account.

Nevertheless, there might be other confounding factors that we do not observe. Whether this is the case in our study is inherently non-testable. Thus, we do not and cannot say anything in this section concerning the likelihood that our assumption is fulfilled. Rather, we want to ask how crucial certain plausible deviations are for our results. As the CIA is not an "all or nothing" assumption, different degrees of its violation still allow to bound causal effects and to receive meaningful parameters. We restrict the analysis to full-time in this section.

5.1 Delete certain individuals

Taking up the discussion of Section 4.3, the two reasons for potential failure of the CIA that immediately come to mind are a health shock or a death of a parent or partner between period 0 and period 1. As, in particular, we cannot identify in the data whether a health shock between 0 and 1 was due to caregiving or, the other way around, care responsibilities were not taken up due to a health shock, we cannot account for a health shock, as this is potentially a "bad control".

In a robustness check, however, we identify all individuals who experienced a health shock or a death of a parent and exclude them from the analysis to see whether they affect the findings. Not uncommon to the health economic literature (see e.g., García-Gómez, 2011), we use the self-stated health on a 5-point scale to define a measure of health shock. In order to allow for a wide definition, we define a health shock as a deterioration to either "bad" (category 4) or "very bad (category 5)". This includes 4,537 person-year observations. A stricter condition of a reduction by at least two categories and to either "bad" (category 4) or "very bad (category 5)" is fulfilled by 1,643 person-year observations. Moreover, 1,005 had to suffer from the loss of a parent or spouse. Figure 5 reports the findings where individuals according to the wide definition of a health shock between 0 and 1 and those who have lost a parent are excluded from the sample. The results are statistically indistinguishable from the baseline results.

Figure 5: Exclusion of individuals with potential health shock or death of a parent between 0 and 1

Source: SOEP. Own calculations. Note: The graph shows the point estimates and the 95% confidence intervals.

5.2 Partial identification

In this section 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. This analysis is also in the spirit of the one suggested by Altonji et al. (2005) without the need to make strong parametric assumptions.

5.2.1 General framework

We now assume that the CIA does not hold

$$Y_t^0 \not\perp\!\!\!\perp D_1|X_0, Y_0 \qquad \forall t > 0$$

but that the failure is due to an unobserved variable U_0. Could we condition on it, we had

$$Y_t^0 \perp\!\!\!\perp D_1|X_0, Y_0, U_0 \qquad \forall t > 0.$$

Hence, all the unobserved heterogeneity that leads to assumed endogeneity problems is captured by U_0. For simplicity, we assign U the time indicator 0 but it could also be a variable measured between 0 and 1. To keep things as simple as possible, Ichino et al. (2008) follow Rosenbaum and Rubin (1983) who proposed U_0 to be binary. This is appealing, since the distribution of a binary variable is fully determined by its mean. To describe how U_0 affects both treatment and outcome, four probabilities p_{ij}, $i \in \{0,1\}$; $j \in \{0,1\}$ are defined as

$$\begin{aligned}
p_{01} &= Pr(U_0 = 1|D_1 = 0, Y_t = 1) \\
p_{00} &= Pr(U_0 = 1|D_1 = 0, Y_t = 0) \\
p_{11} &= Pr(U_0 = 1|D_1 = 1, Y_t = 1) \\
p_{10} &= Pr(U_0 = 1|D_1 = 1, Y_t = 0). \qquad (1)
\end{aligned}$$

Treatment status D_1 and binary outcome category Y_t are observed in the data and, hence, individuals can be assigned one of the four probabilities p_{ij} where i denotes treatment status and j the outcome. The four above equations fully define the distribution of the hypothetical confounding variable U_0. Differences in these probabilities will mechanically introduce a correlation between U_0 and both, D_1 and Y_t and, thus, U_0 will be an important confounding factor.

Given values for p_{ij} we simulate U_0 by drawing 200 times from Bernoulli distributions with the respective parameters for each individual and estimate the ATT 200 times, conditioning on X_0 and Y_0 as before, but also on U_0. Taking the average over all results provides us with point estimates as well as standard errors of the average treatment effect where the CIA is relaxed.[18]

We follow Ichino et al. (2008) and set p_{ij} such that we control the "outcome effect" (the relationship with Y_t) and the "selection effect" (the relationship with D_1) of U_0. As an illustration, think of U_0 as a health shock again that both affects the probability to work and to provide care. $U_0 = 1$ indicates a health shock, $U_0 = 0$ means no health shock. This unobserved variable certainly has a negative and strong selection effect such that unhealthy people are less likely to provide care. It may also have a negative outcome effect. More formally, Ichino et al. (2008) define the parameter $s = p_{1.} - p_{0.}$ as the selection effect where

$$p_{i.} = Pr(U_0 = 1|D_1 = i) = p_{i0} \cdot P(Y_t = 0|D_1 = i) + p_{i1} \cdot P(Y_t = 1|D_1 = i) \quad i \in \{0,1\}.$$

The larger this effect, the larger is the effect of U_0 on selection into treatment keeping the outcome fixed. The outcome effect, defined as $d = p_{01} - p_{00}$ reflects the correlation between U_0 and the untreated counterfactual outcome. As an example, an outcome effect of $d = -0.05$ means that, in the group of non-carers, among those who work the likelihood to experience a health shock is 5 percentage points smaller. The higher d the stronger is this correlation. Likewise, a selection effect of $s = -0.05$ implies that among caregivers the likelihood of $U_0 = 1$ is lower than among non-caregivers. Given these settings, U_0 is a variable that is both correlated with treatment and outcome and should be accounted for in the estimations. Once we set values for d and s we can derive the four p_{ij} by solving an equation system and simulate U_0.[19]

5.2.2 Choice of selection and outcome effects

In principle, d and s could be arbitrarily chosen and certainly such that the identified effect of care provision on full-time work turns zero or even positive. This, however, does not deliver any useful information as it is always possible to reduce the set of assumptions so far that zero is included in the identified bounds (see the worst-case bounds by Manski, 1995, that always include a treatment effect of zero). Thus, a major challenge is to find reasonable deviations from the CIA.

We follow the reasoning by Altonji et al. (2005) and argue that we have a high quality panel data set that allows to observe a large amount of variables determining care provision and labor force participation. Among them are baseline health, age, education, preferences, and pre-treatment outcomes. While we certainly do not capture everything

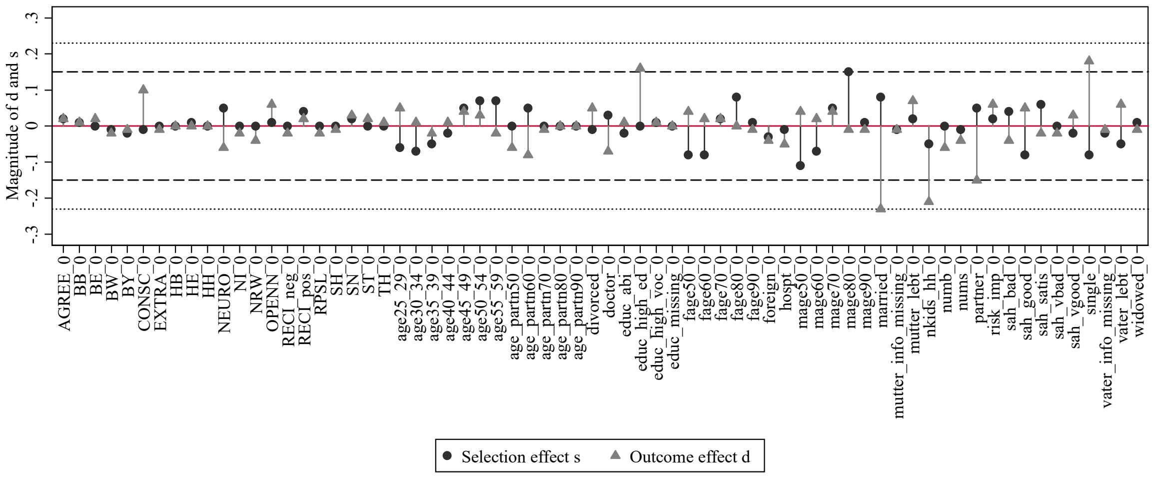
[18]We use a modified version of the Stata command sensatt that is written by Nannicini (2007).

[19]Using the two equations above, also assuming that $p_{11} - p_{10} = 0$ and assuming a value for $P(U_0$ we have four equations and can determine the four unknown p's.

it seems hard to imagine that there a unobservables with a drastically higher impact than the observables have.

Thus, one way to find reasonable values for d and s is to go back to equation system (1) and – starting the other way around – use observed binary variables in the data set, substitute them for the unobserved U_0 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_0 might 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. Results are reported in Figure 6 and, more detailed, in Table S1 in the supplementary materials.

Figure 6: Parameters for calibration of the sensitivity analysis



Source: SOEP. Own calculations. Note: See Table A1 for translations of variable names. All values are reported in Table S1 in the supplementary material. Full-time work is not shown here. Year dummies are not reported for legibility but have values of d and s of virtually zero.

Figure 6 reveals that most of the variables have selection and outcome effects of at most 0.1 in absolute values. The dashed black line marks the interval of selection effects all observable variables fall in (± 0.15), while the grey dotted line (± 0.23) does the same for the outcome effect. The only exception is the pre-treatment outcome full-time work – which is left out in the figure – that has a selection effect of $s = -0.05$ (among the carers, the pre-treatment full-time employment rate is 5 ppts. lower), and an outcome effect of $d = 0.8$ (the probability to have worked full-time in $t = 0$ is 80 ppts. higher among those who work in $t = 1$ than those who do not work in $t = 1$). The huge value for full-time work mainly reflects the path dependence in full-time work and, therefore, this pre-treatment outcome variable is not very helpful to find credible values for s and d.

We test three different combinations of d and s to bound the treatment effects under weaker assumptions than the CIA. First, we calibrate a confounder U_0 to have the same bivariate correlation with treatment and outcome as pre-treatment full-time work. Even though this should not be the most interesting case as this variable is somewhat particular, this provides a first benchmark. Next, we assume a left out variable U_0 that has a correlation much stronger than all other observed variables by using $d = -0.23$ and $s = -0.15$. If there was indeed such a variable and we took this into account, the treatment effect should increase in absolute values (get further away from zero). The example for such a variable discussed before was a health shock or the death of a spouse. It should be noted, however, that even the health shock considered in Section 5.1 has outcome and selection effects of far less than 0.1 and, thus, should be exceeded by this variable. In total, this variable U_0 is linked much stronger to treatment and outcome than any of the observed variables.

As the second specification increases the treatment effect, we want to challenge our results by using $d = -0.23$ and $s = 0.15$. This parameter combination will push the treatment effects towards zero. As it seems to be hard to imagine a variable with such a drastic effect and in a direction opposed to the one discussed by a health shock, this could be seen as a credible lower bound of the true effects. Again, note that a handful of variables either has a larger d or s but none has both parameters at such high levels.

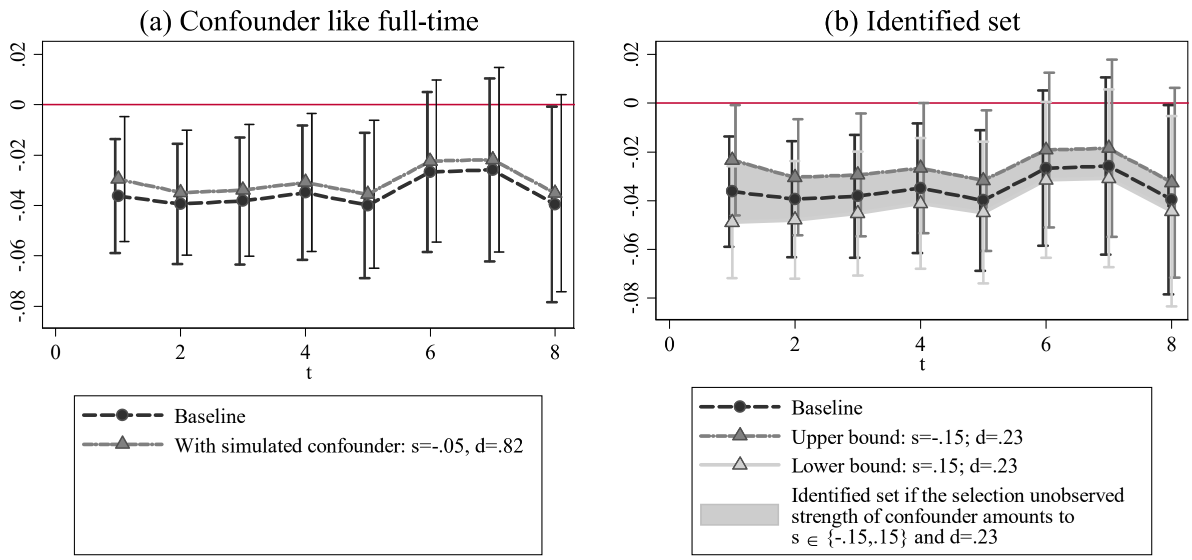
5.2.3 Results

Figure 7(a) shows the resulting effects for a model that includes U_0 calibrated to have the same effect as the pre-treatment outcome (light grey triangles) and relates it to the same model without such a confounding factor (represented by black circles; shown before in the upper right panel of Figure 4). The difference in the effects of both models is statistically indistinguishable and also the magnitude of the effect is fairly similar.

Figure 7(b) shows the set that we can identify if we had knowledge that the impact of the unobserved heterogeneity is bounded between $s \in \{-0.15, 0.15\}$ and $d = 0.23$ (grey-shaded area). The resulting estimates are bounded between -2.4 and -5.0 percentage points. If there was a variable we left out that worked like a health or other life event shock but had a much stronger impact than observed health shocks by our definition above, the true effect would be around -5.0. If the left-out variable had an opposing effect on treatment or outcome – somehing we could not come up with a realistic example for – and, again, it had a very strong effect on both, we would still identify an effect of -2.4 percentage points.

To sum up, even if there was another confounding factor with effects as extreme as the pre-treatment outcome or considerably stronger than any of the other observed controls: conditioning on it would only partly reduce the magnitude of effects. Given the large amount of other variables we control for – and that these variables have much smaller

Figure 7: Sensitivity analysis for full-time employment



Source: SOEP. Own calculations. Note: The graph shows the point estimates and the 95% confidence intervals.

selection and outcome effects even though these are variables as important as age, education, parental characteristics and personality traits – it seems hard to imagine unobserved variables with even more drastic effects that would, if we conditioned on them, destroy the results. Thus, even if the CIA were not to hold, our effects would remain fairly stable and all our conclusions from Section 4 sustain.

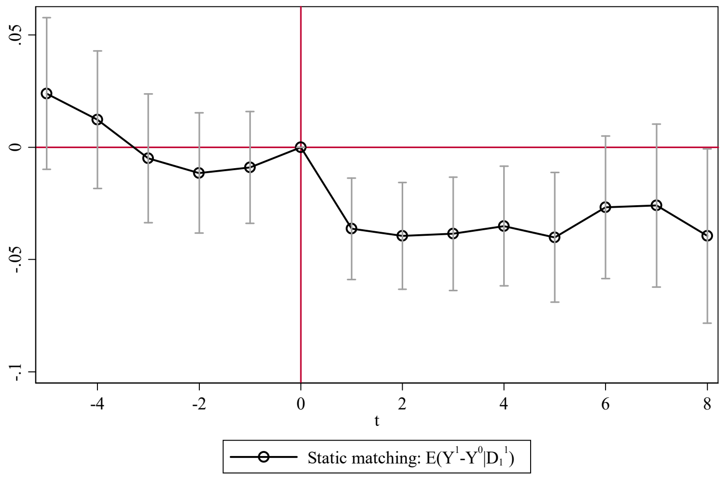
5.3 Pre-treatment Trends

In the following we test whether there are potential anticipatory effects of future care provision. To do so, we repeat the baseline static matching procedure from Figure 4 also to outcomes in the years before treatment. More specifically, we include placebo estimates that, as an example, use care in $t = 1$ as the treatment, full-time work in $t = -2$ as an outcome and all other controls in $t = 0$ to match on. The results are shown in Figure 8. Apparently, there is no significant pre-treatment change in full-time employment due to later care provision. The only significant drop here takes place in period one, which is the already familiar and persistent 4 ppts. reduction.

5.4 Assuming adverse measurement error

A further scenario in which we would falsely attribute the observed correlation between labor supply and informal care to the effect can arise under presence of non-classical measurement error. Assume a situation where individuals that suffer from unemployment

Figure 8: Pre-treatment trends for full-time employment



Source: SOEP. Own calculations. Note: The graph shows the point estimates and the 95% confidence intervals.

falsely report a positive amount of hours spent caring in order to justify their unemployment. This would inflate our estimates.

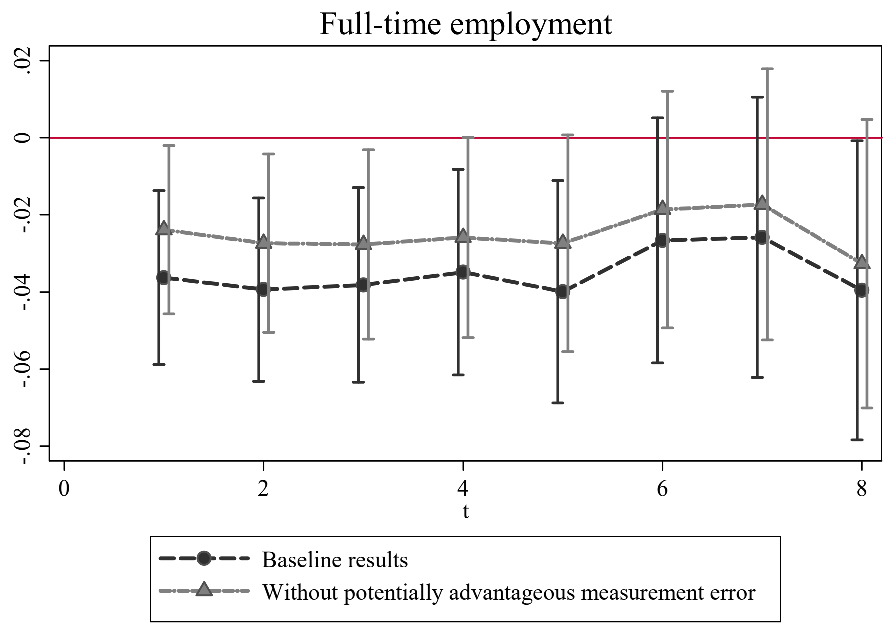
By assuming a worst case scenario, we reassign all individuals who stop working between $t = 0$ and $t = 1$ to the control group of non-carers (independent of their reported care status). As the majority of individuals does not state to provide care anyway, this effectively only changes the treatment status of 84 women who stopped working and report a positive amount of hours cared. Yet, this change will mechanically drive our estimates towards zero as we absorb some of the observed correlation that adds up to our baseline effects. The major question is just how strong. Figure 9 shows this impact on the results (gray triangles). With such a drastic measurement error where each individuals that gave up her job falsely reported to also provide care, the effects of care provision on full-time work would change to the region of the previously seen lower bound from Section 5.2 but remained statistically and economically important.

6 A dynamic design

6.1 Empirical strategy II

The previous approach answers a relevant question: given that a women provides care today, what effects can she expect for her labor force status today, in one year, in eight years? Given that the treatment is defined in year 1 only, this effect is a mixture of different care provision paths later on. The treatment group consists of individuals that provide only one year of care but also of those who care for two consecutive years, three years

Figure 9: Impact of assumed adverse measurement error, full-time work



Source: SOEP. Own calculations. Note: The graph shows the point estimates and the 95% confidence intervals.

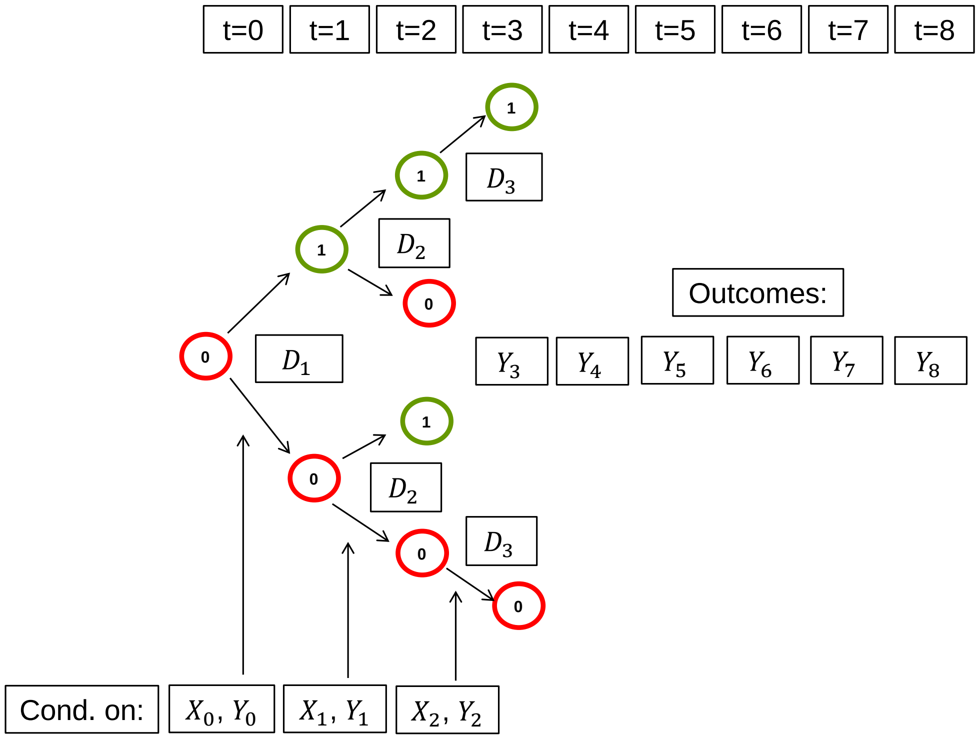
and any other care spell (like care, no care, care, no care,....). Likewise, the control group includes individuals who take up care provision later on.[20]

A potentially more interesting ideal experiment would be to assign women randomly to different paths of care. By this means, not only the start of caregiving is randomized, as in the static setting. Also, the selection out of care is controlled for by dynamic attributes. The main advantage of such an approach is that we can relate effects of different care paths to one another, for instance, in order to see whether the static effect is dominated by one particular path. Thus, it seems natural to ask whether the (long-run) effect of providing more consecutive years of care differs from providing (at least) one year. This, however, considerably complicates estimation as time-varying control variables and outcomes along the care provision path potentially affect the decision to stay caregiver or to cease. Lechner (2009) suggests an approach that is able to capture the effects of different treatment paths where it is decided sequentially at different nodes on a decision tree whether the treatment is continued or stopped (or, more generally, another treatment is taken). In an example, Figure 10 shows an excerpt of potential paths (D_1, D_2, D_3), where $D_t \in \{0, 1\}$ for $t = 1, 2, 3$, that can be taken on in three years. Dynamic matching/reweighting means that we add a time dimension to the matching/reweighting process. As static matching aims to control for any differences in observed characteristics just prior to the caregiv-

[20]However, see a robustness check in Figure A3 in the Appendix where we restrict the control group to individuals that never provide care throughout the full observation period. The results do not differ. This is not the preferred specification as the definition of the control group depends on future caregiver status.

ing decision, dynamic matching also balances time differences in the controls that may influence any particular care path.

Figure 10: Dynamic design



Own illustration.

We argue that this partly dynamic modelling of the care dynamics – up to three consecutive years – is sufficiently interesting. First of all, most careprovision spells in our data set have a short term nature. 60% of care spells in the sample last for one period, 17% for two periods, and 7% for three (see Section 3). Even a strong effect differential between care and no care in $t > 3$ would, thus, not have a significant average impact on the overall effect of caring for at least three years as only a couple of individuals has longer care episodes. Yet, it is potentially relevant to model the dynamic decision for three years (instead of fully sticking to the static model) at least for the following reason. Often, when individuals enter their care spell, they do not have enough information to build an expectation on the duration of the spell and on the burden they take on. Thus, it is conceivable that many individuals keep their labor market participation unchanged in the first period. It is probably fair to assume that those who have already provided care for two years have a fairly good (though certainly not perfect) idea of their ability to take on the double burden of working and caring and at least more information (if only vague) on the potential future duration of the care spell. Thus, while at the first node, many individuals potentially make an explicit short-term decision on their labor market participation during the care episode, this is more likely to be a longer term decision – meaning a decision for the full care episode – at the second or third node and it appears to be interesting to explicitly look at the effects of caring at least three consecutive years on labor market outcomes.

In the following we fully draw on Lechner and Miquel (2010) and only very briefly sketch their ideas for the dynamic model, restricting the outline to two periods. The interested reader is referred to Lechner (2008, 2009) and Lechner and Miquel (2010) to find derivations and in-depths discussions of the full model (including identification and estimation). The observational rule for two periods of potential care provision reads

$$Y_t = D_1 D_2 Y_t^{11} + (1-D_1)D_2 Y_t^{01} + D_1(1-D_2)Y_t^{10} + (1-D_1)(1-D_2)Y_t^{00}, \qquad t \geq 2$$

where D_2 is careprovision in period 2 and Y^{11} the potential outcome of caring two consecutive periods and analogously for the three other potential outcomes.[21]

We could be interested in the effect of caring for two consecutive years as opposed to not caring two consecutive years for the group of individuals who provide care in the first year. This is a kind of an average treatment effect on the treated (for those treated in period 1) and can directly be compared to the ATT from the static version. In technical terms, we want to know

$$DATT_t = E(Y_t^{11} - Y_t^{00} | D_1 = 1)$$

where $DATT_t$ is called the dynamic treatment effect on the treated. Of course, effects for other differences in potential outcomes (that is, other treatment paths) and other subpopulations (e.g. the full population of individuals who did not provide care in $t = 0$) can, in principle, be calculated as well.

Estimation of this effect, again, amounts to finding observable outcomes that can be used to estimate the unobservable counterfactual outcomes. In essence, this is finding individuals that took on exactly the two paths ($D_1 = 1, D_2 = 1$) and ($D_1 = 0, D_2 = 0$) but share – except for the treatment, or parts of the treatment – the same characteristics as the subpopulation we want to calculate the $DATT$ for, here, the caregivers in period one, $D_1 = 1$. This amounts to the "weak dynamic conditional independence assumption" (Lechner and Miquel, 2010):

1. $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp D_1 | X_0, Y_0$
2. $Y_2^{00}, Y_2^{10}, Y_2^{01}, Y_2^{11} \perp\!\!\!\perp D_2 | X_1, X_0, Y_1, Y_0, D_1$

This means that conditional independence is assumed to hold at each node and is achieved by sequentially modelling all transitions between two years (e.g., the one from $t = 0$ to

[21]Note, when $t > 2, \quad Y_t^{k,l}, \quad k,l \in \{0,1\}$ is a mixture of all those potential outcomes that follow in the care path after the sequence $[k,l]$.

$t = 1$, then the one from $t = 1$ to $t = 2$ and so on) and, thereby, conditioning on each node for the full set of pre-treatment control variables. For instance, at the transition from $t = 1$ to $t = 2$ we control for X_0 and X_1 and, again, previous outcomes Y_0 and Y_1. This explicitly allows for individuals who started to provide care in $t = 1$ and then, in $t = 2$, stopped caregiving due to effects of care provision on either control variables (for instance a drop in own health) or on labor market outcomes. Given that characteristics of potential care recipients are also in the set of controls, this also allows for stopped care provision because there was no need anymore (e.g., because the care recipient passed away). Thus, we explicitly take into account changes in control variables and outcomes over time to explain why individuals take on different treatment paths. The remaining reasons to choose different paths are assumed to not be systematically related to the individual's potential outcomes.

Estimation, thus, involves several steps that we outline here. In contrast to the static version we fully restrict the analysis to inverse probability weighting and do not use matching.[22]

1. Estimate the propensity score for the decision on the first node ($Pr(D_1 = 1|X_0, Y_0)$) and the two propensity scores (depending on the decision in the last period) for the second node ($Pr(D_2 = 0|D_1 = 0, X_1, X_0, Y_1, Y_0), Pr(D_2 = 1|D_1 = 1, X_1, X_0, Y_1, Y_0)$)

2. Define the relevant dynamic treatment and control group

$$D = \begin{cases} 1 & \text{if } (D_1 = 1) \cdot (D_2 = 1) = 1 \\ 0 & \text{if } (D_1 = 0) \cdot (D_2 = 0) = 1 \end{cases}$$

3. Compute the inverse probability weights:

$$W = \begin{cases} \dfrac{1}{Pr(D_2 = 1|D_1 = 1, X_1, X_0, Y_1, Y_0) \cdot Pr(D_1 = 1|X_0, Y_0)} & \text{if } D = 1 \\ \dfrac{1}{Pr(D_2 = 0|D_1 = 0, X_1, X_0, Y_1, Y_0) \cdot (1 - Pr(D_1 = 1|X_0, Y_0))} & \text{if } D = 0 \end{cases}$$

4. In order to make our estimator less sensitive towards very high or very low propensity scores we only keep observations within the 5^{th} and 95^{th} percentile of the $Pr(D_1 = 1|X_0, Y_0)$ distribution. Furthermore, we condition on the common support of $Pr(D_1 = 1|X_0, Y_0)$, $Pr(D_2 = 1|D_1 = 1, X_1, X_0, Y_1, Y_0)$, and $Pr(D_2 = 1|D_1 = 1, X_1, X_0, Y_1, Y_0)$ respectively.[23]

[22] This is mainly to fully follow Lechner (2009) who only uses IPW for the dynamic model.

[23] See Figure S3 in the supplementary materials for an exemplary visual overview of the propensity score distributions as well as the exact number of observations that are dropped for every single restriction.

5. Then the dynamic average treatment effect amounts to:
$DATT_t = (D'WD)^{-1}D'WY_t$

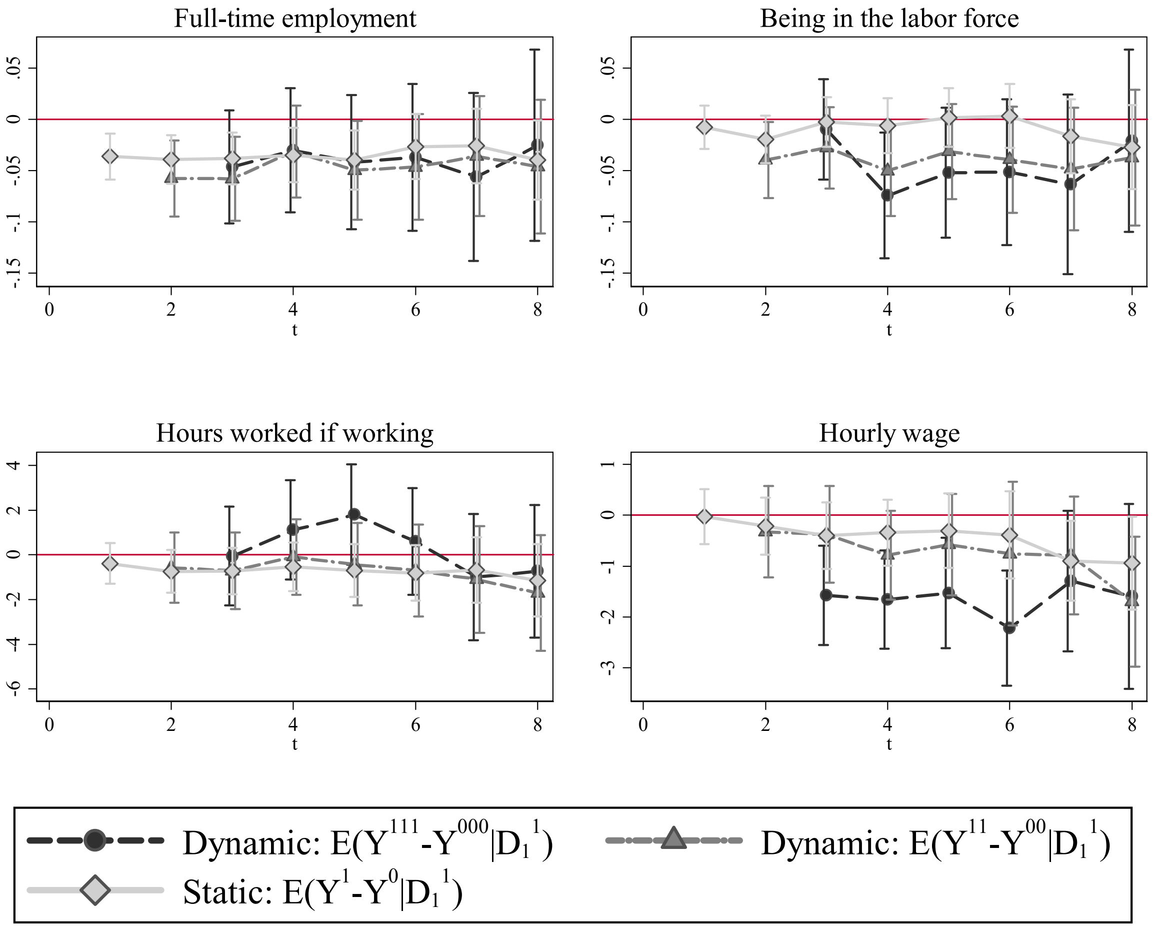
6.2 Estimation Results – Dynamic model

Figure 11 adds to the static effects the estimated effects of providing care in both year 1 and 2 compared to not providing care in both years ($E(Y_t^{11} - Y_t^{00}|D_1 = 1)$) as well as providing care in all first three years compared to not caring then ($E(Y_t^{111} - Y_t^{000}|D_1 = 1)$). A first general and main result is that, for full-time employment and conditional hours worked, it does not make a difference whether one looks at the effect of caregiving for at least one year or to caregiving for at least two or three consecutive years. Most point estimates do not differ significantly. While this is partly due to larger standard errors in the dynamic estimations, the point estimates are mostly also quite similar in magnitude, too. Individuals who have been providing care for three periods do not have a significantly lower probability to work-full time due to care provision in the third year than those who provided care for one year – the effect is a 5 ppt. reduction compared to 4 ppts. for one year.

Thus, contrary to what we expected, individuals seem to directly decide in the first year of care provision about their short- and medium-run labor market participation. Or, put differently, there seem to be no dynamic effects – at least for the period of three years – of care provision on the likelihood to work full-time in Germany. Moreover, this picture does not change in the long-run perspective until several years after care provision. For full-time work and conditional hours, the dynamic effects are somewhat more pronounced, but the differences are rather small.

The findings are somewhat different for the probability to be in the labor force and for wages where we find differences between different care paths. Here it seems that longer care spells translate into higher effects. As opposed to the static model, the probability of being in the labor force is reduced by around 3–6 ppts. for longer-lasting care spells (after 7 and 8 years). For wages, caring at least three consecutive years goes along with a significant wage penalty of nearly 2€ /hour (around 14% in relative terms). However, considerably smaller sample sizes and less degrees of freedom for the dynamic specifications also add more noise to the results. Given the general comparability of both approaches, we do not repeat the robustness checks of Section 5 here, and, for alternative specifications that are also data demanding, turn back to the static version.

Figure 11: Labor market effects of informal caregiving for females – Dynamic version

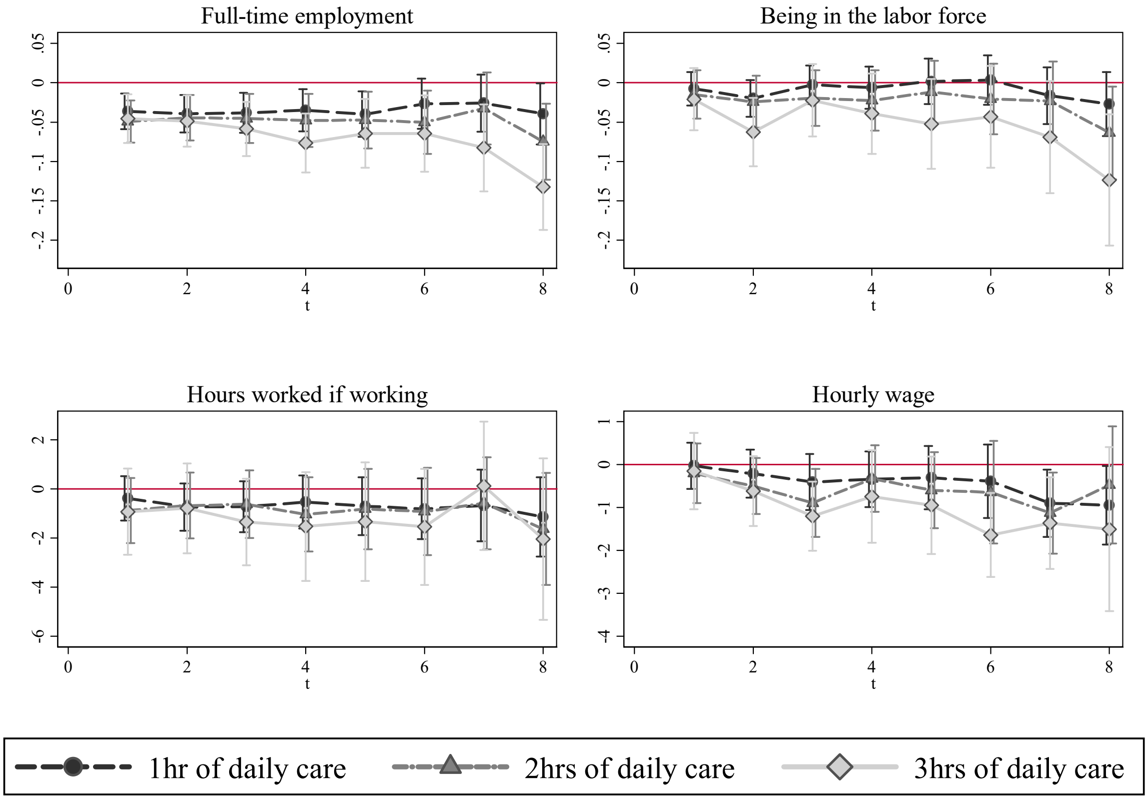


Source: SOEP. Own calculations. Note: The graph shows the point estimates and the 95% confidence intervals.

7 Alternative specifications

An important question is how sensitive the effects are with respect to the care intensity. In the baseline specifications, we defined the treatment to be at least one hour of care per day. In the following we vary this definition by restricting the treatment to at least two hours or three hours per day. Figure 12 – which returns to the static version due to sample size reasons as well as comparably small differences between static and dynamic approach – compares the results for these definitions with the baseline results. Apparently, there are hardly any differences between one and two hours of care per day, both in the short- and the longer-run. Moreover, short-run effects (effects in $t = 1$) also do not differ between three daily hours and one hour as a treatment definition for any of the four outcome variables. However, longer run effects for full-time work and employment are stronger if we use the cut-off of three hours. This is remarkable as the strongest effects seem to materialize when, in most cases, the care episode has already ceased. Those who provided care of at least three hours per day are, 8 years later, around 15 ppts. less likely to work full-time and to be employed. Moreover, those who stay in the labor force earn, on average, 2 Euro per hour less, which is a considerable wage penalty.

Figure 12: Results of the static version – Variations in treatment definition



Source: SOEP. Own calculations. Note: The graph shows the point estimates and the 95% confidence intervals.

One issue in estimating longer run effects is that, due to our sample construction, the average age of women in $t = 8$ is 48 and, thus, higher than in $t = 1$ (44). As long as effects of care provision are not heterogeneous in age, this should not be a problem. Nevertheless, the results might be affected by women who anyway leave the labor force when they get older. As an example, a 60 year old women in $t = 1$ automatically drops out of the sample in $t = 6$ and longer-run effects can only be estimated for women who are at most 57 years old when they start to provide care. Figure A4 in the Appendix repeats the analysis but restricts women to be up to 55 years in $t = 1$. The results are largely unchanged for the whole set of outcome variables. Thus, our estimates do not seem to be affected by either the mandatory retirement threshold nor other effects related to aging. We provide more analyses on age differences in the supplementary materials where we split the sample at median age of 44 and look at differences between younger and older women. We find that differences are quite small.

8 Conclusion

In this paper we assessed labor market outcomes as an important part of the implicit costs of informal care provision. In order to identify these costs we use matching techniques and inverse probability weighting. We exploit the panel information and a large set of individual controls (including measures of personality traits) to justify the identifying assumptions but also relax the assumptions in sensitivity analyses. We compare effects of providing care in a certain year on contemporaneous and later outcomes to effects of up to three consecutive years of care provision. Thereby, we contribute to the literature by both analyzing longer run effects as well as explicitly taking into account the dynamics of care provision.

An overview of our results is given in Table 3. Most importantly, we find significant initial negative effects of informal care provision on the probability to work full-time. The 4 percentage points reduction in the probability to work full-time after caring for at least one year is persistent over time. These effects are largely comparable for women who provide care for at least three consecutive years. Providing care for a higher intensity (at least three hours per day) has a stronger long-term effect on full-time work. Conditional working hours are, on average, only slightly affected. Long-run effects are reductions around 1 hour per week, which are also statistically insignificant. There are no short-run effects on the likelihood of being in the labor force but quite considerable negative effects for both longer care episodes and higher care intensities. Hourly wages are not affected in the short-run but we find a long-run wage penalty of around 1 to 1.5 Euro for women who provide care (irrespective of duration and intensity). Alternative specifications show that the effects are not only driven by older women who provide care.

We scrutinize our results by versatile tests to check whether they still hold even if there are deviations from our identifying assumption. For example, by simulating an additional confounder with a selection effect stronger than all observed ones, we are able to credibly bound the effect on full-time between 2.4 and 5.0 percentage points but argue, that if any, 5.0 should be more likely than 2.4.

The reduction in full-time work seems to be mostly driven by the intensive margin of labor supply. Women do not leave the labor market – at least for shorter durations and moderate care intensities – but switch to part-time work. Yet, after the care spell has ceased, these women do not seem to switch back to full-time work. From a social planner's point of view, these effects would directly translate into costs and would weaken at least one argument in favor of informal care as opposed to other modes of care (informal care is usually assumed to be cheaper for the society). The following back-of-the-envelope calculation may elucidate this argument by showing that the estimated labor market responses due to informal care go along with fiscal costs. See Table S2 in the supplementary

Table 3: Summary of results

Outcome	Care episode	Care provision / day	Short-run effect	Long-run effect
Full-time employment	$\geq$ 1 year	$\geq$ 1 hour	$-$ 4 ppts*	$-$ 4 ppts*
	$\geq$ 3 years	$\geq$ 1 hour	$-$ 4 ppts*	$-$ 4 ppts*
	$\geq$ 1 year	$\geq$ 3 hours	$-$ 5 ppts*	-9 to $-$ 15 ppts*
Conditional working hours	$\geq$ 1 year	$\geq$ 1 hour	≈ 0	-1
	$\geq$ 3 years	$\geq$ 1 hour	≈ 0	-1
	$\geq$ 1 year	$\geq$ 3 hours	-1	-2
Being in the labor force	$\geq$ 1 year	$\geq$ 1 hour	≈ 0	≈ 0
	$\geq$ 3 years	$\geq$ 1 hour	≈ 0	$-$ 3 to $-$ 6 ppts
	$\geq$ 1 year	$\geq$ 3 hours	≈ 0	-8 to $-$ 15 ppts*
Hourly wages	$\geq$ 1 year	$\geq$ 1 hour	≈ 0	€ -1*
	$\geq$ 3 years	$\geq$ 1 hour	€ -1.5*	€ -2*
	$\geq$ 1 year	$\geq$ 3 hours	≈ 0	€ -2*

Source: SOEP, own calculations. Summary of the results of different specifications as reported in Sections 4, 6, and 7. Short-run effect is one year after the start of a care spell (or after three years for care episodes of at least three years). Long-run effect is 7 - 8 years after start of a care spell. * indicates significance at the 5% level.

materials for details on the following derivations. Over the time span of eight years, on average, the females reduced their hours worked from 32.12 to 31.4 on the intensive margin. Together with the reduced employment probability (reduction from 74% to 73% on average over eight years) and the average wage penalty of female caregivers, caregiver's total labor market income would decrease from 16,998.88€ to 15,894.84€. For these incomes, income taxes amount to 1,817€ or 1,489€. Assuming a constant average consumption rate for caregivers and non-caregivers, one can calculate a resulting differential in the absolute amount of paid value-added-tax[24] which amounts to 108.37€. In total, according to this simplified calculation and based on our estimates, informal caregivers pay 436€ less taxes each year. For 2 million female caregivers currently in Germany, the resulting total tax differential due to informal care is estimated to be 873 million € per year.

A potential way to keep women in full-time work in the long run could be the expansion of the system of parental leave benefits to the informal care sector. Currently, German parents can leave their job for up to 14 months to care for their children and receive 60 per cent of their income (up to € 1800). Expanding this to informal caregivers could fulfill two goals. First, caregivers could take a one year leave and do not need to take on the double burden of care provision and full-time work which probably has negative health effects (Schmitz and Stroka, 2013). Second, women are prevented from switching to part-time jobs to circumvent the double burden – apparently once women switched to part-time work they often do not switch back later. As long as caregivers have a legal claim to

[24] The value added tax is calculated as: $(16,998.88 - 1,817) - (15,894.84 - 1,489)$€ times the average value-added tax rate (weighted average between 19 and 7%, assumed to be 15%).

return to their previous job after one year (as is the case with the parental leave system) chances would probably be improved that informal care provision only has short-run but no long-run labor market consequences.

References

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113(1):151–184.

Augurzky, B., Hentschker, C., Krolop, S., and Mennicken, R. (2013). Pflegeheim Rating Report 2013 – Ruhiges Fahrwasser erreicht. Hannover: Vincentz network., Essen.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *Review of Economic Studies*, 81(2):608–650.

Bolin, K., Lindgren, B., and Lundborg, P. (2008). Your next of kin or your own career?: Caring and working among the 50+ of Europe. *Journal of Health Economics*, 27(3):718–738.

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

Carmichael, F. and Charles, S. (1998). The labour market costs of community care. *Journal of Health Economics*, 17(6):747–765.

Carmichael, F. and Charles, S. (2003). The opportunity costs of informal care: does gender matter? *Journal of Health Economics*, 22(5):781–803.

Casado-Marín, D., Pilar García-Gómez, P., and Ángel López-Nicolás (2011). Informal care and labour force participation among middle-aged women in Spain. *SERIEs*, 2(1):1–29.

Crespo, L. and Mira, P. (2014). Caregiving to elderly parents and employment status of european mature women. *The Review of Economics and Statistics*, 96(3):693–709.

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

Ettner, S. L. (1995). The impact of "parent care" on female labor supply decisions. *Demography*, 32(1):63–80.

Ettner, S. L. (1996). The opportunity costs of elder care. *The Journal of Human Resources*, 31(1):189–205.

Fevang, E., Kverndokk, S., and Roed, K. (2012). Labor supply in the terminal stages of lone parents' lives. *Journal of Population Economics*, 25(4):1399–1422.

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

Geyer, J. and Korfhage, T. (2015a). Long-term Care Insurance and Carers' Labor Supply – A Structural Model. *Health Economics*, 24(9):1178–1191.

Geyer, J. and Korfhage, T. (2015b). Long-Term Care Reform and the Labor Supply of Household Members: Evidence from a Quasi-Experiment. Technical report.

Heger, D. (2014). Work and Well-Being of Informal Caregivers in Europe. Ruhr Economic Papers 0512, Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Ruhr-Universität Bochum, Universität Dortmund, Universität Duisburg-Essen.

Heitmueller, A. (2007). The chicken or the egg?: Endogeneity in labour market participation of informal carers in england. *Journal of Health Economics*, 26(3):536 – 559.

Heitmueller, A. and Inglis, K. (2007). The earnings of informal carers: Wage differentials and opportunity costs. *Journal of Health Economics*, 26(4):821–841.

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.

Johnson, R. W. and Sasso, A. T. L. (2000). The trade-off between hours of paid employment and time assistance to elderly parents at midlife. Working paper, The Urban Institute.

Lechner, M. (2008). Matching estimation of dynamic treatment models: Some practical issues. In D. Millimet, J. S. and Vytlacil, E., editors, *Modelling and Evaluating Treatment Effects in Econometrics*, volume 21 of *Advances in Econometrics*, pages 289–333.

Lechner, M. (2009). Sequential Causal Models for the Evaluation of Labor Market Programs. *Journal of Business & Economic Statistics*, 27:71–83.

Lechner, M. and Miquel, R. (2010). Identification of the effects of dynamic treatments by sequential conditional independence assumptions. *Empirical Economics*, 39(1):111–137.

Manski, C. F. (1995). *Identification problems in the social sciences*. Harvard University Press, Cambridge and Mass.

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.

Meng, A. (2012a). Informal Caregiving and the Retirement Decision. *German Economic Review*, 13(3):307–330.

Meng, A. (2012b). Informal home care and labor-force participation of household members. *Empirical Economics*, DOI 10.1007/s00181-011-0537-1.

Michaud, P.-C., Heitmueller, A., and Nazarov, Z. (2010). A dynamic analysis of informal care and employment in england. *Labour Economics*, 17(3):455–465.

Müller, R., Unger, R., and Rothgang, H. (2010). Reicht eine zweijährige Familien-Pflegezeit für Arbeitnehmer? Wie lange Angehörige zu Hause gepflegt werden. *Soziale Sicherheit. Zeitschrift für Arbeit und Soziales*, 10(6–7):230–237.

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

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.

Rothgang, H. (2010). Social insurance for long-term care: An evaluation of the German model. *Social Policy & Administration*, 44(4):436–460.

Schmidt, M. and Schneekloth, U. (2011). Abschlussbericht zur Studie "Wirkungen des Pflege-Weiterentwicklungsgesetzes".

Schmitz, H. and Stroka, M. A. (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. and Westphal, M. (2015). Short- and medium-term effects of informal care provision on female caregivers' health. *Journal of Health Economics*, 42(C):174–185.

Schneekloth, U. and Leven, I. (2003). Hilfe und Pflegebedürftige in Privathaushalten in Deutschland 2002.

Schulz, E. (2010). The Long-Term Care System for the Elderly in Germany. DIW Discussion Paper 1039, DIW Berlin.

Skira, M. M. (2015). Dynamic wage and employment effects of elder parent care. *International Economic Review*, 56(1):63–93.

SOEP Group (2014). SOEP 2013 – Documentation of Person-related Status and Generated Variables in PGEN for SOEP v30.

Statistisches Bundesamt (2015). Pflegestatistik 2013. Pflege im Rahmen der Pflegeversicherung. Deutschlandergebnisse.

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

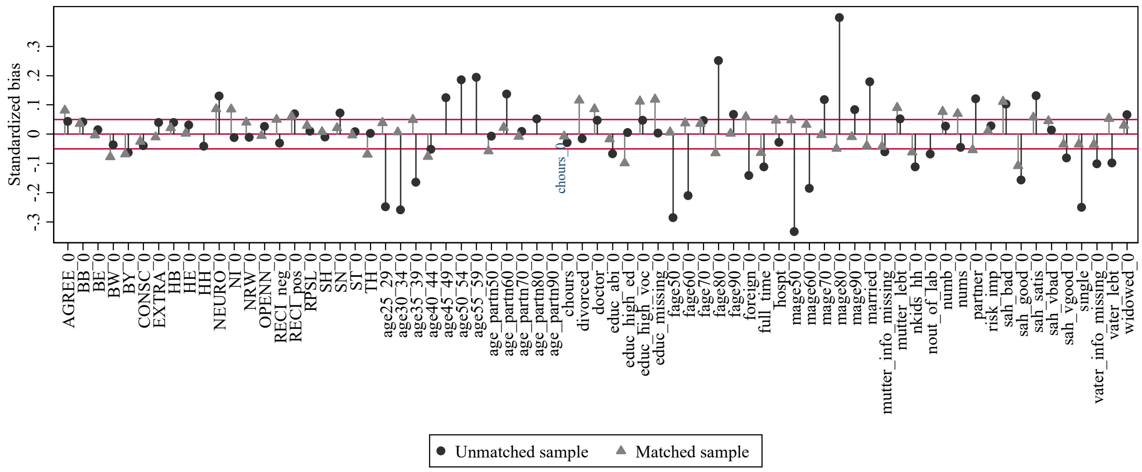
Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP): scope, evolution, and enhancements. *Journal of Applied Social Science Studies (Schmollers Jahrbuch: Zeitschrift für Wirtschafts- und Sozialwissenschaften)*, 127(1):139–169.

Wolf, D. A. and Soldo, B. J. (1994). Married women's allocation of time to employment and care of elderly parents. *The Journal of Human Resources*, 29(4):1259–1276.

Appendix

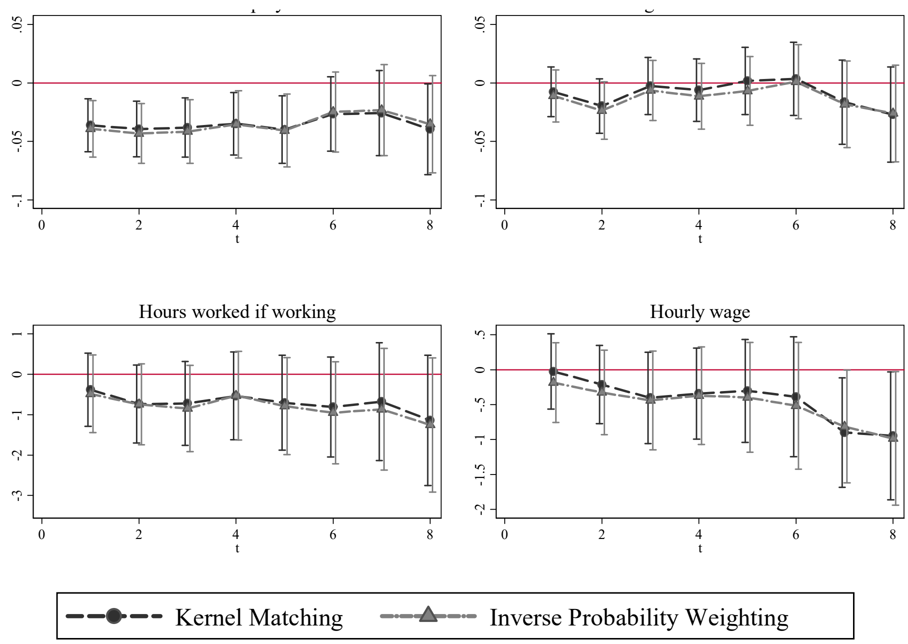
Additional figures

Figure A1: Matching quality for full-time work



Source: SOEP. Own calculations. Note: The figure shows the normalized differences for both the unmatched and the matched sample. The normalized difference between treatment group (1) and control group (0) is calculated according to: $Diff = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{\frac{1}{2}(\sigma_1^2+\sigma_0^2)}}$ where $\bar{x}$ is the sample mean and σ^2 the sample variance. Here, we report differences for all variables that are potentially included. Recall that only the subgroup of variables chosen by the double selection procedure (Belloni et al., 2014) is used in the propensity score estimations. See Table A1 for translations of variable names. An Epanechnikov kernel with of bandwidth 0.0018 is used. The two red lines mark a standardized bias of $\pm 5\%$. While a couple of variables falls outside this range in the unmatched sample this is only marginally the case for two of the variables in the matched sample. Pre-treatment full-time work as the most important control is highlighted in the figure. Year dummies are not reported for legibility but are well within the red lines and included in the estimations.

Figure A2: Static version, Kernel matching vs. IPW estimators



Source: SOEP. Own calculations. Note: The graph shows the point estimates and the 95% confidence intervals.

Figure A3: Difference between baseline results and same estimation with restriction to never carers in the control group

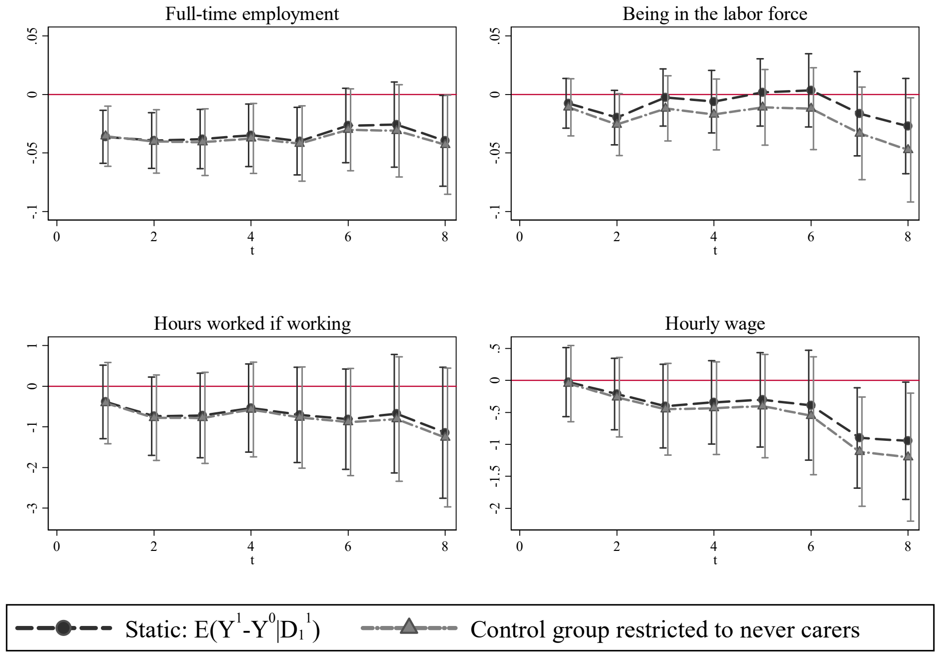
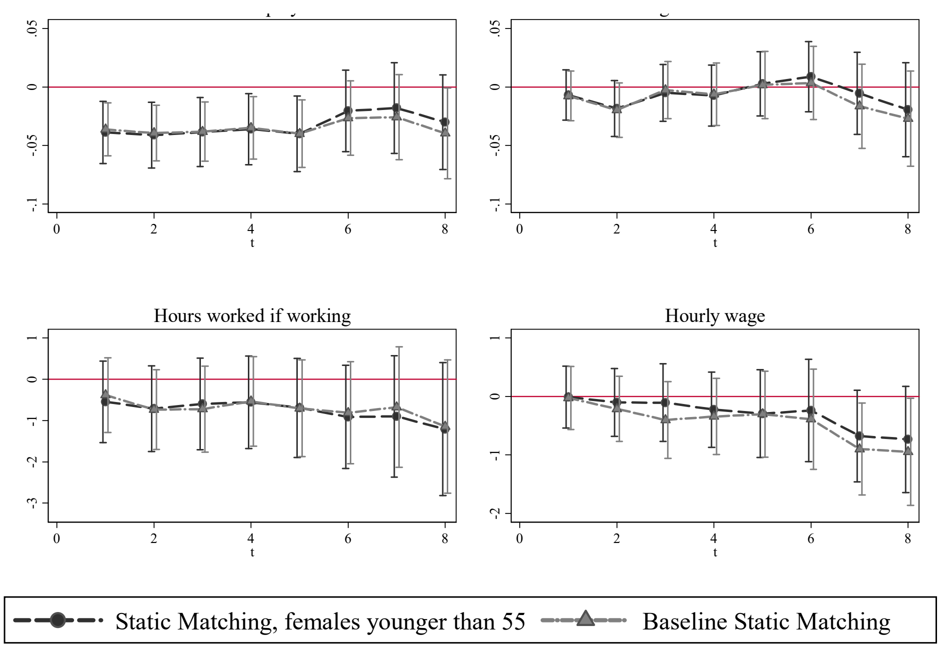


Figure A4: Results for females younger than 55 in $t = 1$



Source: SOEP. Own calculations. Note: The graph shows the point estimates and the 95% confidence intervals.

Additional tables

Table A1: Variable description

Variable	Description	Mean	SD	Min	Max	Variable name
Outcome variables:						
Full time	Binary indicator of working full-time	0.35	0.48	0	1	full_time_0
Hourly wage	Gross monthly wage/(number of hours worked * 4.3)	14.11	8.75	2.0	149.5	hourly_wage_0
Hours	Number of actual hours worked per week (here: unconditional)	21.90	18.66	0	80	hours_0
Employed	Binary indicator, working full-time, part-time, vocational training, or marginal and irregular part-time employment	0.24	0.42	0	1	nout_of_lab_0
Care obligations:						
Age of mother						
$\in [30, 39]$		0.00	0.00	0	0	mage30_0
$\in [40, 49]$		0.03	0.17	0	1	mage40_0
$\in [50, 59]$		0.15	0.36	0	1	mage50_0
$\in [60, 69]$		0.21	0.41	0	1	mage60_0
$\in [70, 79]$		0.18	0.39	0	1	mage70_0
$\in [80, 89]$		0.08	0.28	0	1	mage80_0
$\in [90, 99]$		0.02	0.12	0	1	mage90_0
Mother alive		0.67	0.47	0	1	mutter_lebt_0
Age of father						
$\in [30, 39]$		0.00	0.01	0	1	fage30_0
$\in [40, 49]$		0.01	0.11	0	1	fage40_0
$\in [50, 59]$		0.11	0.31	0	1	fage50_0
$\in [60, 69]$		0.17	0.38	0	1	fage60_0
$\in [70, 79]$		0.14	0.35	0	1	fage70_0
$\in [80, 89]$		0.06	0.24	0	1	fage80_0
$\in [90, 99]$		0.02	0.13	0	1	fage90_0
Father alive		0.49	0.50	0	1	vater_lebt_0
Partner existent		0.27	0.44	0	1	partner_0
Age of partner		34.91	23.27	0	89	age_partn_0
Number of siblings		1.86	1.73	0	18	nums_0
Socio-economics and willingness to provide care:						
Own age						
$\in [25, 29]$		0.10	0.30	0	1	age25_29_0
$\in [30, 34]$		0.12	0.32	0	1	age30_34_0
$\in [35, 39]$		0.14	0.34	0	1	age35_39_0

Continued on next page

Table A1 – *continued*

Variable	Description	Mean	SD	Min	Max	Variable name
$\in [40,44]$		0.16	0.36	0	1	age40_44_0
$\in [45,49]$		0.15	0.36	0	1	age45_49_0
$\in [50,54]$		0.13	0.34	0	1	age50_54_0
$\in [55,59]$		0.12	0.32	0	1	age55_59_0
Education:						
A-Levels	Completed academic track	0.08	0.27	0	1	educ_abi_0
Voc. train.	Higher education and vocational training	0.07	0.25	0	1	educ_high_voc_0
Higher	Higher education	0.22	0.41	0	1	educ_high_ed_0
Missing	missing	0.01	0.12	0	1	educ_missing_0
Marital status:						
Single		0.17	0.38	0	1	single_0
Married		0.66	0.47	0	1	married_0
Divorced		0.10	0.30	0	1	divorced_0
Widowed		0.03	0.17	0	1	widowed_0
Foreign		0.08	0.27	0	1	foreign_0
Kids	Number of kids in the household	0.90	1.06	0	12	nkids_hh_0
BIG-5 Inventory						
Neuroticism	Average of answers on 7-point scales	4.32	0.70	1	7	NEURO_0
Conscientiousness	Average of answers on 7-point scales	4.75	0.51	1	7	CONSC_0
Agreeableness	Average of answers on 7-point scales	4.74	0.58	1	7	AGREE_0
Openness	Average of answers on 7-point scales	4.58	1.08	1	7	OPENN_0
Extraversion	Average of answers on 7-point scales	4.96	0.64	1	7	EXTRA_0
Reciprocity						
positive	Average of answers on 7-point scales	5.54	1.00	1	7	RECI_pos_0
negative	Average of answers on 7-point scales	2.91	1.24	1	7	RECI_neg_0
Risk aversion	Self-stated measure between 0 (very risk averse) and 10 (risk willing)	4.29	2.18	0	10	risk_imp_0
Ability to provide care:						
Self-assessed health (SAS)						
– Very Good	Binary: SAS = very good	0.09	0.29	0	1	sah_vgood_0
– Good	Binary: SAS = good	0.45	0.50	0	1	sah_good_0
–Satisfactory	Binary: SAS =satisfactory	0.32	0.47	0	1	sah_satis_0
– Bad	Binary: SAS = bad	0.12	0.33	0	1	sah_bad_0
– Very bad	Binary: SAS =very bad	0.02	0.16	0	1	sah_vbad_0
Doctor visits	Number of doctor visits previous 3 months	2.56	3.85	0	99	doctor_0

Continued on next page

Table A1 – *continued*

Variable	Description	Mean	SD	Min	Max	Variable name
Hospital stays	Number of hospital stays previous year	0.15	0.53	0	48	hospt_0
Year and Federal state dummies:						
Year						
=2002		0.10	0.30	0	1	y2002_0
=2003		0.09	0.29	0	1	y2003_0
=2004		0.09	0.29	0	1	y2004_0
=2005		0.08	0.28	0	1	y2005_0
=2006		0.09	0.28	0	1	y2006_0
=2007		0.08	0.27	0	1	y2007_0
=2008		0.08	0.26	0	1	y2008_0
=2009		0.08	0.27	0	1	y2009_0
=2010		0.07	0.25	0	1	y2010_0
=2011		0.08	0.27	0	1	y2011_0
=2012		0.08	0.27	0	1	y2012_0
Federal state:						
BE	Berlin	0.04	0.19	0	1	BE_0
SH	Schleswig-Holstein	0.03	0.17	0	1	SH_0
HH	Hamburg	0.01	0.12	0	1	HH_0
NI	Lower Saxony	0.09	0.28	0	1	NI_0
HB	Bremen	0.01	0.09	0	1	HB_0
NRW	North-Rhine Westphalia	0.21	0.40	0	1	NRW_0
HE	Hesse	0.07	0.26	0	1	HE_0
RPSL	Rhineland-Palatinate and Saarland	0.06	0.24	0	1	RPSL_0
BW	Baden-Württemberg	0.12	0.33	0	1	BW_0
BY	Bavaria	0.15	0.36	0	1	BY_0
BB	Brandenburg	0.04	0.19	0	1	BB_0
ST	Saxony-Anhalt	0.04	0.20	0	1	ST_0
TH	Thuringia	0.04	0.20	0	1	TH_0
SN	Saxony	0.07	0.25	0	1	SN_0

Notes: Source SOEP

Table A2: Matching results corresponding to Figure 4

Outcome	Year	ATT	Std. err.	t-statistic	Observations
Full-time employment	1	−0.036	0.012	−3.147	63,532
	2	−0.039	0.012	−3.246	54,816
	3	−0.038	0.013	−2.986	47,608
	4	−0.035	0.014	−2.575	40,640
	5	−0.040	0.015	−2.724	34,057
	6	−0.027	0.016	−1.646	27,741
	7	−0.026	0.019	−1.400	21,994
	8	−0.039	0.020	−1.992	16,874
Conditional working hours	1	−0.390	0.462	−0.846	38,360
	2	−0.738	0.492	−1.501	32,518
	3	−0.723	0.531	−1.363	27,903
	4	−0.541	0.553	−0.978	23,676
	5	−0.714	0.598	−1.193	19,734
	6	−0.813	0.631	−1.289	16,041
	7	−0.669	0.743	−0.900	12,658
	8	−1.137	0.823	−1.381	9,687
Hourly wages	1	−0.027	0.275	−0.097	37,673
	2	−0.213	0.285	−0.749	31,914
	3	−0.402	0.334	−1.205	27,397
	4	−0.338	0.332	−1.017	23,254
	5	−0.297	0.375	−0.791	19,376
	6	−0.386	0.437	−0.882	15,747
	7	−0.899	0.399	−2.252	12,444
	8	−0.943	0.467	−2.018	9,527
Being in the labor force	1	−0.008	0.011	−0.692	63,565
	2	−0.020	0.012	−1.646	54,845
	3	−0.002	0.012	−0.195	47,637
	4	−0.006	0.014	−0.442	40,667
	5	0.002	0.015	0.124	34,037
	6	0.004	0.016	0.221	27,726
	7	−0.016	0.018	−0.881	21,979
	8	−0.027	0.021	−1.306	16,867

Source: SOEP, own calculations. Employed bandwidths: Full-time employment: 0.0016855; Conditional working hours: 0.001999; Hourly wages: 0.0020078; Being in the labor force: 0.00181.

Supplementary material not intended to be published

Selected variables using the double selection procedure

Exemplary case: full-time work, static version.

full_time_0married_0sah_vbad_0CONSC_0mage50_0fage50_0age25_29_0Xsingle_0age25_29_0Xsah_vbad_0
age25_29_0Xmage70_0age25_29_0Xfage70_0
age25_29_0Xvater_lebt_0age30_34_0Xsah_good_0
age30_34_0Xdoctor_0age30_34_0XCONSC_0age30_34_0Xy2009_0age30_34_0Xmage80_0
age30_34_0Xfage40_0age30_34_0Xfull_time_0age35_39_0Xwidowed_0age35_39_0Xrisk_imp_0
age35_39_0XHH_0age35_39_0XHB_0age40_44_0Xfage50_0age40_44_0Xfull_time_0age45_49_0Xfull_time_0
age50_54_0Xfage60_0age50_54_0Xfull_time_0age55_59_0Xmutter_lebt_0age55_59_0Xfull_time_0
educ_abi_0XST_0educ_abi_0Xmage90_0educ_abi_0Xfage40_0educ_high_voc_0Xfage90_0educ_high_ed_0XAGREE_0
educ_high_ed_0Xmage40_0educ_high_ed_0Xmutter_lebt_0educ_missing_0XBE_0educ_missing_0XSH_0
educ_missing_0XRPSL_0
married_0XNRW_0divorced_0Xsah_good_0divorced_0Xmage40_0widowed_0XRECI_pos_0widowed_0Xfage90_0
foreign_0XNEURO_0 foreign_0XSH_0foreign_0XHH_0foreign_0Xmage90_0foreign_0Xfage90_0nkids_hh
_0Xsah_good_0 nkids_hh_0XNEURO_0
nkids_hh_0XBW_0nkids_hh_0Xmutter_lebt_0single_0Xpartner_0single_0XCONSC_0single_0XHH_0single_0XBY_0
partner_0Xsah_good_0
partner_0Xvater_lebt_0partner_0Xmutter_lebt_0age_partn_0Xmutter_lebt_0age_partn_0Xfull_time_0
sah_good_0Xfull_time_0sah_satis_0Xmutter_lebt_0sah_vbad_0XHB_0sah_vbad_0XBB_0doctor_0XRECI_pos_0
NEURO_0Xmage80_0CONSC_0Xfage60_0CONSC_0Xvater_lebt_0CONSC_0Xmutter_lebt_0CONSC_0Xfull_time_0
AGREE_0Xfage50_0OPENN_0Xmage50_0OPENN_0Xmage60_0EXTRA_0Xmage60_0RECI_pos_0Xfull_time_0
SH_0Xmage40_0HH_0Xy2012_0HH_0Xmage50_0HH_0Xmage60_0HB_0Xmage40_0HE_0Xfage60_0RPSL_0Xmage90_0
BW_0Xfull_time_0BB_0Xmage40_0ST_0Xmage90_0ST_0Xfage40_0TH_0Xfage90_0y2003_0Xmage40_0
y2003_0Xfage30_0
y2004_0Xmage40_0y2006_0Xfage40_0y2007_0Xfull_time_0y2011_0Xmage40_0mage50_0Xfage40_0
mage50_0Xmutter_lebt_0
mage60_0Xfage90_0mage60_0Xmutter_lebt_0mage70_0Xfage50_0mage70_0Xfull_time_0mage80_0Xmutter_lebt_0

Note: _0 indicate the time periods. X stands for interactions between two variables. See Table A1 for translations of variable names.

Table S1: Parameters for calibration of the sensitivity analysis

Variable	$p11$	$p10$	$p01$	$p00$	s	d
$full\_time\_0$	0.89	0.09	0.89	0.89	−0.05	0.82
$age25\_29\_0$	0.05	0.03	0.13	0.13	−0.06	0.05
$age30\_34\_0$	0.04	0.05	0.13	0.13	−0.07	0.01
$age35\_39\_0$	0.08	0.09	0.13	0.13	−0.05	−0.02
$age40\_44\_0$	0.14	0.14	0.16	0.16	−0.02	0.01
$age45\_49\_0$	0.24	0.18	0.17	0.17	0.05	0.04
$age50\_54\_0$	0.24	0.19	0.15	0.15	0.07	0.03
$age55\_59\_0$	0.18	0.18	0.10	0.10	0.07	−0.02
$educ\_abi\_0$	0.07	0.06	0.09	0.09	−0.02	0.01
$educ\_high\_voc\_0$	0.09	0.08	0.07	0.07	0.01	0.01
$educ\_high\_ed\_0$	0.35	0.17	0.32	0.32	0.00	0.16
$educ\_missing\_0$	0.01	0.02	0.01	0.01	0.00	0.00
$married\_0$	0.62	0.79	0.51	0.51	0.08	−0.23
$divorced\_0$	0.15	0.07	0.13	0.13	0.00	0.05
$widowed\_0$	0.03	0.05	0.02	0.02	0.01	−0.01
$foreign\_0$	0.03	0.05	0.06	0.06	−0.03	−0.04
$nkids\_hh\_0$	0.36	0.52	0.39	0.39	−0.05	−0.21
$single\_0$	0.16	0.06	0.29	0.29	−0.08	0.18
$partner\_0$	0.75	0.85	0.68	0.68	0.05	−0.15
$age\_partn50\_0$	0.62	0.64	0.60	0.60	0.00	−0.06
$age\_partn60\_0$	0.12	0.19	0.07	0.07	0.05	−0.08
$age\_partn70\_0$	0.01	0.01	0.01	0.01	0.00	−0.01
$age\_partn80\_0$	0.00	0.00	0.00	0.00	0.00	0.00
$age\_partn90\_0$	0.00	0.00	0.00	0.00	0.00	0.00
$nums\_0$	0.20	0.24	0.22	0.22	−0.01	−0.04
$numb\_0$	0.21	0.26	0.21	0.21	0.01	−0.06
sah_vgood_0	0.09	0.06	0.11	0.11	−0.02	0.03
sah_good_0	0.36	0.38	0.49	0.49	−0.08	0.05
sah_satis_0	0.41	0.37	0.30	0.30	0.06	−0.02
sah_bad_0	0.13	0.16	0.09	0.09	0.04	−0.04
sah_vbad_0	0.01	0.03	0.01	0.01	0.00	−0.02
$doctor\_0$	0.32	0.39	0.30	0.30	0.02	−0.07
$hospt\_0$	0.09	0.11	0.09	0.09	−0.01	−0.05
$NEURO\_0$	0.56	0.61	0.51	0.51	0.05	−0.06
$CONSC\_0$	0.59	0.53	0.62	0.62	−0.01	0.10
$AGREE\_0$	0.57	0.53	0.53	0.53	0.03	0.02
$OPENN\_0$	0.55	0.51	0.55	0.55	0.01	0.06
$EXTRA\_0$	0.56	0.56	0.56	0.56	0.00	−0.01
$RECI\_pos\_0$	0.57	0.57	0.55	0.55	0.04	0.02
$RECI\_neg\_0$	0.50	0.52	0.50	0.50	0.00	−0.02
$risk\_imp\_0$	0.61	0.52	0.57	0.57	0.02	0.06
BE_0	0.07	0.03	0.05	0.05	0.00	0.02

Continued on next page

Table S1 – *continued*

Variable	$p11$	$p10$	$p01$	$p00$	s	d
SH_0	0.02	0.03	0.02	0.02	0.00	−0.01
HH_0	0.01	0.01	0.02	0.02	0.00	0.00
NI_0	0.06	0.09	0.07	0.07	0.00	−0.02
HB_0	0.00	0.01	0.01	0.01	0.00	0.00
NRW_0	0.17	0.21	0.18	0.18	0.00	−0.04
HE_0	0.06	0.09	0.07	0.07	0.01	0.00
$RPSL\_0$	0.07	0.06	0.05	0.05	0.00	−0.02
BW_0	0.07	0.13	0.11	0.11	−0.01	−0.02
BY_0	0.1	0.14	0.14	0.14	−0.02	−0.01
BB_0	0.07	0.04	0.05	0.05	0.01	0.01
ST_0	0.07	0.03	0.06	0.06	0.00	0.02
TH_0	0.05	0.04	0.05	0.05	0.00	0.01
SN_0	0.13	0.07	0.09	0.09	0.02	0.03
$y2002\_0$	0.09	0.10	0.09	0.09	0.00	−0.01
$y2003\_0$	0.09	0.1	0.09	0.09	0.00	0.00
$y2004\_0$	0.06	0.08	0.09	0.09	−0.02	−0.01
$y2005\_0$	0.09	0.09	0.08	0.08	0.01	0.00
$y2006\_0$	0.08	0.08	0.09	0.09	−0.01	0.00
$y2007\_0$	0.11	0.10	0.08	0.08	0.02	0.00
$y2008\_0$	0.07	0.06	0.08	0.08	−0.01	0.01
$y2009\_0$	0.09	0.08	0.08	0.08	0.01	0.00
$y2010\_0$	0.06	0.06	0.07	0.07	−0.01	0.01
$y2011\_0$	0.11	0.09	0.08	0.08	0.02	0.00
$y2012\_0$	0.08	0.08	0.08	0.08	0.00	0.00
$mage50\_0$	0.09	0.07	0.22	0.22	−0.11	0.04
$mage60\_0$	0.16	0.15	0.24	0.24	−0.07	0.02
$mage70\_0$	0.27	0.24	0.23	0.23	0.05	0.04
$mage80\_0$	0.23	0.24	0.08	0.08	0.15	−0.01
$mage90\_0$	0.02	0.03	0.01	0.01	0.01	−0.01
$fage50\_0$	0.05	0.04	0.15	0.15	−0.08	0.04
$fage60\_0$	0.13	0.11	0.21	0.21	−0.08	0.02
$fage70\_0$	0.19	0.18	0.17	0.17	0.02	0.02
$fage80\_0$	0.16	0.14	0.07	0.07	0.08	0.00
$fage90\_0$	0.02	0.03	0.01	0.01	0.01	−0.01
$vater\_lebt\_0$	0.55	0.51	0.61	0.61	−0.05	0.06
$vater\_info\_missing\_0$	0.05	0.05	0.07	0.07	−0.02	−0.01
$mutter\_lebt\_0$	0.78	0.74	0.78	0.78	0.02	0.07
$mutter\_info\_missing\_0$	0.05	0.04	0.05	0.05	−0.01	−0.01

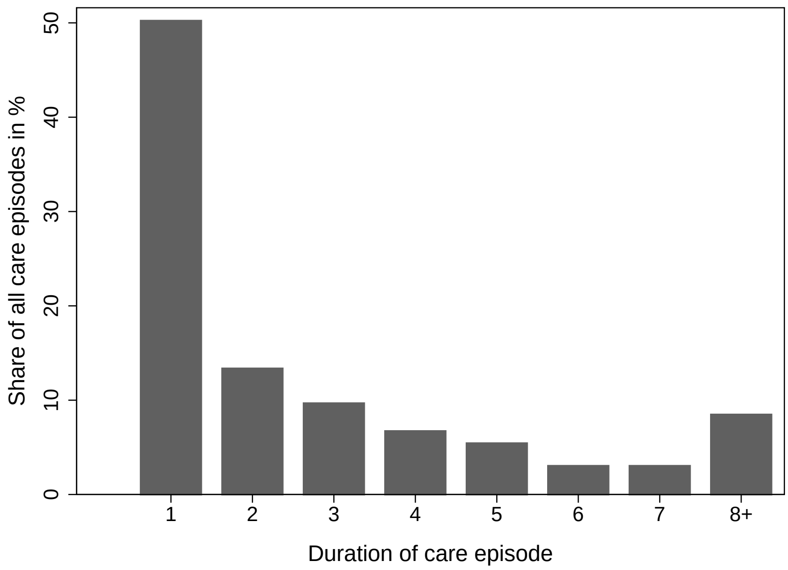
Notes: Source SOEP

Table S2: Back-of-the-envelope calculation of fiscal effects

Mean employment of never-carers		Estimated effects due to informal care (averaged over all eight periods)		
Hourly wage	13.86	Wage effect:	-0.44	13.42
Cond. hours	32.12	Eff. on hours:	-0.72	31.40
Employed:	0.74	Employment effect:	-0.0092	0.73
Fiscal effects:				
	Prior to informal care	Due to informal care		
Uncond. hours:	23.77	22.95		
Wage differential:			Difference:	
Labor income per year	16,998.88	15,894.84	-1,104.04	
Tax differential:				
thereof income tax:	1,817	1,489	-328	
Average consumption rate:	0.931			
Value-added-tax (VAT) rate (weighted average between 7 and 19%)	0.15			
Total amount VAT:	2,120.15	2,011.78	-108.37	
		Total	-436.37	
		# female informal caregiver:	2m	
			-872.75m €	less tax revenue p.a.

In order to calculate back-of-the-envelope fiscal effect, we start with the counterfactual average level of labor supply (Employment probability and conditional hours) and the respective mean wage of non-carer and add the estimated effects to get levels in labor supply and the wage for caregivers. Now we can calculate gross incomes for caregivers and non-carers. Now we can roughly estimate tax differential between both groups. In order to compute the income tax, we make use of the average tax rate for both incomes. Additionally, the VAT also contributes to the tax differential and we calculate this based on an average consumption rate. All in all caregivers annually pay €436 less taxes. Multiplied with 2 million female caregivers in Germany, this would translate into an annual fiscal loss of €873m due to informal caregiving.

Figure S1: Distribution of imputed care episodes

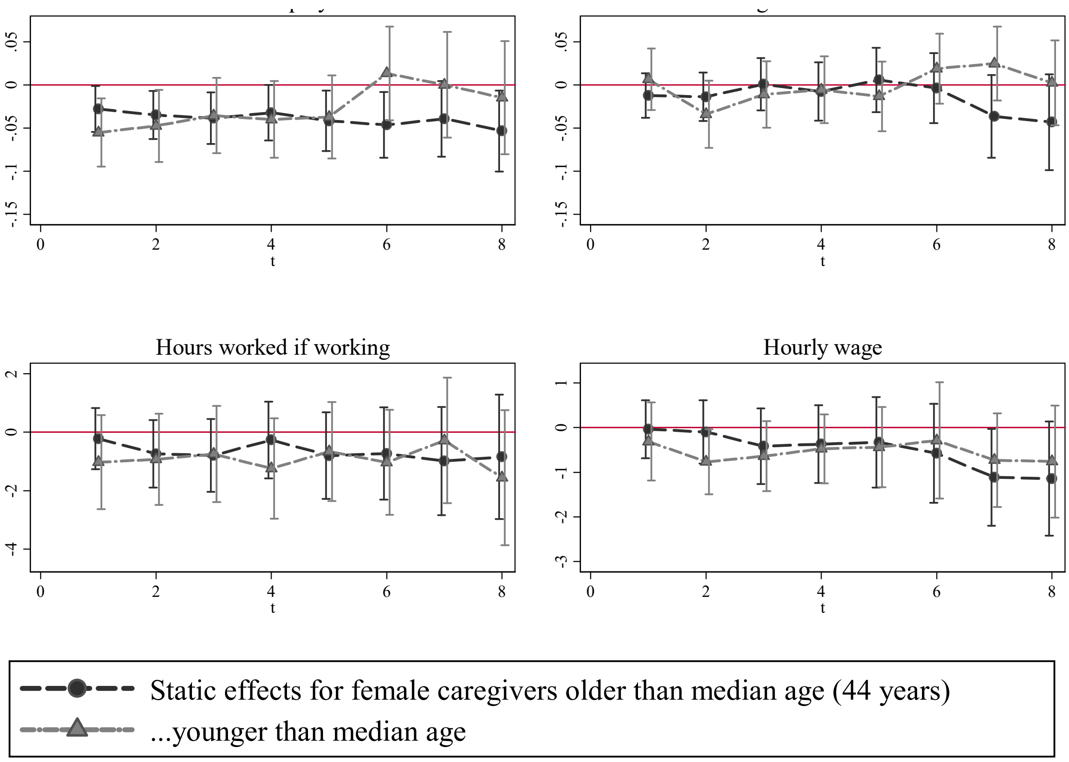


Source: SOEP, own calculations. This graph shows the distribution of episodes of consecutive care of at least one hour per day. The data are restricted to starting waves 1 to 5 as defined in Figure 1 to ensure that every spell can last for at least 8 years. Imputation is according to the following scheme: whenever a women provided care in a certain year and two years later, the care indicator one year later is set to 1 even if no care provision is stated in the questionnaire.

Alternative specifications

Figure S2 reports results when we split the sample at the median age of 44 in $t = 1$ and carry out analyses for the two groups younger and older than 44. Both for legibility and sample size reasons we restrict the analyses to the static case of providing care in year 1. This can be justified by the small differences between caring for at least one year compared to at least two or three years.

Figure S2: Results of the static version – Younger vs. older than median age



Source: SOEP. Own calculations. Note: The graph shows the point estimates and the 95% confidence intervals.

The results are remarkably similar for both groups, with minor differences. The short-run effect on full-time employment is roughly the same for both groups. Yet, after year 5 younger individuals who provided care before are no less likely to work full-time anymore than their no caring counterparts. It is, however, surprising that this drop back to zero appears between year 5 and 6 and we do not have an explanation why this should take place exactly at this point in time. Conditional hours evolve quite similar while the effects on employment are slightly smaller for younger individuals. Finally, short-term wage effects are slightly larger for younger individuals, yet, not significant either.

Common support restrictions

Figure S3 shows the distribution of the propensity scores on each node of the decision tree (see Figure 10) by actual care state (light gray for caregivers, dark gray for non-carers) – exemplarily for full-time work as an outcome and the first two periods. The red vertical lines indicate the region of common support (the smallest set of overlap in the support between both groups). The upper panel depicts this observed probability of starting caregiving after the initial period ($t = 0$). The projected probabilities are small, reflecting the low fraction of caregivers in period $t = 1$. The middle panel shows the same for the second period (caregiving in $t = 2$) conditional on having cared in $t = 1$, here the odds are balanced. The bottom panel plots the propensity score of continuing not to care for the second period $t = 0$. All those propensity scores are used to construct the inverse probability weights for the dynamic estimates. The overall conclusion from this graph is, that the overlap between treatment and control group is good and that the restriction on the common is not crucial. Out of the 63,372 observations we drop 710 at the first node, 138 at the second, and 641 at the third which are off the common support. The restriction on observations lying within the 5th and 95th quantile is more binding where we drop 6,337 observations.

Figure S3: Common support