The Journal of the Economics of Ageing 21 (2022) 100343

Contents lists available at ScienceDirect

The Journal of the Economics of Ageing

journal homepage: www.elsevier.com/locate/jeoa

Full length article

Revisiting longer-term health effects of informal caregiving: Evidence from the UK

Jannis Stöckel *, Judith Bom

Erasmus School of Health Policy & Management, Erasmus University Rotterdam, The Netherlands

ARTICLE INFO

JEL classification:
I10
I18
C21
J14

Keywords:
Informal care
Mental health
Physical health
Propensity score matching
Regression adjusted matching
Dynamic sequential matching

ABSTRACT

We estimate the longer-term and dynamic effects of providing informal care on caregivers' health in the United Kingdom. Using propensity score matching to address the endogeneity of informal care provision, we estimate static and sequential matching models exploring health effects at the extensive and intensive margin of informal caregiving and their persistence for up to five years. Our results suggest substantial negative health effects confined to the mental domain and asymmetrically experienced by caregivers providing more than 20 hours of weekly care. Further, our dynamic sequential matching results indicate that for caregivers providing multiple years of higher intensity care the negative effects persist.

Introduction

Ageing populations pose a serious challenge to health care systems in developed economies. The United Kingdom (UK) is an exemplary case, by 2050 more than a quarter of its population is expected to be 65+ and over 10% is predicted to be 80+ (OECD, 2019), drastically increasing the long-term care (LTC) demand (de la Maisonneuve & Martins, 2015). One solution to meet this demand is to rely on informal care, care provided by friends or relatives. Informal care is often preferred by the care-recipient and is from a governmental perspective a low-cost alternative to formal care. In addition, there is evidence that (partially) substituting formal by (unskilled) informal care does not jeopardize care-recipients' health. Receiving informal care can lower medical expenditures (van Houtven & Norton, 2004, 2008), decrease the likelihood of infections and bedsores (Coe et al., 2019) and improve recipients' mental health (Barnay & Juin, 2016). In the UK, informal care already plays a crucial role in meeting current care demand with more than 18% of the 50+ population providing informal care in contrast to the OECD average of 13.5% (OECD, 2019).

Despite these benefits there are concerns regarding the impact of informal care on caregivers' labor market and health outcomes. To make informed decisions on adapting current policies to future demands a thorough understanding of such effects is necessary. Previous studies either found no or negative effects of informal care provision on labor market outcomes (see Lilly et al. (2007) and Bauer and Sousa-Poza (2015) for reviews) but considerable health effects for the caregiver due to the mental and physical strain (see Bom et al., 2019a for a review).

Up to now most literature has focused on the immediate impact of care provision. However, it is also important to understand how these effects develop over time as many caregivers provide several years of care. According to the 2011 UK Census men and women at age 50 can expect to spend 4.9 and 5.9 years of their remaining life providing care (ONS, 2017). In light of this prospect, it is important to focus on health outcomes as conflicting hypotheses regarding the impact of duration of caregiving on health exist.[1] There are three opposing hypotheses regarding the association between the duration of informal care provision and the impact of care provision (see for overviews: Townsend et al.

* Corresponding author.
E-mail address: stockel@eshpm.eur.nl (J. Stöckel).

[1] A related literature focuses on the longer-term impact of caregiving on labor market outcomes. Schmitz & Westphal (2017) studied the German context and found informal care provision to have a longer-term impact on labor market outcomes, this effect did not differ dependent on the duration of care provision (e.g. individuals that provided 1 year of care compared to multiple years of care provision). Rellstab et al. (2020) studied the Dutch context and did not find any impact of care provision on labor market outcomes, which they argue might be attributable to the generous formal support system in the Netherlands.

https://doi.org/10.1016/j.jeoa.2021.100343

Available online 13 October 2021

1989, Haley & Pardo, 1989; Pinquart & Sorensen, 2003a). The first hypothesis is called the wear-and-tear hypothesis implying the impact of care provision will worsen over time as coping resources decline and care needs increase. For example, a prolonged exposure to stress arising from care tasks might deplete caregivers' resources to deal with the care strain. The trait hypothesis, on the other hand, suggests that the caregiving burden is constant. Even when health of the care-recipient deteriorates, caregivers maintain a constant level of adaptation. The care providers namely have pre-existing coping skills and resources which remain present during the care task. Lastly, the adaptation hypothesis argues that individuals learn to adapt to the situation. Following this theory, the negative impact of care provision will decline when care tasks are prolonged as caregivers develop new coping strategies or become less affected by the stress involved in their care tasks.

Some studies already investigated how longer-term or high intensity informal caregiving is associated with health. In the UK for example, Hirst (2005), Legg et al. (2013), Vlachantoni et al. (2016) and Lacey et al. (2019) found a correlation between either long-term or high intensity care and negative health outcomes. However, these studies have difficulties in ascertaining the direction of causality: is poor health caused by the act of providing informal care or do individuals in poor health, e.g. due to old age, more often provide informal care? To study the causal impact of care provision on health over time one must account for endogeneity concerns resulting from the selection of individuals into informal caregiving. Bom and Stöckel (2021) used Dutch and UK panel data to explore whether health effects differ across long-term care systems of varying generosity in public care provision. However, they focused on the immediate effects of providing informal care while ignoring the temporal dimension of these effects and the dynamic nature of providing informal care over multiple years.

To our knowledge thus far only Schmitz and Westphal (2015) and De Zwart et al. (2017) have studied longer-term health effects of informal caregiving in a causal framework. Using German panel data and focusing on female caregivers, Schmitz and Westphal (2015) find negative mental health effects persisting for up to three years after care provision. De Zwart et al. (2017) used panel data from multiple continental European countries to explore the effect of spousal caretaking among the elderly population. They report negative effects on mental health and increased medical consumption in the first year after care provision. The disappearance of health effects over time could mean that caregiving effects do not last or that individuals find ways to cope with them, however, it might also result from selective attrition as individuals with demanding caregiving tasks are more likely to drop out of the panel surveys.

To better understand the longer-term health effects of care provision we explore the health effects of providing informal care in the UK context using data from the Understanding Society (USoc) longitudinal survey. We estimate both (i) the immediate and longer-term health effects of providing informal care for up to 5 years after the initial caregiving decision and (ii) the effect of providing additional years of care. These effects, and their relation to care intensity and caregiver characteristics, help policymakers to gauge the potential consequences of informal care provision and to identify those subgroups in largest need of support.

This study contributes to the literature on the longer-term health effects of providing informal care for the caregiver in several ways. The detailed individual-level information on caregivers and recipients allows us, unlike most previous studies focusing often on female or spousal caregivers, to explore the heterogeneity of caregiving effects across different groups of caregivers (e.g. by gender, care-recipient and intensity of care). Further, we estimate the health effects of multiple years of care provision using a dynamic matching approach (Lechner, 2009). The added benefit of this approach is that we can investigate the impact of additional years of care provision to determine how health effects evolve with continued caregiving. To our knowledge we provide the first causal estimates of caregiving effects in the UK that take the temporal dimension into account by exploring short- and longer-term health effects while explicitly addressing the dynamic nature of caregiving across multiple periods. While the UK is similar to Germany (studied by Schmitz & Westphal, 2015) with regards to the prominent role of informal caregivers in delivering social care services (Comas-Herrera et al., 2010), the countries differ in their generosity of caregiver allowances and formal care alternatives (Curry et al., 2019). Our results thereby also provide new evidence on the existence of caregiving effects and their magnitude from a different institutional context.

We find strong negative effects on mental health that are concentrated among high-intensity caregivers and remain persistent for multiple years. Additionally, the estimates from our dynamic matching procedure indicate that the mental health effect of care provision seems to persist for individuals providing care over multiple years. Using alternative outcome measures we confirm the consistency of our results and their economic relevance.

The paper proceeds as follows. Section 2 provides an overview of the UK long-term care system. The empirical strategy is outlined in Section 3, followed by a discussion of the underlying dataset in Section 4. Section 5 presents our results, starting with the baseline findings from a static model before proceeding to the dynamic matching approach. Robustness checks are described in Section 6, followed by a discussion and conclusion in Section 7.

Institutional background

Formal LTC in the UK is organized in a mixed-system combining universal and means-tested benefits. Health services provided by the National Health Service (NHS) are free at the point of delivery and predominantly financed from general taxation. The health-related components of LTC, which mostly entail nursing services, are funded via the NHS when granted by the GP (Comas-Herrera et al., 2010). Other types of LTC, such as residential care and help with personal tasks at home, are the responsibility of local authorities (Glendinning, 2013). Access to these services is dependent on locally determined needs-assessments. This care is offered via a safety-net structure requiring individuals to deplete their wealth before becoming eligible for publicly funded care (Colombo et al., 2011).[2] This system ensures that publicly funded LTC services are only provided to those with severe needs and unable to pay themselves (Fernández et al., 2009).[3] In 2015 the UK spent about 1.5% of its GDP on LTC with 23% of these expenditures related to social care (ONS, 2015).

As public LTC services are means-tested, a large part of LTC is provided informally with more than 18% of the UK 50+ population providing care (OECD, 2019). Additionally, more than a third of all caregivers do so for more than 20 hours per week according to data from the 2011 UK Census (ONS, 2013). In response to this large dependence on informal care, various policies to support informal caregivers (e.g. by providing information or support groups) are in place. The 2014 Care Act gave caregivers the right to receive a needs-assessment and corresponding support services (European Commission, 2018). However, reaching caregivers with this support is difficult as only six percent of caregivers receive any form of local authority support (Yeandle, 2016). Financial support is offered to informal caregivers via a "carer's allowance" (Carers UK, 2016). This allowance, amounting to £66.15 a week (approximately $86) in 2020 (UK Government, 2020) is paid to

[2] Income and assets (including under certain circumstances housing wealth) are considered. Individuals with assets above GBP 23.250 are ineligible for support. Those with assets between GBP 14.250 - GBP 23.250 (approximately $18.448 - $30.100) are required to contribute to the costs while individuals with assets below GBP 14.250 have their costs completely covered (NHS, 2018).

[3] In case of self-funding expected costs are about £15/hour (approximately $19) for home care (Age UK, 2019a) and £600 and £800/week (approximately $777 and $1036) for care homes and nursing homes (Age UK, 2019b).

caregivers who meet restrictive conditions.[4] As the take-up of the allowance and its monetary value is low it is not a potent incentive to take up informal care for the related monetary gains (Colombo et al., 2011).

Methods

The decision to provide informal care is not random. Individuals 'select into' informal caregiving, thereby creating endogeneity when studying its impact on health. We aim to overcome this problem by matching individuals on observable characteristics affecting health outcomes and the decision to provide informal care. To do so, we follow the intuition regarding the caregiving decision as proposed by Schmitz and Westphal (2015) who define three areas affecting the transition into informal care. The first are care obligations, as the most important determinant of informal care provision is the presence of a family member in need and the presence of alternative potential caregivers. The second category, willingness to provide care, refers to personality traits and socio-economic characteristics, as these affect individuals' inclination towards providing care. Lastly, the ability to provide care refers to individuals' own health status.

Our empirical strategy builds upon the potential outcomes framework by Rubin (1974) and addresses the endogeneity of providing informal care using regression adjusted propensity score matching (Rubin, 1979). The main assumption underlying propensity score matching is the conditional independence assumption (CIA). The CIA in our context states that after conditioning on a set of observable variables the potential health outcomes for both caregivers and non-caregivers are the same in the absence of informal care provision at all considered time periods. This implies that differences in health outcomes between caregivers and non-caregivers can be attributed to the provision of informal care. While it is only possible to match upon observed differences, this is not necessarily a problem as often unobserved characteristics are correlated to observed differences (Stuart, 2010). As we are able to match upon a broad range of variables related to informal care and health, we assume that we are able to capture unobserved covariates via their correlations with variables included in our propensity score regression. Additionally, following Lechner (2009), we exploit the panel structure of our data to match individuals upon information from the period directly preceding informal care provision to make this assumption more credible. The advantages of this strategy are that (i) providing care cannot affect the covariates and (ii) the previous caregiving status likely captures much of the unobserved heterogeneity affecting treatment assignment. For example, elements related to health status might be affected by past care provision while at the same time affecting current treatment assignment.

Static matching

Our first aim is to estimate the longer-term impact of becoming an informal caregiver, abstracting from the question of the number of years someone provides care for. This static approach means that we match starting caregivers with non-caregivers and follow these two groups over time. We identify individuals as treated when we observe their transition into caregiving, everyone who does not report any care-episode is included in the control group (untreated).[5]

Propensity scores of providing informal care are estimated using probit models. We estimate the propensity of providing low, medium or high intensity informal care conditional on the variables affecting the transition into care provision at t_{-1}. We use these propensity scores to match treated to untreated individuals. To increase the quality of the matching we estimate propensity scores separately by intensity group: all individuals providing care, irrespective of the reported intensity, and separately for the different intensity levels of low, medium and high intensity as defined in the data section. To make use of the large amount of information available in the dataset we use a kernel matching approach that uses weighted averaging on the untreated sample to form the counterfactual group.[6] We assess the common support, whether there is sufficient overlap in characteristics between the treated and untreated individuals, as the risk of kernel matching lies in the increased chance of including "bad matches", untreated individuals that are highly dissimilar to the treated group, in the estimation (Caliendo & Kopeining, 2008). Furthermore, as we do not match on actual covariates but on propensity scores, we assess whether balance of covariates is achieved after the matching procedure. We do so by using the standardized bias (Rosenbaum & Rubin, 1985). In our baseline specification we only match based on the estimated propensity scores, allowing us to include nearly all untreated individuals, therefore using more information and lowering variance (Caliendo & Kopeining, 2008). However, matching only on propensity scores itself has some drawbacks especially when a large range of control variables is used (Iacus et al., 2012). In the robustness checks section, we explore this issue in greater detail by changing our matching approach using precise one-to-one matching and coarsened exact matching (Blackwell et al. 2009)

Finally, the average treatment effect on the treated (ATT) is estimated by regressing health outcomes on the treatment indicator (providing care) and all control variables used in the propensity score estimation with individuals in the control group weighted by their estimated kernel weights. By regressing on the control variables alongside the treatment indicator we aim to correct for remaining residual differences in the covariate distributions between the treatment and control group (Lechner, 2009; Rubin 1973). We do not use the covariates from later waves as these might be affected by the treatment. The health impact of providing care is estimated for the immediate time after first reported care provision and up to five years afterwards. Fig. 1 provides a graphical representation of the static and dynamic matching designs.

Dynamic sequential matching

The static matching approach aims to answer the question *"If an individual starts to provide informal care in period t_0 (for an undefined time spell) does it change his or her health outcomes thereafter?"*. The treatment group hence contains individuals who stopped providing care in t_1 and

[4] Individuals can receive the carer's allowance when they (i) are aged 16 or over (ii) provide at least 35 hours (h) of care a week; (iii) earn less than £123 per week (approximately $152); (iv) are not full-time students or studying for more than 21 h a week; (v) normally live in the UK and have been in the UK for at least two of the last three years (UK Government, 2020).

[5] Future informal caregivers, although an ideal pool of suitable control group members, are not included in the control group due to the way our data is structured. To maximize the number of observable treated individuals we pool starting caregivers from across waves. To assess the robustness of this decision we also considered a situation where only caregivers starting to provide care within their first three years of survey participation are included, effectively moving half of the caregivers into the control group with t_{-1} being their respective entry wave. The results are depicted in Figure A2.1 and are highly similar to our baseline results.

[6] We use the Stata command psmatch2 (Leuven & Sianesi, 2003) using an Epanechnikov kernel with a 0.03 bandwidth. The bandwidth choice is a trade-off between a small variance and an unbiased estimate of the true density function (Caliendo & Kopeining, 2008). While not reported in detail we have tested varying bandwidths, e.g. 0.01 and 0.06, with negligible impact on our results.

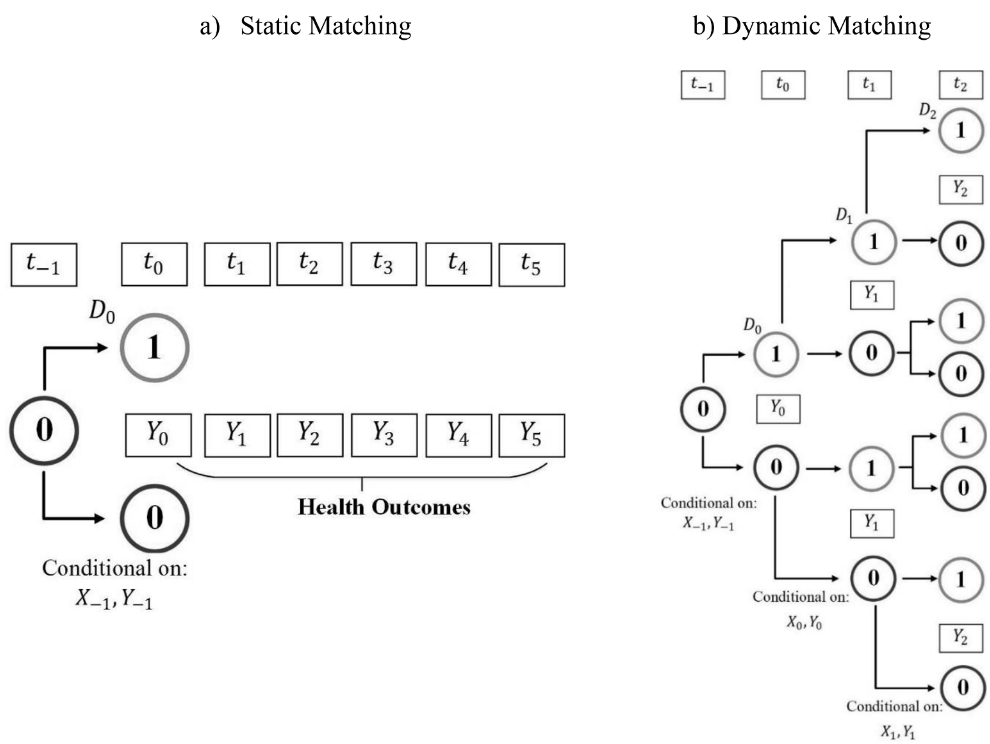


Fig. 1. Matching Designs (Own illustration based on illustration by Schmitz & Westphal 2017). Note: D refers to the decision to either: (1) provide informal care or (0) not to provide informal care at a certain decision node. Y refers to the health outcomes, X refers to the included covariates.

those who continued caregiving for various years. This might bias the treatment effect estimates for periods past t_0 as they are partially based on individuals that no longer provide care. To explore to what extent the longer-term treatment effects are driven by multi-period caregivers we use a dynamic matching approach following the work of Lechner & Miquel (2010) and Schmitz & Westphal (2017).

In contrast to the static approach, the dynamic matching explicitly estimates the effect of providing a second (third) year of care while considering a potential effect of caregiving in t_0 (t_1) on health and other endogenous covariates that influence the decision to care provision in subsequent years, such as labour market status. It therefore helps to understand how the health effect of care provision is affected by duration of care and whether the static treatment effects over time are representative for the population of multi-period caregivers. Further, it allows us to answer the question whether caregivers adapt to their caregiving responsibilities over time.

The dynamic matching is computationally demanding. It requires the estimation of treatment probabilities at all possible decision nodes, thereby leading to 2^T possible treatment paths where T is the maximum possible treatment duration. Further, it requires the availability of all health outcomes and covariates at the time-period prior to (continued) caregiving as the matching is repeated at all decision nodes. We limit ourselves to the case of $T = 3$ (see Fig. 1). This is motivated by two considerations. First, for our sample this time-window seems sufficient as the broad majority of caregivers provides a maximum of 3 years of consecutive care.[7] Second, in the dynamic framework time t is not defined relative to the first individual caregiving episode but fixed to allow for all potential treatment pathways to be observed, leading to less observable starting caregivers at t_0.

To illustrate the approach in more detail we provide an example showing the steps undertaken to estimate the marginal effect of providing two years of care instead of one. The treatment group in this example comprises everyone that provided informal care in both waves (t_0 and t_1), whereas the control group consist of everyone that provided care in the first wave (t_0) but not in the second (t_1). In the dynamic matching design in Fig. 1 this refers to comparing the group that followed the path *0–1-1* with the group following the route *0–1-0*.

Consider a binary indicator D_t encoding care provision in period t. As in the static estimations, we start our analysis by estimating the propensity of providing informal care at the first node ($D_0 = 1$) conditional upon not providing care in the period before, and pre-treatment health outcomes and other covariates using a probit model. The propensity of providing informal care at the first node is: $\Pr(D_0 = 1|X_{-1}, Y_{-1})$. Therefore the resulting estimate is equivalent to the immediate effect (t_0) estimated in the static matching framework.

In extension we also estimate the decision taken at the second node ($D_1 = 1$) were caregivers decide to (dis-)continue caregiving. We estimate the propensity scores of both options conditional upon already being a caregiver and on health and the other observables both at the first and the second node. The propensity of providing informal care at the second node after providing care in the first period is: $\Pr(D_1 = 1|D_0 = 1, X_{-1}, Y_{-1}, X_0, Y_0)$. The propensity of discontinuing care provision is:$\Pr(D_1 = 0|D_0 = 1, X_{-1}, Y_{-1}, X_0, Y_0)$.

Just like in the static model, we estimate the propensity of providing informal care by care intensity, here medium and high intensity informal caregiving are combined due to sample restrictions. We use the propensity scores to calculate inverse probability weights (IPW). As IPW estimates might be sensitive to very high or low weights from individuals with very high or low propensity scores (Robins et al., 2000), we check whether our results are robust to removing observations with

[7] Among the caregivers in our sample approximately 46.8% provide one year of care, 20.6% provide two and 13.0% provide three years of care. This leaves 19.6% of the caregivers providing more than three years of consecutive care. These numbers are based on caregivers starting in USoc waves 2–4 allowing all included respondents, in theory, to be able to have a caregiving spell of five or more years. Caregivers discontinuing their survey participation but continue to provide informal care are by definition not observed.

extreme weights as proposed by Lechner (2009).[8] Furthermore, for all scores we condition upon common support: in case no untreated counterparts with a similar propensity score for our treated respondents are present, these treated observations are excluded from the analysis.

Based on the estimated propensity scores we calculate inverse probability weights for both the treatment and the control group. These are defined as follows:

$$\frac{1}{(\Pr(D_0=1|X_{-1},Y_{-1}))*(\Pr(D_1=1|D_0=1,X_{-1},Y_{-1},X_0,Y_0))} \textit{for the treated group}$$

$$\frac{1}{(\Pr(D_0=1|X_{-1},Y_{-1}))*(\Pr(D_1=0|D_0=1,X_{-1},Y_{-1},X_0,Y_0))} \textit{for the control group}$$

We estimate the dynamic average treatment effect on the treated (those who provide two years of care) by regressing health on the treatment while controlling for remaining differences by adding all covariates from the previous waves and weighting the data using the calculated inverse probability weights. We hence estimate, in this example, the health effects at t_1 of providing care in t_0 and t_1, compared to only providing care at t_0.

This sequential matching strategy was proposed by Lechner (2009) to estimate treatment effects in settings with dynamic treatment durations. While it follows a similar intuition as the static matching procedure identification is based on an augmented version of the CIA: the weak dynamic conditional independence assumption. Consider the case above when comparing outcomes of two and one years of informal care. The weak conditional independence assumption combines two parts. Firstly, the initial conditional independence assumption stating that potential outcomes in t_0 and t_1 are independent of treatment status in t_0 once we match upon observables at t_{-1}. Secondly, that potential outcomes in t_0 and t_1 are independent of continued treatment in t_1 once we condition on control variables and outcomes at both t_{-1} and t_0 and treatment status at the initial node t_0.

Data

We use data from the Understanding Society (USoc) dataset, also known as the UK Household Longitudinal Study (UKHLS; University of Essex, 2019); an annually conducted representative panel survey of the adult UK population (aged 16+). It started in 2009 with approximately 40,000 respondents across 30,000 households as the successor of the British Household Panel Survey (BHPS), which ended in 2008. In 2010, members of the last BHPS-wave were invited to join the USoc after which an additional 8,000 individuals joined. This paper uses all nine completed waves conducted between 2009 and 2019.[9]

Informal caregivers are identified using the question *"Is there anyone living with you who is sick, disabled or elderly whom you look after or give special help to (for example a sick, disabled or elderly relative/husband/wife/friend etc.)?"*. Individuals providing care outside their own household are identified based on the question *"Do you provide regular service or help for any sick, disabled or elderly person not living with you? [Exclude help provided in course of employment]"*. Apart from being able to identify individuals providing care inside and outside their own household the questionnaire also covers care intensity (hours per week) and the relationship between the caregiver and care-recipient.

We explore differences in the impact of caregiving dependent on the reported hours of care in a typical week as it is covered in the caregiving module of the annual USoc individual surveys. Based on these reported hours (h) of care we split our sample of caregivers in three; low intensity (<10h of care per week), medium intensity (between 10 and 20h) and high intensity caregivers (more than 20h). When evaluating our results, one however must be aware of a potential downward bias in our estimates due to an underrepresentation of caregivers in the upper end of the intensity distribution. The share of high intensity caregivers in our sample (12.8%) is lower compared to the UK Census of 2011 which indicates that nation-wide about 37% of the caregivers provide care for more than 20h a week (ONS, 2013) or the 17% reported in the 2014 European Social Survey (ESS, 2014).

Health outcomes

Various studies report the impact of care provision on mental and physical health (e.g. Pinquart & Sörensen, 2003b). To identify potential changes in both health domains we use the SF-12 health questionnaire in which individuals self-report on 12 questions related to various aspects of their own health in the past four weeks. From the survey we derive the physical (PCS) and mental (MCS) component summary scores which are constructed using different subscales related to physical and mental health.[10] The two health scales are validated for the UK context and range from 0 to 100, where a higher score represents better health. By construction MCS and PCS scores have a mean of 50 and standard deviation of 10 (Ware et al., 1995).

Time structure

For the static matching procedure, we define a relative time variable depending on an individual's first reported care-episode as observed in the sample, meaning the first time an individual reports to provide care as part of his or her individual observation period as a survey participant.[11] Fig. 2 provides a visualization of this time structure for the case of an individual entering the survey in wave 1 reporting their first care episode in wave 4. Among caregivers t_{-1} is defined as the period before the first reported caregiving episode. For everyone in the control group t_{-1} is the individual's first appearance as a survey participant in absence of any care episode during their participation. This time structure is chosen to maximize the number of observable treated individuals.

The analysis sample for the dynamic specification uses an augmented time-structure to allow for the modeling of all decision nodes between t_{-1} and t_2 and the comparison of various care trajectories. The time variable is normalized to t_{-1} being the entry wave of an individual into the panel for those who provide no care at any time-point and caregivers who start providing care within the first four participation waves. To increase the number of observable caregivers in the different caregiving trajectories we additionally include individuals whose caregiving trajectory starts after at least four periods of not providing informal care. For these the fifth participation is defined as t_{-1}.We therefore emulate the time-structuring in the static design by pooling caregivers from different starting waves

[8] We check our results using two approaches: (1) dropping scores for the first decision that are smaller than 5% or larger than 95% of the estimated propensity score distribution (2) dropping scores for the first decision that are smaller than 1% or larger than 99% of the estimated propensity score. Qualitatively our results are robust to these different specifications (results available upon request).

[9] It follows an overlapping panel structure where waves cover two years but overlap for one and individual respondents are surveyed every 12 months. Therefore, nine waves are available for the 10-year time period.

[10] The PCS comprises the subscales: Physical functioning, Role-Physical, Bodily Pain and General Health. The MCS comprises the subscales: Vitality, Social Functioning, Role-Emotional and Mental Health.

[11] Ideally, we would like to ensure that individuals observed as becoming first-time caregivers did not provide informal care previous to their participation in USoc which by definition we cannot observe. Unobserved previous caregiving status and its potential effect on certain covariates such as labour market participation or health outcomes would result in a violation of the conditional independence assumption. This cannot be checked without explicitly asking respondents about previous caregiving. As also among the control group individuals might have provided care before participating in USoc we do not think that this leads to a relevant bias in practice.

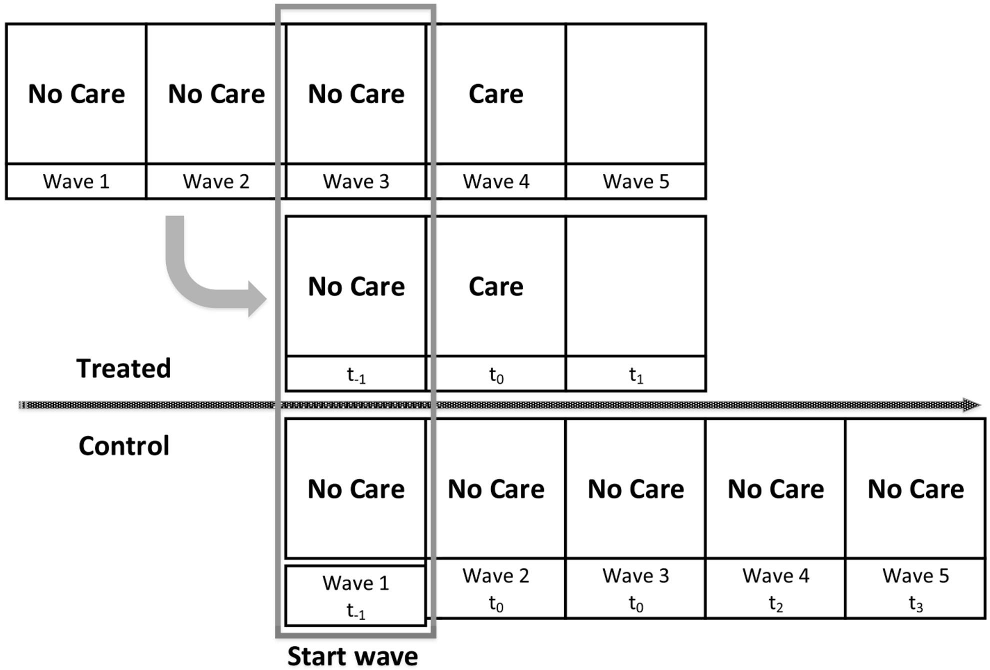


Fig. 2. Static Dataset – Time structure example.

but allowing these to have caregiving spells of up to three years. The important difference is that not all care-giving trajectories start at t_0 and at all decision nodes the control group contains individuals who will transition into a caregiving spell in future periods.

Sample selection

We construct two distinct datasets to implement the static and dynamic matching procedures. Individuals who identified as caregivers in their first observation period are excluded as transition into caregiving is not observed. For the static estimation we include all individuals that provide information for at least two time points on their health outcomes (t_{-1} and t_0) and provide full information on all covariates used in the propensity score estimation at t_{-1}. Individuals here predominantly drop out of the sample because they are proxy respondents or fail to provide sufficient information on their health outcomes or existing family members.[12] Individuals remain in the sample during the subsequent time points t_1 to t_5 in case information on the outcome-variables is available. For the dynamic estimation procedure, data requirements are more restrictive as we re-estimate propensity scores at each decision node. For this analysis, only individuals with complete information on all control variables for three waves (t_{-1} to t_1) and full information on outcome variables for four consecutive periods are included in the sample. Table A1.1 in the appendix provides an overview of the exclusion criteria applied to the analysis samples.

Results

Matching quality

The descriptive statistics for the static matching sample are depicted in Table 1. Before the propensity score matching there is strong covariate imbalance between the control and treatment groups. For example, the individuals in the treatment group are on average almost nine years older and in worse health than their counterparts in the control groups across all health measures. We calculate the standardized bias for each covariate by taking the difference in means between the treatment and control group and dividing it by the square root of the average sample variance control and treatment groups (Rosenbaum & Rubin, 1985). This provides a standardized measure of the difference between both groups expressed as percentage points[13]. We follow the rule of thumb suggested by Caliendo & Kopeinig (2008) which states that there is sufficient balance when the bias is below 3–5%. Prior to the matching there is considerable imbalance between caregivers and the control pool. Matching corrects this imbalance, as can be seen in the post-matching differences between the groups. Details on the matching can be found in Appendix A3. Table A3.1 compares the pre- and post-treatment covariate balance and provides evidence for the large imbalance between treatment and control groups but also within the treatment group across the different levels of caregiving intensity. The matching by intensity group balances the covariates in all groups. Fig. A3.1 depicts for each intensity group the overlap in propensity scores as well as the imbalance pre- and post-matching graphically, providing evidence that the matching is successful. In each specification a small number of individuals is identified as off-support and dropped from the analysis: four in the any-care specification, none for low-intensity caregivers only in the treatment group and three and four for medium- and high-intensity caregivers. The corresponding information for the dynamic matching is displayed in Appendix A4.

Static matching results – Treatment effects by care intensity

The results of the baseline static matching procedure by intensity will be presented graphically. All underlying estimates are reported in Appendix Table A1.2. While the graphs depict the full sample results by

[12] Co-habiting family members are observed at every wave but family networks outside of respondents' households are only inquired about every second wave. If living family members are reported in a future wave this is applied backwards where information is missing. Likewise, if a family member is reported as deceased this is applied forward.

[13] The exact formula is $(\bar{x}_{D=1} - \bar{x}_{D=0})/\sqrt{0.5^*(V(x_{D=1}) + V(x_{D=0}))}$ where x is the control variable of interest, $V(x)$ is the sample variance for the treated ($D = 1$) and control ($D = 0$) groups.

Table 1
Descriptive Statistics and Matching Quality.

	Summary Statistics					Post-Matching Differences			
	Control Pool		Treated			Bias in %			
	Mean	SD	Mean	SD	Bias in %	Any Care	<10 h	10–20 h	>20 h
Mother alive	0.70	(0.46)	0.59	(0.49)	**−22.00**	0.40	0.60	−2.80	−3.40
Mother age	59.68	(11.28)	68.12	(12.71)	−4.00	0.80	1.10	−1.80	−2.70
Father alive	0.63	(0.48)	0.45	(0.50)	**−36.50**	1.30	1.40	−1.00	−2.90
Father age	60.99	(10.76)	68.01	(12.08)	**−23.70**	1.90	2.30	0.60	−2.30
Both parents alive	0.59	(0.49)	0.39	(0.49)	**−40.90**	1.00	1.30	−2.10	−2.70
Siblings alive	0.88	(0.32)	0.87	(0.34)	**−5.60**	0.10	0.60	−0.10	−3.60
Partner existing	0.63	(0.48)	0.70	(0.46)	**13.90**	0.50	0.30	0.20	1.70
Partner age	43.67	(14.13)	52.75	(14.61)	**35.90**	1.00	0.10	2.00	4.50
Own age	41.37	(17.14)	50.53	(16.14)	**55.00**	0.10	−0.10	4.00	5.20
Female	0.53	(0.50)	0.59	(0.49)	**13.10**	0.00	0.00	2.20	0.40
Children in HH	0.39	(0.49)	0.31	(0.46)	**−16.60**	−1.10	−0.40	−1.70	0.30
Children < 14 in HH	0.35	(0.48)	0.28	(0.45)	**−16.40**	−0.90	−0.20	−1.40	0.20
Highest Education: Primary/other	0.18	(0.38)	0.21	(0.41)	**9.20**	−1.60	−1.50	−1.10	−0.30
Highest Education: Secondary	0.44	(0.50)	0.43	(0.50)	−1.40	−0.30	−0.70	1.80	0.80
Highest Education: Tertiary	0.39	(0.49)	0.36	(0.48)	**−6.10**	1.70	2.00	−0.90	−0.50
Any paid work	0.63	(0.48)	0.57	(0.50)	**−14.00**	2.80	2.80	0.90	0.20
Full time work	0.50	(0.50)	0.40	(0.49)	**−21.10**	2.90	2.90	−0.70	−1.60
Self employed	0.07	(0.25)	0.08	(0.27)	3.90	0.20	0.60	0.00	0.30
Unemployed	0.05	(0.22)	0.04	(0.21)	−2.70	−1.20	−0.90	0.70	0.40
Homecarer	0.05	(0.22)	0.06	(0.24)	**5.10**	−2.30	−1.40	−1.40	−1.70
Disabled	0.02	(0.15)	0.04	(0.19)	**8.20**	−1.90	−2.20	−2.90	−2.70
Retired	0.14	(0.35)	0.26	(0.44)	**29.30**	−0.70	−1.20	2.30	3.70
Equivalent HH income	1647.76	(1168.87)	1686.65	(1219.98)	3.30	3.30	2.80	1.60	1.00
Big 5: Openness	4.59	(1.27)	4.58	(1.32)	−1.10	1.10	1.10	−1.10	−0.10
Big 5: Conscientiousness	5.45	(1.09)	5.57	(1.10)	**11.20**	2.20	1.90	2.00	1.40
Big 5: Extroversion	4.59	(1.30)	4.62	(1.29)	2.80	−0.70	−0.50	−1.30	−1.50
Big 5: Agreeableness	5.57	(1.03)	5.69	(1.02)	**11.60**	1.00	0.60	2.20	1.30
Big 5: Neuroticism	3.56	(1.43)	3.53	(1.45)	−1.90	0.40	−0.30	−0.20	1.20
Self Assessed Health	2.42	(1.06)	2.58	(1.05)	**15.10**	−2.80	−2.50	−1.50	−2.40
Mental Component Score	51.09	(9.10)	49.98	(9.85)	**−11.80**	1.70	1.80	0.70	1.80
Physical Component Score	51.56	(9.97)	49.46	(10.86)	**−20.10**	3.00	2.70	1.80	2.20
Long Standing Illness/Disability	0.29	(0.45)	0.37	(0.48)	**18.00**	−2.20	−2.10	0.10	−1.60
Number of Functional Limitations	0.44	(1.22)	0.55	(1.29)	**8.50**	−3.00	−2.60	−1.70	−3.20
Satisfaction with own Health	5.09	(1.60)	4.73	(1.72)	**−21.60**	2.10	1.60	0.60	1.00
Satisfaction with Income	4.64	(1.64)	4.51	(1.69)	**−7.80**	2.90	2.00	1.50	1.20
Satisfaction with Leisure Time	4.77	(1.62)	4.79	(1.68)	0.90	0.50	−0.10	0.80	0.50
Satisfaction with Life	5.37	(1.37)	5.20	(1.47)	**−11.60**	2.00	1.70	1.00	−0.10
Inversed GHQ Score	10.41	(2.73)	10.26	(3.01)	**−5.20**	0.80	0.90	1.00	−0.30
Unique Individuals		**13141**		**7106**		**7102**	**5249**	**812**	**934**

Source: USoc Waves 1–9, own calculations. Note: The number of informal caregivers (treated) in the different intensity groups do not add up to 7106 as 104 caregivers provide unclear information regarding the amount of care provided. Individuals off-support are not included in the post-matching groups. Regional dummies are ommitted. Big 5 variables range from 1 to 7 (low–high); SAH ranges from 1 to 5 (excellent-poor); satisfaction questions range from 1 to 7 (completely dissatisfied-completely satisfied); SF-12 Mental and Physical Component Scores range from 0 to 100 (lowest – highest level of health); inversed General Health Questionnaire (GHQ) score ranges from 0 to 12 (most distressed - least distressed).

caregiving intensity, we will in text also discuss the results by gender which can be found in Table A1.2. In the baseline analysis we estimate the effect of any informal care provision irrespective of the reported intensity. Fig. 3 depicts the estimated ATTs on both the (a) mental and (b) physical health scores across time and their 95% confidence intervals. To show that matching resolved any pre-treatment differences in health between both groups, the graphs depict the pre-treatment estimate at t_{-1}. As can be seen in all graphs, before commencing of care provision no differences in physical or mental health are present between the matched groups.

To ease interpreting estimated coefficients with respect the magnitude of reported ATT estimates we report the corresponding effect size in percentage points of the standard deviation in brackets.[14] In the first years after the start of care provision, we observe small immediate negative effects in the mental domain of −0.476 ($p < 0.01$; 5.07% SD) at t_0 and −0.687 ($p < 0.001$; 7.32% SD) at t_1. In later periods these effects remain negative although not significant at the 5% level. In the physical domain baseline estimates indicate a small positive effect of 0.441 ($p < 0.01$; 4.26% SD) at t_0 but no effects thereafter. The separate analyses by gender show that the results are driven by female caregivers. Female caregivers experience small negative mental health effects of −0.507 ($p < 0.01$; 5.17% SD) at t_0 and −0.915 ($p < 0.01$; 9.34% SD) at t_1 and −0.595 ($p < 0.05$; 6.07% SD) at t_3. Male caregivers, do not experience consistent mental health effects but we observe a small positive impact on the PCS of 0.507 ($p < 0.05$; 5.14% SD) at t_0 and 0.870 ($p < 0.05$; 8.83% SD) and 0.764 ($p < 0.05$; 7.76% SD) at t_4 and t_5.

To explore heterogeneities in the estimated treatment effects we subdivide caregivers into treatment groups according to the reported weekly hours of care. Fig. 4 plots the results by care intensity for (a) mental and (b) physical health. Low intensity caregiving (<10h per week) is depicted in light grey, medium intensity (between 10 and 20h per week) in dark grey, and high intensity caregiving (>20h per week) in

[14] We use the unweighted pre-treatment standard deviation estimates for the mental and physical component scores for the respective analysis samples.

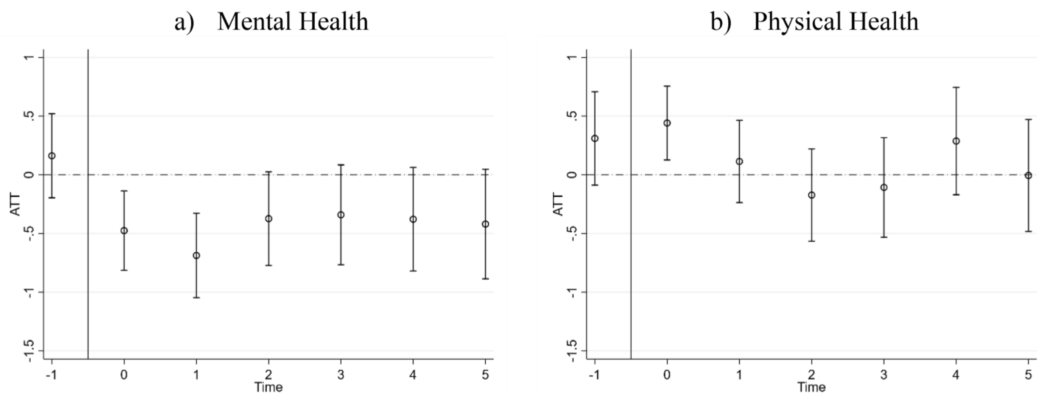


Fig. 3. Health Effects of Any Care Provision. *Source:* USoc Waves 1–9, own calculations. *Note:* ATT estimates pre-treatment (t_{-1}) is the mean difference between treatment and control groups after the matching procedure is applied.

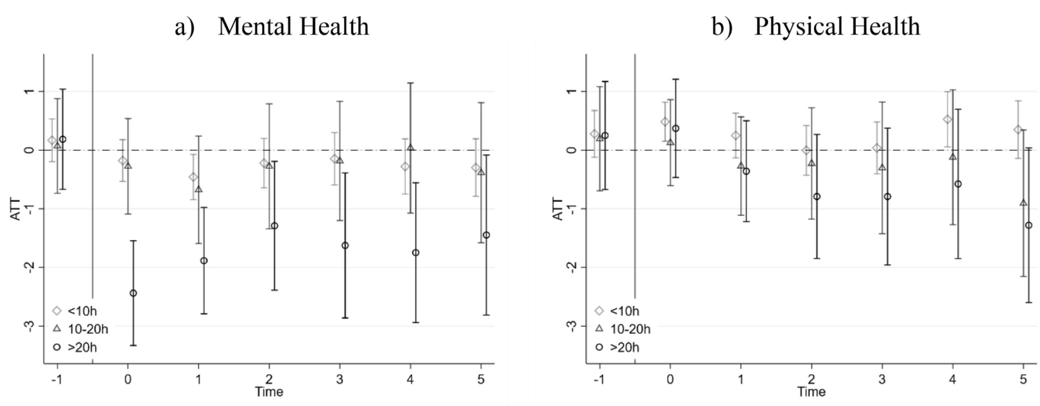


Fig. 4. Health Effects by Intensity of Care Provision. *Source:* USoc Waves 1–9, own calculations. *Note:* ATT estimates pre-treatment (t_{-1}) is the mean difference between treatment and control groups after the matching procedure is applied.

black. Low intensity care is the most commonly observed with 5,249 individuals (74%), followed by 934 high (13%) and 812 (11%) medium intensity caregivers.[15] By construction, effects in later periods are less precisely estimated as not all caregivers are observed for all years. At t_5 only 1,881 or 36% of the initial low intensity caregivers are still observable, for medium intensity this is 296 (36%) but only 293 (31%) high-intensity caregivers.

The large heterogeneity in the estimated treatment effects underline the importance of care intensity. In the mental domain we do not find effects for low or medium intensity caregivers, although all estimates point towards the negative direction. Among individuals providing care for more than 20h per week we observe strong initial negative effects of -2.438 ($p < 0.001$; 26.24% SD) at t_0 and -1.884 ($p < 0.001$; 20.27% SD) at t_1. While these effects decrease for subsequent periods, they remain largely persistent with -1.290 ($p < 0.05$; 13.86% SD), -1.625 ($p < 0.05$; 17.49% SD), -1.748 ($p < 0.01$; 18.82% SD) and -1.448 ($p < 0.01$; 15.59% SD) at t_2 to t_5. The previously apparent differences in caregiving effect by gender decrease when stratifying the samples by care-intensity. Both male and female high-intensity caregivers experience comparable negative mental health effects following transition into caregiving; -2.471 ($p < 0.001$; 25.40% SD) for females and -2.396 ($p < 0.001$; 27.66% SD) for males at t_0. However, for female respondemts the negative effects are more persistent than for males.

In the physical domain the pattern across care intensity levels is different. For low intensity caregivers we find a small positive immediate effect of 0.441 ($p < 0.01$; 4.36% SD) at t_0 while for the other intensity groups the coefficient is similar but insignificant. At subsequent periods the estimated effects vary considerably. For low intensity caregivers the estimated coefficients are insignificant while varying around zero, except for t_4 in which we observe a positive effect of 0.561 ($p < 0.05$; 5.55% SD). For medium and high intensity caregivers' coefficients are negative throughout except for the positive estimates at t_0 but not significantly different from zero. The coefficients for high intensity care seem to follow a downward trend. In the Appendix we provide evidence that this pattern seems driven by age-dependent physical health trends captured inadequately in the propensity score based matching (see Robustness Checks section and Figure A2.2 and A2.3).[16]

Static matching results – Treatment effects by Caregiver-Recipient relationship

While the results in Figs. 3 and 4 suggest that the negative effects of providing informal care are mediated by the intensity of care provided an alternative mediating factor could also be the relationship between caregiver and recipient (Bobinac et al., 2010). To explore this, we

[15] Caregivers providing no intensity information are excluded (104). Detailed sample sizes by time-period t can be found in Appendix Table A1.2.

[16] As propensity scores are a summary measure estimated using many covariates, these age-related trends are not guaranteed to be perfectly captured, e.g. when including age continuously instead of age-group dummies that capture different trends across age-groups. For example, a younger individual might receive a high propensity score due to his/her physical health being low and/or other strong predictors but would be faced with an entirely different physical health trajectory in the short and medium term compared to an older individual. As illustrated in the Appendix this age-dependent trend is not present for mental health.

Table 2
Treatment effects including care-relationship interaction terms.

Any Care												
	t = 0		**t = 1**		**t = 2**		**t = 3**		**t = 4**		**t = 5**	
	MCS	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**
Any Care	−0.207	0.655	−0.458	0.196	−0.294	−0.122	0.043	−0.133	−0.135	0.447	−0.385	0.152
	(0.201)	(0.183)	(0.221)	(0.212)	(0.246)	(0.236)	(0.258)	(0.262)	(0.279)	(0.289)	(0.306)	(0.314)
Spousal Care	**−1.640*****	**−0.754****	**−1.543*****	**−0.766***	**−1.193****	−0.561	**−1.679****	−0.502	**−1.712****	−0.573	**−1.508***	−0.888
	(0.301)	(0.291)	(0.346)	(0.324)	(0.379)	(0.389)	(0.451)	(0.445)	(0.501)	(0.521)	(0.617)	(0.583)
Parental Care	−0.082	−0.249	−0.025	0.060	0.198	0.062	−0.355	0.213	−0.032	−0.187	0.374	−0.089
	(0.216)	(0.194)	(0.246)	(0.230)	(0.273)	(0.253)	(0.300)	(0.287)	(0.342)	(0.329)	(0.383)	(0.376)
Control		**13141**		**12497**		**10575**		9753		**8701**		**8185**
Treatment		**7102**		**5863**		**4901**		**4200**		**3239**		**2503**
Spousal Care		1042		830		677		546		429		322
Parental Care		2937		2477		2088		1806		1409		1112

High Intensity Care												
	t = 0		**t = 1**		**t = 2**		**t = 3**		**t = 4**		**t = 5**	
	MCS	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**
Any Care	**−2.486*****	0.133	**−1.724***	−0.635	−1.038	−1.100	−1.051	−1.531	−1.430	−0.765	−1.714	−1.872
	(0.604)	(0.537)	(0.676)	(0.603)	(0.763)	(0.673)	(0.831)	(0.787)	(0.846)	(0.871)	(0.975)	(0.942)
Spousal Care	−0.056	0.224	−0.509	0.108	−0.888	−0.082	−1.273	0.944	−1.143	0.370	−1.153	1.065
	(0.689)	(0.617)	(0.752)	(0.702)	(0.809)	(0.790)	(1.003)	(0.905)	(1.083)	(1.130)	(1.408)	(1.249)
Parental Care	0.291	0.623	0.178	0.992	0.404	1.428	−0.349	1.612	0.534	0.189	3.095	0.817
	(0.836)	(0.687)	(0.903)	(0.801)	(1.022)	(0.876)	(1.190)	(1.007)	(1.225)	(1.197)	(1.513)	(1.379)
Control		**13141**		**12497**		**10575**		**9753**		**8701**		**8185**
Treatment		934		722		606		488		384		293
Spousal Care		374		286		239		185		149		110
Parental Care		207		167		149		115		94		68

Source: USoc Waves 1–9, own calculations. Note: *$p < 0.05$, ** $p < 0.01$, ***$p < 0.001$, standard errors in parentheses. MCS columns depict the estimated coefficients for the SF12 Mental Component Scores, PCS the SF12 Physical Component Scores.

include an interaction term for the caregiver care-recipient relationship. In a first step we do so for the specification in which all caregivers are in the treatment group, irrespective of the reported intensity. As shown in the upper half of Table 2 in this specification it seems that while informal care itself has no effects on either mental or physical health, the provision of care to one's partner does have a strong, negative effect on mental health. The interaction term for parental care on the other hand is small and insignificant across all time-periods.

However, when doing the same exercise but only on the group of caregivers for which our baseline specification finds consistent and substantial negative mental health effects, those providing high-intensity care, this is not observed. These results are shown in the lower half of Table 2. The reason for this pattern is likely found in the different distribution of intensity by informal caregiver-recipient relationship. While care to parents makes up 41% of all caregivers it is only 22% among high intensity caregivers. Care to a partner, which predominantly happens within the own household, makes up only 15% of the overall population of caregivers but 40% of the high-intensity caregiving sample.

Dynamic matching results – Treatment effects by care intensity

Next to estimating the impact of at least one year of care provision, we aim to investigate the impact of providing additional years of informal care and the extent to which the static ATT estimates are driven by ignored changes in treatment status over time. For this dynamic matching approach, we estimate the propensity of (not) providing informal care at every decision node and drop scores in case the observation is off support. In Appendix A4 we report the propensity scores for the different care-trajectories as well as an overview of the excluded individuals. Further, we estimate the treatment effects using the static matching approach for the same sample used in the dynamic matching to provide a comparison between both estimation strategies. For the dynamic matching we merged the groups of medium and high intensity caregivers due to concerns about statistical power. Further we excluded individuals with unstable care trajectories to not wrongfully capture the impact of increasing care intensity among continuing caregivers. The results for both mental and physical health are depicted in Table 2. Due to data limitations, we are unable to estimate the propensity of providing medium or high intensity care after the second node. Most individuals continue care provision when already done so for two years in a row and with a higher intensity. This limited variation in care-continuation among this groups obstructs us from running these models which are highly data-demanding as they include covariates from all previous waves. In the below presented results we hence only present the estimates for any and low-intensity caregivers at the 3rd year.

The first column of Table 3 depicts the estimated caregiving effect at t_0 for the first period of informal care. For mental health both dynamic and static matching indicate a small and insignificant negative coefficient for any care provided of −0.312. When separating the different intensity levels there are no significant differences for low intensity caregivers while for medium or high intensity caregivers the estimates are negative and significant with −1.374 ($p < 0.01$; 14.66% SD).

At t_1 the static matching, which pools both continuing and discontinuing caregivers together, again indicates a continuing negative mental health effect of −1.413 ($p < 0.001$; 15.08% SD). The dynamic matching estimates, which account for the impact of previous-period caregiving on covariates and health outcomes, indicate that the health effects among continuing caregivers are larger. When focusing on the group of continuing caregivers the difference is −2.254 ($p < 0.001$; 24.05% SD) when using the never caregivers as a control group. Explicitly comparing one against two years of medium/high intensity care results in an estimated health impact of the second year of care provision that is again larger at −1.836 ($p < 0.05$; 19.59% SD) although less precisely estimated.

For physical health no effects are found at t_0, nor among the any and low estimates at any of the other timepoints. Later on, we however observe mixed results. Using the static approach, a negative impact at t_2 is observed. Focusing on the dynamic results, the estimates at the second

Table 3
Dynamic Matching Estimates.

	Mental Health						
	t = 0	t = 1			t = 2		
	Static/Dynamic	Static	Dynamic (2v0)	Dynamic (2v1)	Static	Dynamic (3v0)	Dynamic (3v2)
Any	−0.312	−0.394*	−0.350	−0.151	−0.190	0.0195	0.610
	(0.194)	(0.201)	(0.315)	(0.405)	(0.202)	(0.405)	(0.771)
Low Intensity	−0.007	−0.116	−0.103	0.197	0.090	−0.215	0.582
	(0.209)	(0.218)	(0.323)	(0.451)	(0.224)	(0.518)	(0.829)
Medium/High Intensity	**−1.374****	**−1.413****	**−2.254*****	**−1.836***	**−1.148***		
	(0.462)	**(0.470)**	**(0.546)**	**(0.810)**	**(0.452)**		
	Physical Health						
	t = 0	t = 1			t = 2		
	Static/Dynamic	Static	Dynamic (2v0)	Dynamic (2v1)	Static	Dynamic (3v0)	Dynamic (3v2)
Any	0.178	−0.148	0.012	0.144	0.059	0.014	−0.147
	(0.173)	(0.191)	(0.283)	(0.383)	(0.200)	(0.361)	(0.574)
Low Intensity	0.340	0.016	0.080	0.094	0.355	0.477	−0.073
	(0.193)	(0.217)	(0.303)	(0.437)	(0.224)	(0.375)	(0.622)
Medium/High Intensity	−0.334	−0.707	**1.379****	1.267	**−1.010****		
	(0.387)	(0.390)	**(0.506)**	(0.800)	**(0.424)**		
Treatment (Control)	1672 (18812)	1672 (18812)	700 (17268)	700 (967)	1672 (18812)	406 (15756)	406 (274)
Low Intensity	1285 (18812)	1285 (18812)	536 (17260)	536 (738)	1285 (18812)	296 (15865)	296 (223)
Medium/High Intensity	313 (16081)	313 (16081)	146 (16660)	146 (186)	313 (16081)		

Note: The static and dynamic results at t = 0 slightly differ at the second decimal as either matching or inverse probability weighting is used, the depicted results are the static results. The table presents the dynamic ATT, the effect of providing an additional year of care compared to an individual following an alternative care trajectory. A second/third year of informal care is compared to: not providing informal care (2v0 or 3v0); providing care for only one year (2v1); or providing care for two years (3v2). It compares the health of treated and matched controls based on the information from the directly preceeding wave. * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors are in parentheses. Please note that the displayed static results are based on a different sample than the baseline static results as a differently conditioned sample is used. See the Appendix for more details.
Source: USoc, own calculations.

node however point into the positive direction, and turn significant when comparing individuals observed for two years of care compared to their non-caregiving counterparts. While this might indicate a positive impact of continuing care provision, we would like to be cautious in interpreting these results in particular. Among this small group matching at all nodes becomes difficult. This seems specifically the case among results in the physical health dimension for high intensity caregivers. As presented in the robustness checks section these effects are most sensitive to alternative model specifications. Additionally, please note for all of these estimates, the standard errors rather large. This makes it for many of these estimates, impossible to differentiate between a zero-effect or a non-significant effect.

Robustness checks

There are two primary concerns related to the estimation of causal effects using matching-based estimators. Firstly, a violation of the main identifying assumption, the conditional independence assumption (CIA), and secondly the model dependence of estimated treatment effects to the specific way in which matches are obtained. The CIA is an inherently untestable assumption and hence one has to rely on data-driven methods to explore the extent to which the observed treatment effects could be reasonably explained by an omitted variable that effects both selection into treatment and (health) outcomes in absence of treatment. We applied the simulation-based method proposed by Ichino et al. (2008) to explore the sensitivity of our results to such an omitted variable having both a selection effect (s) and an outcome effect (d). An example for such an omitted variable could be the latent health of a family member leading to both an increased likelihood to provide care (selection effect) as well as lower mental health irrespective of treatment uptake (outcome). In Appendix Section A2 we provide a detailed description of the procedure. In short, the simulation re-estimates the

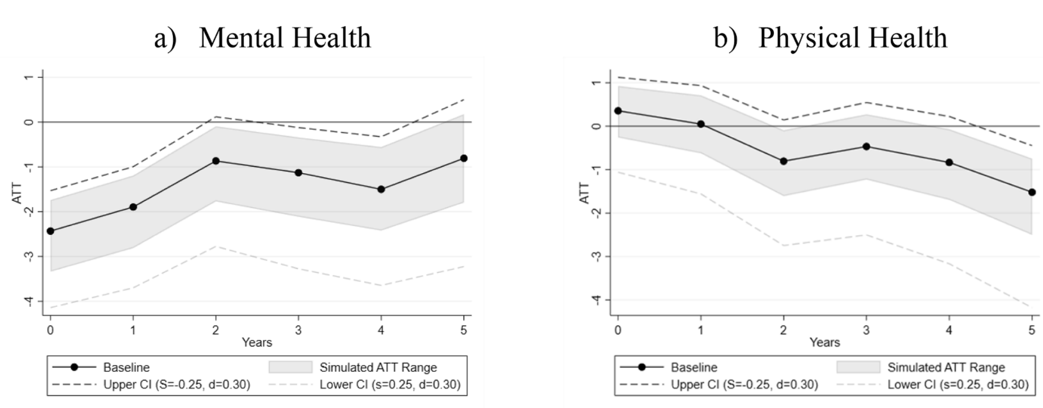


Fig. 5. Simulated Violation of the CIA. *Source:* USoc Waves 1–9, own calculations.

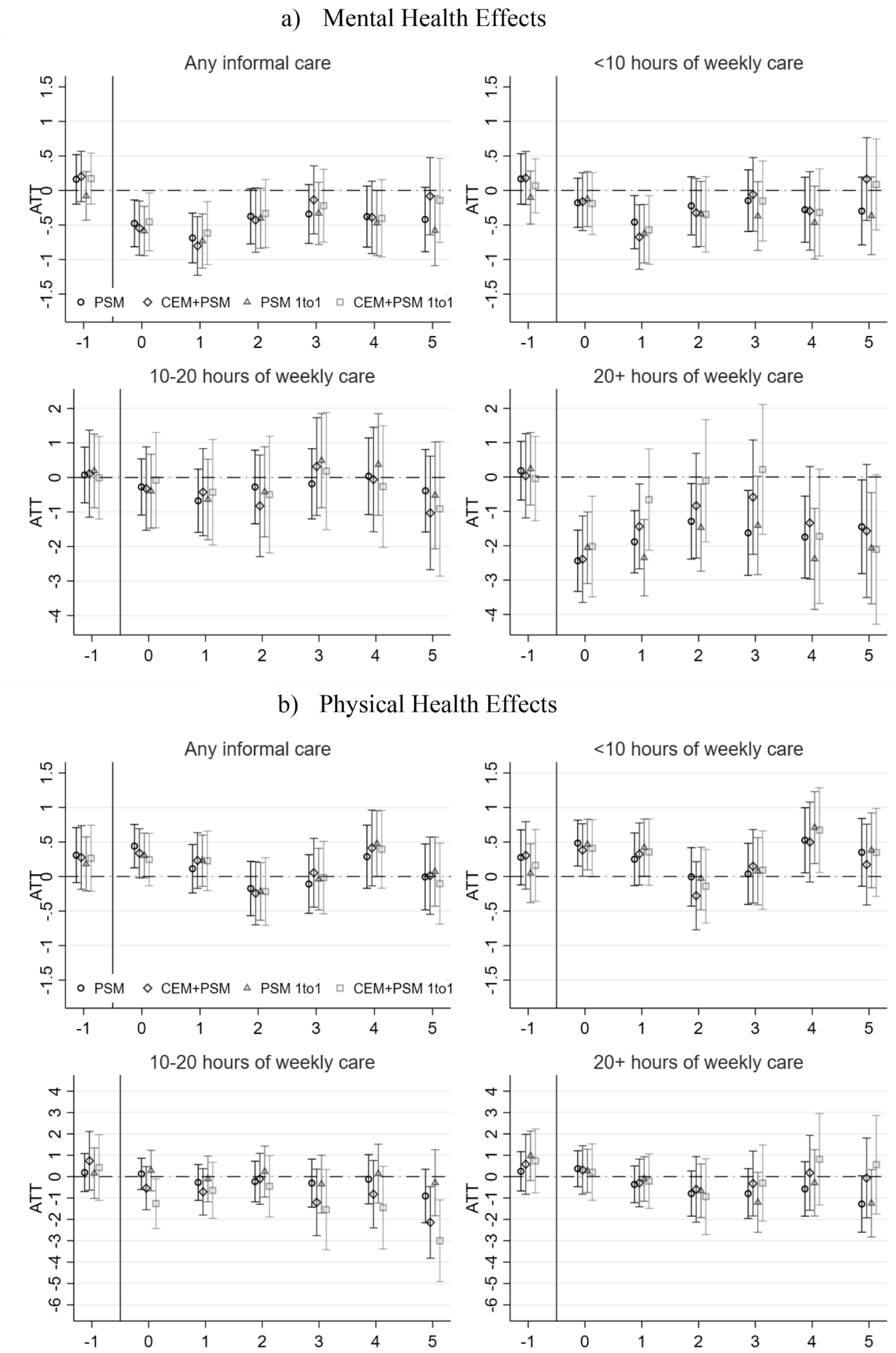


Fig. 6. Results by Matching Procedure. *Source:* USoc Waves 1–9, own calculations. *Note:* ATT estimates pre-treatment (t_{-1}) is the mean difference between treatment and control groups after the matching procedure is applied.

propensity scores leaving out one of our covariates to estimate the space of selection and outcome effects among our pool of covariates to obtain a range for s and d that is plausible (see Table A2.1). From this range we construct a confounder that has a combination of selection and outcome effects that would bias our estimated ATTs severely into either direction by choosing a combination of the highest observed *s* and *d* among our included control variables. Based on this we then calculate a range of ATTs and corresponding confidence intervals to obtain a bound on the treatment effects and their sensitivity to omitted variables. Fig. 5 depicts both of these for mental and physical health for the high-intensity caregiving group.

While there is clear evidence for a simulated confounder to shift the overall range of estimated treatment effects especially the mental health effects for the initial periods after providing informal care decrease in size but remain strongly significant.

To address the second primary concern with respect to the sensitivity of our results to our matching procedure we also explore different ways to obtain matched treatment and control groups. As discussed by Iacus et al. (2012) one primary concern with propensity score based methods is the tendency that high dimensionality in the chosen covariates in can result in increasing imbalance in covariate values, even if the one-dimensional propensity score differences are minimal. We address this by using coarsened exact matching (Blackwell et al., 2009) to pre-process the data. Coarsened exact matching pre-processes the treatment and control group by creating strata from both samples which are highly similar based on a selection of covariates that should be balanced under all circumstances before conducting the propensity score estimation to weigh observations within these strata.

The covariates we choose for the coarsened exact matching are the age and living status of the parents, gender, the presence of a long-standing illness or disability, the respondents age and lastly mental and physical component scores at t_{-1}. This reduces the number of included control and treatment individuals considerably as they are not within the strata in which these covariates would be balanced.[17] After reducing the sample, the regular propensity score matching is performed in the same way as in our baseline specification. Next to this coarsened exact matching with subsequent propensity score matching we additional perform a one-to-one matching using propensity scores only using the reduced sample after the coarsened exact matching and the baseline specification. Fig. 6 depicts for both mental and physical health the ATT estimates for all specifications and by treatment intensity. The matching quality for each is similar across specifications while details can be found in Appendix A3 Figs. A3.2 to 3.4. The results depicted in Fig. 6 suggests that across specifications the mental health effects are largely similar especially for the initial periods after care-provision has started. However, with regards to the longer-term effects the matching procedure seems to have a stronger influence on the resulting ATT estimates.

Besides these two main concerns we conduct a range of additional analysis to address other related concerns about our identification strategy. For the sake of brevity, we only shortly discuss these with detailed results available in the Appendix. The observed mental health effects could be explained by an ongoing trend that started before informal care provision. Due to our time-structure we observe for many caregivers their mental and physical health for multiple periods prior to providing care. Appendix Fig. A2.3 plots the mean MCS and PCS for all intensity levels for up to four years prior to providing informal care. There is little evidence that the observed results are driven by a distinct negative trend in caregivers' mental health before the actual onset of care provision.

Further, we explore the existence of a potential downward bias underlying our results due to selective attrition. We follow De Zwart et al. (2017) by splitting our sample into two groups and re-estimating the initial treatment effects. For the first group we observe health states past t_1, while for the second group we can only observe health immediately after providing informal care due to permanent survey attrition. Fig. A2.4 plots the treatment effects for both groups and indicates that the attrition sample experiences more persistent negative effects directly before discontinuing their participation, but no physical health effects. These results indicate some evidence for a downward bias in our estimated treatment effects for mental health in later periods.

While the SF-12 component scores allow us to measure mental and physical health, the interpretation of effect sizes is not straightforward. To do so we use an alternative outcome measure, the general health questionnaire (GHQ), which is a mental health screening instrument with defined thresholds identifying individuals at risk of developing a mental illness. Appendix Fig. A2.5 present the corresponding results. When considering this alternative outcome measures our results remain generally the same, indicating an asymmetric effect on mental health especially among high-intensity caregivers. However, the results for GHQ scores depict a pronounced dose–response relationship, not observed when using MCS as the mental health measure. Further, these results also indicate that the decrease in mental health is economically relevant as the share of individuals with scores surpassing screening thresholds increases substantially by 4.3 to 7.7 percentage points (about 17 to 34%) depending on the comparison being made within caregivers (before/after) or using the matched control individuals (see Appendix A2 for details).

A concern for our dynamic matching approach stems from the fact that for the later waves we condition on a large set of covariates as all intermediate covariates at each node are included. To check whether our propensity score estimates are suffering from overfitting we follow Lechner (2008) and condition on a smaller set of covariates capturing the most recent information and limited information (socio-economic status and health outcomes) from the previous decision nodes. The results from this alternative specification which are presented in Table A2.2 are similar to our main analysis. Additionally, we check whether our results are sensitive to more stringent regression adjustment by conditioning on the covariates from all previous waves. This does not substantially alter our estimates for the mental health effects (Table A2.3). In the alternative models, however, we observe that the high-intensity PCS estimates are most sensitive to model specifications leading to slightly different effect sizes but qualitatively similar results.

Discussion & conclusion

Providing informal care can have negative health effects for informal caregivers. Based on the current literature there is an insufficient understanding of how these effects persist over time, and differ by care-intensity and duration. We try to explore these questions by estimating the long-term and dynamic effects of caregiving on caregivers' health using a general population panel survey from the UK.

While early studies on cross-sectional data commonly report caregivers to have low physical health (Carretero et al., 2009), we only find mixed evidence for a causal relationship. Our estimates indicate that informal care leads to a small and short-lived increase in physical health among caregivers providing <20h of weekly care. A potential alternative explanation for this finding could however be that self-reported physical health is prone to bias as caregivers might change their opinion about their own health by taking the health of the care-recipient as a reference point (Di Novi et al., 2015). For caregivers' mental health outcomes, we find immediate and persisting negative effects of providing care. These effects are heterogeneous and mostly incurred by individuals providing more than 20h of care per week. The initial negative effects on mental health slowly decrease in size throughout the years but remain persistent up to four and five years after initial care provision depending on the specification. These effects are, potentially due to limited attrition, more persistent than estimates from previous

[17] The remaining number of individuals in the groups by treatment intensity are for control/treated: 7549/4162 for any care, 6797/3149 for low intensity, 2205/446 for medium intensity, and 2364/534 for high intensity.

studies that only found direct effects (De Zwart et al., 2017) or effects up to the first three years of care provision (Schmitz & Westphal, 2015). Additionally, our estimates may be downward biased as high-intensity caregivers are underrepresented in our sample.

For these high intensity caregivers (individuals providing $\geq$20h a week) the estimated negative health effects are similar in magnitude compared to earlier results by Schmitz & Westphal (2015) who focus on individuals providing at least three hours of care on a weekday ($\geq$15 h). For intensive caregivers the results hence seem robust across different countries with different care systems. This finding also supported by a recent study from Bom and Stöckel (2021) who explore the health effects of informal caregiving for a sample of older UK and Dutch caregivers. For low intensity caregivers this is however not the case: Schmitz & Westphal already find a strong negative effect of -1.9 on the MCS for individuals providing one hour of care per weekday, whereas we do not observe health effects for individuals providing $<$20 hours of care per week in the UK. There might be several explanations for this difference. The intensity levels of care provision are, first, not completely similar and the composition of caregivers within these groups might differ. Second, country differences in the long-term care system and support options might drive changes in the size of the caregiving effect as they influence both the selection into care as well as the caregiving experiences within these groups. Courtin et al. (2014) for example describe financial support policies for high-intensity informal caregivers to be available to a wider group in Germany than the UK, mainly due to the strict eligibility criteria in the latter. At the same time there exists a broader range of non-financial support open to all caregivers in the UK compared to Germany. However, given the consistent absence of reliable information on social care consumption in many panel surveys we cannot explore this directly.

Lastly, our dynamic matching results provide insights into the extent to which the static results, indicating decreasing mental health effects over time, are representative for the population of individuals that provide care for multiple consecutive years. There is evidence that the static results do not sufficiently capture that among individuals who provide care for more than one year the mental health effects do not improve over time. Rather for these multi-year caregivers' mental health remains to be negatively affected.

Our study also has several limitations. First, one might question the use of self-reported health measures and prefer, in our case unavailable, administrative information like medical claims or admission data. We believe that given the population we are studying, informal caregivers, these self-reported health measurements better capture changes in health than information regarding health care usage. For mental health this is especially the case as often not all individuals suffering from mental health problems receive or seek treatment. Additionally, administrative information can only capture actual consumption but highly burdened caregivers might forego medical care. Foregoing care could be directly caused by the intensity of caregiving as well as the potential stigma associated with seeking help as a caregiver itself. In addition, our results remain unchanged when using alternative outcome measures (see Appendix Fig. A2.5) and indicate that the reported effects are economically relevant from the individuals' perspective. This leaves us confident that the reported mental health effects are of interest to policymakers wishing to assess the extent of spillover effects arising from the reliance on informal care to meet social care demands.

While our dataset allowed us to explore the health effects of informal care provision along multiple dimensions not all desired information is available which is a limitation of this study. Firstly, our measure of care intensity is self-reported caregiving hours. Increased hours are likely to reflect a larger overall caregiving burden, however, the tasks performed by caregivers are highly disease-specific and play an important role in the experienced caregiving burden (Pearlin et al., 1990). In addition, Urwin et al. (2021), who use data from USoc from 2015 to 2017, provide evidence that hours of care provided are reported differently by caregivers and recipients raising questions with respect to what is counted in these self-reported hours. Therefore, reported hours are an incomplete measure. A related cause for uncertainty is the absence of information on why informal care was taken up and discontinued, a process that itself could affect especially mental health outcomes. Another concern refers to our focus on informal caregiving irrespective of whether this occurs alongside formal care as the USoc does not capture such services consistently. Therefore, we cannot explore to what extent these services might serve as a complement or substitute to informal care or help to mitigate the negative health effects in the medium and long run.[18] Ideally future research would have insight into the type of caregiving tasks, formal care use and information on reasons for care take up and discontinuation.

Moreover, despite using a rather large dataset, sample size issues still remained a problem for identification of various additional models we would have liked to run. First of all, the dynamic models showed to be highly data-demanding and we observed too little variation among high-intensity caregivers at later caregiving nodes to be able to identify health effects of providing care for a third year. Second, our dataset lacked information on the health outcomes of many of the care recipients, thereby making our sample too small to separately estimate the family effect, the impact of a family member becoming ill on one's own health. Capturing health outcomes of family members within the survey or the possibility to link data to administrative records to observe the health of family members (as done in Bom et al. (2019b)) would aid in disentangling both effects.

Lastly, an important limitation of our study is its reliance on a matching-based identification strategy and the underlying assumptions. While we do test the robustness of our results to violations of the main identifying assumption and the choice of matching approaches an ideal strategy would rely on exogenous variation in informal caregiving that addresses concerns about time-varying unobserved confounders. Earlier studies exploring the short-term health effects of informal caregiving often relied on parental health shocks or the number of siblings as an instrumental variable. However, the exogeneity of parental health shocks is at least questionable (Schmitz & Westphal, 2015). Recently, Eibich (2021) evaluated the validity of many commonly used instruments confirming these suspicions. Another general concern with respect to alternative IV-based methods is whether the estimated local average treatment effects can be generalized for the entire treatment population (Angrist & Imbens, 1995). Recent studies such as Bakx et al. (2020) or Fischer & Müller (2020) exploited institutional rules and reforms in countries' LTC-sectors as a source of credible exogenous variation in the uptake of formal and informal care use. In our case, however, such an identification strategy was not feasible.

To conclude, our results confirm previous studies reporting negative mental health effects of informal care provision and show that the effects persist up to four or five years after initial care provision. Our estimates suggest that most UK caregivers do not experience substantial adverse health outcomes after providing informal care. However, especially high-intensity caregivers show to be most strongly affected by informal caregiving. This group of caregivers is also most likely to provide care for multiple years which our results suggest to lead to continuing lower mental health. Given the increasing reliance on informal care, these results provide useful insights for policymakers facing difficult trade-offs regarding the allocation of limited resources to support caregivers. Our results indicate that especially high-intensity and long-term caregivers should be targeted to offset the substantial negative mental health effects. While informal care provides undisputable benefits to public

[18] USoc wave 7 did include a detailed survey module on informal and formal care sources for recipients; 47% report informal care as the only source of care with 44% reporting a mix of formal and informal care and 8% formal care only. Hours of informal care received are highly similar irrespective of whether it is provided alongside formal care or not. Detailed results are available upon request.

health care systems and care-recipients the consequences for those providing the care need to be accounted for.

CRediT authorship contribution statement

Jannis Stöckel: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Judith Bom:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization.

Declaration of Competing Interest

None of the authors has any conflict of interest to declare.

Acknowledgements

We gratefully acknowledge comments provided by Cameron Birchall, Joaquim Vidiella-Martin, Sara Rellstab, Pieter Bakx, Eddy van Doorslaer, Erik Schut, Job van Exel, and Werner Brouwer, participants at the LSE Care Policy Evaluation Centre seminar, the KU Leuven Public Economics seminar, the 2019 lowlands Health Economic Study Group, the 2019 UK Health Economists' Study Group summer meeting, the 2019 International Health Economics Congress and the ESE/ESHPM PhD Research Symposium. This study received financial support from the Network for Studies on Pensions, Ageing and Retirement Grant *optimal saving and insurance for old age: The role of public-long term care insurance* (Bom) and the Erasmus Initiative *Smarter Choices for Better Health* of the Erasmus University Rotterdam (Stöckel). All remaining errors are our own.

A1 – Dataset conditioning and detailed results

Table A1.1
Dataset Conditioning.

Description	Observations (Individuals)
Merged USoc waves 1–9 (2009–2019)	407722 (85908)
Panel A: Static Matching Data Conditioning	
Excluding individuals not providing caregiving information or proxy respondents.	374917 (79378)
Excluding respondents with single or non-consecutive initial first two observations and starting caregivers.	255742 (48138)
Conditioning on full set of control variable at t = -1 and health outcomes in t>=0	211098 (45777)
Excluding individuals without consecutive observation at t = -1 and t = 0	126067 (20247)
Analysis Dataset:	126067 (20247)
Panel B: Dynamic Matching	
Excluding individuals not providing caregiving information or proxy respondents, and individuals without information on health outcomes	345533 (77750)
Excluding individuals not participating for at least four consecutive waves.	236338 (35263)
Excluding individuals without complete information on control variables at all periods.	116400 (20884)
Excluding individuals without complete information on control variables.	114232 (10485)
Excluding individuals to ensure that individuals are not included twice (as individuals starting care provision are pooled forward)	82940 (20485)
Analysis Dataset:	82940 (20485)

Table A1.2
Static Estimation Results Stratified by Gender and Care Intensity.

	Full Sample											
	t = 0		**t = 1**		**t = 2**		**t = 3**		**t = 4**		**t = 5**	
	MCS	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**	**MCS**	**PCS**
Any Care	**−0.476****	**0.441****	**−0.687*****	0.113	−0.374	−0.173	−0.341	−0.108	**−0.378**	0.288	**−0.420**	−0.006
	(0.173)	(0.160)	(0.184)	(0.179)	(0.204)	(0.201)	(0.217)	(0.216)	(0.225)	(0.233)	(0.238)	(0.243)
Low Intensity	−0.177	**0.484*****	−0.458	0.250	−0.223	−0.004	−0.147	0.038	−0.279	**0.526***	−0.297	0.350
<10 h weekly care	(0.182)	(0.170)	(0.196)	(0.194)	(0.215)	(0.216)	(0.228)	(0.226)	(0.240)	(0.240)	(0.250)	(0.250)
Medium Intensity	−0.275	0.128	−0.676	−0.272	−0.276	−0.228	−0.185	−0.303	0.036	−0.121	−0.385	−0.905
10–20 h weekly care	(0.415)	(0.374)	(0.467)	(0.428)	(0.544)	(0.484)	(0.518)	(0.573)	(0.566)	(0.587)	(0.609)	(0.637)
High Intensity	**−2.438*****	0.371	**−1.884*****	−0.360	**−1.290***	−0.791	**−1.625***	−0.792	**−1.748****	−0.577	**−1.448***	−1.280
>20 h weekly care	(0.456)	(0.428)	(0.462)	(0.438)	(0.560)	(0.539)	(0.631)	(0.595)	(0.608)	(0.650)	(0.696)	(0.673)
Control		**13141**		**12497**		**10575**		**9753**		**8701**		**8185**
Treatment		**7102**		**5863**		**4901**		**4200**		**3239**		**2503**
Low		5,249		4,357		3,662		3,164		2,433		1,881

(*continued on next page*)

Table A1.2 (*continued*)

Medium		812		701		570		492		380		296
High		934		722		606		488		384		293
Females Only												
	t = 0		t = 1		t = 2		t = 3		t = 4		t = 5	
	MCS	PCS	MCS	PCS	MCS	PCS	MCS	PCS	MCS	PCS	MCS	PCS
Any Care	**−0.579****	0.400	**−0.915*****	0.030	−0.321	−0.264	−0.516	**−0.595****	−0.583	−0.093	−0.608	−0.514
	(0.235)	(0.211)	(0.252)	(0.233)	(0.274)	(0.274)	(0.301)	(0.292)	(0.310)	(0.310)	(0.328)	(0.313)
Low Intensity	−0.258	**0.550****	**−0.592****	0.220	−0.135	−0.056	−0.200	−0.319	−0.426	0.166	−0.288	−0.021
<10 h weekly care	(0.248)	(0.224)	(0.272)	(0.257)	(0.294)	(0.295)	(0.316)	(0.302)	(0.329)	(0.327)	(0.344)	(0.327)
Medium Intensity	−0.314	−0.122	−1.038	−0.531	−1.277	0.284	−0.640	−1.071	−0.725	−0.353	−1.319	−1.173
10–20 h weekly care	(0.557)	(0.510)	(0.595)	(0.559)	(0.688)	(0.609)	(0.691)	(0.735)	(0.777)	(0.771)	(0.788)	(0.760)
High Intensity	**−2.471******	0.012	**−2.442******	−0.493	−1.049	−1.362	**−2.252*****	**−1.853****	−1.420	−0.970	**−2.055****	**−2.446****
>20 h weekly care	(0.600)	(0.539)	(0.603)	(0.538)	(0.676)	(0.710)	(0.832)	(0.774)	(0.794)	(0.766)	(0.958)	(0.869)
Control		**6934**		**,624**		**5632**		**5207**		**4616**		**4369**
Treatment		**4211**		**3488**		**2916**		**2507**		**1954**		**1516**
Low		3,061		2,551		2,135		1,853		1,435		1118
Medium		486		426		345		296		236		180
High		600		463		397		321		257		196
Males Only												
	t = 0		t = 1		t = 2		t = 3		t = 4		t = 5	
	MCS	PCS	MCS	PCS	MCS	PCS	MCS	PCS	MCS	PCS	MCS	PCS
Any Care	−0.319	**0.507****	−0.283	0.165	−0.375	−0.096	0.058	0.590	−0.058	**0.870****	−0.057	**0.764****
	(0.250)	(0.242)	(0.260)	(0.275)	(0.302)	(0.288)	(0.298)	(0.313)	(0.323)	(0.355)	(0.340)	(0.378)
<10 h weekly care	(0.265)	(0.258)	(0.276)	(0.714)	(0.309)	(0.308)	(0.019)	(0.334)	(0.345)	(0.363)	(0.369)	(0.392)
Low Intensity	−0.069	0.395	−0.227	−0.009	−0.329	−0.034	0.019	0.508	−0.099	**1.048*****	−0.222	**0.871****
Medium Intensity	0.059	0.424	0.233	0.039	0.946	−0.813	0.696	0.618	**1.637****	0.243	1.122	−0.138
10–20 h weekly care	(0.593)	(0.528)	(0.295)	(0.631)	(0.788)	(0.676)	(0.747)	(0.864)	(0.706)	(0.835)	(0.845)	(0.978)
High Intensity	**−2.396******	0.829	−0.871	−0.216	−1.659	0.096	−0.084	0.690	**−1.855****	0.236	0.257	0.779
>20 h weekly care	(0.691)	(0.665)	(0.683)	(0.728)	(0.922)	(0.758)	(0.859)	(0.803)	(0.920)	(1.092)	(0.779)	(0.958)
Control		**6207**		**5873**		**4943**		**4546**		**4085**		**3816**
Treatment		**2890**		**2374**		**1983**		**1691**		**1284**		**987**
Low		2190		1808		1,529		1,313		999		763
Medium		325		274		224		195		143		115
High		333		258		208		166		127		96

Source: USoc Waves 1–9, own calculations. *Note*: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, standard errors in parentheses. MCS columns depict the estimated coefficients for the SF-12 Mental Component Scores, PCS the SF-12 Physical Component Scores.

A2 – Additional analyses and robustness checks

Alternative time structure and composition of treatment and control groups

To explore whether our choice to pool caregivers from later USoc waves together to increase the number of observed caregivers has an impact on our results we here re-estimate the static treatment effects with an augmented time-structure. Instead of defining t_{-1} to be the wave preceding informal caregiving for all caregivers this is only done for those individuals starting to provide care within their first three participation waves. This effectively splits the caregiver sample in half, one of which remains to be the treatment group and a second group of future caregivers that joins the control group and for which t_{-1} is defined as the first participation wave as for the never-caregivers in the control group. As can be seen in Fig. A2.1 the resulting ATT estimates are highly similar to those presented in Figs. 3 and 4.

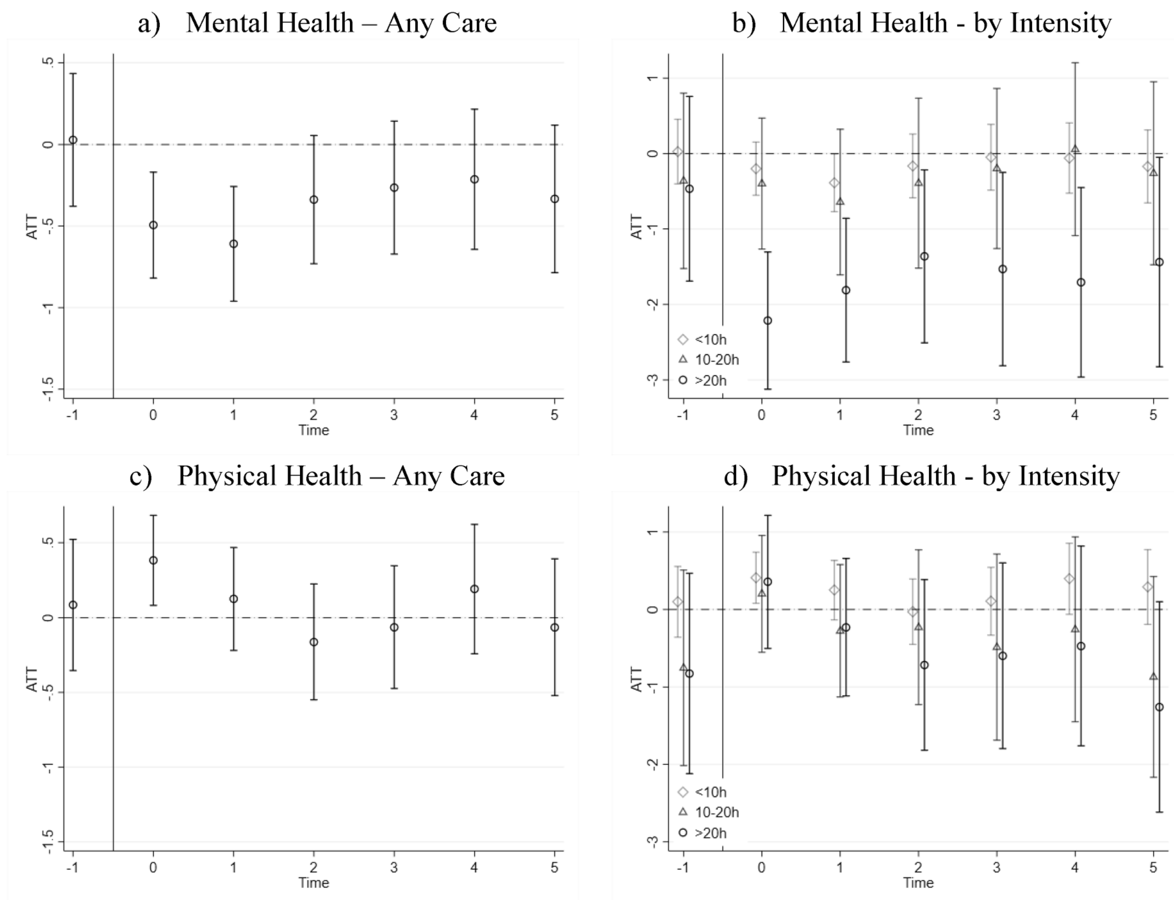


Fig. A2.1. Baseline Results with future Caregivers in Control Group. *Source:* USoc Waves 1–9, own calculations. *Note:* ATT estimates pre-treatment (t_{-1}) is the mean difference between treatment and control groups after the matching procedure is applied.

Simulated violation of the conditional independence assumption

In this sensitivity analysis we focus on high-intensity caregivers and mental health outcomes to test the robustness of our main result. Our estimation strategy relies on the conditional independence assumption (CIA) stating that selection into treatment is driven only by observable variables. We use a rich set of covariates covering multiple dimensions relevant to the selection into informal caregiving and mental health in the propensity score estimation, hence we argue that the CIA is likely to hold. However, the CIA is untestable and unobserved variables might be influencing the selection into informal caregiving and mental health outcomes, thereby biasing our estimates. To assess the robustness of our estimates to such a violation we follow Ichino et al. (2008) who propose a simulation-based sensitivity analysis for matching estimators. We only roughly sketch the underlying procedure and intuition behind the procedure. A more elaborate discussion can be found in Ichino et al. (2008).

Consider the conditional independence assumption in our context:

$$Y_t^0 \perp\!\!\!\perp T \mid X_{-1}, Y_{-1} \forall\, t$$

After conditioning on a set of pre-treatment control (X_{-1}) and outcome variables (Y_{-1}) the potential outcomes in a given period t in absence of informal care provision (Y_t^0) across treatment and control groups are the same. Now consider that this assumption is violated due to a confounder U. If we additionally condition on this confounder the CIA would be satisfied:

$$Y_t^0 \perp\!\!\!\perp T \mid X_{-1}, Y_{-1}, U \;\forall\, t$$

Ichino et al. (2008) outline a sensitivity analysis that simulates a binary U in the context of a binary outcome variable. A binary U is attractive as its distribution can be expressed by four parameters p_{ij} where i indicates the treatment status (0,1) and j indicates the outcome status (0,1). For continuous outcomes they propose a transformation of the continuous outcome:

$$\widehat{Y} = \begin{cases} 1, Y > \bar{Y} \\ 0, else \end{cases}$$

In our case we set $\bar{Y}$ equal to the sample mean of MCS and PCS across high-intensity caregivers and the overall control group. The four p_{ij} that determine the distribution of U are defined as:

$$p_{01} = \Pr(U = 1|T = 0, \widehat{Y} = 1)$$

$$p_{00} = \Pr(U = 1|T = 0, \widehat{Y} = 0)$$

$$p_{11} = \Pr(U = 1|T = 1, \widehat{Y} = 1)$$

$$p_{10} = \Pr(U = 1|T = 1, \widehat{Y} = 0)$$

These p_{ij}describe the distribution of U:

$$\begin{aligned} Pr(U = 1) = \quad & p_{11}\Pr\left(\widehat{Y} = 1|T = 1\right)\Pr(T = 1) + p_{10}\Pr\left(\widehat{Y} = 0|T = 1\right)\Pr(T = 1) \\ & + p_{01}\Pr\left(\widehat{Y} = 1|T = 0\right)\Pr(T = 0) + p_{00}\Pr\left(\widehat{Y} = 0|T = 0\right)\Pr(T = 0) \end{aligned}$$

Ichino et al. (2008) propose to choose the values of p_{ij} in such a way to deliberately control for the selection effect s of U on treatment uptake and the outcome effect d of U on the probability to observe $\widehat{Y} = 1$. These effects are defined as:

$$s = p_{1\cdot} - p_{0\cdot}$$

where

$$p_{i\cdot} = \Pr(U = 1|T = i) = p_{i0}{*}\Pr\left(\widehat{Y} = 1|T = i\right) + p_{i1}{*}\Pr\left(\widehat{Y} = 1|T = i\right) with i \in \{0, 1\}$$

The outcome effect is defined as

$$d = p_{01} - p_{00}$$

The sensitivity analysis is then conducted by choosing a set of p_{ij} that model a confounder with specific selection and outcome effects. Given these values for s and d one has to assume a value for $\Pr(U = 1)$ and the relationship between p_{11} and p_{01} in order to be able to solve the equations for $\Pr(U)$, d, and s for all p_{ij}. We follow the example given by Ichino et al. (2008) by assuming $\Pr(U = 1) = 0.5$ and $p_{11} - p_{01} = 0$ to solve for all of p_{ij} given the empirically observed $\Pr\left(\widehat{Y} = i|T = i\right)$ and $\Pr(T = i)$.

As an example, consider an unobserved confounder with negative selection ($s < 0$) and positive outcome effect ($d > 0$). This might be the unobserved availability of resources to purchase formal care services by the recipient itself. As informal care is largely done by family-members the availability of resources is likely to decrease the likelihood of (high-intensity) informal care provision while increasing mental health in the absence of treatment. By not accounting for such a confounder our estimation would therefore underestimate the effect of informal care on mental health. By selecting the magnitude and direction of selection and outcome effects and the corresponding p_{ij} the effect of U is simulated by drawing repeatedly from a Bernoulli distribution with the desired distributional properties. Robust estimates of both the ATT and the corresponding standard errors are then given by their averages across these simulations.

To calibrate the sensitivity analysis, we need a starting point of selection and outcome effects to obtain the parameters p_{ij} for the distribution of a realistic confounder. Ichino et al. (2008) recommend inspecting the selection and outcome effects of important covariates in the propensity score estimation in order to find reasonable values for the simulation. Table A2.1 depicts the estimated effects (d and s) and the parameters (p_{ij}) for all covariates used in the estimation of propensity scores and using the mental component scores as the outcome of interest. These are obtained by using a customized version of the user-written command for Stata by Nannicini (2007)[19] implementing the sensitivity analysis proposed by Ichino et al. (2008). Please note that continuous variables were adapted to comply with the binary nature of U by transforming them into categorical variables.

[19] Nannicini, T. (2007). Simulation-based sensitivity analysis for matching estimators. *The Stata Journal*, *7*(3), 334–350.

Table A2.1
Estimated selection and outcome effects.

Variable	p11	p10	p01	p00	p1.	p0.	s	d
Care Obligations								
Mother alive	0.48	0.56	0.66	0.75	0.53	0.70	−0.17	−0.09
Mother aged < 49	0.06	0.06	0.12	0.15	0.06	0.13	−0.07	−0.03
Mother aged 50–59	0.13	0.15	0.21	0.24	0.14	0.22	−0.08	−0.03
Mother aged 60–69	0.12	0.17	0.20	0.22	0.15	0.21	−0.06	−0.02
Mother aged 70–79	0.08	0.12	0.10	0.10	0.11	0.10	0.01	0.00
Mother aged 80–89	0.08	0.05	0.03	0.03	0.06	0.03	0.03	0.00
Mother aged > 90	0.91	0.94	0.87	0.84	0.93	0.86	0.07	0.03
Father alive	0.36	0.48	0.60	0.68	0.43	0.63	−0.20	−0.08
Father aged < 49	0.03	0.03	0.08	0.10	0.03	0.09	−0.06	−0.02
Father aged 50–59	0.10	0.13	0.17	0.21	0.12	0.19	−0.07	−0.04
Father aged 60–69	0.11	0.14	0.20	0.22	0.13	0.21	−0.08	−0.02
Father aged 70–79	0.05	0.12	0.11	0.11	0.09	0.11	−0.02	0.00
Father aged 80–89	0.05	0.06	0.03	0.02	0.05	0.03	0.02	0.01
Father aged > 90	0.95	0.96	0.90	0.89	0.96	0.90	0.06	0.01
Both parents alive	0.32	0.41	0.56	0.64	0.38	0.60	−0.22	−0.08
Living siblings	0.84	0.86	0.87	0.90	0.85	0.88	−0.03	−0.03
Living partner	0.73	0.75	0.67	0.59	0.74	0.64	0.10	0.08
Partner aged < 29	0.04	0.05	0.10	0.11	0.05	0.10	−0.05	−0.01
Partner aged 30–39	0.11	0.15	0.18	0.19	0.13	0.19	−0.06	−0.01
Partner aged 40–49	0.10	0.16	0.15	0.15	0.14	0.15	−0.01	0.00
Partner aged 50–59	0.13	0.11	0.09	0.07	0.12	0.08	0.04	0.02
Partner aged 60–69	0.13	0.11	0.10	0.05	0.12	0.08	0.04	0.05
Partner aged 70–79	0.15	0.12	0.04	0.02	0.13	0.03	0.10	0.02
Partner aged 80–89	0.08	0.05	0.00	0.00	0.06	0.00	0.06	0.00
Partner aged > 90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Willingness to Care								
Own age < 29	0.12	0.12	0.26	0.33	0.12	0.29	−0.17	−0.07
Own age 30–39	0.15	0.23	0.23	0.24	0.19	0.23	−0.04	−0.01
Own age 40–49	0.14	0.18	0.18	0.20	0.16	0.18	−0.02	−0.02
Own age 50–59	0.16	0.16	0.12	0.11	0.16	0.12	0.04	0.01
Own age 60–69	0.18	0.15	0.13	0.07	0.16	0.11	0.05	0.06
Own age 70–79	0.19	0.11	0.07	0.04	0.15	0.06	0.09	0.03
Own age 80–89	0.06	0.04	0.02	0.01	0.05	0.02	0.03	0.01
Own age > 90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Female	0.59	0.68	0.50	0.58	0.64	0.53	0.11	−0.08
Education: Primary/Other lower	0.32	0.33	0.18	0.17	0.33	0.18	0.15	0.01
Education: Secondary	0.41	0.40	0.43	0.45	0.41	0.45	−0.04	−0.02
Education: Tertiary	0.27	0.26	0.39	0.38	0.27	0.39	−0.12	0.01
Self employed	0.06	0.04	0.07	0.06	0.05	0.07	−0.02	0.01
Unemployed	0.07	0.07	0.04	0.06	0.07	0.05	0.02	−0.02
Employed	0.35	0.35	0.58	0.55	0.35	0.57	−0.22	0.03
Working full-time	0.28	0.24	0.51	0.49	0.25	0.50	−0.25	0.02
Retired	0.36	0.27	0.17	0.10	0.31	0.14	0.17	0.07
In education/other	0.04	0.02	0.09	0.12	0.03	0.10	−0.07	−0.03
Homecarer	0.09	0.15	0.04	0.06	0.13	0.05	0.08	−0.02
Disabled	0.03	0.10	0.01	0.05	0.07	0.02	0.05	−0.04
Income Quintile 1 (Lowest)	0.24	0.27	0.18	0.23	0.25	0.20	0.05	−0.05
Income Quintile 2	0.26	0.25	0.19	0.20	0.25	0.20	0.05	−0.01
Income Quintile 3	0.21	0.23	0.21	0.19	0.22	0.20	0.02	0.02
Income Quintile 4	0.16	0.16	0.21	0.20	0.16	0.20	−0.04	0.01
Income Quintile 5 (Highest)	0.14	0.10	0.22	0.19	0.12	0.21	−0.09	0.03
HH Income Fraction > 0.5	0.51	0.50	0.50	0.48	0.50	0.49	0.01	0.02
Married	0.62	0.62	0.51	0.42	0.62	0.48	0.14	0.09
Single	0.14	0.12	0.22	0.29	0.13	0.25	−0.12	−0.07
Separated/Divorced	0.09	0.10	0.07	0.08	0.10	0.07	0.03	−0.01
Widowed	0.04	0.03	0.04	0.04	0.03	0.04	−0.01	0.00
Partnership	0.11	0.13	0.16	0.17	0.12	0.16	−0.04	−0.01
Children in HH	0.30	0.40	0.38	0.42	0.36	0.39	−0.03	−0.04
Young children in HH	0.28	0.38	0.34	0.38	0.33	0.36	−0.03	−0.04
Region: North-East	0.04	0.05	0.04	0.04	0.05	0.04	0.01	0.00
Region: North-West	0.12	0.13	0.11	0.12	0.12	0.11	0.01	−0.01
Region: Yorkshire	0.06	0.08	0.08	0.08	0.08	0.08	0.00	0.00
Region: East Midlands	0.08	0.09	0.07	0.08	0.08	0.08	0.00	−0.01
Region: West Midlands	0.05	0.09	0.07	0.08	0.07	0.07	0.00	−0.01
Region: East England	0.11	0.07	0.10	0.09	0.08	0.09	−0.01	0.01
Region: London	0.08	0.08	0.09	0.11	0.08	0.10	−0.02	−0.02
Region: South East	0.09	0.10	0.14	0.13	0.10	0.14	−0.04	0.01
Region: South West	0.09	0.07	0.09	0.08	0.08	0.09	−0.01	0.01
Region: Wales	0.11	0.10	0.06	0.07	0.11	0.07	0.04	−0.01

(*continued on next page*)

Table A2.1 (*continued*)

Variable	p11	p10	p01	p00	p1.	p0.	s	d
Care Obligations								
Region: Scotland	0.10	0.07	0.10	0.08	0.08	0.09	−0.01	0.02
Region: Northern Ireland	0.06	0.07	0.05	0.05	0.07	0.05	0.02	0.00
Living in urban area	0.72	0.78	0.74	0.78	0.75	0.76	−0.01	−0.04
Big 5: Openness	0.55	0.48	0.55	0.53	0.51	0.55	−0.04	0.02
Big 5: Conscientiousness	0.62	0.49	0.57	0.45	0.55	0.52	0.03	0.12
Big 5: Extraversion	0.56	0.48	0.56	0.49	0.52	0.54	−0.02	0.07
Big 5: Agreeableness	0.67	0.62	0.60	0.55	0.64	0.58	0.06	0.05
Big 5: Neuroticism	0.36	0.64	0.38	0.68	0.52	0.50	0.02	−0.30
Ability to Care								
Self-Assessed Health	0.46	0.67	0.37	0.52	0.59	0.43	0.16	−0.15
MCS	0.72	0.33	0.78	0.41	0.49	0.63	−0.14	0.37
PCS	0.52	0.43	0.72	0.66	0.46	0.69	−0.23	0.06
LSI	0.35	0.48	0.25	0.34	0.43	0.29	0.14	−0.09
Number of Functional Limitations	0.23	0.36	0.14	0.23	0.30	0.18	0.12	−0.09
Satisfaction with Health	0.52	0.30	0.65	0.44	0.39	0.57	−0.18	0.21
Satisfaction with Income	0.59	0.38	0.67	0.49	0.47	0.60	−0.13	0.18
Satisfaction with Leisure	0.70	0.49	0.69	0.50	0.58	0.62	−0.04	0.19
Life Satisfaction	0.69	0.38	0.73	0.46	0.51	0.63	−0.12	0.27
GHQ Score	0.80	0.50	0.83	0.52	0.62	0.72	−0.10	0.31

The largest outcome and selection effects in absolute terms are estimated for the set of pre-treatment health and well-being outcomes which is also precisely why we condition on pre-treatment outcomes. The second largest absolute selection and outcome effects are estimated for working full-time ($s = -0.25$) and scoring above four on the neuroticism seven-point scale ($d = -0.30$). For the sensitivity analysis we select two pairs of s and d to obtain upper and lower bounds for our ATT estimates. These are $s \in \{-0.25, 0.25\}$ and $d = 0.30$.[20]

Physical health effects and by Age-related trends

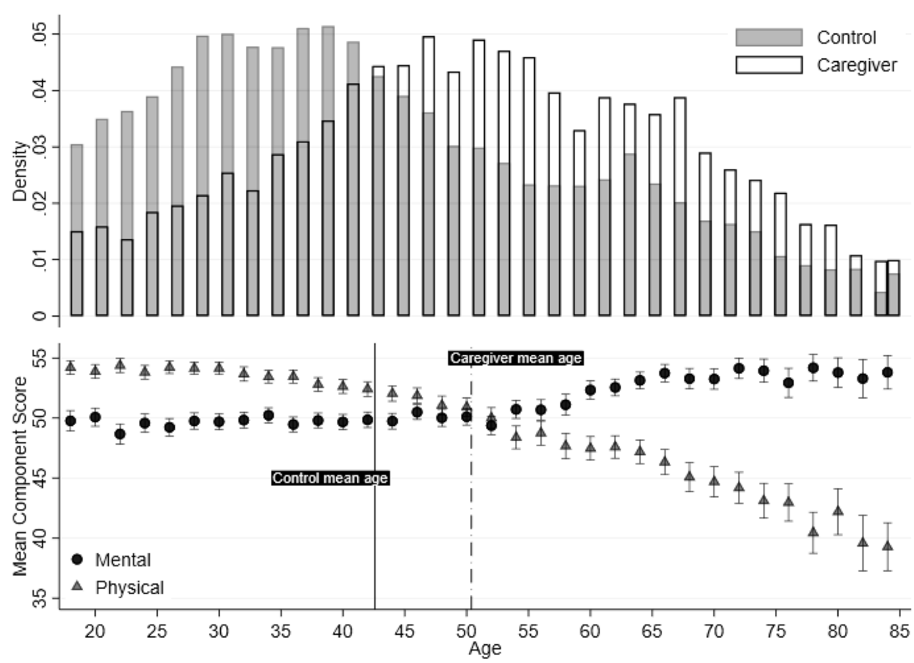


Fig. A2.2. Age Distribution by Caregiving Status & Health Trends. Note: This figure presents cross-sectional differences in mental and physical health by age-bins of two years using data from the analysis sample provided at t_{-1} before any care is provided.

[20] The selection effects for physical health are analogous. The outcome effects differ in an expected manner with the largest effects occurring among own health outcomes and age-related covariates, representing the fact that physical health is more age-dependent than mental health. For both the maximum outcome and selection effects depict a similar range, hence we use the same simulated effects for both outcomes.

Pre-Treatment trends

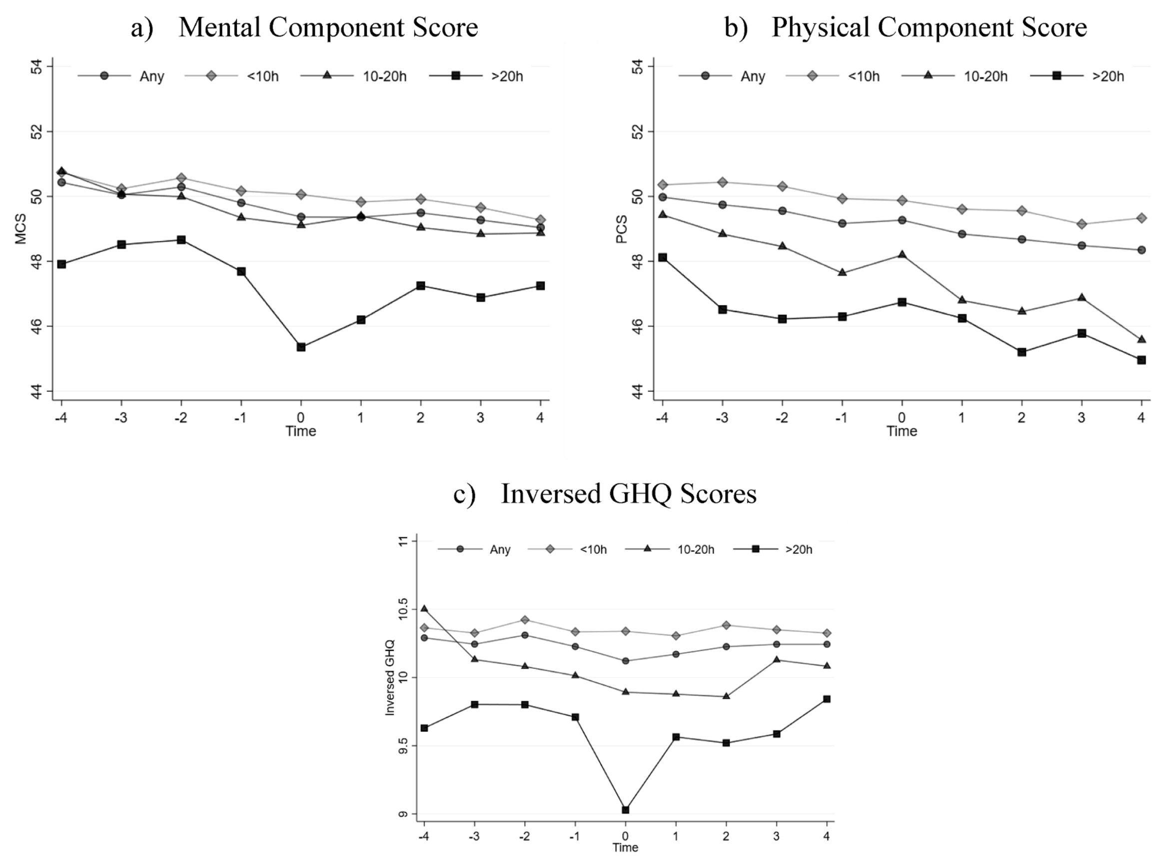


Fig. A2.3. Mental & Physical Health Pre-Treatment Trends. *Source:* USoc Waves 1–9, own calculations. *Note:* For this graph an unbalanced panel of caregivers is used that provide information on health outcomes at least at t_{-2} corresponding to 75% of all caregivers. MCS columns depict the estimated coefficients for the SF-12 Mental Component Scores, PCS the SF-12 Physical Component Scores. GHQ is the General Health Questionnaire score on the 12-point scale and inversed for ease of interpretation.

Selective attrition

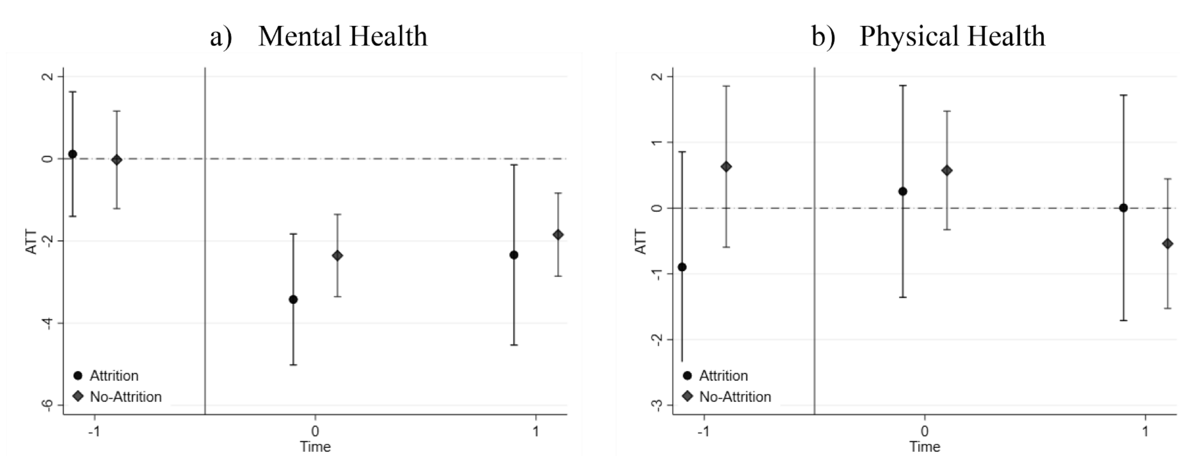


Fig. A2.4. Initial Treatment Effect (High Intensity) by Attrition. *Source:* USoc Waves 1–9, own calculations. *Note:* ATT estimates pre-treatment (t_{-1}) is the mean difference between treatment and control groups after the matching procedure is applied.

Alternative mental health measure

We explore the existence and magnitude of treatment effects on mental health using an alternative measure, the General Health Questionnaire (GHQ), to assess the robustness of our results and their economic relevance. The GHQ is a screening device to identify individuals at high risk of developing a non-psychotic minor psychiatric disorder such as anxiety or depression in general population surveys or outside of clinical environments (Goldberg et al., 1997)[21]. Respondents' answers to the 12 items are transformed to a single score on a 0 (best) to 12 (worst) scale. To ease the visual interpretation of results in line with the other outcomes measures we have inverted the scale to range from 12 (best) to 0 (worst) so that negative coefficients indicate negative mental health effects. Fig. A2.5 depicts the estimated treatment effects across the care-intensity levels.

The overall structure of results across time periods and care intensity is generally in line with the overall results based on the SF-12 MCS. Interestingly, our results using the GHQ as an outcome measure indicate a clearer dose–response relationship between informal care intensity in hours per week and mental health effects than our results using the MCS to measure mental health. To assess the practical implications of these results we convert the GHQ scores into a 'caseness' dummy. Following this definition, individuals scoring a 4 and above are identified as being a "case", meaning that these individuals experience high mental strain and are at risk of developing a mental disease. This definition does not indicate the definitive presence of a minor psychiatric disease, however, individuals scoring a 4 and above on the GHQ survey in a primary care environment should be referred to a mental health specialist for further investigation due to concern for their long-term mental health (Jackson, 2007)[22].

Across non-caregivers about 16.4% of individuals cross this threshold at the pre-treatment period while the share is with 17.0 similar among low-intensity caregivers. Among medium-intensity caregivers the share is already considerably higher before care-provision (21.9%) and even more so for the high-intensity group (25.8%). In the period of first informal care provision these shares increase with caregiving intensity; for low intensity caregivers to 17.3%, for medium intensity to 23.8% and to 30.1% for high intensity caregivers. In contrast, among comparable not treated individuals these shares were respectively: 17.4% (low), 20.8% (medium) and 22.4% (high).

Robustness checks dynamic sample

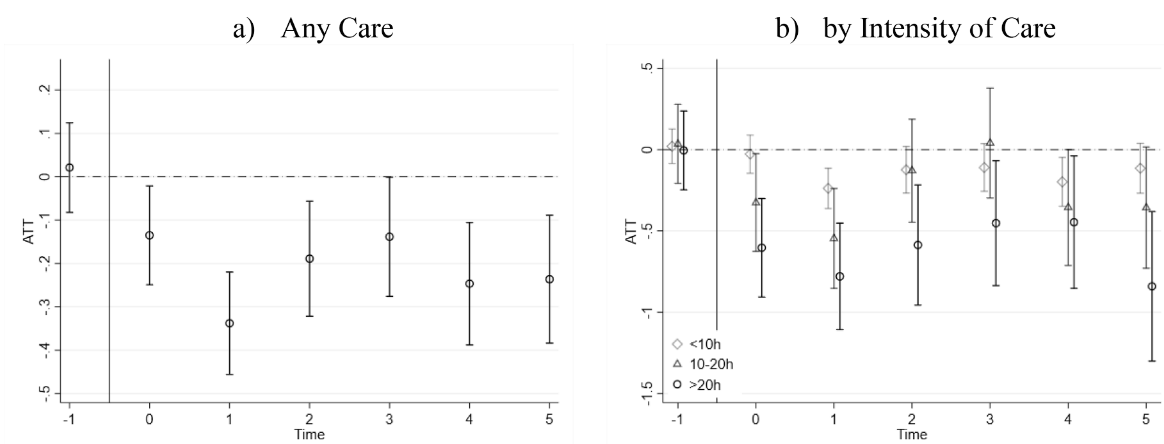


Fig. A2.5. Mental Health Effects measured using the GHQ. *Source:* USoc Waves 1–9, own calculations. *Note:* ATT estimates pre-treatment (t_{-1}) is the mean difference between treatment and control groups after the matching procedure is applied. GHQ is the General Health Questionnaire score on the 12-point scale and inversed for ease of interpretation.

Table A2.2
Robustness check: Propensity score estimations with limited covariates.

	Mental Health						
	t = 0	t = 1			t = 2		
	Static/Dynamic	Static	Dynamic (2v0)	Dynamic (2v1)	Static	Dynamic (3v0)	Dynamic (3v2)
Any	−0.312	−0.394*	−0.396	−0.045	−0.190	0.17	0.477
	(0.194)	(0.201)	(0.307)	(0.400)	(0.202)	(0.391)	(0.724)
Low Intensity	−0.007	−0.116	−0.16	0.192	0.090	0.044	0.882
	(0.209)	(0.218)	(0.3331)	(0.452)	(0.224)	(0.505)	(0.852)
Medium/High Intensity	**−1.374****	**−1.413****	**−2.257*****	**−1.994***	**−1.148***		
	(0.462)	**(0.470)**	**(0.561)**	**(0.842)**	**(0.452)**		

(*continued on next page*)

[21] Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological medicine*, *27*(1), 191–197.

[22] Jackson, C. (2007). The general health questionnaire. *Occupational medicine*, *57*(1), 79–79.

Table A2.2 (*continued*)

	Physical Health						
	t = 0	t = 1			t = 2		
	Static/Dynamic	Static	Dynamic (2v0)	Dynamic (2v1)	Static	Dynamic (3v0)	Dynamic (3v2)
Any	0.178	−0.148	0.034	0.28	0.059	−0.198	−0.313
	(0.173)	(0.191)	(0.280)	(0.373)	(0.200)	(0.346)	(0.556)
Low Intensity	0.340	0.016	155	0.121	0.205	0.218	−0.3
	(0.193)	(0.217)	(0.302)	(0.429)	(0.432)	(0.396)	(0.670)
Medium/High Intensity	−0.334	−0.707	1.053	1.321	**−1.010****		
	(0.387)	(0.390)	(0.499)	(0.799)	**(0.424)**		
Treatment (Control)	1672 (18812)	1672 (18812)	700 (17265)	700 (967)	1672 (18812)	418 (15821)	418 (278)
Low Intensity	1285 (18812)	1285 (18812)	534 (17255)	534 (745)	1285 (18812)	296 (15943)	296 (230)
Medium/High Intensity	313 (16081)	313 (16081)	150 (16800)	150 (189)	313 (16081)		

Note: The static and dynamic results at t = 0 slightly differ at the second decimal as either matching or inverse probability weighting is used, the depicted results are the static results. The table presents the dynamic ATT, the effect of providing an additional year of care compared to an individual following an alternative care trajectory. A second/third year of informal care is compared to: not providing informal care (2v0 or 3v0); providing care for only one year (2v1); or providing care for two years (3v2). It compares the health of treated and matched controls based on the information from the directly preceeding wave. * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors are in parentheses. Please note that the displayed static results are based on a different sample than the baseline static results as a differently conditioned sample is used. See the Appendix for more details.
Source: USoc, own calculations.

Table A2.3
Robustness check: Regression adjustment at all waves.

	Mental Health						
	t = 0	t = 1			t = 2		
	Static/Dynamic	Static	Dynamic (2v0)	Dynamic (2v1)	Static	Dynamic (3v0)	Dynamic (3v2)
Any	−0.312	−0.394*	−0.401	−0.285	−0.190	−0.004	0.477
	(0.194)	(0.201)	(0.306)	(0.403)	(0.202)	(0.392)	(0.724)
Low Intensity	−0.007	−0.116	−0.106	0.324	0.090	−0.118	0.686
	(0.209)	(0.218)	(0.323)	(0.438)	(0.224)	(0.495	(0.801)
Medium/High Intensity	**−1.374****	**−1.413****	**−2.314****	**−1.768***	**−1.148***		
	(0.462)	**(0.470)**	**(0.525)**	**(0.793)**	**(0.452)**		
	Physical Health						
	t = 0	t = 1			t = 2		
	Static/Dynamic	Static	Dynamic (2v0)	Dynamic (2v1)	Static	Dynamic (3v0)	Dynamic (3v2)
Any	0.178	−0.148	0.0123	0.256	0.059	−0.089	−0.313
	(0.173)	(0.191)	(0.279)	(0.366)	(0.200)	(0.349)	(0.556)
Low Intensity	0.340	0.016	0.08	0.121	0.355	0.24	−0.239
	(0.193)	(0.217)	(0.303)	(0.429)	(0.224)	(0.372)	(0.637)
Medium/High Intensity	−0.334	−0.707	**1.235****	1.464	**−1.010****		
	(0.387)	(0.390)	**(0.463)**	(0.754)	**(0.424)**		
Treatment (Control)	1672 (18812)	1672 (18812)	700 (17268)	700 (9 6 7)	1672 (18812)	406 (15756)	406 (2 7 4)
Low Intensity	1285 (18812)	1285 (18812)	536 (17260)	536 (7 3 8)	1285 (18812)	296 (15865)	296 (2 2 3)
Medium/High Intensity	313 (16081)	313 (16081)	146 (16660)	146 (1 8 6)	313 (16081)		

Note: The static and dynamic results at t = 0 slightly differ at the second decimal as either matching or inverse probability weighting is used, the depicted results are the static results. The table presents the dynamic ATT, the effect of providing an additional year of care compared to an individual following an alternative care trajectory. A second/third year of informal care is compared to: not providing informal care (2v0 or 3v0); providing care for only one year (2v1); or providing care for two years (3v2). It compares the health of treated and matched controls based on the information from the directly preceeding wave. * p < 0.05, ** p < 0.01, *** p < 0.001. Standard errors are in parentheses. Please note that the displayed static results are based on a different sample than the baseline static results as a differently conditioned sample is used. See the Appendix for more details.
Source: USoc, own calculations.

A3 – Static matching

Table A3.1
Pre- and Post-Matching Covariate Imbalance (baseline specification).

	Pre-Matching Differences				Post-Matching Differences			
	Standardized Bias in %							
	Any Care	<10 h	10–20 h	>20 h	Any Care	<10 h	10–20 h	>20 h
Mother alive	−22.00	−19.60	−21.90	−32.90	0.40	0.60	−2.80	−3.40
Mother age	−4.00	−0.40	−4.70	−22.20	0.80	1.10	−1.80	−2.70
Father alive	−36.50	−36.30	−32.20	−39.40	1.30	1.40	−1.00	−2.90
Father age	−23.70	−22.90	−18.70	−31.00	1.90	2.30	0.60	−2.30
Both parents alive	−40.90	−40.60	−38.90	−42.60	1.00	1.30	−2.10	−2.70
Siblings alive	−5.60	−4.90	−4.30	−9.30	0.10	0.60	−0.10	−3.60
Partner existing	13.90	13.00	9.10	21.20	0.50	0.30	0.20	1.70
Partner age	35.90	33.80	31.10	47.70	1.00	0.10	2.00	4.50
Own age	55.00	54.80	53.00	52.90	0.10	−0.10	4.00	5.20
Female	13.10	11.10	14.60	23.40	0.00	0.00	2.20	0.40
Children in HH	−16.60	−17.70	−19.50	−5.10	−1.10	−0.40	−1.70	0.30
Children < 14 in HH	−16.40	−18.00	−19.70	−2.70	−0.90	−0.20	−1.40	0.20
Highest Education: Primary/other	9.20	3.00	18.50	33.70	−1.60	−1.50	−1.10	−0.30
Highest Education: Secondary	−1.40	−0.20	−1.00	−6.80	−0.30	−0.70	1.80	0.80
Highest Education: Tertiary	−6.10	−2.10	−14.80	−23.50	1.70	2.00	−0.90	−0.50
Any paid work	−14.00	−6.10	−23.30	−46.50	2.80	2.80	0.90	0.20
Full time work	−21.10	−14.40	−28.70	−51.60	2.90	2.90	−0.70	−1.60
Self employed	3.90	6.80	−3.40	−6.80	0.20	0.60	0.00	0.30
Unemployed	−2.70	7.10	8.70	8.40	−1.20	−0.90	0.70	0.40
Homecarer	5.10	0.20	3.90	26.70	−2.30	−1.40	−1.40	−1.70
Disabled	8.20	3.20	17.80	22.00	−1.90	−2.20	−2.90	−2.70
Retired	29.30	26.70	28.90	39.20	−0.70	−1.20	2.30	3.70
Equivalent HH income	3.30	8.60	−11.30	−19.60	3.30	2.80	1.60	1.00
Big 5: Openness	−1.10	1.60	−8.00	−10.50	1.10	1.10	−1.10	−0.10
Big 5: Conscientiousness	11.20	12.40	8.30	7.00	2.20	1.90	2.00	1.40
Big 5: Extroversion	2.80	5.20	0.30	−5.90	−0.70	−0.50	−1.30	−1.50
Big 5: Agreeableness	11.60	9.40	19.00	14.60	1.00	0.60	2.20	1.30
Big 5: Neuroticism	−1.90	−3.90	0.50	8.70	0.40	−0.30	−0.20	1.20
Self Assessed Health	15.10	8.80	25.30	39.50	−2.80	−2.50	−1.50	−2.40
Mental Component Score	−11.80	−7.40	−18.30	−31.60	1.70	1.80	0.70	1.80
Physical Component Score	−20.10	−13.30	−32.10	−43.40	3.00	2.70	1.80	2.20
Long Standing Illness/Disability	18.00	14.40	26.60	28.00	−2.20	−2.10	0.10	−1.60
Number of Functional Limitations	8.50	2.70	18.30	26.70	−3.00	−2.60	−1.70	−3.20
Satisfaction with own Health	−21.60	−17.90	−25.60	−38.80	2.10	1.60	0.60	1.00
Satisfaction with Income	−7.80	−2.60	−18.20	−29.90	2.90	2.00	1.50	1.20
Satisfaction with Leisure Time	0.90	2.00	0.50	−7.50	0.50	−0.10	0.80	0.50
Satisfaction with Life	−11.60	−8.10	−19.90	−25.50	2.00	1.70	1.00	−0.10
Inversed GHQ Score	−5.20	−0.90	−14.10	−22.70	0.80	0.90	1.00	−0.30
Unique Individuals	**7106**	**5253**	**812**	**937**	**7102**	**5249**	**812**	**934**

Source: USoc Waves 1–9, own calculations. Note: The number of treated individuals depicted in Panel B excludes individuals identified as off-support. GHQ is the General Health Questionnaire score on the 12-point scale and inversed for ease of interpretation.

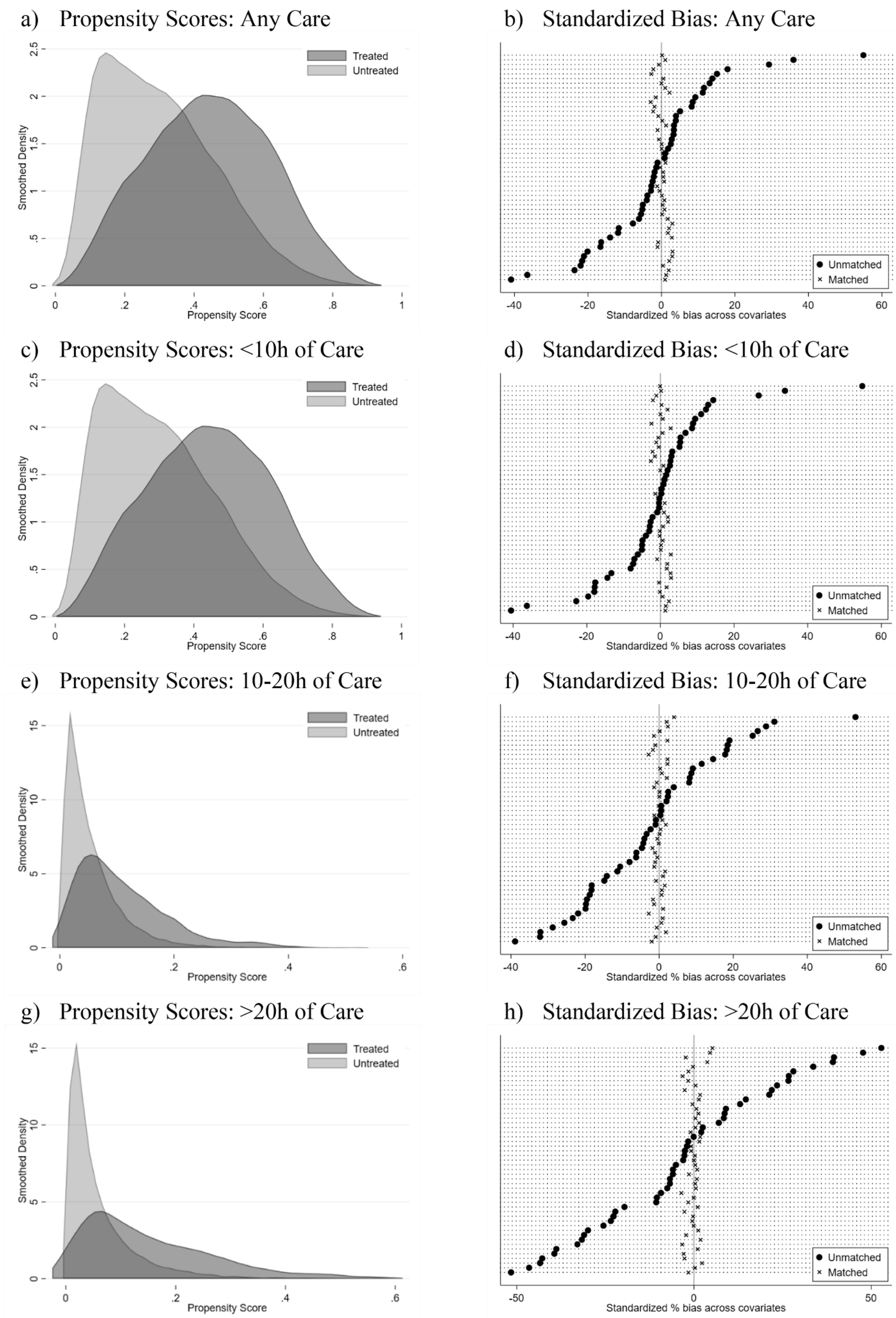


Fig. A3.1. Baseline Specification – Propensity Score Based Matching. *Source:* USoc Waves 1–9, own calculations. *Note:* Standardized bias in the matched group is the sample differences in means after applying the propensity score-based weights.

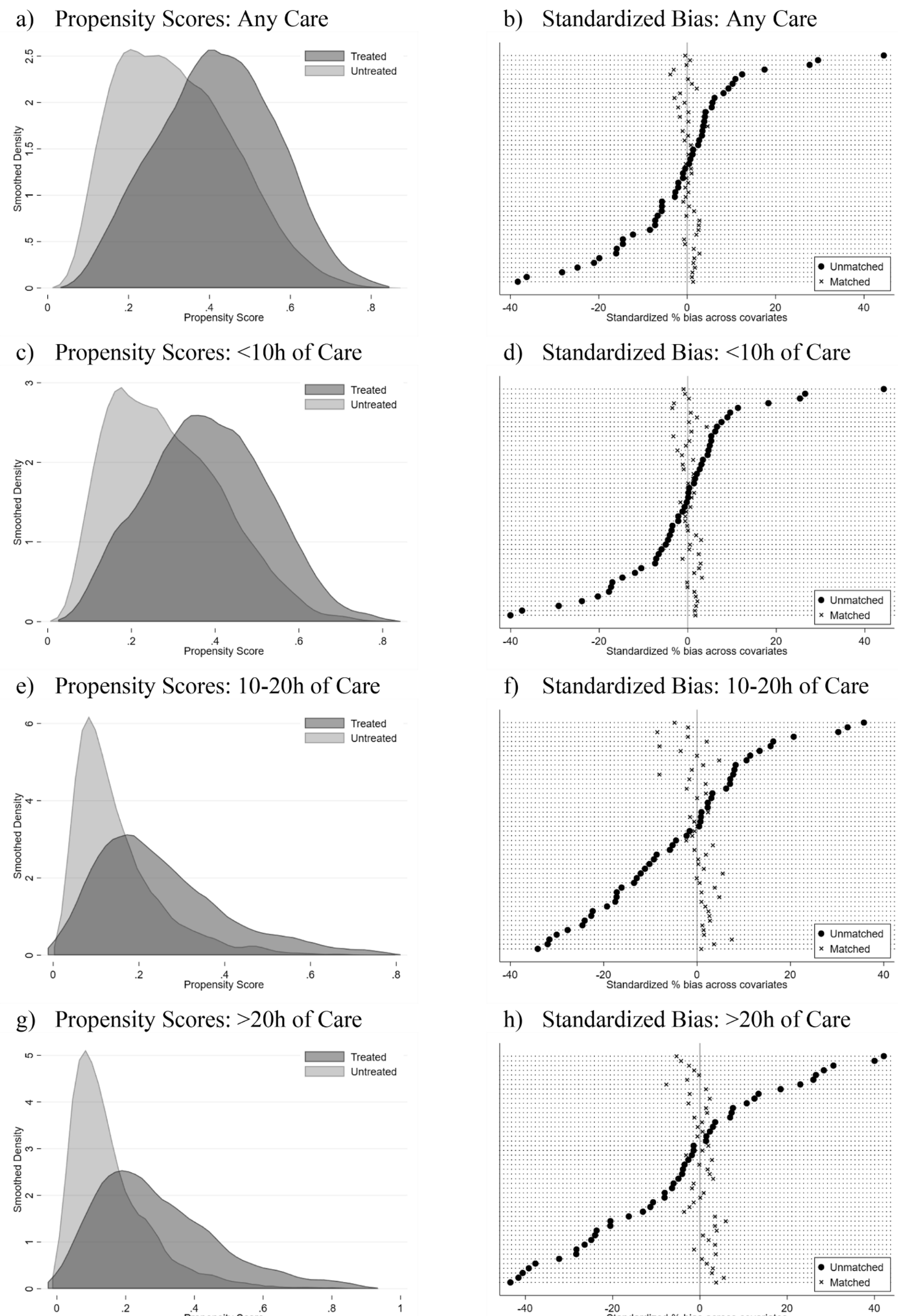


Fig. A3.2. Coarsened Exact Matching before Propensity Score Based Matching. *Source:* USoc Waves 1–9, own calculations. *Note:* Standardized bias in the matched group is the sample differences in means after applying the propensity score-based weights.

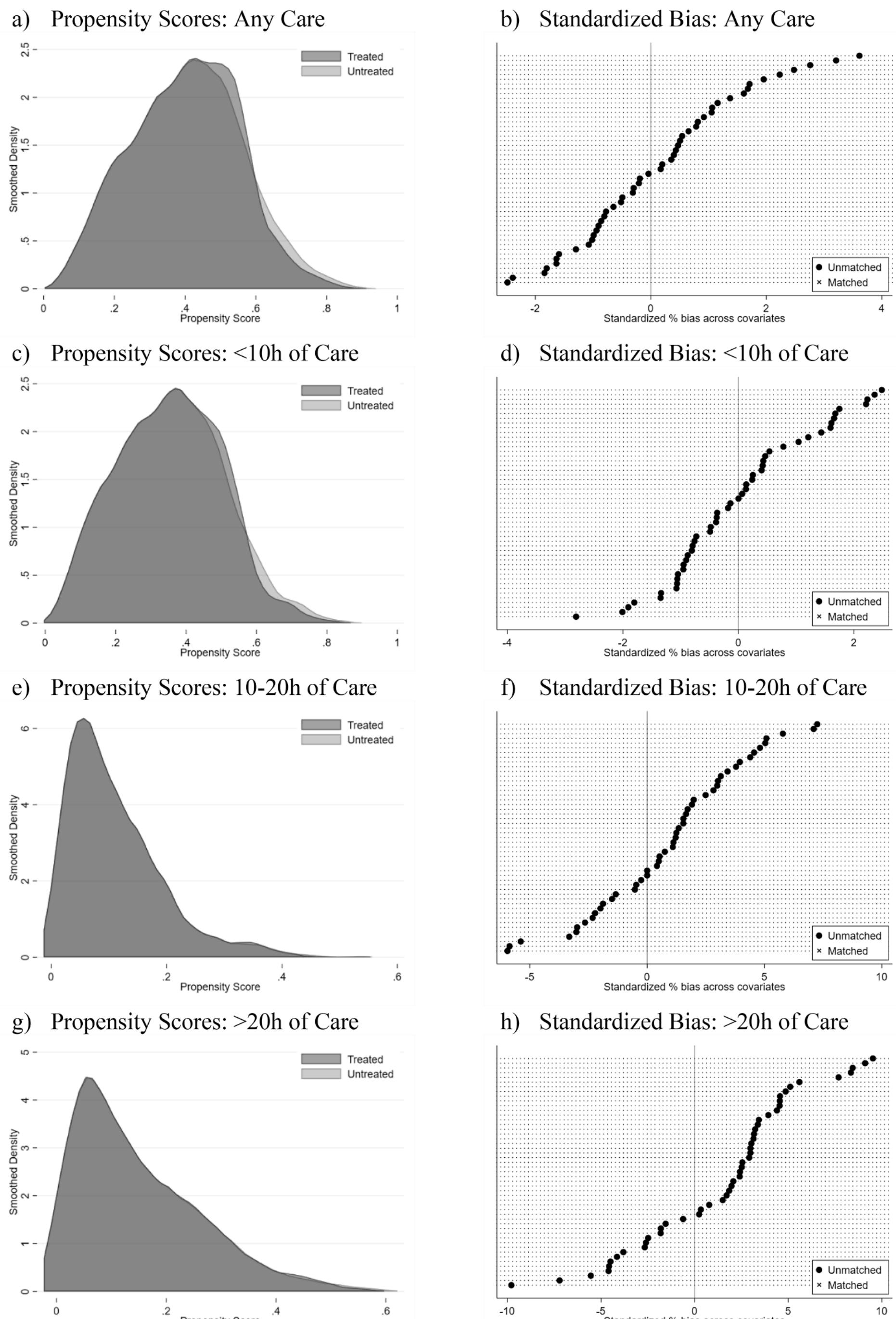


Fig. A3.3. Propensity Score Based Matching 1-to-1. *Source:* USoc Waves 1–9, own calculations. *Note:* As individuals are matched 1-to-1 the standardized bias depicts no difference between treated and control group after matching as no re-weighting occurs and individuals are paired based on propensity scores directly.

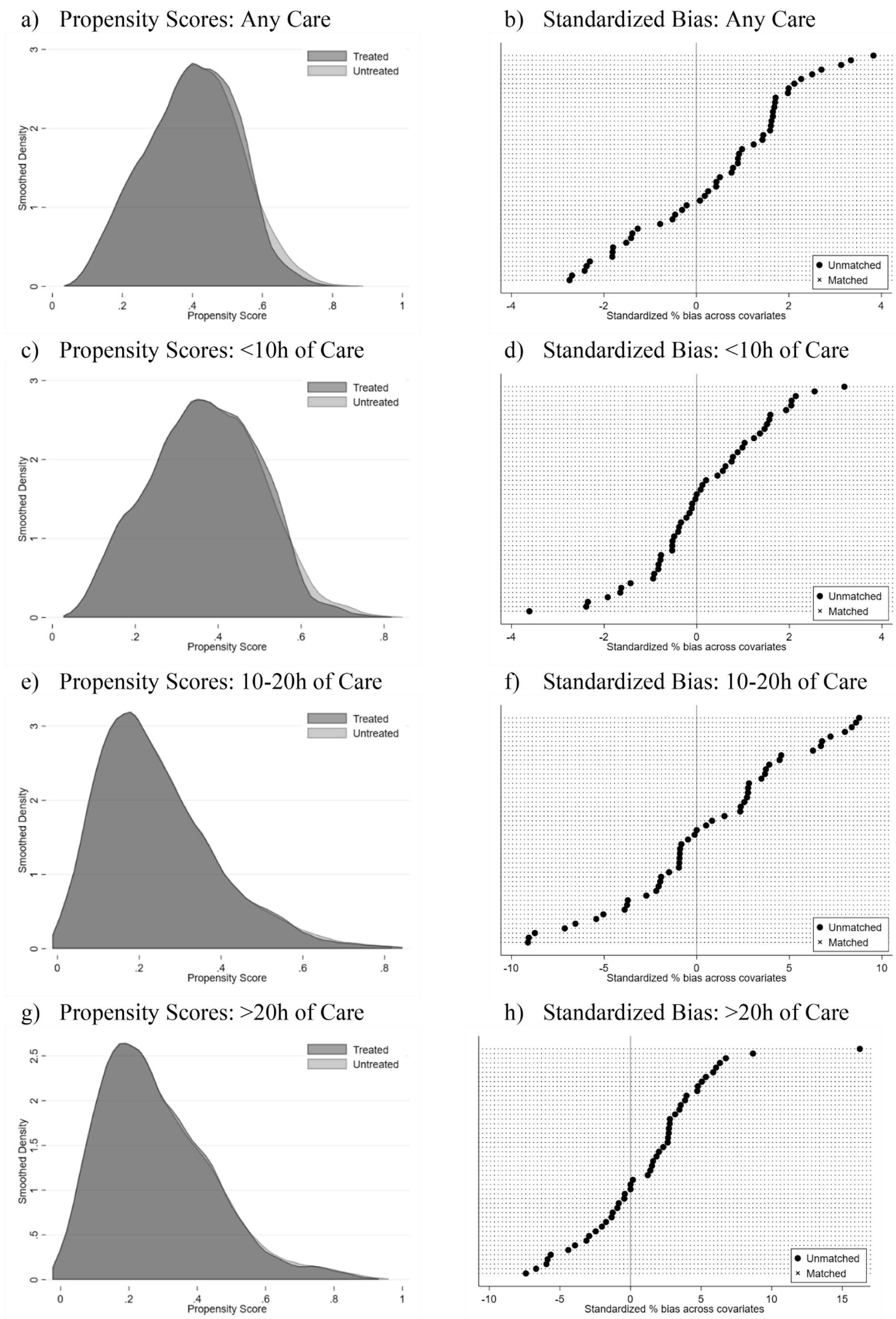
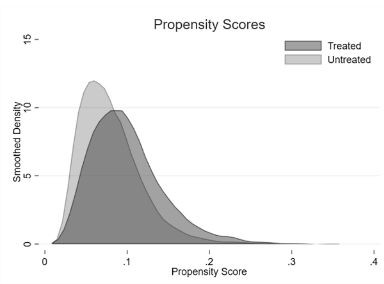


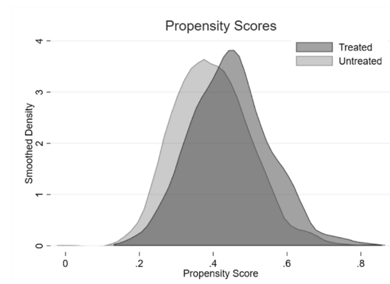
Fig. A3.4. Coarsened Exact Matching before Propensity Score Based Matching 1-to-1. *Source:* USoc Waves 1–9, own calculations. *Note:* As individuals are matched 1-to-1 the standardized bias depicts no difference between treated and control group after matching as no re-weighting occurs and individuals are paired based on propensity scores directly.

A4 – Dynamic sequential matching

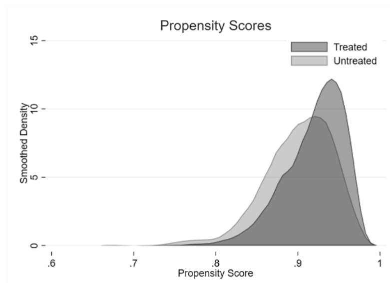
a) Propensity to start IC (Pr (D_0=1|X_{-1}, Y_{-1}))



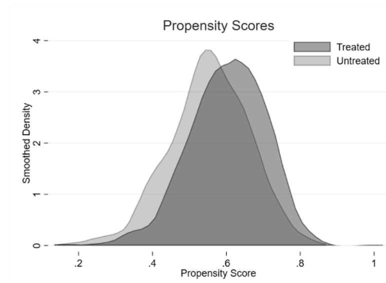
b) Propensity of IC after IC in year 1 (Pr(D_1=1 | D_0=1, X_{-1}, Y_{-1}, X_0, Y_0)



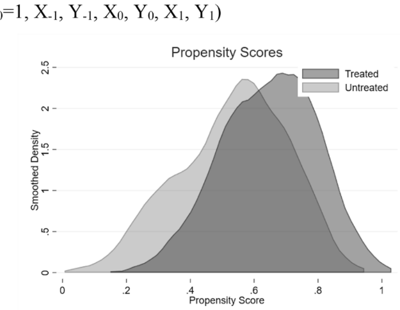
c) Propensity of no IC after no IC in year 1 (Pr(D_1=0 | D_0=0, X_{-1}, Y_{-1}, X_0, Y_0)



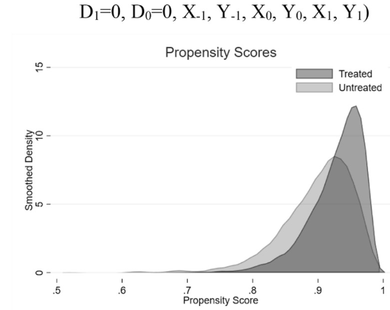
d) Propensity of no IC after IC in year 1 (Pr(D_1=0 | D_0=1, X_{-1}, Y_{-1}, X_0, Y_0)



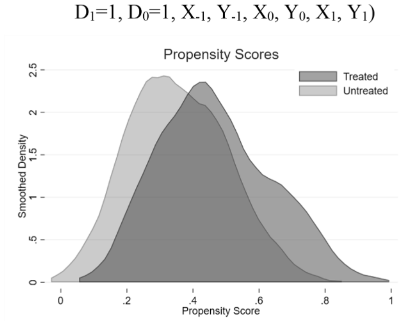
e) Propensity of IC after 2 years of IC (Pr(D_2=1 | D_1=1, D_0=1, X_{-1}, Y_{-1}, X_0, Y_0, X_1, Y_1)



f) Propensity of no IC after 2 years of no IC (Pr(D_2=0 | D_1=0, D_0=0, X_{-1}, Y_{-1}, X_0, Y_0, X_1, Y_1)



g) Propensity of no IC after 2 years of IC (Pr(D_2=0 | D_1=1, D_0=1, X_{-1}, Y_{-1}, X_0, Y_0, X_1, Y_1)



h) Propensity of no IC after IC in t=0 and no IC in t=1 (Pr(D_2=0 | D_1=0, D_0=1, X_{-1}, Y_{-1}, X_0, Y_0, X_1, Y_1)

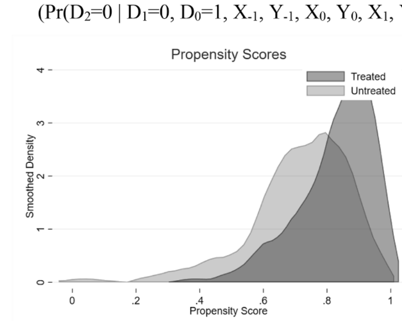


Fig. A4.1. Estimated Propensity Scores, any sample.

Table A4.1
Overview excluded propensity scores.

Informal Care Trajectory	Off support	Off support low	Off support high
Initial informal care (Pr ($D_0 = 1 \| X_{-1}, Y_{-1}$))	1	0	0
2nd year of care provision (Pr($D_1 = 1 \mid D_0 = 1, X_{-1}, Y_{-1}, X_0, Y_0$)	0	0	11
2nd years of no care provision (Pr($D_1 = 0 \mid D_0 = 0, X_{-1}, Y_{-1}, X_0, Y_0$)	5	13	613
Informal care at t_0 but not t_1 (Pr($D_1 = 0 \mid D_0 = 1, X_{-1}, Y_{-1}, X_0, Y_0$)	17	9	18
3rd year of care provision (Pr($D_2 = 1 \mid D_1 = 1, D_0 = 1, X_{-1}, Y_{-1}, X_0, Y_0, X_1, Y_1$)	15	5	–
3rd year of no care provision (Pr($D_2 = 0 \mid D_1 = 0, D_0 = 0, X_{-1}, Y_{-1}, X_0, Y_0, X_1, Y_1$)	230	118	–
Informal care at t_0 and t_1 but not t_2 (Pr($D_2 = 0 \mid D_1 = 1, D_0 = 1, X_{-1}, Y_{-1}, X_0, Y_0, X_1, Y_1$)	5	10	–
Informal care at t_0 but not t_1 and t_2 (Pr($D_2 = 0 \mid D_1 = 0, D_0 = 1, X_{-1}, Y_{-1}, X_0, Y_0, X_1, Y_1$)	59	95	–

References

Age UK, 2019a. Paying for homecare. https://www.ageuk.org.uk/information-advice/care/paying-for-care/paying-for-homecare/.

Age UK, 2019b. Paying for a care home. https://www.ageuk.org.uk/information-advice/care/paying-for-care/paying-for-a-care-home/.

Angrist, J.D., Imbens, G.W., 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. J. Am. Stat. Assoc. 90 (430), 431–442.

Barnay, T., Juin, S., 2016. Does home care for dependent elderly people improve their mental health? J. Health Econ. 45, 149–160. https://doi.org/10.1016/j.jhealeco.2015.10.008.

Bauer, J.M., Sousa-Poza, A., 2015. Impacts of informal caregiving on caregiver employment, health, and family. J. Popul. Ageing 8 (3), 113–145.

Blackwell, M., Iacus, S., King, G., Porro, G., 2009. cem: Coarsened exact matching in Stata. Stata J. 9 (4), 524–546.

Bobinac, A., Van Exel, N.J.A., Rutten, F.F., Brouwer, W.B., 2010. Caring for and caring about: disentangling the caregiver effect and the family effect. J. Health Econ. 29 (4), 549–556. https://doi.org/10.1016/j.jhealeco.2010.05.003.

Bakx, P., Wouterse, B., van Doorslaer, E., Wong, A., 2020. Better off at home? Effects of nursing home eligibility on costs, hospitalizations and survival. J. Health Econ. 102354 https://doi.org/10.1016/j.jhealeco.2020.102354.

Bom, J., Bakx, P., Schut, F., van Doorslaer, E., 2019a. The Impact of Informal Caregiving for Older Adults on the Health of Various Types of Caregivers. Gerontologist 59 (5), e629–e642. https://doi.org/10.1093/geront/gny137.

Bom, J., Bakx, P., Schut, F., van Doorslaer, E., 2019b. Health effects of caring for and about parents and spouses. J. Econ. Ageing 14. https://doi.org/10.1016/j.jeoa.2019.100196.

Bom, J., Stöckel, J., 2021. Is the grass greener on the other side? The health impact of providing informal care in the UK and the Netherlands. Soc. Sci. Med. 269, 113562.

Carretero, S., Garcés, J., Ródenas, F., Sanjosé, V., 2009. The informal caregiver's burden of dependent people: Theory and empirical review. Arch. Gerontol. Geriatr. 49 (1), 74–79. https://doi.org/10.1016/j.archger.2008.05.004.

Caliendo, M., Kopeinig, S., 2008. Some practical guidance for the implementation of propensity score matching. J. Econ. Surveys 22 (1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x.

Carers UK (2016). State of Caring 2016. https://www.carersuk.org/for-professionals/policy/policy-library/state-of-caring-2016.

Coe, N.B., Guo, J., Konetzka, R.T., Van Houtven, C.H., 2019. What is the marginal benefit of payment-induced family care? Impact on Medicaid spending and health of care recipients. Health Econ. 28 (5), 678–692. https://doi.org/10.1002/hec.3873.

Colombo, F., Llena-Nozal, A., Mercier, J., Tjadens, F., 2011. Help Wanted? Providing and paying for long-term care. In: OECD Health Policy Studies. OECD Publishing. https://doi.org/10.1787/9789264097759-en.

Comas-Herrera, C., Pickard, L., Wittenberg, R., Malley, J., & King, D., 2010. The long-term care system for the elderly in England. European Network of Economic Policy Research Institutes. ENEPRI Research Report No. 74 Contribution to WP1 of the Ancien Project.

Courtin, E., Jemiai, N., Mossialos, E., 2014. Mapping support policies for informal carers across the European Union. Health Policy 118 (1), 84–94.

Curry, N., Schlepper, L., Hemmings, N., 2019. What can England learn from the long-term care system in Germany? Research report. Nuffield Trust.

Di Novi, C., Jacobs, R., Migheli, M., 2015. The quality of life of female informal caregivers: From Scandinavia to the Mediterranean Sea. Eur. J. Popul. 31, 309–333. https://doi.org/10.1007/s10680-014-9336-7.

Eibich, P., 2021. Care or self-care? The impact of informal care provision on health behaviour (No. WP-2021-005). Max Planck Institute for Demographic Research, Rostock, Germany.

European Commission (2018). Informal care in Europe: Exploring Formalisation, Availability and Quality. https://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=8106&type=2&furtherPubs=no.

ESS Round 7: European Social Survey Round 7 Data (2014). Data file edition 2.2. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

Fernández, J.-L., Forder, J., Truckeschitz, B., Rokosova, M., & McDaid, D., 2009. How can European states design efficient, equitable and sustainable funding systems for long-term care for older people?, Policy Brief No. 11, World Health Organisation Europe, Copenhagen.

Fischer, B., Müller, K.U., 2020. Time to care? The effects of retirement on informal care provision. J. Health Econ. 102350.

Glendinning, C., 2013. Long Term Care Reform in England: A Long and Unfinished Story in Reforms. In: Ranci, C., Pavolini, E. (Eds.), Long-Term Care Policies in Europe. Springer, New York.

Haley, W.E., Pardo, K.M., 1989. Relationship of severity of dementia to caregiving stressors. Psychol. Aging 4, 389–392. https://doi.org/10.1037/0882-7974.4.4.389.

Hirst, M., 2005. Carer distress: a prospective, population-based study. Soc. Sci. Med. 61 (3), 697–708.

Iacus, S.M., King, G., Porro, G., 2012. Causal inference without balance checking: Coarsened exact matching. Polit. Anal. 1–24.

Ichino, A., Mealli, F., Nannicini, T., 2008. From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? J. Appl. Econ. 23 (3), 305–327. https://doi.org/10.1002/jae.998.

Lacey, R.E., McMunn, A., Webb, E., 2019. Informal caregiving patterns and trajectories of psychological distress in the UK Household Longitudinal Study. Psychol. Med. 49, 1652–1660. https://doi.org/10.1017/S0033291718002222.

Lechner, M., 2009. Sequential causal models for the evaluation of labor market programs. J. Business Econ. Stat. 27 (1), 71–83. https://doi.org/10.1198/jbes.2009.0006.

Lechner, M., Miquel, M., 2010. Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions. Empirical Economics 39 (1), 111–137. https://doi.org/10.1007/s00181-009-0297-3.

Legg, L., Weir, C.J., Langhorne, P., Smith, L.N., Stott, D.J., 2013. Is informal caregiving independently associated with poor health? A population-based study. J. Epidemiol. Community Health 67, 95–97.

Leuven, E., Sianesi, B., 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Boston College Department of Economics, Statistical Software Components. Downloadable from http://ideas.repec.org/c/boc/bocode/s432001.html.

Lilly, M.B., Laporte, A., Coyte, P.C., 2007. Labor market work and home care's unpaid caregivers: A systematic review of labor force participation rates, predictors of labor market withdrawal, and hours of work. Milbank Q. 85, 641–690. https://doi.org/10.1111/j.1468-0009.2007.00504.x.

de la Maisonneuve, C., Martins, J.O., 2015. The future of health and long-term care spending. OECD J.: Econ. Stud. 2014 (1), 61–96. https://doi.org/10.1787/19952856.

NHS, 2018.When the council might pay for your care. https://www.nhs.uk/conditions/social-care-and-support-guide/money-work-and-benefits/when-the-council-might-pay-for-your-care/.

OECD, 2019. In: Health at a Glance 2019: OECD Indicators. OECD Publishing, Paris. https://doi.org/10.1787/4dd50c09-en.

ONS, 2013. 2011 Census analysis: Unpaid care in England and Wales, 2011 and comparison with 2001. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/articles/2011censusanalysisunpaidcareinenglandandwales2011andcomparisonwith2001/2013-02-15#national-comparisons.

ONS, 2015. UK Health Accounts reference tables. - Long-term care expenditure by function and financing scheme in 2015. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/ukhealthaccounts/2015.

ONS, 2017. Unpaid carers provide social care worth £57 billion. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/articles/unpaidcarersprovidesocialcareworth57billion/2017-07-10.

Pearlin, L.I., Mullan, J.T., Semple, S.J., Skaff, M.M., 1990. Caregiving and the stress process: An overview of concepts and their measures. Gerontologist 30 (5), 583–594. https://doi.org/10.1093/geront/30.5.583.

Pinquart, M., Sörensen, S., 2003a. Associations of Stressors and Uplifts of Caregiving With Caregiver Burden and Depressive Mood: A Meta-Analysis. J. Gerontol.: Ser. B 58 (2), 112–128. https://doi.org/10.1093/geronb/58.2.P112.

Pinquart, M., Sörensen, S., 2003b. Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. Psychol. Aging 18 (2), 250–267. https://doi.org/10.1037/0882-7974.18.2.250.

Rellstab, S., Bakx, P., García-Gómez, P., van Doorslaer, E., 2020. The kids are alright - labour market effects of unexpected parental hospitalisations in the Netherlands. J. Health Econ. 69 https://doi.org/10.1016/j.jhealeco.2019.102275.

Robins, J.M., Hernan, M.A., Brumback, B., 2000. Marginal Structural Models and Causal Inference in Epidemiology. Epidemiology 11, 550–560. https://doi.org/10.1097/00001648-200009000-00011.

Rosenbaum, P.R., Rubin, D.B., 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. Am. Statist. 39 (1), 33–38. https://doi.org/10.2307/2683903.

Rubin, D.B., 1973. The use of matched sampling and regression adjustment to remove bias in observational studies. Biometrics 29, 185–203. https://doi.org/10.2307/2529685.

Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66 (5), 688–701. https://doi.org/10.1037/h0037350.

Rubin, D.B., 1979. Using multivariate matched sampling and regression adjustment to control bias in observational studies. J. Am. Stat. Assoc. 74, 318–328. https://doi.org/10.2307/2286330.

Schmitz, H., Westphal, M., 2015. Short-and medium-term effects of informal care provision on female caregivers' health. J. Health Econ. 42, 174–185. https://doi.org/10.1016/j.jhealeco.2015.03.002.

Schmitz, H., Westphal, M., 2017. Informal care and long-term labor market outcomes. J. Health Econ. 56, 1–18. https://doi.org/10.1016/j.jhealeco.2017.09.002.

Stuart, E., 2010. Matching methods for causal inference: A review and a look forward. Stat. Sci. 25 (1), 1.

Townsend, A., Noelker, L., Deimling, G., Bass, D., 1989. Longitudinal impact of interhouse caregiving on adult children's mental health. Psychol. Aging 4, 393–401. https://doi.org/10.1037/0882-7974.4.4.393.

UK Government (2020). Carer's allowance. https://www.gov.uk/carers-allowance.

University of Essex. Institute for Social and Economic Research. (2019). Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 12th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-13.

Urwin, S., Lau, Y.S., Grande, G., Sutton, M., 2021. The extent and predictors of discrepancy between provider and recipient reports of informal caregiving. Soc. Sci. Med. 113890.

Van Houtven, C.H., Norton, E.C., 2004. Informal care and health care use of older adults. J. Health Econ. 23 (6), 1159–1180. https://doi.org/10.1016/j.jhealeco.2004.04.008.

Van Houtven, C.H., Norton, E.C., 2008. Informal care and Medicare expenditures: testing for heterogeneous treatment effects. J. Health Econ. 27 (1), 134–156. https://doi.org/10.1016/j.jhealeco.2007.03.002.

Vlachantoni, A., Robards, J., Falkingham, J., Evandrou, M., 2016. Trajectories of informal care and health. *SSM—Population*. Health 2, 495–501.

Ware, J.E., Kosinski, M., Keller, S.D., 1995. SF-12: How to score the SF-12 physical and mental health summary scales, 2nd. ed. The Health Institute, New England Medical Center, Boston, Massachusetts.

Yeandle, S., 2016. Caring for our carers: An international perspective on policy developments in the UK. Juncture 23 (1), 57–62. https://doi.org/10.1111/j.2050-5876.2016.00896.x.

de Zwart, P.L., Bakx, P., van Doorslaer, E.K.A., 2017. Will you still need me, will you still feed me when I'm 64? The health impact of caregiving to one's spouse. Health Econ. 26, 127–138. https://doi.org/10.1002/hec.3542.