The effect of informal caregiving on caregivers' sleep quality and mental health

Comparing caregivers to non-caregivers using propensity score matching

Author: Jorien Kempers

Student number: 620149

Supervisor: Jannis Stöckel

Second reader: Pilar Garcia Gomez

Rotterdam, 12-07-2022

Word count: 13,028

Abstract

The ageing population in industrialized countries is becoming an increasing concern for governments, as it is associated with reduced population health and an increasing demand for care. Informal care plays an important role in these countries to meet this increasing demand. However, informal caregivers are burdened by providing informal care. Several studies found a negative effect on the mental health of caregivers. This does not only impact the individual itself, but also firms and the society as a whole. Informal care can also affect the sleep quality of caregivers. Sleep is not only essential for the physical health of individuals, but also for the mental health of individuals as it regulates emotions. Therefore, it could be that sleep quality is an explanation for the mental health effects found by previous studies. This research will analyse the impact of informal care provision on the sleep quality and the mental health of individuals for any type of care and the different intensities of care, using data from the UK Household Longitudinal study. Propensity score matching is used to match non-informal caregivers to informal caregivers to calculate the average treatment effect on the treated (ATT). This study finds significant negative effects of providing informal care on the sleep quality and mental health of individuals that provide any type of care, but also for the different intensities. In addition, a negative relationship is found between the number of hours of caregiving and sleep quality and mental health. Indicating that the most severe negative effects on sleep quality and mental health are found among high intensity caregivers. The results in this research are economically relevant as the observed effects imply an increase in the number of individuals experiencing significant sleep problems and the number of depressed individuals based on a clinically validated survey screening measure. These results are important for policymakers as they need to rely on informal caregivers to care for the ageing population.

Keywords: informal care, sleep quality, mental health, propensity score matching, regression adjustment

Table of contents

1. Introduction

The population is ageing dramatically in most industrialized countries. This is caused by a combination of decreased fertility rates and increased life expectancy (Bauer & Sousa-Poza, 2015). Across countries of the Organisation for Economic Co-operation and Development (OECD), which are predominantly high-income economies, the share of individuals aged over 65 years increased from less than 9% in 1960 to 17% in 2015 and is projected to rise to 28% by 2050 (OECD, 2017). An example of an OECD country with an ageing population is the United Kingdom (UK). More than a quarter of its population is expected to be over 65 years and even over 10% is expected to be over 80 years by 2050 (OECD, 2019). This ageing population is facing more long-term disabilities that often require not only short-term curative care, but also long-term care (LTC) services.

LTC services exist of personal care services provided to individuals with a limited capacity to care for themselves. These services do not only increase the well-being of disabled elderly, but also of nonelderly who face difficulties in performing daily life activities. LTC services can be delivered by institutions, but also by informal caregivers (Kaye, Harrington, & LaPlante, 2010). Informal caregiving can be broadly defined as providing unpaid care to friends and close relatives (Colombo, Llena-Nozal, Mercier, & Tjadens, 2011). In the UK, components of LTC that are health related and granted by the GP are funded via the National Health Service (NHS) (Comas-Herrera, Pickard, Wittenberg, Malley, & King, 2010). However, local authorities are responsible for other types of LTC, such as help with personal care tasks and residential care. The local authorities determine who has severe needs and is not able to pay the care themselves. These patients get access to funded LTC services (Fernandez, Forder, Truckeschitz, Rokosova, & McDaid, 2009). However, a large part of the patients that need LTC services do not get access to these funded services. Therefore, a major share of this increasing demand for elderly care is being met informally (Bauer & Sousa-Poza, 2015). In the UK more than 18% of the 50+ population is providing informal care (OECD, 2019). From a governmental perspective, informal caregiving is a convenient source to meet the increasing care demand and to reduce the cost-intensive use of formal care. In 2011 the economic value of unpaid care in the UK was estimated at £132 billion, representing an even higher amount than the annual cost of the NHS (Buckner & Yeandle, 2011). Informal care can also be seen as a benefit for the care recipients, because often they prefer to live in their own homes and receive care from someone they know (Brouwer, van Exel, van den Berg, van den Bos, & Koopmanschap, 2005; Carmichael & Charles, 2003). Furthermore, informal caregiving can be fulfilling for the caregivers themselves, because they enjoy caring for someone in their social environment (Brouwer et al., 2005).

Informal caregiving can be seen as a solution for both the government and the society as a whole, due to the increasing demand for elderly care. However, at the same time several researchers have found that informal caregiving can cost a lot of time and can be mentally stressful as well as physically exhausting for the informal caregivers. Subsequently, this can result in negative effects on both physical and psychological wellbeing and the career of the informal caregiver (Bauer & Sousa-Poza, 2015). Especially caregivers who provide care at a high intensity, are the main caregiver and are providing informal care to someone that is not living in the same household, are experiencing these burdens (Hirst, 2005). The research of Meng (2012) found that informal caregivers in Germany keep working almost the same number of hours after becoming an informal care provider. Suggesting that the labour market participation remains high. At the same time, researchers did find an effect on the mental health of care providers. An explanation could be that by combining work and informal caregiving, the sleep quality of informal caregivers is affected. The quality of sleep is important for people's mental health, because during periods of sleep our brain is regulating emotions (Crivello, Barsocchi, Girolami, & Palumbo, 2019). The research of Gadie, Shafto, Leng, and Kievit (2017) found that in the UK poor sleep quality is strongly associated with poorer mental health. This can suggest that lower sleep quality may be an explanation for the decreasing mental health of informal caregivers.

Hence, both sleep quality and mental health are important outcome measures. Sleep quality is important, because it affects both the mental and physical health of individuals (Clement-Carbonell, Portilla-Tamarit, Rubio-Aparicio, & Madrid-Valero, 2021). The research of Baglioni, Spiegelhalder, Lombardo, and Riemann (2010) found that individuals with poor sleep quality experience a substantial amount of negative emotions, because sleep is needed in order to regulate emotions. Therefore, poor sleep quality is associated with anxiety and depression. Sleep quality is also important for the physical functioning of individuals as it is needed for physical recovery (Seow et al., 2020). Mental health is also important, because it has a substantial impact on both the individual itself as well as on firms and the society as a whole. Individuals suffering from mental illnesses imply great economic costs for governments. For instance, the estimated cost of mental illness in the UK was £105.2 billion for the year 2009-2010. This consists of both direct costs of mental illness (e.g., the cost of healthcare or disability costs) and indirect costs (e.g., the care the individual now needs from family). This economic cost also includes costs of foregone productivity at work as a result of presenteeism (i.e., being at work, but being less productive) and absenteeism (Bubonya, Cobb-Clark, & Wooden, 2017).

Understanding the consequences of sleep quality and mental health is important for policymakers in order to assess whether it is necessary to change support policies for informal caregivers to be able to continue to rely on informal caregivers as a society. As the real ageing pressure is to come when the large post-war birth cohorts retire and the need for LTC services increases, but at the same time fertility rates decreases. Resulting in an increasing pressure on a smaller proportion of the population that needs to provide informal care and pay the taxes to finance the social security (Bloom et al., 2015). Are the benefits worth the costs of providing informal care? Therefore, the aim of this research is to answer to following questions:

1a: What is the effect of informal caregiving on the sleep quality and mental health of caregivers?

1b: What is the role of care intensity on the sleep quality and mental health of caregivers?

To answer these questions, data from the UK Household Longitudinal Study is used. The empirical evidence is obtained from a sample of 30,122 individual-year observations over a time period of 4 waves (wave 1, 4, 7 and 10). Propensity score matching is used to find the average sleep quality and mental health differences between informal caregivers and non-informal caregivers. The sleep quality and mental health outcomes are explored across different matching estimators. Furthermore, the mechanism with respect to sleep quality and mental health is investigated.

The remainder of this thesis is divided into the following sections. The next section provides a literature review on the association of providing informal care and sleep quality as well as providing informal care and mental health. The third section describes which method will be used and section four presents which data and variables will be used in this research. Section five contains the results and is followed by section 6 which will test the robustness of these results. The last section entails a discussion and conclusion.

2. Theoretical framework

First of all, understanding what can be seen as informal caregiving is of utmost importance in this research. An individual is identified as an informal caregiver according to the European Commission when: the caregiver and the receiver have a close relationship with each other, the caregiver provides a lot of different care giving duties, there are no official working hours, the caregiver is not trained to provide care, it does not have a contract and does not get paid (Triantafillou et al., 2010). Most of the identified informal caregivers are middle-aged women who are providing personal care services (Navaie-Waliser, Spriggs, & Feldman, 2002). This section will focus on the theoretical impact of informal care provision on labour market outcomes, sleep quality outcomes and mental health outcomes of the informal caregiver.

2.1 Effect of informal caregiving on labour market outcomes

The impact of informal caregiving on the labour market outcomes are important for governments as caregivers are often of working age. This means that for most individuals the time to provide informal care competes with time to be in paid employment, because time is scarce (van Houtven, Coe, & Skira, 2013). Caregivers are experiencing difficulties with finding the right balance between their responsibilities as an informal caregiver and their work responsibilities (Navaie-Waliser et al., 2002). Therefore, caregivers may reduce the hours that they work or even stop working (van Houtven et al., 2013). The research of Berecki-Gisolf, Lucke, Hockey, and Dobson (2008) for example found that women in their 50s are twice as likely to lower their participation in paid employment after they transitioned into informal caregiving compared to non-caregivers. This is in line with the research of Lilly, Laporte, and Coyte, (2007) who found that caregivers are more likely to reduce their work hours compared to non-caregivers. They also found that caregivers who provide intensive care are more likely to leave the labour market compared to non-caregivers. These reductions could cause a problem for governments to finance their social security. Therefore, the trade-off for governments is between individuals taking up the care work or paying taxes to contribute to social security.

Caregivers could also reduce their leisure time or/and the hours of sleep to keep their working hours the same, because of financial reasons or to get a break from providing informal care (van Houtven et al., 2013). The research of Meng (2012) found that if German women provide 8 hours of care per week, they will reduce their hours of work by 35 minutes per week. For male caregivers this reduction in working hours will be 48 minutes per week. Meaning that both men and women who are providing informal care are almost working the same number of hours as before. However, on top of these hours they also provide 8 hours of care per week. This may suggest that the labour market participation remains high, which can imply that the sleep quality of informal caregivers is affected.

However, it could also be that there is a difference in what caregivers report in care-time and what the actual care-time is. The research of Meng (2012), but in general most of informal care research, relies upon information that the caregiver itself reports other than on information that the receiver reports (Urwin, Lau, Grande, & Sutton, 2021). Therefore, the research of Urwin et al. (2021) analysed if there is a difference between care providers and receivers with regard to reporting on informal care. In the UK, they found that if the informal care provider and receiver both agree on the fact that informal care is given, they do not agree on the number of hours of informal care. Providers of informal care report on average 10.55 (37%) more hours of care per week compared to care receivers. Therefore, it could also be that the labour market participation remains high, but that the actual number of hours individuals spend on providing informal care are much lower in reality than reported. If this is the case, there is also the possibility that the sleep quality is not affected by providing informal care.

2.2 Sleep quality and mental health

As mentioned, previous findings on the labour market outcomes could suggest worse sleep quality among informal caregivers. Sleep quality can be assessed by subjective and objective measures. The Pittsburgh Sleep Quality Index (PSQI) is a self-reported commonly used measure which consists of 19 items in seven domains which are combined in one overall sleep quality score. The higher the score, the lower the sleep quality (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989). However, sleep quality can also be measured by a collection of objective measures that are being taken from polysomnography (PSG). Examples of these objective measures are measures reflecting the architecture of sleep such as the temporal amounts or percentage of stage 1 sleep, stage 2 sleep, rapid eye movement (REM) or slow wave sleep which cannot be measured with the PSQI (Krystal & Edinger, 2008). However, according to the research of O'Donnel et al. (2009), who studied 24 healthy subjects between 55 and 74 years old, these objective measures taken from PSG correlate with the subjective domains used for PSQI.

The research of Gao, Chapagain, and Scullin (2019) performed a systematic review and meta-analysis of 35 studies which included data of 3268 informal caregivers. In their research, they included studies which measured sleep quality with the PSQI and studies which used data from PSG. They found that informal caregivers experience lower sleep quality and sleep duration compared to non-caregivers. Their results showed that caregivers lose 2.42 to 3.50 hours of sleep per week, because they face difficulties with falling asleep and remaining asleep. These results confirm the expected negative association between informal care provision and sleep quality. This decrease found in sleep quality among informal caregivers is likely to impact the mental health of informal caregivers as previous research showed that sleep is really important for people's health. The research of Crivello et al. (2019) found that sleep is not only needed to regulate emotions, but is also important for people's general physical health. According to the research of Pilcher and Ott (1998), quality of sleep is even more important than the quantity of sleep in relation to measures of well-being and health. Therefore, studies often use quality of sleep as a determinant for well-being outcomes (Weinberg, Noble, & Hammond, 2016). Several studies investigated the relation between sleep quality and mental health. The research of Gadie et al. (2017) showed that poor sleep quality is strongly associated with poor mental health in the UK and that it remains stable across the lifespan of people. Furthermore, O'Leary, Bylsma, and Rottenberg (2017) found that poor sleep quality leads to more depressive symptoms due to impairment in the emotion regulation of individuals. Therefore, the negative effects of informal caregiving on the sleep quality of caregivers could be associated with poor mental health among caregivers.

2.3 Effect of informal caregiving on mental health

Previous studies found evidence for a negative effect of sleep quality on mental health. This could be an explanation for the mental health effects found among informal caregivers by Bom, Bakx, Schut, and van Doorslaer (2019). They performed a systematic literature review, including 15 articles, and found a causal negative effect of informal caregiving to an elderly person on the mental health of the caregiver.

Several studies also found a negative effect on the mental health of caregivers using propensity score matching. This method is used to match caregivers to non-caregivers. For instance, de Zwart, Bakx, and van Doorslaer (2017) analysed the mental health of caregivers aged 50 years and older who provide informal care to their own partner using data from the Survey of Health, Ageing and Retirement in Europa (SHARE). They found that both male and female caregivers experience more depressive symptoms compared to non-caregivers. For male caregivers the score increases with 0.45 and for female caregiver with 0.57, representing relatively large increases compared to the sample mean of 1.948. This negative relationship is also supported by the research of Stroka (2014), which compared the drug intake of both male and female informal caregivers aged 35 and older in Germany using propensity score matching. This study found that both male and female caregivers use more antidepressants, tranquilizers, analgesics and gastrointestinal agents which indicates that caregivers experience worse mental health than non-caregivers.

Furthermore, several papers also studied the impact of different care intensities on the mental health of caregivers using propensity score matching. The research of Schmitz and Westphal (2015) examined if differences in hours of caregiving have a different impact on the mental health of German women. They compared the mental health effects across three groups based on hours of care provision per weekday: at least 1 hour (≥ 5h per week), at least 2 hours (≥ 10h per week) or providing at least 3 hours of care per weekday (≥ 15h per week). They found that there are barely differences in mental health effects between providing care for at least 1 hour per weekday and 2 hours per weekday (-1.90 and -2.00 respectively). However, providing care for at least 3 hours of per weekday results in a stronger decrease (-3.02) in the mental health of informal caregivers compared to the other two intensity groups. All these effects are significant at a 0.01 significance level**.** Therefore, they concluded that a higher care intensity results in a higher burden on the mental health of caregivers. The study of Stöckel and Bom (2022) divided caregivers into different subgroups based on the hours of care provided per week. They grouped the individuals slightly differently compared to Schmitz and Westphal (2015). They divided caregivers into low (less than 10 hours per week), medium (between 10 and 20 hours per week) and high intensity caregivers (more than 20 hours per week). Nevertheless, the did not find any significant mental health effects for individuals who provide informal care at a low or medium intensity. This is different compared to the results found by Schmitz and Westphal (2015) who did find significant negative mental health effects for individuals who provided care for less than 20 hours. However, they did find a negative mental effect for individuals who provide care at high intensity. They observed that individuals who provide care at a high intensity experience a lower mental health of 2.44 points. For females, they found a mental health score which is 2.47 points lower. This is in line with the results found by Schmitz and Westphal (2015) for high intensity female caregivers. So, both studies find evidence that providing care at a higher intensity is associated with worse mental health.

To conclude this section, informal care provision has a big impact on the daily life of caregivers, because it is very time consuming. Therefore, informal caregivers may decide to reduce their working hours or even stop working. However, researchers also found evidence of caregivers that do not reduce their labour force participation. This can result in a lower sleep quality, which is of vital importance to regulate emotions. Therefore, a reduction in the sleep quality of caregivers could be an explanation for the negative effects found on the mental health of caregivers by previous research.

3. Methodology

In this research we will be focusing on two outcome variables: sleep quality and mental health, denoted as $Y_{i,t}$. They are a function of informal caregiving (IC) and other covariates. Normally, a regression equation that looks like this would be estimated:

$$Y_{i,t} = \beta 0 + \beta 1 * ICi_{,t} + g * X_{i,t} + \mu_{i,t}$$

Where $Y_{i,t}$ is our dependent variable which is sleep quality or mental health. β0 the regular constant. $IC_{i,t}$ is a dummy for providing informal care with β1 being the coefficient of interest. In this equation is $X_{i,t}$ a vector of covariates and g the vector of coefficients for these. The covariates included in this research are relevant for both treatment selection and the outcomes of interest.

This regression equation could give causal estimates with a simple means test if we would have access to data from a randomized control trial (RCT). However, this is not the case and therefore such an approach is not appropriate to use in this research. Also estimating this equation by running an Ordinary Least Squares regression would give us biased estimates of our parameter of interest, because in our data there is a fundamental problem of selection regarding the decision to provide informal care. This decision is not random, because individuals select themselves into providing informal care. Meaning that people who are providing informal care are not comparable to people that are not providing informal care. Also, we cannot observe the same individual providing informal care and not providing informal care at the same time. Therefore, we have selection bias which can result in biased estimates and this causes a risk of drawing the wrong conclusions (Caliendo & Kopeinig, 2008).

However, there are quasi-experimental research methods that try to mimic a RCT. An example of such a method is the difference in difference (DID) approach. Multiple groups and time periods are needed to conduct such a method. In which the treatment group is exposed to some intervention, for example an exogenous policy shock, at some point in time and the control group is not exposed to this intervention (Wing, Simon, & Bello-Gomez, 2018). However, we do not have an exogenous policy shock in our data and we are only able to use data from wave 1, 4, 7 and 10. Therefore, we will not be able to estimate a causal impact with a DID approach in our research.

Another method that tries to deal with selection bias is the instrumental variable (IV) approach. A requirement in order to use this method is the availability of a valid IV in the dataset. Such an IV affects the decision to provide informal care, has no direct effect on sleep quality or mental health and is uncorrelated with any other covariate (Baiocchi, Cheng, & Small, 2014). In our research a good IV would be an exogenous health shock to the parents, e.g. if the parents get a stroke. However, we do not have access to this data. Therefore, the IV approach will also not be appropriate to use in our research.

3.1 Introduction on statistical matching

To address the problem of selection bias, the propensity score matching approach is appropriate to use. The underlying identifying assumption here is the conditional independence assumption (CIA), which means that conditioning on multiple observable variables (covariates) the potential outcomes on sleep quality or the mental health outcomes will be the same for both caregivers and non-caregivers in absence of providing informal care. Meaning that if there is a difference in outcomes after providing informal care it is due to providing informal care (Caliendo & Kopeinig, 2008).

In this approach, propensity scores are used in order to match the untreated/control individuals (non-informal caregivers) to the treated individuals (informal caregivers). Propensity scores are the estimated probability of being an informal caregiver given a set of covariates. Now all the covariates are being collapsed into one value, the propensity score (Benedetto, Head, Angelini, & Blackstone, 2018). Without propensity scores, matching would be very difficult, because of the high dimensionality of the observable characteristics. High dimensionality results in few matches as every covariate would need to be identical among treated and control individuals. Therefore, matching based on propensity scores will result in more matches. Where the propensity scores are used as weighting scheme to construct the counterfactual group (Dehejia & Wahba, 2002). However, bad matches are a concern when using propensity score matching, because individuals could have the same propensity score without having the exact same value on each covariate (Benedetto et al., 2018).

3.2 The four steps of using propensity score matching

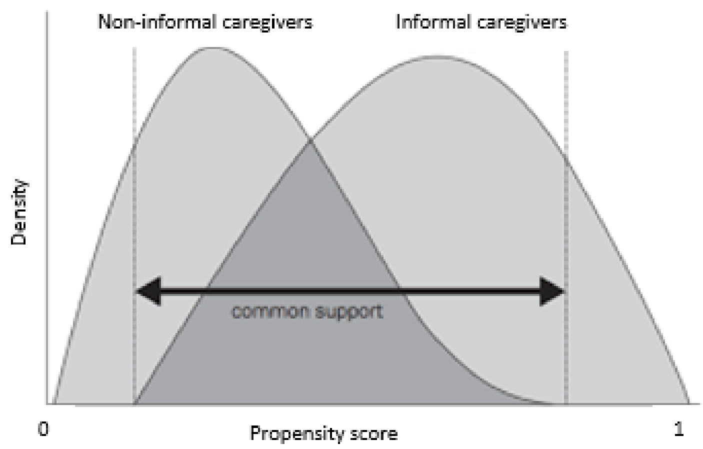
In this research we will follow the steps of implementing propensity score matching according to Jalan and Ravallion (2003). First a propensity score for all individuals needs to be calculated, then the treatment and control group will be matched according to some matching strategy, after this the matching quality will be assessed and lastly the average difference in outcome between the treatment and control group will be calculated.

Calculating propensity scores require a choice on the model used in their estimation and the variables included in this. We use a probit model to predict treatment uptake. Covariates included in this should be all those affecting selection into treatment (informal caregiving) and the outcome in absence of treatment to satisfy the CIA. While we can only include observed set of relevant covariates, the correlation of these with unobserved confounders (Stuart, 2010) gives us confidence that the chosen specification is valid. Propensity scores are estimated using a broad definition of informal caregiving and three separate groups of caregiving intensity. See the data section for a detailed list of covariates and definitions.

After calculating the propensity score for each individual, a propensity score matching approach is required in order to match untreated with treated individuals. Multiple propensity score matching approaches are available which differ in the number of untreated individuals that are included in the matching and the way these untreated individuals are weighted. In this research, we will use the Kernel based matching approach for our baseline results, with a bandwidth of 0.03. This kernel gives all treated individuals a weight of one. All untreated individuals with a propensity score that lies within a range of 0.03 of the propensity score of a treated individual get a high weight and all untreated individuals with a propensity score that lies outside this range get a low weight. An advantage of this approach is that more information will be used which results in lower variance. However, biased estimates are a concern with this approach due to potential bad matches (Caliendo & Kopeinig, 2008).

Therefore, the performance of the propensity score matching approaches will be assessed to make sure that we will not include the bad matches in this research. First, the common support will be assessed. The common support area is the area where there is overlap in observed characteristics between treatment and control group (Caliendo & Kopeinig, 2008). Figure 1 shows what the common support area looks like. The common support area shows that there are individuals that do not provide informal care, but whose likelihood to be giving informal care is very comparable to the group that actually provides informal care. Our sample will be restricted to treated and untreated individuals, whose propensity scores are located in the common support area. The second approach to assess the matching quality is to test if balance is achieved by calculating the standardized bias for all covariates. This approach will compare the situation before and after the matching for every covariate. For every covariate the standardized bias will be calculated by taking the difference in means between the treatment and control group and dividing this difference by the square root of the average variance of the sample of the treatment and the control group (Rosenbaum & Rubin, 1985). When the standardized difference between the treatment and control group is below 3% or 5%, balance is achieved and matching is being done successfully according to Caliendo and Kopeinig (2008).

Figure 1. Common support area



Note: Adapted from “Impact evaluation in practice”, Gertler, P. J., Martinez, S., Premand, P., Rawlings, L.B. and Vermeersch, C.M.J, 2011, Washington DC: World Bank, 110

In the last stage the average treatment effect on the treated (ATT) is estimated using the mean-difference between informal caregivers and the weighted control group individuals identified as being on the common support (Caliendo & Kopeinig, 2008).

4. Data

4.1 Data description

In this research, data form the UK Household Longitudinal Study is used (UKHLS; University of Essex, 2019). The survey population is a representative sample of the adult UK population (16+). The survey started in 2009 with 40.000 respondents across 30.000 households and has an overlapping panel structure. Respondents are interviewed annually allowing us to provide a longitudinal sample of individuals followed overtime. The UKHLS collects regularly information on socio-economic background, caregiving, demographics, etc. However, questions with regard to sleep quality do not appear in every wave. Therefore, we cannot use data of all waves, but only the data that is collected in waves 1, 4, 7 and 10.

4.2 Variables measurement

4.2.1 Dependent variables

Sleep quality

The first dependent variable in this analysis is sleep quality. Sleep quality is measured using the PSQI. This measure is assessing the sleep quality of an individual during the previous month. Normally the PSQI consists of 19 items in seven domains. The domains are: subjective sleep quality (one item), sleep latency (two items), sleep duration (one item), sleep efficiency (three items), sleep disturbance (nine items), sleep medication (one item) use and daytime dysfunction (two items). Each domain has a range from 0 to 3 and the PSQI is conducted by summing up all domains. Meaning that normally the conducted PSQI score has a range from 0 to 21, with a higher score indicating a lower sleep quality (Buysse et al., 1989). However, the items used to construct the domain sleep efficiency are not available in the survey we use. Therefore, an adjusted version of the PSQI is conducted with a range from 0 to 18. The other items that are missing in the survey, will not cause a problem for constructing the adjusted PSQI since at least one of the items is available in the survey. It is not expected that this adjusted PSQI will cause bias with measuring the sleep quality, because only one domain is missing and the PSQI is not a weighted index. In addition, the research of Ong, Carde, Gross, and Manber (2011) found that on average poor sleepers report the highest scores on the following domains: daytime dysfunction, subjective sleep quality, sleep latency, sleep disturbance and sleep duration. The missing domain is only a small contribution to the high PSQI score. Table 1 shows the domains and items that are included in our adjusted PSQI.

The PSQI is designed to identify if an individual is a good or a bad sleeper. Normally, if an individual has a PSQI score that is higher than 5, which is about one fourth of the maximum score, it indicates that this individual is having significant sleep difficulties (Buysse et al., 1989). Since our PSQI is adjusted we will assume that this threshold will be one fourth of our maximum score, which is about 4. However, the PSQI is inversed in this research, meaning a higher score indicates a better sleep quality. Therefore, the threshold, which indicates the significant sleep difficulties, will also be inversed. Now, individuals who have a score which is lower than 14 on the PSQI are identified as bad sleepers.

Table 1. Items used to conduct the adjusted version of the PSQI

Items of PSQI and questions asked	Answering options
Domain 1: Subjective sleep quality	
Quality of sleep overall: During the past month, how would you rate your sleep quality overall?	Very good (0), fairly good (1), fairly bad (2) and very bad (3)
Domain 2: Sleep latency	
Cannot get to sleep within 30 minutes: During the past month, how often have you had trouble sleeping because you cannot get to sleep within 30 minutes?	Not during last month (0), less than once a week (1), once or twice a week (2) and three or more times a week (3)
Domain 3: Sleep duration	
Hours of actual sleep: How many hours of actual sleep did you usually get per night during the last month?	Number of hours recoded: less than 7 (0), between 6 and 7 (1), between 5 and 6 (2) and less than 5 hours (3).
Domain 5: Sleep disturbance	
Wake up in the night: During the past month, how often have you had trouble sleeping because you woke up in the middle of the night or early in the morning?	Not during last month (0), less than once a week (1), once or twice a week (2) and three or more times a week (3)
Cough or snore loudly: During the past month, how often have you had trouble sleeping because you cough or snore loudly?	The sum of both is recoded: 0 (0), 1-2 (1), 3-4 (2), 5-6 (3)
Domain 6: Use of sleep medication	
Taken medicine to help to sleep: During the past month, how often have you taken medicine (prescribed or "over the counter") to help you sleep?	Not during last month (0), less than once a week (1), once or twice a week (2) and three or more times a week (3)
Domain 7: Daytime dysfunction	
Trouble staying awake during the day: During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?	Not during last month (0), less than once a week (1), once or twice a week (2) and three or more times a week (3)

Note: The domains and items used in this research to conduct the adjusted PSQI. Information on the domain sleep efficiency is missing. The adjusted PSQI is an unweighted measure to identify good and bad sleepers.

Mental health

The second dependent variable is mental health. To measure changes in mental health, the mental component summary scores of the 12-item Short Form Health Survey (MCS-12) is used. This questionnaire consists of 12 questions that are related to different aspects of the respondents' health in the past four weeks. The conducted score has a range from 0 to 100. The higher the score, the better the mental health of an individual. Meaning that a score of 100 is indicating the most desirable mental health state of an individual (Ware, Kosinski, & Keller, 1996). A MCS-12 score of 45.6 or lower is seen as the optimal threshold to detect 30-day depressive disorders in Europe (Vilagut et al., 2013).

In addition, the General Health Questionnair-12 (GHQ-12) is used in order to see if this measure gives the same results as the MCS-12. The conducted score is based on 12 items and the range goes from 0 to 36. A score which is higher than 15 is indicating distress and a score of 20 or higher is indicating severe problems with psychological distress (Imran, Tariq, Pervez, Jawaid, & Haider, 2016). A difference compared to the MCS-12 is that a higher score on the GHQ-12 is associated with worse mental health. Meaning that a score of 0 is indicating the best mental health state an individual can have and a score of 36 the worst mental health state (McCabe, Thomas, Brazier, & Coleman, 1996). In order to make a good comparison between the effect of informal caregiving on the different mental health measures, an inversed GHQ-12 score is created. Therefore, the thresholds should also be inversed. Meaning that a score of lower than 21 indicates distress and a score of 16 or lower than 16 indicates severe problems with psychological distress.

4.2.2 Independent variables

Providing informal care

The independent treatment indicator in this research captures if an individual is providing informal care or not. Informal caregivers are identified based on two questions on care provided inside or outside the household. The first question asked: "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.)?". The second question is: "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]"

In addition, we will also focus on the different care intensities of providing informal care. The care intensity is measured in hours spent per week on caregiving. In this paper we follow the definition of intensity groups used in previous research (Stöckel & Bom, 2022) resulting in three categories of caregivers: low intensity caregivers (<10 hours of care per week), medium intensity caregivers (between 10 and 20 hours per week) and high intensity caregivers (more than 20 hours per week).

4.2.3 Covariates

We identify covariates relevant for treatment uptake and outcomes in absence of treatment based on existing research (relevant evidence is depicted in table 2). These covariates are current labour force status, sex, age, marital status, education, household income, number of children under the age of 16 living in the household, region, lagged sleep quality and lagged mental health (Zwart et al., 2017; Schmitz & Westphal, 2015; Stöckel & Bom, 2022). The reason to include lagged sleep quality and mental health is that the sleep quality and mental health of an individual in the past are heavily correlated with the sleep quality and mental health of that same individual today. Therefore, it could be that someone has already poor sleep quality in the past and is now observed being an informal caregiver who also has poor sleep quality.

Table 2. Evidence found on the reasons to include variables as covariates

Variables	Evidence found on taking up informal care	Authors	Evidence found on sleep quality	Authors	Evidence found on mental health	Authors
Labour force status	↓ working more hours	He & Mchenry (2016)	↑ after being retired	Myllyntausta & Stenholm (2018)	↓ unemployed	Llena-Nozal (2009)
Sex	↑ females	Navaie-Waliser et al. (2002)	↓ females	Li et al. (2018)	↓ female	Rosenfield & Smith (2010)
Age	↑ middle-aged people	Dooghe (1992)	↓ older people	Gadie et al. (2017)	↓ older people	Jokela et al. (2013)
Marital status	↑ married people	Borg & Hallberg (2006)	↓ unmarried people	Hale (2005)	↓ unmarried people	Palner & Mittelmark (2002)
Education	↑ higher educated	Tur-Sinai et al. (2020)	↓ lower educated	Assari et al. (2013)	↓ lower educated	Ahn et al. (2019)
Household Income	↑ higher income level	Mentzakis et al. (2009)	↑ higher income level	Friedman et al. (2007)	↓ lower income level	Sareen et al. (2011)
Children <16 in HH	↓ more children	Mentzakis et al. (2009)	↓ more (young) children	Meltzer & Montgomery-Downs (2011)	↓ more (young) children	Meltzer & Montgomery-Downs (2011)
Region	↑ rural areas	Cohen et al. (2022)	↑ rural areas	Haseli-Mashhadi et al. (2009)	↑ rural areas	Kovess-Masféty et al. (2005)
Lagged sleep quality	↑ poor sleep quality	Kotronoulas et al. (2013)	↑ poor sleep quality	Cunnington et al. (2013)	↓ poor sleep quality	Gadie et al. (2017)
Lagged mental health	↑ poor mental health	Kaschowitz & Brandt (2017)	↑ better mental health	Lu & Liu (2021)	↓ poor mental health	Jones (2013)

Note: Studies done by more than 2 authors are indicated with “et al.” to make the table easier to read.

4.3 Construction and descriptive statistics of final dataset

4.3.1 Sample selection

After merging the data conditioning on all the variables mentioned above for wave 1, 4, 7 and 10, the dataset consists of 173,508 observations and 78,312 individual-year observations. All individuals will be included if they provide information for at least one time period on their sleep quality, mental health, if they do or do not provide informal care and on all the covariates. Observations are excluded from the dataset if there were missing values or proxy respondents. Respondents that refused to answer the question or answered the question with "I don't know" are also excluded from the dataset. This results in a dataset with 30,122 observations and 17,548 individual-year observations that will be used in this research. Table A1 in the appendix gives an overview on the how the dataset is constructed.

4.3.2 Descriptive statistics

In this paper approximately 15% of the individual-year observations correspond to be an informal caregiver of which 27% provide care inside and 78% outside the household with an overlap of 5% reporting both types of care (Table A2). The hours that these individual-year observations are spending on caring are divided into three categories: low intensity, medium intensity and high intensity. Figure 2 shows for each type of informal care the percentage shares of individuals-year observations by the different care intensities. For all types of informal care provision, most individual-year observations provide it at a low intensity. This is the case for 82.1% of the individual-year observations that provide care outside the household. However, only 42.8% of the individual-year observations that provide care inside the household and 40.6% of the individual-year observations that provide both inside and outside the household care report that they are providing care at a low intensity. Furthermore, 40.2% of the individual-year observations that provide care inside the household and both types of care are providing care at a high intensity. However, this is only the case for 6.2% of the individual-year observations that provide care outside the household.

Figure 2. Percentage share of individual-year observation providing informal care by intensity

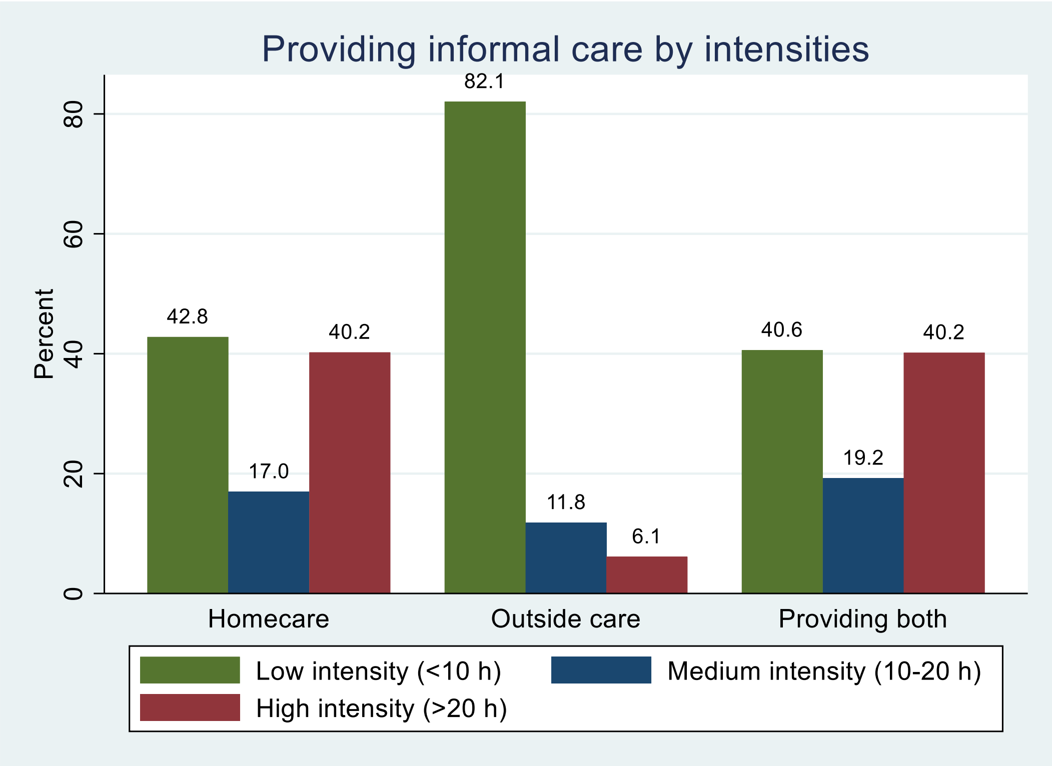


Table 3 provides summary statistics for non-informal caregivers and informal caregivers. The average score on the PSQI is a little bit higher for individuals that do not provide informal care compared to individuals that do provide informal care. Meaning that their sleep quality is better. Figure A3 in the appendix shows the distribution of the PSQI score for both groups. This graph shows that the distribution of the PSQI score for informal caregivers compared to non-informal caregivers is very similar. Furthermore, the average score on the MCS-12 questionnaire and the GHQ-12 questionnaire is also a little bit higher for individuals that do not provide informal care compared to the average score of individuals that do provide informal care. The distribution of the MCS-12 score and the GHQ-12 score for informal caregivers compared to non-informal caregivers are also very similar (figure A3).

Labour force status is divided into: part-time work, full-time work, unemployed and retired. Table 3 shows that both non-informal caregivers and informal caregivers are mostly working part-time or full-time (98% and 91% respectively). However, informal caregivers are working more part-time than full-time compared to non-informal caregivers. Table 3 also shows that a higher percentage of the informal caregiver sample is female compared the sample of non-informal caregivers and also the average age is higher. The average age of an informal caregiver is approximately 49 years and the average age of non-informal caregivers is 44. The variable education is divided into three categories: primary, secondary and tertiary education. We classified “other qualification" and "no qualification" as primary education, "A level etc" and "GCSE etc" as secondary education and "other high" and "degree" as tertiary education. Table 3 shows that the education level and the marital status of both samples are almost the same. For both samples most of the individuals reported that they have secondary or tertiary education and most are being married. The average amount of household income of informal caregivers is a little lower compared to the household income of non-informal caregivers (£3,792 vs £3,919). The number of children living in the household under the age of 16 varies from 0 to 7 across different households. Approximately 58% of the non-informal caregivers and 67% of the informal caregivers sample have no children under the age of 16 living in the household.

Table 3. Summary statistics of non-informal caregivers and informal caregivers

	Non-informal caregiver		**Informal caregiver**	
Variables	**Mean**	**Std. dev.**	**Mean**	**Std. dev.**
Sleep quality	13.353	2.925	12.686	3.167
Mental health (MCS-12)	49.255	9.217	48.282	9.832
Mental health (GHQ-12)	25.355	4.919	24.570	5.308
Full-time work	.701	.458	.601	.490
Part-time work	.278	.448	.375	.484
Unemployed	.002	.042	.001	.030
Retired	.004	.064	.008	.087
Female	.533	.499	.637	.481
Own age	44.375	11.279	48.719	10.110
Single	.257	.437	.204	.403
Married	.606	.489	.639	.480
Partnership	.006	.079	.006	.075
Separated or Divorced	.117	.322	.135	.342
Widowed	.014	.116	.017	.129
Primary education	.081	.273	.096	.295
Secondary education	.397	.489	.450	.498
Tertiary education	.521	.500	.454	.498
Equivalent HH income	3,919.359	2,584.792	3,791.556	2,199.480
Children <16 in HH	.705	.964	.553	.909

Note: This table provides summary statistics for the entire sample. Regional dummies are omitted. Sleep quality is measured with the inverted adjusted PSQI ranging from 0 to 18. The MCS-12 is the mental component score of the SF-12 questionnaire and ranges from 0 to 100. The GHQ-12 score is inversed and has a range from 0 to 36. The higher the sleep quality and mental health scores, the better. The household (HH) income is measured per month.

5. Results

In this section the results will be discussed that are obtained by applying the 4 steps of propensity score matching as discussed in section 3. The results will be discussed for any type of care and for the different intensities of informal caregiving.

5.1. Matching quality

The predicted propensity scores for providing any type of care and for the different intensities are calculated based on the probit estimates presented in table A4 in the appendix. These estimates show that the probability to provide any type of care and to provide different hours of care is significantly associated with some observed characteristics. Figure A5 in the appendix shows the propensity scores for the treated and the untreated group for the broad definition of caregiving and by care intensity. It shows that the propensity scores for all individual-year observations are located in the common support area. Meaning that untreated and treated individual-year observations are comparable based on their propensity scores and no individual-year observations are excluded from the matching procedure. Meaning that for any type of care provision 25,567 untreated individual-year observations are matched to 4,555 treated individual-year observations. For low intensity 26,863 untreated individual-year observations are matched to 3,250 treated individual-year observations and for medium intensity 29,467 untreated individual-year observations to 606 treated individual-year observations. Lastly, for high intensity 29,432 untreated individual-year observations are matched to 690 treated individual-year observations.

Table 4 assesses the quality of the matching by checking whether there is balance in the distribution of the covariates in the matched treatment and control group. For any type of care, before the matching a lot of the covariates are not in balance between the treatment and control group, because the standardized bias for these covariates is higher than 5%. However, after the matching the standardized bias for all the covariates are below 5%, which means that the matching for any type of care is being done successfully. For low intensity of informal care provision there are still a few covariates with a standardized bias above the 5% after the matching. For both medium and high intensity there are 20 covariates that are imbalanced after the matching. Therefore, it is arguable if the matching for the different care intensities is being done successfully.

Table 4. Standardized bias in percentage before and after matching

	Standardized bias in %							
			By intensity					
	Any care		<10 h		10-20 h		>20 h	
	Pre-matching	Post-matching	Pre-matching	Post-matching	Pre-matching	Post-matching	Pre-matching	Post-matching
Lagged PSQI	-17.3	-3.4	-12.3	-4.3	-15.8	**-15.0**	-31.2	**-25.8**
Lagged MCS-12	-6.7	-1.3	-2.0	-0.5	-5.8	**-5.3**	-23.3	**-19.4**
Lagged GHQ-12	-11.3	-2.0	-8.0	-2.3	-7.8	**-7.1**	-22.9	**-19.0**
Full-time work	-21.3	-4.0	-13.7	**-5.3**	-25.9	**-24.7**	-39.6	**-32.4**
Part-time work	20.8	3.8	13.7	**5.3**	24.5	**23.0**	37.4	**30.6**
Unemployed	-2.4	0.2	-2.2	0.0	-5.8	0.0	-0.5	-0.4
Retired	4.7	1.4	3.6	1.4	6.4	**6.6**	5.2	4.3
Female	21.3	3.2	15.7	4.4	26.6	**24.5**	31.6	**25.5**
Own age	40.6	4.0	39.1	**9.1**	39.4	**36.6**	29.8	**23.7**
Single	-12.7	-1.1	-13.9	-2.8	-6.2	**-5.5**	-7.6	**-5.9**
Married	6.8	1.4	9.4	3.3	-0.4	-0.7	-1.1	-1.2
Partnership	-0.7	-0.3	-0.5	-0.4	2.5	2.4	-4.9	-3.8
Separated or Divorced	5.3	-0.4	3.8	-0.7	4.7	4.0	9.6	**7.8**
Widowed	2.7	-0.7	0.3	-1.2	7.9	**8.2**	5.9	**5.0**
Primary education	5.1	0.2	0.8	-1.0	4.9	4.5	19.5	**16.3**
Secondary education	10.7	1.6	9.0	3.4	16.5	**15.5**	7.6	**5.9**
Tertiary education	-13.5	-1.7	-9.4	-2.8	-19.3	**-18.0**	-20.0	**-16.2**
Equivalent HH income	-5.3	-0.8	-0.8	-0.6	-12.1	**-11.3**	-16.8	**-13.4**
Children <16 in HH	-16.2	-3.1	-20.5	**-5.3**	-12.6	**-12.3**	4.6	3.4
Northeast	6.7	1.7	4.3	1.8	9.7	**9.2**	10.3	**8.4**
Northwest	-3.2	-0.3	-4.4	-1.1	4.3	3.9	-3.3	-2.8
Yorkshire and the Humber	-2.0	-0.7	-3.8	-1.4	-4.5	-4.5	8.3	**6.7**
East Midlands	3.3	0.6	3.3	1.2	5.4	4.7	-0.3	-0.5
West Midlands	5.9	1.0	5.1	1.9	6.7	**5.9**	6.0	4.8
East of England	-3.1	-0.5	-0.8	-0.3	-2.2	-2.4	-13.2	**-10.5**
London	-5.7	-0.8	-4.4	-1.2	-7.5	**-6.9**	-7.3	**-5.9**
Southeast	-7.2	-1.0	-2.4	-0.8	-19.9	**-17.7**	-15.6	**-12.2**
Southwest	0.1	-0.6	1.4	0.1	-8.7	**-8.2**	1.9	1.3
Wales	1.6	0.1	1.2	0.4	1.9	1.5	2.4	1.9
Scotland	1.7	0.1	1.5	0.1	0.9	0.3	2.6	2.0
Northern Ireland	6.6	1.2	1.6	0.1	17.5	**17.6**	12.8	**10.4**

Note: For every covariate the standardized bias in percentage is calculated before and after the matching. The threshold is 5%. The covariates of which the distribution between the treatment and control group is very different after the matching have a standardized bias above 5% which are shown in red and bold in the figure.

5.2 Effect of caregiving on sleep quality and mental health

Table 5 shows that all the estimated treatment effects for sleep quality are significant. We find that people who provide any type of informal care score 0.41 lower on the PSQI compared to individuals that do not provide informal care. Meaning that they experience worse sleep quality. The findings also indicate that providing care at a higher intensity is associated with a bigger negative impact on the PSQI. Individuals who provide less than 10 hours of care per week, report a score on the PSQI which is 0.27 points lower compared to individuals who do not provide less than 10 hours of care per week. However, individuals who provide more than 20 hours of care per week, report a score on the PSQI which is 1.27 points lower compared to individuals who do not provide more than 20 hours of care per week.

For both mental health measures, all coefficients depicted in table 5 are significant. Indicating that providing any type of informal care and informal care at the different intensities result in a difference in mental health outcomes between the treatment and control group. Caregivers that provide any type of informal care report a score that is 0.92 points lower on the MCS-12 and 0.58 points lower on the GHQ-12. Indicating that informal caregivers experience worse mental health compared to non-caregivers. Again, individuals who provide care at a higher intensity, report a significantly lower score on both mental health outcomes compared to individuals who provide informal care at a lower intensity. Caring at a low intensity is associated with a 0.40 lower score on the MCS-12 and providing informal care at a high intensity is associated with a 3.19 lower score on the MCS-12. For the GHQ-12, providing informal care at a low intensity is associated with a score that is 0.36 lower and for high intensity this score is 1.80 lower.

Concluding, informal care provision results in worse sleep quality and mental health. The magnitude of this effect increases when the intensity of care provision increases. Meaning, that the intensity of informal care provision has an impact on both the sleep quality and mental health effects of informal caregivers.

Table 5. Average treatment effect of informal caregiving on sleep quality and mental health

ATT		**By intensity**		
Outcome variables	**Any care**	**<10 h**	**10-20 h**	**>20 h**
Sleep quality	-.405*	-.271*	-.647*	-1.275*
	(.051)	(.057)	(.136)	(.130)
Mental health (MCS-12)	-.917*	-.400*	-.948*	-3.188*
	(.159)	(.176)	(.412)	(.420)
Mental health (GHQ-12)	-.578*	-.359*	-.609*	-1.795*
	(.086)	(.100)	(.244)	(.223)

Note: Sleep quality is measured with the inversed PSQI. The MCS-12 is the mental component score of the SF-12 questionnaire and the GHQ-12 is inversed. *T > 1.96 or T < -1.96, standard errors in parentheses.

6. Robustness check

In this section we will check the robustness of our baseline results. The goal of a robustness check is to increase the reliability and validity of our baseline results, presented in the previous section. These results are based on the kernel based matching approach. In this section the results will be explored using a regression adjustment and the nearest neighbour matching approach. To ensure that our baseline results are not dependent on the chosen matching approach. In addition, we will check if there could be reversed causality between the outcome variables.

6.1 Dealing with the remaining covariate imbalance

After assessing the balance in section 5.1 it appeared that there is still some remaining covariate imbalance between the treatment and the control group after the kernel based matching. Suggesting that the matching is not being done successfully and the treatment and control group are not comparable for low, medium and high intensity of caregiving. Resulting in uncertainty about these results. Therefore, we will explore the results using a regression adjustment and the nearest neighbour matching approach. Both deal with the concern of remaining covariate imbalance, but in a different way.

Regression adjustment, also called double-robust method, is used to control for this remaining covariate imbalance between the treatment and control group after the matching (Stuart, 2010). Regression adjustment uses the propensity scores as weights in a regression model in order to control for the residual confounding effect that is caused by the covariates that are imbalanced within the matched sample (Nguyen et al., 2017). Meaning that the covariate imbalance presented in table 6 remains the same, but that regression adjustment will account for this imbalance.

The second approach we used to deal with the concern of remaining covariate imbalance is the one-to-one nearest neighbour matching approach. This approach selects for each treated individual an untreated individual that has a propensity score that is closest to the propensity score of the treated individual. By matching on the nearest neighbour, this approach tries to ensure that there is no covariate imbalance existing after the matching. There are two variants of nearest neighbour matching, i.e. with and without replacement. The first variant means that an untreated individual can be used more than once in order to match to multiple treated individuals. An advantage of this variant is that the average quality of the matching will be higher and the bias will be lower. Especially when the propensity scores of the untreated and treated individuals are very different from each other. However, as shown in figure A5 in the appendix is this not the case with our data. Furthermore, a disadvantage of matching with replacement is that a lot of individuals will be excluded from the analysis when the treatment and control group differ in size. Therefore, this research uses nearest neighbour matching without replacement. This means that an untreated individual can only be used one time in the matching procedure (Caliendo & Kopeinig, 2008).

In the nearest neighbour matching process, the same probit estimates depicted in table A4 and calculated propensity scores are used as in our baseline results. The propensity scores depicted in figure A6 in the appendix show almost complete overlap between treated and untreated individual-year observations included in the matching process for the broad definition of caregiving and the different intensities. The number of included treated individual-year observations for any type of care are 3,326 (from the 4,555), for low intensity 2,654 (from the 3,259), for medium intensity 580 (from the 606) and lastly for high intensity 656 (from the 690). Meaning that some treated individuals-year observations are excluded from the matching procedure due to a lack of good matches. Table A7 in the appendix assesses the standardized bias for every covariate after doing the nearest neighbour matching approach. Unlike the baseline matching using a kernel, the nearest neighbour matching achieves what is intended to decrease the bias. Although, some covariates are still imbalanced, the number is much lower compared to the kernel based matching approach.

6.2 Effect of caregiving on sleep quality and mental health

The results of the robustness checks are compared to the baseline results in order to see if the results are changing when dealing with covariate imbalance (table 6). Overall, the magnitudes we find for any type of caregiving and low intensity caregiving on the PSQI, the MCS-12 and the GHQ-12 barely change when comparing the results of the robustness checks with the baseline results. This is not surprising since there was no or only little covariate imbalance for these types of care after doing the Kernel based matching. All these effects have a significant t-value or are significant at a 10% significance level and most of them even at a 1% significance level.

For medium and high intensity of care the imbalance provided in table 4 was substantial. Therefore, it is expected that the results provided for these groups do change when dealing with the problem of covariate imbalance. For medium intensity of care, table 6 shows that the effect on sleep quality provided by the robustness checks are smaller compared to the baseline result. All effects are significant. However, when comparing the MCS-12 results across the different methods for medium intensity it stands out that the magnitude indicated by the kernel based matching approach is similar to the magnitude indicated by the regression adjustment (-0.95 and -0.86 respectively). Both effects are significant. However, the magnitude on the MCS-12 provided with the nearest neighbour matching approach is clearly lower (-0.35) compared to those methods. This effect is insignificant at all significance levels. It is surprising that the magnitudes between the robustness checks differ that much as both are dealing with covariate imbalance. In addition, it is also surprising that this magnitude estimated by the nearest neighbour matching approach for medium intensity is even lower than for low intensity of care (-0.45). The effects on the GHQ-12 for medium intensity caregivers provided by the robustness checks are smaller compared to the baseline results. This is in line with the effects on the sleep quality of medium intensity caregivers. However, the magnitude provided by the nearest neighbour matching approach is insignificant at every significance level and is again even lower than magnitude for low intensity of care estimated by the nearest neighbour matching approach (-0.378 vs -0.408). Furthermore, for high intensity of care we find smaller magnitudes on the PSQI, the MCS-12 and the GHQ-12 with the robustness checks compared to the baseline results. Therefore, we can conclude that dealing with substantial covariate imbalance matters for the sizes of the magnitudes on sleep quality and mental health.

Although the magnitudes change when dealing with covariate imbalance for the intensity groups which faced the most covariate imbalance, the overall conclusion on sleep quality and mental health remain the same. Meaning that informal care provision is associated with worse sleep quality and mental health of individuals. Furthermore, we find a negative association between the hours of care an individual provides and the sleep quality and mental health of an individual.

6.3 Economic relevance

So far, we only looked at the magnitudes and the significance of the sleep quality and mental health effects of caregivers. However, it is also important to explore the economic relevance of our results. This deals with the question if providing informal care pushes caregivers into significant sleep problems or clinical depressions. Table 7 compares the sleep quality and mental health of the average caregiver (treated) with that of the average individual in the counterfactual group (untreated). It shows that the average untreated individual is already experiencing significant sleep problems, because their PSQI score is lower than the threshold of 14. Therefore, providing informal care will result in reduced sleep quality given the effect sizes. Especially among caregivers that provide care for more than 20 hours per week, as these individuals experience a larger impact. Furthermore, table 7 shows that the average informal caregiver does not have a MCS-12 score below 45.6, which is indicating clinical depression. This means that providing informal care does not push the average informal caregivers into a depression. However, the average caregiver that provides informal care at a high intensity is getting close to this threshold (shown by the orange colour). Lastly, individuals who have a score of lower than 21 on the GHG-12 face distress. Table 7 shows that an average informal caregiver is not experiencing distress. However, again providing informal care for more than 20 hours per week is bringing people closer to the threshold, but not as close as the result found on the MCS-12. Concluding, the sleep quality and mental health of the average informal caregiver is lower compared to the average non-informal caregiver. For the sleep quality of individuals this means that informal caregiving even worsens the sleep quality which implies that the number of individuals experiencing significant sleep problems increases. Especially among caregivers that provide care for more than 20 hours per week. Furthermore, the mental health score of the average caregiver is getting closer to the thresholds compared to non-informal caregiver. This means that the average informal caregiver is not experiencing a clinical depression or distress, but as the mental health score decreases it implies that the number of individuals experiencing these mental health problems increases. Again, this is the most common among high intensity caregivers.

Table 6. Average treatment effect of informal caregiving on sleep quality and mental health across different methods

	Any care			**By intensity**								
				<10 h			**10-20 h**			**>20 h**		
Outcome variable	Kernel matching	Regression adjustment	Nearest neighbour	Kernel matching	Regression adjustment	Nearest neighbour	Kernel matching	Regression adjustment	Nearest neighbour	Kernel matching	Regression adjustment	Nearest neighbour
Sleep quality	-.405*	-.347***	-.348***	-.271*	-.193***	-.178**	-.647*	-.356***	-.460**	-1.275*	-.808***	-.777***
	(.051)	(.044)	(.076)	(.057)	(.048)	(.084)	(.136)	(.118)	(.186)	(.130)	(.111)	(.183)
Mental health (MCS-12)	-.917*	-.869***	-.632***	-.400*	-.421***	-.449*	-.948*	-.865**	-.350	3.188*	-2.301***	-2.140***
	(.159)	(.139)	(.235)	(.176)	(.155)	(.259)	(.412)	(.360)	(.597)	(.420)	(.365)	(.576)
Mental health (GHQ-12)	-.578*	-.529***	-.416***	-.359*	-.314***	-.408***	-.609*	-.409**	-.378	-1.795*	-1.277***	-1.297***
	(.086)	(.077)	(.127)	(.095)	(.086)	(.140)	(.244)	(.205)	(.311)	(.223)	(.200)	(.311)

Note: Sleep quality is measured with the inversed PSQI. The MCS-12 is het mental component score of the SF-12 questionnaire and the GHQ-12 is inversed. For kernel based matching: *T > 1.96 or T < -1.96. For regression adjustment and nearest neighbour matching: *p < 0.1, ** p < 0.05, ***p < 0.01. For both standard errors in parentheses.

Table 7. The scores on the outcome measures across different methods

	Any care						**By intensity**											
							<10 h						**10-20 h**						**>20 h**					
	Kernel matching		Regression adjustment		Nearest neighbour		Kernel matching		Regression adjustment		Nearest neighbour		Kernel matching		Regression adjustment		Nearest neighbour		Kernel matching		Regression adjustment		Nearest neighbour	
Outcome variable	U	T	U	T	U	T	U	T	U	T	U	T	U	T	U	T	U	T	U	T	U	T	U	T
PSQI	**13.09**	**12.69**	**13.03**	**12.69**	**13.11**	**12.76**	**13.14**	**12.87**	**13.06**	**12.87**	**13.08**	**12.91**	**13.25**	**12.60**	**12.95**	**12.60**	**13.06**	**12.60**	**13.18**	**11.91**	**12.72**	**11.91**	**12.69**	**11.91**
MCS-12	49.20	48.28	49.15	48.28	48.96	48.33	49.22	48.82	49.25	48.83	49.27	48.82	49.11	48.16	49.02	48.16	48.47	48.12	49.02	**45.83**	48.13	**45.81**	47.98	**45.84**
GHQ-12	25.15	24.57	25.10	24.57	25.06	24.64	25.17	24.81	25.13	24.81	25.24	24.83	25.22	24.61	25.02	24.61	24.96	24.59	25.17	23.38	24.65	23.38	24.66	23.36

Note: The PSQI is inversed and measures sleep quality. The MCS-12 and the inversed GHQ-12 measure the mental health of individuals. "U" stands for untreated which is the score on the outcome measures for individuals in absence of the treatment. "T" stands for treated which is the score on the outcome measures for individuals with the treatment. The scores which are bold and red of colour are below the threshold and scores which are bold and orange are close to the threshold. A PSQI score below 14 is indicating significant sleep problems, a MCS-12 score below or equal to 45.6 is indicating clinical depression and a GHG-12 score below 21 is indicating distress.

6.4 Testing for reversed causality

In our research we find consistent negative effects of informal caregiving on sleep quality and mental health. Therefore, it could be that poor sleep quality as a result of informal care provision is a mechanism for the poor mental health of informal caregivers. However, it could also be that poor mental health explains the poor sleep quality of informal caregivers. Therefore, regressions are performed in which current mental health is included as explanatory variable in the sleep quality analysis for any type of care and the different intensities. Meaning that we match individuals on the same covariates as before and use the propensity scores as weights in a regression in which current mental health as explanatory variable is added for the outcome variable PSQI. In this regression the same covariates as used in the matching process and a treatment indicator (any care, low intensity, medium intensity or high intensity) are included as control variables.

In this regression we do not use the inversed GHQ-12 as explanatory variable, because one item deals with the question if an individual recently lost much sleep because they were worried. Therefore, GHQ-12 would be highly correlated with the PSQI and as a result will not pick up the individual variation. However, the MCS-12 does not have an item on sleep. Meaning that this is an appropriate mental health measure to use as an explanatory variable in the regression. This variable is included as a dummy, which is a 1 if the MCS-12 score of an individual is 45.6 or lower (threshold for clinical depression) and is a 0 if this is higher than 45.6. So, this dummy indicates poor mental health.

Table 9 provides the results of these regressions. It shows that individuals who are providing any type of informal care and who have a poor mental health, score 1.24 points lower on the PSQI. Informal caregivers who provide informal care at a low or at a medium intensity and have a poor mental health score 1.25 points lower on the PSQI and high intensity caregivers with a poor mental health score 1.24 points lower on the PSQI. All these effects are significant at 1% significance level. Meaning that informal caregivers with worse mental health also experience worse sleep quality. These results could indicate that bad mental health of informal caregivers could result in bad sleep quality of informal caregivers instead of the other way around. Therefore, we can conclude that there is probably reversed causality between the outcome variables sleep quality and mental health.

Table 8. Regression poor mental health on sleep quality

Variables	Sleep quality			
		By intensity		
	Any care	<10 h	10-20 h	>20 h
Poor mental health	-1.244***	-1.250***	-1.249***	-1.236***
	(.043)	(.042)	(.048)	(.052)
Any care	-.274***	-	-	-
	(.045)			
Low intensity	-	-.156***	-	-
		(.049)		
Medium intensity	-	-	-.251*	-
			(.137)	
High intensity	-	-	-	-.699***
				(.133)
Covariates	YES			
Observations	30,122	30,122	30,073	30,122
R-squared	0.327	0.325	0.330	0.331

Note: Poor mental health is a dummy variable which is a 1 if the MCS-12 score is 45.6 or lower and a 0 if otherwise. The other covariates are omitted from the table. *p < 0.1, ** p < 0.05, ***p < 0.01, standard errors in parentheses

7. Discussion and conclusion

The large and increasing share of elderly in developed countries has become an increasing problem for society, because an ageing population is associated with reduced population health and an increasing demand for care. To meet this increasing demand, society relies on informal caregivers. However, several researchers have found negative effects of providing informal care on the mental health of informal caregivers. Sleep quality could be an important determinant for the mental health effects that are found among caregivers. For policymakers, it is important to understand the consequences that informal caregiving has on the sleep quality and mental health of caregivers as the need for LTC services will further increase in the near future. Therefore, this thesis explores the effect of providing informal care on the sleep quality and mental health of caregivers and how these differ by care intensity, using data from the UK Household Longitudinal Study. In order to find causal estimates, propensity score matching is used to deal with the concerns of endogeneity resulting from selection bias.

We find that informal caregiving results in worse sleep quality and that these effects are heterogeneous across the caregiving population rising in increasing intensity of care provided. These results are consistent across robustness checks although an overall decrease in the magnitude of effects for medium and high intensity groups is observed. This is likely explained by the remaining covariate imbalance in our baseline specification as opposed to the analyses conducted as robustness checks addressing these concerns.

Furthermore, we also find that informal caregiving results in worse mental health of individuals. Comparing the mental health results to the sleep quality results we find that the same holds with respect to the increased effects for different intensity groups, the results of the robustness check and the explanation for the decrease in magnitudes between the results of the robustness checks and the baseline specification. A remarkable difference compared to our sleep quality results, is that the magnitudes we find for medium intensity caregivers on both mental health measures with the nearest neighbour matching approach are lower compared to the effects we find for low intensity caregivers by the nearest neighbour matching approach. However, these effects are insignificant at all significance levels. Therefore, all our findings on mental health indicate that changes in care intensity negatively influences the mental health of informal caregivers. This is in line with previous literature (Stöckel & Bom, 2022; Bom et al., 2019; Schmitz & Westphal, 2015) who also found the importance of care intensity on mental health of caregivers. Our findings on high intensity caregiving are in line with the findings of Stöckel and Bom (2022). They also explored the mental health of informal care provision on the caregivers using the UK Household Longitudinal Study. Although they also find similar magnitudes for any type of care, low and medium intensity. What is contradictory to their research is that their findings for these types of care are not significant. An explanation for this difference in significance might be that they exploit the panel structure in their research.

Beforehand, we expected that our findings on sleep quality would be an explanation for our findings on the mental health of informal caregivers. As researchers found that sleep quality is an important determinant for the mental health of individuals. For example, the research of Crivello et al. (2019) found that sleep is needed in order to regulate emotions. This would also be in line with previous studies (Gadie et al., 2017; O'Leary et al., 2017) who found that worse sleep quality negatively impacts mental health of individuals. However, as a robustness check we performed a regression in which a dummy for poor mental health is included as explanatory variable in the sleep quality analysis for any type of care and the different intensities. Our findings indicate that informal caregivers with poor mental health experience significantly worse sleep quality. Therefore, we can conclude that there is probably reversed causality between sleep quality and mental health.

In this research, there are a few limitations and drawbacks that can provide a foundation for further research. The first limitation in this research is the problem of reversed causality previously described. Reversed causality is a problem, as it violates the assumption of strict endogeneity which can lead to inconsistent estimates and drawing the wrong conclusions. Therefore, we cannot conclude that our findings on sleep quality are an explanation for the findings on the mental health of caregivers. An IV can overcome the problem of reversed causality by isolating the exogenous variation in the parameter of interest. An appropriate IV in our research would be a variable that exogenously shifts our sleep quality variable, such as number of children aged 0-5 living in the household. However, information on this variable is not available in wave 1 and using this variable as IV would result in a substantial number of individual-year observations that needed to be excluded. Therefore, we considered this IV not appropriate for our study and due to a lack of other appropriate IVs in our dataset, this research is not able to resolve this problem of reversed causality.

Another limitation in this research is that we run a pooled regression. Meaning that we have 4 waves of data and some people appear in multiple waves. However, we treat them as different individuals and do not follow people overtime. Meaning that serial correlation in the error terms is a concern in this research, because the error terms in the previous period will affect the error terms in this period. Since we are not following them overtime, we cannot deal with time invariant serial correlation. In this research we tried to deal with serial correlation in the outcome variables by conditioning on the lagged outcome variables. However, serial correlation remains a problem here, because individuals can already be an informal caregiver for several years. Meaning that the mental health we condition on of the previous period will be a mix of serial correlation and the impact of caregiving realized a couple of years ago. These remaining problems with serial correlation affect the efficiency of the obtained results, because with serial correlation the standard errors are biased which can result false significant regression coefficients and drawing the wrong conclusions (Bahmani-Oskooee, 1987).

A third limitation is the dependence on self-reported measures in this research. For example, one item of the PSQI is the overall quality of sleep of an individual. If someone needs 8 hours of sleep and sleeps less than 8 hours, this person will probably report a bad sleep quality. So, duration of sleep is important for these people. However, other people need less sleep and the duration will be less important. Also, the mental health outcome measures are self-reported. Di Novi, Jacobs, & Migheli (2015) found that people tend to compare their health to the health of the care receiver and therefore report a higher health score than they actually have. This could suggest that informal caregivers overestimate their mental health. Furthermore, the hours spend on caring is also a self-reported measure. The research of Urwin et al. (2021) found that providers of informal care report on average 10.55 (37%) more hours of care per week compared to care receivers. This could suggest that caregivers overestimate the hours of care they provide. These self-reported measures result in self-report bias. This means that people do not give answers that are completely accurate, because they want to make a good impression or they do not fully know the correct answer. This can, for example, result in a downward bias of the mental health effects that are found in this research. Therefore, depending on self-reported measures could cause a problem with the validity of this research.

Lastly, relying on matching methods is associated with a few limitations. The first thing that could cause a problem are unobserved characteristics that simultaneously affect the assignment into proving informal care and the sleep quality or mental health of an individual. Although we assume that the unobserved characteristics are correlated with the observed ones, there is no certainty that this is the case in this research. Also, the matching quality is a concern in this research, because for both the baseline results and the results of the one-to-one nearest neighbour matching, there was still some covariate imbalance. Both limitations can result in biased and inconsistent estimates, which can result in drawing the wrong conclusions.

Future research could improve some of the methodological issues we faced in order to further develop and confirm or refute our findings. Firstly, future research should try to deal with the problem of reversed causality between sleep quality and mental health by having access to a dataset with a valid IV. Secondly, they should also explore the impact of informal caregiving on the sleep quality and mental health of caregivers using none self-reported data. For example, using objective indices taken form PSG to measure sleep quality, using administrative data such as mental health diagnosis to measure the mental health of caregivers and ask care receivers the number of hours that the caregiver provides care per week. It is interesting to see if these results would be similar to the results found in this research as we only used self-reported measures which is associated with self-report bias. Thirdly, in our research there are still some problems with regard to the matching quality even after applying one-to-one nearest neighbour matching. Future research could try to deal with this by using coarsened exact matching. This method is able to balance the treatment and control group on a selection of the chosen covariates before the matching by temporarily coarsen the selected covariates. Education can, for example, be coarsened into: grade school, high school, college and graduate. After this, exact matching is applied on the coarsened data. Finally, the analysis is being done on the un-coarsened matched data (Blackwell, Iacus, King, & Porro, 2009).

Furthermore, it is also important for future research to focus on the long-term sleep quality and mental health effects of informal caregiving. This could be done by exploiting the panel structure in order to see how the sleep quality and mental health effects of caregivers develop overtime. Future research could also explore the difference in sleep quality and mental health effects between inside and outside household caregivers. It could be that the caregiving patterns of individuals who care inside the household are more interrupted compared to individuals who care outside the household. As individuals who care outside the household could probably better schedule their caregiving hours. Future research can compare the different time patterns of caregivers to see how this affects the sleep quality and mental health of caregivers. Future researchers could also investigate the sleep quality and mental health effects separately for men and women. As differences concerning the social or working life of men and women could result in different effects on sleep quality and mental health.

Concluding, significant negative effects on the sleep quality and mental health of caregivers is found in this research. Our results especially highlight the importance of care intensity on both effects. Furthermore, the results are economically relevant as the observed effects imply an increase in the number of individuals experiencing significant sleep problems and the number of depressed individuals, based on a clinically validated survey screening measure. These findings are important for policymakers as the need for LTC services in the near future will further increase, because of the increasing share of elderly in the population.

8. References

Ahn, J., Kim, N. S., Lee, B. K., Park, J., & Kim, Y. (2019). Relationship of occupational category with risk of physical and mental health problems. Safety and Health at Work, 10(4), 504–511.

Assari, S., Lankarani, M. M., Saleh, D. K., & Ahmadi, K. (2013). Gender modifies the effects of education and income on sleep quality of the patients with coronary artery disease. Int Cardiovasc Res J., 7(4), 141–6.

Baglioni, C., Spiegelhalder, K., Lombardo, C., & Riemann, D. (2010). Sleep and emotions: A focus on insomnia. Sleep Med Rev, 14(4), 227–238.

Bahmani-Oskooee, M. (1987). Demand for International Reserves: Corrections for Serial Correlation and Heteroscedasticity. Applied Economics, 19(5), 609-618.

Baiocchi, M., Cheng, J., & Small, D. S. (2014). Instrumental variable methods for causal inference. Stat Med, 33(13), 2297–2340.

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

Benedetto, U., Head, S. J., Angelini, G. D., & Blackstone, E. H. (2018). Statistical primer: propensity score matching and its alternatives. Eur J Cardiothorac Surg, 53(6), 1112–1117.

Berecki-Gisolf, J., Lucke, J., Hockey, R., & Dobson, A. (2008). Transitions into informal caregiving and out of paid employment of women in their 50s. Social Science & Medicine, 67(1), 122–127.

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). CEM: coarsened exact matching in Stata. The Stata Journal, 9(4), 524–546.

Bloom, D. E., Chatterji, S., Kowal, P., Lloyd-Sherlock, P., McKee, M., Rechel, B., … & Smith, J. P. (2015). Macroeconomic implications of population ageing and selected policy responses. The Lancet, 385, 649–57.

Bom, J., Bakx, P., Schut, F., & Van Doorslaer, E. (2019). The impact of informal caregiving for older adults on the health of various types of caregivers: a systematic review. The Gerontologist, 59(5), e629-e642.

Borg, C., & Hallberg, I. R. (2006). Life satisfaction among informal caregivers in comparison with non-caregivers. Scandinavian Journal of Caring Sciences, 20(4), 427–438.

Brouwer, W. B., van Exel, N. J. A., van den Berg, B., van den Bos, G. A., & Koopmanschap, M. A. (2005). Process utility from providing informal care: the benefit of caring. Health Policy, 74(1), 85-99.

Bubonya, M., Cobb-Clark, D. A., & Wooden, M. (2017). Mental Health and Productivity at Work: Does What You Do Matter? Labour Economics, 46, 150–65.

Buckner, L. & Yeandle, S. (2011). Valuing carers 2011 – Calculating the value of carers' support, London: Carers UK.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. Psychiatry Research, 28(2), 193-213.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys, 22(1), 31-72.

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

Clement-Carbonell, V., Portilla-Tamarit, I., Rubio-Aparicio, M., & Madrid-Valero, J. J. (2021). Sleep Quality, Mental and Physical Health: A Differential Relationship. International Journal of Environmental Research Public Health, 18(2), 460.

Cohen, S. A., Ahmed, N., Brown, M. J., Meucci, M. R., & Greaney, M. L. (2022). Rural-urban differences in informal caregiving and health-related quality of life. The Journal of Rural Health, 38(2), 442-456.

Colombo, F., Llena-Nozal, A., Mercier, J., & Tjadens, F. (2011). Help wanted? Providing and paying for long-term care. OECD Health Policy Studies.

Comas-Herrera, A., 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.

Crivello, A., Barsocchi, P., Girolami, M., & Palumbo, F. (2019). The Meaning of Sleep Quality: A Survey of Available Technologies. IEEE Access, 7, 167374–167390.

Cunnington, D., Junge, M. F., & Fernando, A. T. (2013). Insomnia: prevalence, consequences and effective treatment. Med J Aust, 199(8), S36–40.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. Review of Economics and statistics, 84(1), 151–161.

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.

Dooghe, G. (1992). Informal caregiving of elderly people: a European review. Aging and Society, 12, 369–380.

Fernandez, 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.

Friedman, E. M., Love, G. D., Rosenkranz, M. A., Urry, H. L., Davidson, R. J., Singer, B. H., & Ryff, C. D. (2007). Socioeconomic status predicts objective and subjective sleep quality in aging women. Psychosomatic Medicine, 69, 682-691.

Gadie, A., Shafto, M., Leng, Y., & Kievit, R. A. (2017). How are age-related differences in sleep quality associated with health outcomes? An epidemiological investigation in a UK cohort of 2406 adults. BMJ Open, 7(7), e014920.

Gao, C., Chapagain, N. Y. & Scullin, M. K. (2019) Sleep duration and sleep quality in caregivers of patients with dementia: A systematic review and meta-analysis. JAMA Network Open, 2(8), e199891.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L.B. & Vermeersch, C.M.J. (2011). Impact Evaluation in Practice. Washington DC: World Bank.

Hale L. (2005). Who has time to sleep? J Public Health (Oxf), 27(2), 205–211.

Haseli-Mashhadi, N., Dadd, T., Pan, A., Yu, Z., Lin, X., & Fransco, O. H. (2009). Sleep quality in middle-aged and elderly Chinese: distribution, association factors and associations with cardio-metabolic risk factors. BMC Public Health, 9, 130.

He, D. & McHenry, P. (2016). Does formal employment reduce informal caregiving? Health Economics, 25, 829–843.

Hirst, M. (2005). Carer distress: A prospective, population-based study. Social Science & Medicine, 61, 697–708.

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

Imran, N., Tariq, K. F., Pervez, M. I., Jawaid, M., & Haider, I. I. (2016). Medical students' stress, psychological morbidity, and coping strategies: a cross-sectional study from Pakistan. Acad Psychiatry, 40(1), 92–6.

Jalan, J. & Ravallion, M. (2003). Estimating the benefit incidence of an antipoverty program by propensity-score matching. Journal of Business & Economic Statistics, 21(1), 19–30.

Jokela, M., Batty, G. D., & Kivimaki, M. (2013). Ageing and the prevalence and treatment of mental health problems. Psychological Medicine, 43(10), 2037–2045.

Jones, P. B. (2013). Adult mental health disorders and their age at onset. Br J Psychiatry, 202, s5–s10.

Kaschowitz, J., & Brandt, M. (2017). Health effects of informal caregiving across Europe: A longitudinal approach. Social Science & Medicine, 173, 72–80.

Kaye, H. S., Harrington, C., & LaPlante, M. P. (2010). Long-term care: who gets it, who provides it, who pays, and how much? Health Aff (Millwood), 29(1), 11- 21.

Kotronoulas, G., Wengstrom, Y., & Kearney, N. (2013). Sleep patterns and sleep-impairing factors of persons providing informal care for people with cancer: a critical review of the literature. Cancer Nurs, 36(1), E1–15.

Kovess-Masfety, V., Alonso, J., De Graaf, R., & Demyttenaere, K. (2005). A European approach to rural–urban differences in mental health: the ESEMeD 2000 comparative study. The Canadian Journal of Psychiatry, 50(14), 926–936.

Krystal, A. D & Edinger J. D. (2008). Measuring sleep quality. Sleep Medicine, 9(1), s10-s17.

Lilly, M., 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 Quarterly, 95(4), 641–690.

Llena-Nozal, A. (2009). The effect of work status and working conditions on mental health in four OECD countries. National Institute Economic Review, 209(1), 72–87.

Lu, J., & Liu, K. (2021). Sleep quality and mental health of the elderly in China: evidence from longitudinal data. China Population and Development Studies, 5, 378-393.

Lu, L., Wang, S. B., Rao, W., Zhang, Q., Ungvari, G. S., Ng, C. H., ... & Xiang, Y. T. (2019). The prevalence of sleep disturbances and sleep quality in older Chinese adults: a comprehensive meta-analysis. Behavioral Sleep Medicine, 17(6), 683-697.

McCabe, C., Thomas, K., Brazier, J., & Coleman, P. (1996). Measuring the mental health status of a population: A comparison of the ghq–12 and the sf–36 (mhi–5). The British Journal of Psychiatry, 169(4), 517–521.

Meltzer, L. J., & Montgomery-Downs, H. E. (2011). Sleep in the family. Pediatric Clinics of North America, 58, 765–774.

Meng, A. (2012). Informal home care and labor force participation of household members. Empirical Economics, 44(2), 959–979.

Mentzakis, E., McNamee, P., & Ryan, M. (2009). Who cares and how much: Exploring the determinants of co-residential informal care. Review of Economics of the Household, 7(3), 283–303.

Myllyntausta, S., & Stenholm, S. (2018). Sleep before and after retirement. Current Sleep Medicine Reports, 4(4), 278–283.

Navaie-Waliser, M., Spriggs, A., & Feldman, P. H. (2002). Informal caregiving. Differential experiences by gender. Medical Care, 40(12), 1249–1259.

Nguyen, T. L., Collins, G. S., Spence, J., Devereaux, P. J., Daurès, J. P., Landais, P., & Le Manach, Y. (2017). Comparison of the ability of double-robust estimators to correct bias in propensity score matching analysis. A Monte Carlo simulation study. Pharmacoepidemiology and Drug Safety, 26(12), 1513-1519.

O'Donnell, D., Silva, E. J., Munch, M., Ronda, J. M., Wang, W., & Duffy, J. F. (2009). Comparison of subjective and objective assessments of sleep in healthy older subjects without sleep complaints. Journal of Sleep Research, 18, 254–263.

OECD (2017), "Demographic trends", in Health at a Glance 2017: OECD Indicators, OECD Publishing, Paris, https://doi.org/10.1787/health_glance-2017-73-en

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

O'Leary, K., Bylsma, L. M., & Rottenberg, J. (2017). Why might poor sleep quality lead to depression? A role for emotion regulation. Cognition & Emotion, 31, 1698–1706.

Ong, J. C., Carde, N. B., Gross, J. J., & Manber, R. (2011). A two-dimensional approach to assessing affective states in good and poor sleepers. Journal of Sleep Research, 20(4), 606e610

Palner, J., & Mittelmark, M. B. (2002). Differences between married and unmarried men and women in the relationship between perceived physical health and perceived mental health. Norsk Epidemiologi, 12, 55–61.

Pilcher, J. J., & Ott, E. S. (1998). The relationships between sleep and measures of health and well-being in college students: A repeated measures approach. Journal of Behavioral Medicine, 23(4), 170–178.

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.

Rosenfield, S., & Smith, D. (2010). Gender and mental health: do men and women have different amounts or types of problems? In T. Scheid, & T. Brown (Eds.), A handbook for the study of mental health: Social contexts, theories, and systems (pp. 256e267) New York: Cambridge University Press.

Sareen, J., Afifi, T. O., McMillan, K. A., & Asmundson, G. J. (2011). Relationship between household income and mental disorders: findings from a population-based longitudinal study. Arch Gen Psychiatry, 68(4), 419–427.

Schmitz, H. & Westphal, M. (2015). Short-and medium-term effects of informal care provision on female caregivers' health. J. Health Econ, 42, 174–185.

Seow, L. S. E., Tan, X. W., Chong, S. A., Vaingankar, J. A., Abdin, E., Shafie, S., … & Subramaniam, M. (2020). Independent and combined associations of sleep duration and sleep quality with common physical and mental disorders: results from a multi-ethnic population-based study. PLoS One, 15(7), e0235816.

Stöckel, J., & Bom, J. (2022). Revisiting longer-term health effects of informal caregiving: Evidence from the UK. The journal of economics of ageing, 21, 100343.

Stroka, M. A. (2014). The Mental and Physical Burden of Caregiving Evidence from Administrative Data. Ruhr Economic Papers. Retrieved from https://www.econstor.eu/bitstream/10419/93178/1/780314565.pdf

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. Statistical Science, 25(1), 1–21.

Triantafillou, J., Naiditch, M., Repkova, K., Stiehr, K., Carretero, S., Emilsson, T., ... & Vlantoni, D. (2010). Informal care in the long-term care system European overview paper.

Tur-Sinai, A., Teti, A., Rommel, A., Hlebec, V., Lamura, G. (2020). How many older informal caregivers are there in Europe? Comparison of estimates of their prevalence from three European surveys. International Journal of Environmental Research and Public health, 17, 9531.

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 challenges of measuring informal care time: A review of the literature. Pharmacoeconomics, 39(11), 1209–1223.

Vilagut, G., Forero, C. G., Pinto-Meza, A., Haro, J. M., De Graaf, R., Bruffaerts, R., ... & ESEMeD Investigators. (2013). The mental component of the short-form 12 health survey (SF-12) as a measure of depressive disorders in the general population: results with three alternative scoring methods. Value in Health, 16(4), 564-573.

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. Medical Care, 34(3), 220-233.

Weinberg, M. K., Noble, J. M., & Hammond, T. G. (2016). Sleep well feel well: An investigation into the protective value of sleep quality on subjective well-being. Australian Journal of Psychology, 68, 91–97.

Wing, C., Simon, K. & Bello-Gomez, R. A. (2018). Designing diference in diference studies: best practices for public health policy research. Annu. Rev. Public Health, 39, 453–469.

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 Economics, 26(S2), 127–138.

9. Appendix

Table A1. Data conditioning

Description	**Observations (individuals)**
Merged data wave 1, 4, 7 and 10	173,508 (78,312)
<u>Dependent variable:</u>	
Excluding observations of individuals when: - They do not provide information on mental health or sleep quality - The question is inapplicable - They answer: don't know	133,336 (63,471)
<u>Independent variables:</u>	
Excluding observations of individuals when: - They do not provide information on if they provide informal care - They answer: don't know	133,223 (63,443)
Excluding observations of individuals when: - They do not provide information on if they provide informal care inside HH - They answer: don't know	133,205 (63,434)
Excluding observations of individuals when: - They do not provide information about the number of hours they spend at providing informal care - They answer: don't know - The information that they provide is unclear	132,250 (63,258)
<u>Covariates:</u>	
Excluding observations of individuals when: - They do not provide information on the specific covariate - The question is inapplicable - They answer: don't know	63,368 (33,264)
<u>Lagged outcome variables:</u>	
Excluding observations of individuals if they do not provide information in the previous period	30,122 (17,548)

Note: Respondents for who the question on providing informal care inside the household (HH) and the hours spend on providing informal care was inapplicable got a 0. For the other variables the observations were excluded if the question was inapplicable.

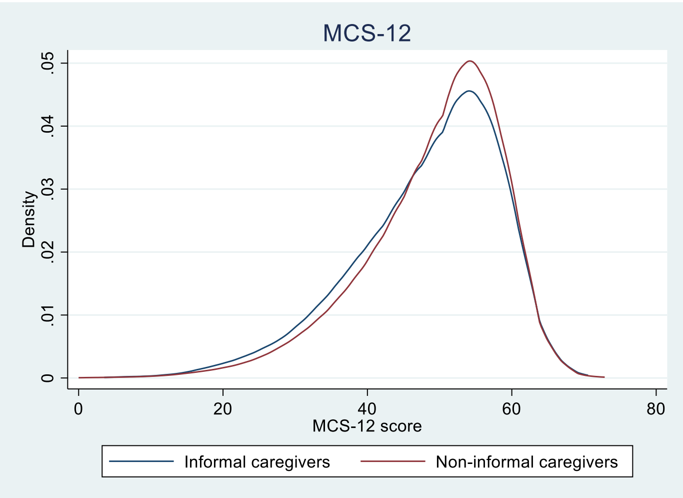
Table A2. Summary statistics for the different types of informal care

Variable	Obs	Mean	Std. dev.
Any type of care	4,555	.151	.358
Care inside HH	1,206	.265	.441
Care outside HH	3,583	.787	.410
Providing both types	234	.051	.221

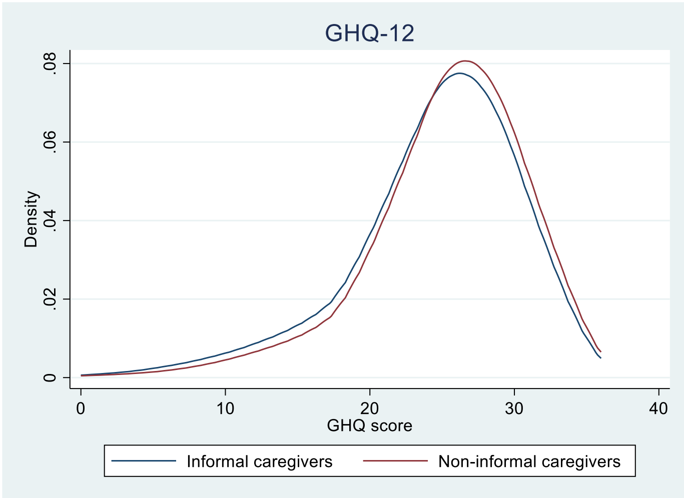
Note: Any type of care represents the share of caregivers of the entire sample. Care inside and outside the household and providing both represents the share of the caregiver's sample.

Figure A3. Distribution outcome measures of entire sample

A) MCS-12 by caregiving status



B) Inversed GHQ-12 by caregiving status



C) Inversed PSQI by caregiving status

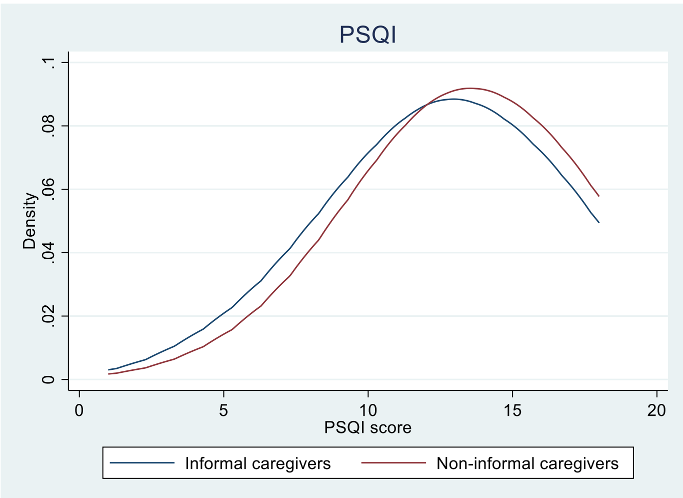


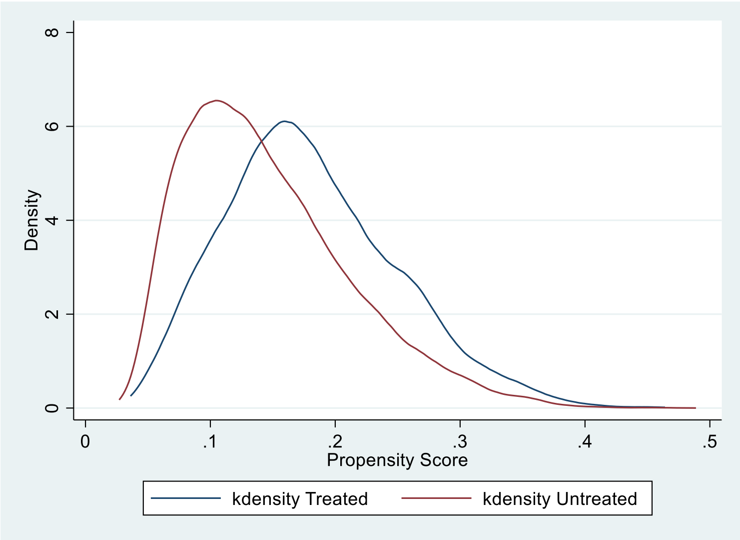
Table A4. Probit estimates

		By intensity		
	Any care	**<10 h**	**10-20 h**	**>20 h**
Variables	Coefficient	Coefficient	Coefficient	Coefficient
Lagged PSQI	-.0224***	-.0145***	-.0158**	-.0305***
	(.0033)	(.0036)	(.0061)	(.0057)
Lagged MCS-12	-.0013	.0012	-.0016	-.0067***
	(.0014)	(.0016)	(.0027)	(.0025)
Lagged GHQ-12	-.0042	-.0056*	.0005	-.0006
	(.0027)	(.0029)	(.0050)	(.0046)
Full-time work	-.0182	.0601	-.0815	-.1868
	(.0728)	(.0817)	(.1289)	(.1175)
Part-time Work	.0880	.1056	.0084	.0032
	(.0733)	(.0823)	(.1294)	(.1176)
Unemployed	-.3583	-.2332	Omitted	-.1823
	(.2790)	(.3039)		(.4448)
Retired	.0493	.0202	.0618	.0582
	(.1368)	(.1512)	(.2293)	(.2241)
Female	.1940***	.1488***	.1800***	.1787***
	(.0203)	(.0220)	(.0393)	(.0387)
Own age	.0199***	.0170***	.0160***	.0133***
	(.0010)	(.0011)	(.0020)	(.0019)
Single	.2367***	.2519***	.0854	.1426
	(.0759)	(.0858)	(.1255)	(.1284)
Married	.2273***	.2924***	-.0026	.0541
	(.0729)	(.0826)	(.1195)	(.1230)
Partnership	.2030	.2403	.1765	-.1407
	(.1371)	(.1510)	(.2308)	(.3023)
Separated or Divorced	.1296*	.1790**	-.0546	.0579
	(.0756)	(.0855)	(.1251)	(.1274)
Primary education	-.0346	-.1100***	-.0147	.2023***
	(.0340)	(.0378)	(.0637)	(.0558)
Secondary education	.1168***	.0888***	.1254***	.0874**
	(.0194)	(.0211)	(.0369)	(.0360)
Equivalent HH income (in 1000 pounds)	-.0059	-.0019	-0.0131	.0130
	(.0041)	(.0042)	(.0085)	(.0088)
Children <16 in HH	-.0212*	-.0570***	.0163	.0871***
	(.0111)	(.0124)	(.0213)	(.0193)
Regional Dummies	YES			

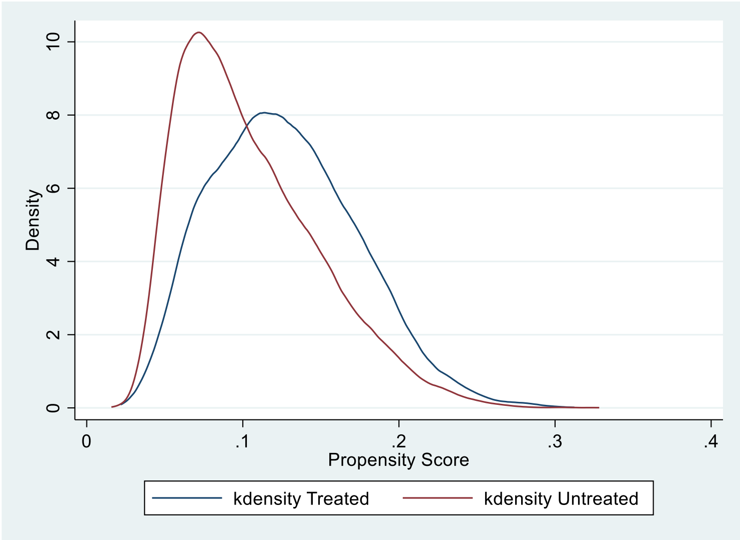
Note: These are the coefficients used to calculate the propensity scores. *p < 0.1, ** p < 0.05, ***p < 0.01, standard errors in parentheses. Widowed and tertiary education are being used as reference category. The variable unemployed is being dropped for medium intensity, because there is no variation in this variable between the treatment and control group.

Figure A5. Graphical representation of propensity scores treated and untreated group

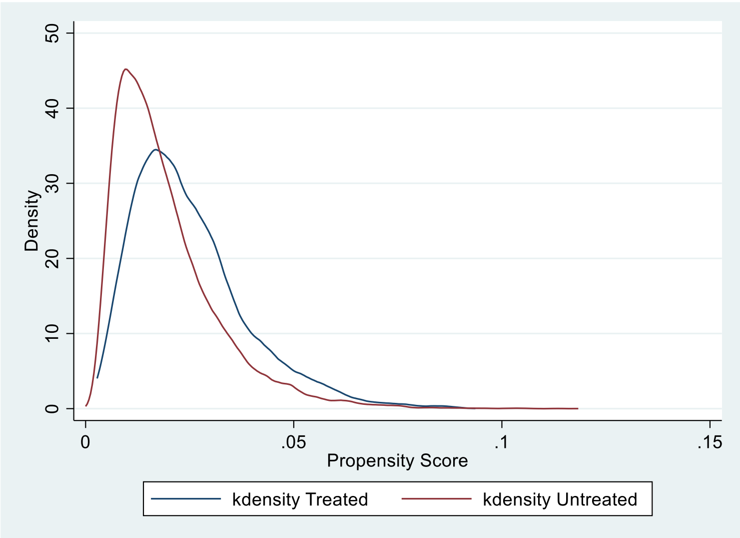
A) Any Care



B) Low Intensity Care (<10h)



C) Medium Intensity Care (10-20h)



D) High Intensity Care (>20h)

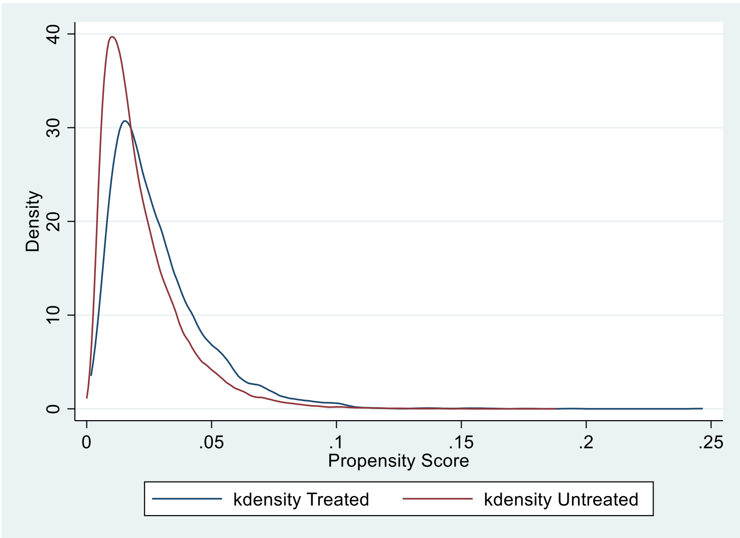
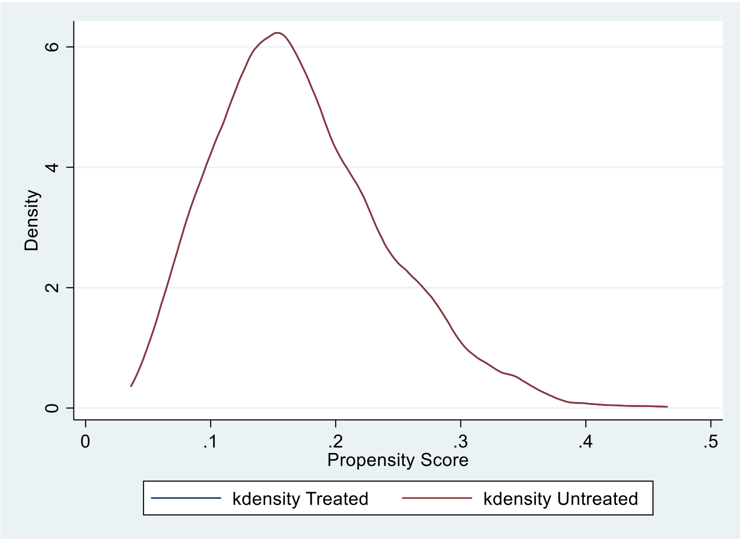
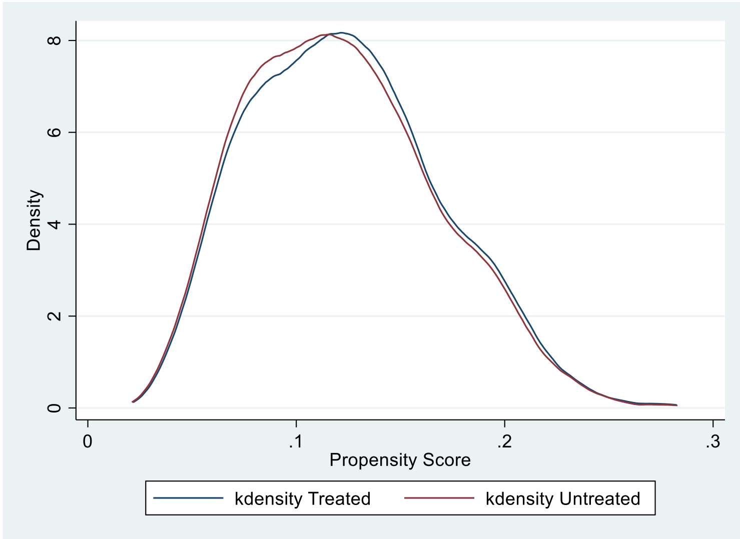


Figure A6. Graphical representation of propensity scores treated and untreated group

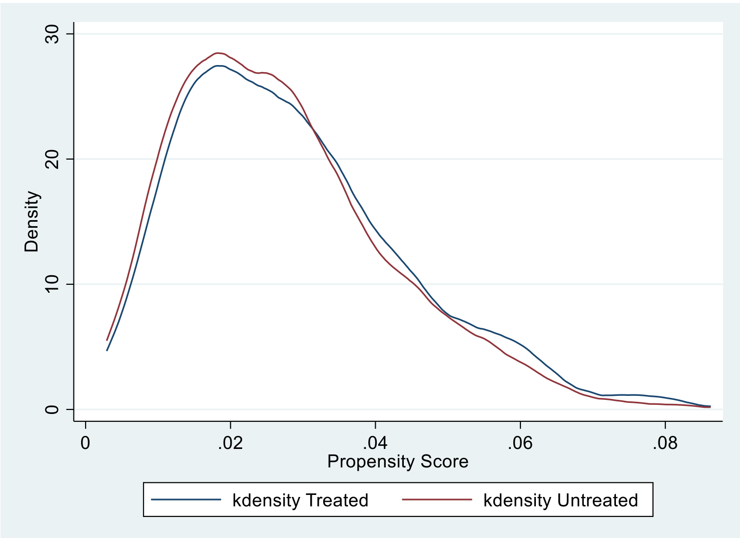
A) Any Care



B) Low Intensity Care (<10h)



C) Medium Intensity Care (10-20h)



D) High Intensity Care (>20h)

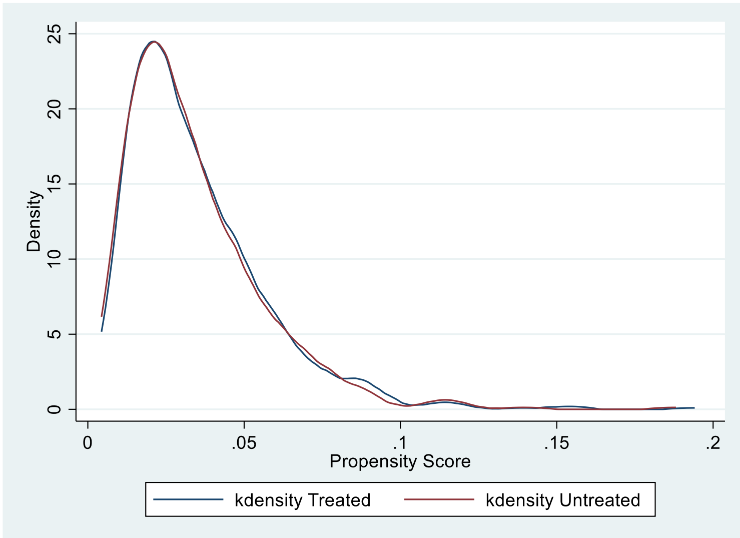


Table A7. Standardized bias in percentage before and after matching

	Standardized bias in %			
		By intensity		
	Any care	**<10 h**	**10-20 h**	**>20 h**
Variables	Pre- and post matching	Pre- and post matching	Pre- and post matching	Pre- and post matching
Lagged PSQI	-0.2	1.2	2.3	-1.8
Lagged MCS-12	2.0	0.4	3.1	-2.3
Lagged GHQ-12	0.8	1.2	2.1	-2.1
Full-time work	-1.9	-2.3	-3.5	4.9
Part-time work	1.8	1.9	4.2	-2.1
Unemployed	2.0	2.7	0.0	-3.2
Retired	-2.0	3.3	-3.2	**-6.7**
Female	-2.2	0.4	-0.4	-2.0
Own age	1.7	-2.0	-1.5	-2.8
Single	0.1	-0.5	**-6.9**	-1.5
Married	1.2	4.3	1.4	-0.6
Partnership	-1.1	2.1	-1.8	**5.5**
Separated or Divorced	-1.8	**-5.0**	**6.9**	3.0
Widowed	0.5	-2.4	1.1	-2.0
Primary education	3.8	-1.9	-2.9	0.4
Secondary education	0.0	1.7	-1.0	**-7.3**
Tertiary education	-2.2	-0.6	2.8	**7.2**
Equivalent HH income	1.2	0.7	-1.5	-0.2
Children <16 in HH	0.7	1.5	4.7	1.8
Northeast	2.4	0.0	-4.9	2.6
Northwest	2.2	-1.0	1.6	-0.5
Yorkshire and the Humber	-2.4	-0.1	2.8	**5.2**
East Midlands	-0.2	1.4	4.3	-2.8
West Midlands	-0.6	1.4	1.7	-1.0
East of England	2.1	2.8	-1.2	-4.8
London	0.6	0.3	**-5.0**	**-5.5**
Southeast	1.4	-2.8	-3.8	4.6
Southwest	0.4	3.4	-4.7	-1.5
Wales	-1.6	-2.0	-2.1	-4.3
Scotland	-1.8	-0.4	**7.1**	-1.6
Northern Ireland	-2.9	-3.1	2.9	**9.6**

Note: For every covariate the standardized bias in percentage is calculated. The threshold is 5%. The covariates of which the distribution between the treatment and control group is very different after the matching have a standardized bias above 5% which are presented in red and bold in the figure. There are now no differences in the standardized bias before and after the matching, because individuals now are matched directly on the propensity scores instead of re-weighting untreated individuals based on the propensity scores. Meaning that the standardized bias now captures no difference between the treated and untreated group