The effect of providing informal care on caregiver's health

Master Thesis
Paul de Zwart

Economics & Business
Health Economics

Erasmus School of Economics
Erasmus University Rotterdam
24-08-2015

The effect of providing informal care on caregiver's health

This thesis has been written under the supervision of
Pieter Bakx
Eddy van Doorslaer
Claudine de Meijer

Abstract

In this study, the causal impact of informal care provision to the partner on caregiver's health is estimated, using a propensity score matching technique to account for the bi-directionality of the relation between providing informal care and caregiver's health. Data from the Survey of Health, Ageing and Retirement in Europe (SHARE) are used to assess the immediate effect of caregiving on health, as well as longer-term effects up to seven years after the care provision. The results suggest that mental health is significantly negatively affected by caregiving, but physical health is not. No evidence is obtained for the persistence of the effect of caregiving on mental health after a time period of 4 or 7 years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Currently, population ageing is a major issue in many European countries. Next to problems regarding the provision of pension money, ageing leads to an increasing demand for long-term care. Informal caregiving by family or friends is often regarded as an important instrument to meet this demand, in particular due to its perceived low cost since informal caregivers are not paid from the formal health care budget. In the Netherlands, for example, the government is aiming to increase its reliance on informal caregivers as opposed to formal caregivers [de Klerk et al., 2014]. However, providing informal care can be mentally and physically challenging [Schmitz and Westphal, 2013]. This has lead to the thought that informal caregiving might be causally related to adverse health effects of the caregiver, which is costly from the view of both the caregiver and care recipient as well as the state. As establishing well-justified policy requires a thorough understanding of all relevant costs, it should be known whether or not these health effects are statistically significant and economically relevant.

In this study, I aim to answer the following research question:

> ***"What is the causal impact of providing informal care to one's partner on the mental and physical health of the caregiver?"***

Not many studies have focused on this causal impact of care-providing on caregiver's health before. Of those studies that do focus on this effect, the one by Coe and Van Houtven [2009] is one of the most influential. In their paper, they use an instrumental variable approach to estimate the causal impact of caregiving on health. This thesis adds on their paper by using another methodology of estimating causal relations between variables (statistical matching), another care recipient (the partner instead of the parent) and another data set, focusing on European countries.

Chapter 2

Literature review

The literature review of this thesis is structured as follows: in the first section, the most important theoretical and empirical considerations relating to studying the stated research question are discussed. The second section deals with empirical findings regarding this question, as obtained by other authors. The reader is referred to section 6.3 for a comparison of these findings to those obtained in this study. Relevant literature concerning the methodology or data used in this thesis will be discussed in chapter 3 (Methodology) and 4 (Variable measurement and data).

2.1 Theoretical and empirical considerations

It is not hard to imagine why the activity of providing informal care could have an effect on physical and mental health. First of all, informal caregiving often involves physical effort which might lead directly to negative physical health effects [Do et al., 2013]. Secondly, there is scientific consensus that caring might be stressful and can contribute to psychiatric morbidity [Schulz and Beach, 1999]. Some authors also postulate a reciprocal relationship between mental and physical health, where psychological distress induces physiological effects and vice versa [Pinquart and Sörensen, 2007]. The theoretical possibility of a causal effect of caregiving on the caregiver's health is taken for granted by most authors. Most of the literature on this subject focuses on empirically establishing the significance and magnitude of this causal effect.

In order to do so, two main problems have to be overcome by any author on this subject. First of all, a proper outcome measure should be constructed that captures the multidimensionality of health. As a result of this multidimensionality, multiple health measures are used in the literature, leading to difficulties in directly comparing study results. Some authors focus on physical health (for example Do et al. [2013]), others include mental health as well (for example Coe and Van Houtven [2009]).

The second main problem is a more econometric one; theoretically, it is possible (or even likely) that the relation between the providing of informal care and the caregiver's health is bi-directional. The decision to provide informal care is not made randomly; it is dependent on, among other things, the ability of the potential caregiver to provide care, suggesting that those in good health are most likely to provide informal care to others. This decision to provide this care is, according to economic theory, also dependent on the opportunity cost of time relative to other caregiver candidates [Do et al., 2013]. This would suggest that those with the lowest socio-economic status, associated with lower health, are most likely to provide informal care. Thus, although the direction of the net effect is not clear on theoretical grounds due to competing theoretical suggestions, the decision to provide informal care is thought to be influenced by one's health.

Because of this endogeneity of caregiving, estimating a regression equation without controlling for the non-randomness of the sample to determine the association between caregiving and health is not a convincing methodology when a causal effect is to be determined. According to Barer and Johnson [1990], the self-selected samples, resulting from a failure to control for the mentioned non-randomness in the samples, were overrepresented in the caregiving literature at that time. This still seems to hold in more recent times; in their meta-analysis integrating the results of 176 studies on the relation between caregiving and physical health, Pinquart and Sörensen [2007] draw conclusions only about associations between these variables, not about a causal relationship.

2.2 Empirical findings

Some relatively recent studies that do address the endogeneity are using an instrumental variable approach to estimate a causal effect between caregiving and caregiver's health. Coe and Van Houtven [2009], who study the health effects on adult children of providing care to their mother, note that not only the decision to start providing informal care, but also the decision to stop providing this care are dependent on one's own health. They use the death of the care recipient as an instrumental variable for the selection out of providing care, while sibling and family characteristics are used for the selection into caregiving. Interestingly, they only find evidence for selection out of caregiving. They discuss their results separately for different subgroups based on sex and marital status, and find significant negative health effects as a result of caregiving activity for all subgroups except single women. Most of these health effects are related to depressive symptoms.

Do et al. [2013] also use an instrumental variable approach to estimate the health effect of caregiving, where they focus on physical health and on female caregivers in South Korea. This country is of special interest, as there are strong traditions of caring for the elderly parents (or parents-in-law), in particular by the daughter-in-law married to the eldest son. They use the limitations in activities of daily life (ADL) of the parents-in-law as instrumental variables, since they presumably have a direct impact on the decision to provide care but not on the daughter-in-law's health. Negative health effects of caregiving are found for all used health measures, such as pain affecting daily activities and a dummy variable representing fair to poor self-rated health.

Two other studies (Van Houtven et al. [2005] and Schmitz and Stroka [2013]) evaluate the relation between (the intensity of the) providing of informal care and the caregiver's drug utilization, which can be regarded as a measure of health. This choice of outcome variable is intriguing, since increased drug utilization due to caregiving can be directly related to increased health care costs by the caregiver. These costs can be compared to the presumed health care savings due to lower formal care demand by the care recipient. Thus, the economic significance of the results can be more easily interpreted than using the approach of this study, in which negative health effects would have to be balanced to health care savings in a non-trivial way in order to conclude about economic significance. However, since costs due to increased drug utilization form only a part of the total costs of impaired health due to caregiving, the net economic benefit of caregiving is still overestimated when these are the only costs taken into account.

Although Van Houtven et al. [2005] use an ordinary least squares regression to evaluate the impact of informal care intensity on caregiver's drug utilization, they test for endogeneity by performing an additional analysis where they use measures of patient functioning as instrumental variables. The care recipients in this study are limited to elderly U.S. veterans with dementia, while no distinction is made between different types of caregivers. The authors find that an increase in the providing of informal care per day by 10% is associated with a 0.71% increase in drug utilization, which is statistically significant but not economically very relevant when these costs are compared to related health care savings.

Schmitz and Stroka [2013] also study the effect of providing informal care on drug utilization. They especially address the dependence of this effect on the working habits of the care provider. In their data set, obtained from a large German sickness fund, they were not able to find both valid and strong instruments. Instead, they rely on fixed-effects methods, accounting for time-invariant but not for time-varying unobserved heterogeneity. The authors do not focus on any particular type of caregiver. They find that those individuals who work full-time and provide informal care significantly consume more drugs used for mental health (antidepressant drugs and tranquilizers) than those who work without providing care and those who provide care but work only part-time. Such a relation is not found for drugs acting on physical health. These results are roughly in line with those of Coe and Van Houtven [2009].

Three further articles discuss the influence of caring on well-being, of which health can be considered to be a part. These studies use a self-reported measure of happiness/life satisfaction as a proxy for well-being. Bobinac et al. [2010] recognize that both a direct caregiving effect (the effect of caring for the care recipient) and a family affect (the effect of caring about the care recipient) are responsible for impaired well-being of the caregiver when care recipient's health declines. They find that both effects are comparable in magnitude. In relation to this thesis, the most important finding is that the caregiving effect remains significant even after controlling for the family effect. Van den Berg and Ferrer-i Carbonell [2007] attempt to find a monetary value for informal caregiving by regressing the well-being of the caregiver on both income and hours of provided informal care. They find a statistically significant negative health effect of caregiving; this effect is largest when the caregiver and care recipient are family-related. The estimate a marginal cost to the caregiver of an hour of informal care to be around 7-10 euro. In contrast to the two studies just mentioned, Leigh [2010] does not find a significant impact of caregiving on well-being in his panel data analysis using fixed effects. A major drawback of the last three articles is that reverse causality is not addressed in these studies.

The study presented in this thesis mainly adds to the studies described in this chapter by using another methodology (statistical matching instead of instrumental variables or other methods) to control for self-selection in the sample. Moreover, this study focuses on the partner as the care recipient, in contrast to most other studies in this field (who either focus on the parents or do not focus on a particular care recipient).

Chapter 3

Methodology

This chapter gives an elaborate description of the methodology used in this study. Mainly, the matching method is introduced, together with its assumptions. The 4-step procedure by Stuart [2010] for performing a matching study is presented in detail and applied to the research question studied in this thesis. Finally, the methodology regarding sample stratification and the exploitation of the panel nature of the data is described.

3.1 Introduction on statistical matching

As mentioned in section 2.1, the decision to provide informal care is not made randomly, but possibly dependent on health. Previous research has primarily used an instrumental variable approach to determine a causal effect of caregiving on health. This thesis adds to the current literature by (among other additions) using another, less widely-used methodology based on statistical matching; for every individual providing informal care (a 'treated' individual), an individual (or a set of individuals, each with their own weight) not providing informal care (the 'non-treated', control individual(s)) with similar observable characteristics is sought. The aim is to find a treated and a control group of individuals with similar covariate distributions. These two groups together then form a reduced sample which is thought to be artificially randomized. Now, a regression analysis can be performed to establish the effect of caregiving on health within this newly constructed sample.

One major advantage of statistical matching as opposed to other methods of causal inference is that by using this method, the researcher will have a better sense of the extent of overlap between the treatment and control group with respect to their covariate distributions [Dehejia and Wahba, 2002]. Even though regression models have been shown to perform inadequately when the overlap is insufficient, actually checking this overlap is not usually done [Stuart, 2010]. In contrast, checking this overlap is part of the standard diagnostics of matching methods. In principle, the matching can be done multiple times; the matched samples with the most similar covariate distribution are then chosen as the final ones. This can be compared to a randomized experiment in which a particular randomization is rejected if the treatment and control groups turn out to be dissimilar with respect to the observable covariates [Stuart, 2010].

3.2 Main assumptions

Two main assumptions have to hold for the matching method to yield an unbiased estimate of the treatment effect [Rosenbaum and Rubin, 1983]. The first is called the Stable Unit-Treatment Value Assumption (SUTVA). Formally, let t represent a certain kind of caregiving activity (treatment), such that $t \in \{$*not caregiving* (0), *caregiving* (1)$\}$. The ith individual under study has a certain health outcome h_{ti} depending on caregiving activity t. The causal effect of caregiving on health for a particular individual i is basically a comparison of h_{0i} and h_{1i}, which cannot be evaluated directly as only one of these two can be measured. In the context of this thesis, the SUTVA states that there is a unique health outcome h_{ti} (for individual i and caregiving activity t) that does not depend on treatment assignment (caregiving activity) of another individual j. In the case of caregiving to partners, the health outcome of one partner is certainly dependent on the caregiving activity (with respect to their partner) of the other partner. Therefore, this assumption is only likely to hold if partners are separated from each other during the analysis. This is performed by sample stratification with respect to gender, which is also useful for other purposes (as described in section 3.4). The only unseparated couples are gay couples, which are small in number (i.e. only 24 lesbian and 17 male gay couples are present in the 2004 wave of the data, out of a total of 31,115 respondents) and thus have a negligible impact on the final result.

The second assumption is that of a Strongly Ignorable Treatment Assignment (SITA). Let $T_i = 1$ if individual i is a caregiver, and $T_i = 0$ if this individual is not (thus T represents the treatment assignment). The assumption of a SITA has two components. The first component is that treatment assignment is independent of the potential health outcomes given the vector of covariates $\vec{x}$: $T \perp (h(0), h(1)) \mid \vec{x}$. This basically means that there should be no relevant unobserved differences between the matched groups of caregivers and non-caregivers, conditional on the observed covariates. As argued by Stuart [2010], this assumption is often more reasonable than it seems, as matching on observed covariates also partly matches on unobserved covariates if they are correlated with the observed ones. Still, the covariate vector used for the matching procedure must be chosen carefully for this assumption to hold. This is elaborated on later in this methodology section; the exact determination of the covariate vector is discussed in section 4.2.3. The second component of SITA states that there is a positive probability of both caregiving and non-caregiving regardless of the values of the covariates: $0 < P(T = 1 \mid \vec{x}) < 1$ for all $\vec{x}$. In other words, no combination of covariates should be fully predictive of either caregiving or non-caregiving. For this reason, those individuals who do not have a care recipient (i.e. those without a living partner) and thus cannot be a care provider are removed from the dataset before matching.

3.3 4-step procedure

Stuart [2010] identified four key steps in performing a matching study, which will be elaborated on in this thesis:

1. Define a distance measure that indicates how close one individual is to another.
2. Implement a matching method, given this distance measure.
3. Assess the quality of the matched samples. These first three steps can be iterated until the samples are well-matched.
4. Analyse the outcome (in this context, the effect of caregiving on health).

3.3.1 Step 1: Define a distance measure

The problem of defining the 'closeness' of one individual to another can be split into two parts. The first part is to determine which covariates are relevant for the matching of individuals. The second part involves the way in which these covariates are combined to form a distance measure.

The assumption of the strongly ignorable treatment assignment has to be kept in mind when deciding on the covariates to be included in the matching procedure. To satisfy this assumption, all variables theoretically related to both health and caregiving should be included. If unobserved variables are theoretically expected to be related to both health and caregiving, one should preferably include observed variables that are likely to be correlated to those; only the part of unobserved variables that is unrelated to the observed ones is of concern. Variables that should not be included are those that may have been affected by the caregiving status of the particular respondent and those that predict caregiving status perfectly. To assure that this is the case, all variables used for the matching are measured in 2004, two years before the treatment variable (caregiving or non-caregiving, which is measured in 2006) is determined. Also, variables that are hypothesized to be associated with caregiving status but not with health outcomes are preferably not used in the matching procedure. However, with respect to these last variables, Stuart [2010] advises to be liberal regarding their inclusion as excluding a potentially important confounder might lead to increased bias. This is in accordance with Schmitz and Westphal [2013], who choose their covariates for the matching procedure based on their association with caregiving status, without considering their relation with health outcomes. They group variables influencing the decision to provide informal care as variables that influence (i) the need to provide care, (ii) the willingness to provide care, and (iii) the ability to provide care. Although there is some overlap between these groups, this framework helps to structurally identify all covariates to be used in the matching procedure. Therefore, these groups will also be used when the included variables are chosen in this study. For the exact definition of all variables, the reader is referred to chapter 4 (Variable measurement and data).

After identification of the included covariates, a decision has to be made how these covariates are combined to form a distance measure. There are different ways to define the distance D_{ij} between two individuals i and j; for example, the 'exact' distance measure indicates that $D_{ij} = 0$ if $\vec{x}_i = \vec{x}_j$ and $D_{ij} = \infty$ if $\vec{x}_i \neq \vec{x}_j$. Thus, individuals are considered to be 'infinitely' different if they differ on at least one of the relevant covariates. Another distance measure is the absolute difference in propensity scores: $D_{ij} = \mid e_i - e_j \mid$, where e_k indicates the propensity score for individual k. It is, in this context, defined as the probability of being a caregiver conditional on the relevant covariates: $e_k(\vec{x}_k) = P(T_k = 1 \mid \vec{x}_k)$. Propensity scores have two key properties [Stuart, 2010]. First, the propensity score is a balancing score. This means that the conditional distribution of the observed covariates given the propensity score is the same for caregivers ($T = 1$) as for non-caregivers ($T = 0$). Thus, matching individuals with similar propensity scores is like a randomized experiment with respect to at least the observed covariates. Second, if the assumption of the strongly ignorable treatments assignment holds given the covariate vector $\vec{x}$, then it also holds given the propensity score. Thus, matching on the propensity score instead of on the full set of covariates $\vec{x}$ is justified. These two properties are mathematically proven by Rosenbaum and Rubin [1983].

In this thesis, the distance measure based on the absolute propensity score difference is used. The main advantage of this distance measure is that propensity scores, in a sense, summarize all relevant covariate information in a single value (i.e. formally, the propensity score is a many-to-one function of $\vec{x}$), while not compromising on the necessary assumptions (as discussed above). Matching directly on observable covariates is feasible only if the number of observable covariates is low [Dehejia and Wahba, 2002]. For example, with two binary variables, one can place all individuals in one out of four groups. In this study, however, $\vec{x}$ is high-dimensional. Thus, exact matches on all covariates are practically non-existent, and some weighting scheme is necessary to be able to match individuals to one another. The propensity score matching method provides such a natural weighting scheme.

Matching based on the propensity score does not exclude the possibility of exact matching on a subset of the covariates. In this study, the caregiving and non-caregiving individuals are matched in strata based on their previous caregiving activity (in 2004, the first wave of the used data). This idea was originally suggested by Lechner [2009], and it is useful since previous caregiving activity is likely to capture an appreciable amount of unobserved heterogeneity that affects later caregiving activity (and thus treatment assignment) as well. Thus, the SITA assumption is much more likely to hold if matched individuals have the exact same previous caregiving activity.

Since the true probability of being a caregiver given all one's covariates (i.e. the true propensity score) is unknown, it has to be estimated using a model relating the covariates to caregiving activity. Then, the matching is based on the estimations of the propensity score. In principle, any model can be used for the estimation. Since multiple predictors are used to estimate a binary outcome (caregiving or not caregiving), logistic or probit regression models are most commonly used. In this study, both of these regression models are tested. The probit regression model seems to give a slightly better covariate balance than the logistic regression model (results not shown); therefore, a probit regression model is used to generate the results presented in chapter 5. For some individuals, the propensity score cannot be estimated since one or more of the covariates necessary for doing so is missing for those individuals. These respondents are therefore removed from the analysis.

3.3.2 Step 2: Implement a matching method

Step 2 of Stuart's four-step matching procedure [Stuart, 2010] comprises the choice and implementation of the particular matching method. Multiple methods are available; the most often-used matching methods seem to be nearest neighbour matching and kernel weighting matching. As the name suggests, nearest neighbour matching selects for each caregiving individual i the non-caregiving individual j with the smallest distance from i (i.e. the individual j with the smallest absolute difference in propensity score compared to individual i). This method has the disadvantage that many individuals are excluded from the analysis if the sizes of the treated and control groups are significantly different, as in this study.

Alternatively, kernel weighting matching is a method in which multiple non-caregiving individuals are matched to each caregiving individual i, with weights determined by their distance from individual i regarding their propensity score and the particular kernel function used. The kernel function relates the absolute propensity score difference to the matching weight. In general, the kernel function can have any shape. The major advantage of this method is that (almost) no individuals need to be excluded from the analysis, thus their information is not lost. A disadvantage is that nontrivial choices have to made regarding the type of kernel function and the value of a bandwidth parameter, a measure of how similar the propensity scores of two individuals should be for them to be regarded as a match. Thus, this choice of bandwidth quantifies the trade-off between bias (a higher bandwidth results in less accurate matches, increasing bias) and variance (a higher bandwidth results in more exploited information of control individuals, leading to a smaller variance) of treatment effect estimates (in this case the effect of caregiving on health).

Most authors seem to apply a pragmatic approach regarding the choice of the matching method. For example, Starks and Garrido [2014] argue that there is no universal 'best' matching method, and that the best method in any particular situation is the one that yields the best covariate balance and still meets the analytic goal. A particularly common matching method among scholars studying the effect of caregiving on health is kernel matching with a quadratic kernel called the Epanechnikov kernel (e.g. Schmitz and Westphal [2013]; Di Novi and Brenna [2013]). To maximise comparability, propensity score matching with Epanechnikov kernel weighting will be the method of first choice in this study. Only when this method does not succeed in obtaining similar covariate distributions for the treatment and control groups, will other matching methods be applied. As is described in section 5.2, this turns out to be not necessary as a proper covariate balance is obtained. The bandwidth will be chosen as small as is possible without significantly excluding individuals from the analysis. It is found that this strategy leads to a bandwidth of 0.03.[1]

[1]For one particular subsample (the long-term effect sample containing only males; the different samples are discussed in section 3.4 and 3.5), it was not possible to avoid removing individuals from the analysis while keeping the matching of sufficient quality due to a

In principle, it is possible that there is not sufficient overlap between the propensity scores of the caregivers and the non-caregivers. This region of overlap is called the common support. As is usual in the literature (e.g. Schmitz and Westphal [2013]), the analysis described in this thesis is restricted to all individuals whose propensity score is located on the common support.

3.3.3 Step 3: Assess the quality of the matched samples

Diagnosing the quality of the matched samples is possibly the most important step in using matching methods [Stuart, 2010]. The matched treated and control groups should be tested for the similarity of their covariate distributions. Only if the covariate distributions from these groups are sufficiently similar, the matching is successful. Although the definition of 'sufficiently similar' is admittedly somewhat subjective, some balance measures and guidelines on how to use them do exist. Rubin [2001] argues that two parameters are especially important for determining if the matching is sufficient: Rubin's B, which is the absolute standardized difference of the means of the propensity score between the treated and matched control group, and Rubin's R, which is the ratio of variance of the propensity score of the treated group to that of the matched control group. According to Rubin [2001], B should be smaller than 0.25 and R should be between 0.5 and 2 in order to consider the samples as sufficiently balanced.

Next to these measures regarding the distribution of the estimated propensity scores, the matched samples can also be compared on the level of individual covariates. Again, standardized differences in means and variance ratios are used to compare the distribution of covariates between the treated and matched control group.[2] In this thesis, the matching is considered sufficient

limited region of common support (see main text). In this case, it is chosen to remove the individuals and retain proper matching, the implications of which are discussed in section 6.1. For this (and only this) particular subsample, a bandwidth of 0.02 was used.

[2]Regarding the comparison of individual covariate distributions between the treated and matched control group, Rubin [2001] provides a useful criterion for matching quality. In particular, he argues that, for both the treated and the matched control group, each of the covariates should be regressed on the estimation of the propensity score. The ratio (between the two groups) of the variance of the residual of this regression (i.e. the part of the covariate that is orthogonal to the propensity score estimation) should be between 0.8 and 1.25 for well-matched covariates. Covariates for which this ratio is in the range [0.5, 0.8) or (1.25, 2] are classified as 'of concern' ,while 'bad' variables have a ratio <0.5 or >2. This ratio is a standard output of the matching diagnostics in Stata, and it is used in this thesis to evaluate matching quality for individual covariates (see table 5.4).

if the distributions for the propensity score and all key covariates are similar (defined by the above-mentioned criteria). The key covariates are considered to be former health, caregiving activity (with the partner as care recipient) and partner's health, with 'health' comprising all three different kinds of health as defined in section 4.2.2. In this way, a particular matching does not have to be rejected if there are differences in the distribution of covariates considered to be of minor importance.[3] Still, as described in section 5.2, the matching performed as described here results in similar covariate distributions between the treated and matched control group for almost all covariates.

3.3.4 Step 4: Analyse the outcome

The outcome analysis stage represents the final part of the analysis, in which the effect of caregiving on health is determined using the matched samples. There are now two ways to determine this effect. Since the matched samples are now artificially randomized (at least with respect to all observable covariates and those parts of unobservable covariates that are correlated to the observable ones), one can now simply compare the health in the treatment group (of caregivers) and the matched control group (of non-caregivers) to calculate the average treatment effect on the treated (ATT) and its statistical significance.

However, as Stuart [2010] argues, matching methods are not designed to compete with regression methods but in fact these methods work better in combination. This is shown by Rubin [1973] as well. The idea is that regression adjustment after matched sampling will take small residual variation in covariate distributions between the matched groups into account when determining the ATT (Bang and Robins [2005] call this a doubly robust estimation). Therefore, in this thesis, after performing the matching an OLS regression model is specified with (the various kinds of) health as the dependent variable(s).[4] The caregiving activity and all covariates used for

[3] It is important to note that, as the number of covariates in the covariate vector increases, the probability that at least one of the covariates differs significantly in distribution between the matched group and the control group increases as well. In turn, this generally leads to worse matching diagnostics but not necessarily to lower matching quality. This discrepancy forms the rationale behind limiting the number of covariates in this thesis, for example by grouping the different countries into broader regions.

[4] As the various health outcomes as defined in this thesis (see section 4.2.2 for details regarding health outcome definitions) are limited in their range (i.e. they have theoretical minima and maxima), an OLS regression model able to predict health outcomes outside that range is clearly not an optimal model. However, it is still preferred over other models for reasons of simplicity.

the matching are used as the independent variables. Naturally, the weight given to all control individuals (specified during the kernel matching) is taken into account in this regression model and standard errors are calculated robustly. Obviously, this step cannot be performed for individuals whose health outcomes are unknown. These individuals are therefore removed from the analysis altogether before matching.

According to Stuart [2010], there is some debate in the literature on matching about the variance estimation. In particular, the question whether to take into account the fact that the propensity score (on which the matching is based) is estimated rather than known is disagreed on. However, it has been shown by Rubin and Thomas [1996] and Rubin and Stuart [2006] that, under quite general conditions, the usage of estimated rather than true propensity scores results in an overestimate of variance. Thus, this leads to conservative estimates of the significance of the ATT, a conclusion drawn by Stuart [2010] as well. Therefore, in this thesis, the fact that propensity scores are estimated rather than known is not further taken into account.

3.4 Sample stratification

The analysis described in this thesis is performed separately for each gender. This has not just the effect of enforcing exact matching on this key covariate; it also yields separate ATTs for both genders. The reason for this choice is threefold. First, men and women differ on caregiving propensity [Coe and Van Houtven, 2009], so separation of these groups would, to some extent, naturally occur when matching purely on propensity scores; making this separation explicit yields information about the separate effect of caregiving on health for each gender while the consequences for the matching are limited. Second, separating by gender leads to a result that can be more easily compared to the other literature on this subject; essentially all similar studies either study health effects on men and women separately (e.g. Coe and Van Houtven [2009]) or they are restricted to women only (e.g. Do et al. [2013]; Schmitz and Westphal [2013]; Di Novi and Brenna [2013]). Third, separating by gender leads to a better acceptability of the SUTVA assumption as most couples are separated by this stratification (as explained in section 3.2).

3.5 Exploitation of panel nature of the data

Since the used data contains four different waves (the reader is referred to section 4.1 for details), this allows for the possibility of determining intermediate- to long-term effects (4 and 7 years, respectively, after the treatment assignment) of caregiving on health in addition to the immediate effect. To do so, a similar strategy is used as the one used by Schmitz and Westphal [2013]; the treatment ("being a caregiver") is defined as caregiving activity at the second wave (measured in 2006), referred to as $t = 0$. Individuals are assigned to the control group if they give care in none of the four waves, or only in the first wave (before the treatment assignment). Those who do not provide care at $t = 0$, but do so at any later wave, are excluded from the analysis as the control group should consist of non-caregivers from $t = 0$ onwards. Treatment and control individuals are matched based only on information of the first wave (measured in 2004, referred to as $t = -2$) in order to meet the SITA assumption (see section 3.2). Those defined as caregivers (at $t = 0$) may or may not be caregivers in any following wave. This is beneficial, as selection out of caregiving due to bad health (in turn possibly caused by caregiving activity) is not of concern as it does not have any effect on the treatment assignment. The determination of the immediate effect of caregiving on caregiver's health is not affected by this particular definition of a caregiver. However, due to this definition, as Schmitz and Westphal [2013] argue, any significant medium- or long-term effect can no longer be ascribed to caregiving in just the year of the treatment assignment. Instead, one can consider the determination of these longer-term effects as similar to an intention-to-treat (ITT) analysis, in which the intention is to be a caregiver at $t = 0$ and stop being a caregiver from the next wave onwards. In this case, some caregivers do not comply and keep on caregiving. An overview of the time structure is presented in figure 3.1, while the group assignment strategy is summarized in table 3.1.

As is clear from the choice of matching covariates (measured at the first wave) and treatment definition (measured at the second wave), individuals not present in either the first or the second wave of the data are excluded from the analysis. As there is some panel attrition, the subset of usable respondents becomes smaller at the third and fourth wave. To maximise the sample size at any wave, the matching is performed separately for each wave (i.e. separate matches to determine the immediate, medium-term and long-term effects). After each matching, an OLS regression (regressing caregiving activity on health) is performed for each outcome measure (the reader is referred to section 4.2.2 for an elaboration on the measures of health outcome), using the same covariates as those used for the matching.

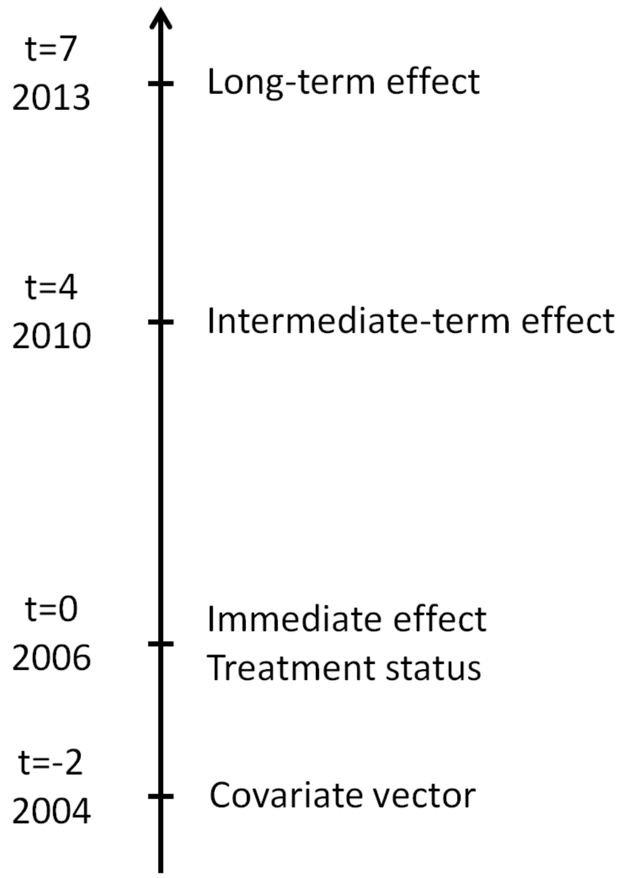

Figure 3.1: An overview of the time structure used in this study is shown.

	2004 **t = -2**	**2006** **t = 0**	**2010** **t = 4**	**2013** **t = 7**
Treatment group	X	1	X	X
Control group	X	0	0	0
Removed	X	0	1	X
	X	0	X	1

Table 3.1: Group assignment strategy. 1 = providing care to the partner; 0 = not providing care to the partner; X indicates that the caregiver status is not specified.

Chapter 4

Variable measurement and data

This chapter, discussing the variable measurement and data, is structured as follows: first, the dataset used in this study is introduced. It is followed by an elaborate description of the measurement of all variables of concern in this thesis: caregiving status, health outcomes and all covariates. Finally, the construction of the final dataset used for the matching is described, and some descriptive statistics for this final dataset are shown.

4.1 SHARE data set

The data on which the analysis presented in this thesis is based is obtained from the Survey of Health, Ageing and Retirement in Europe (SHARE). In this survey, focusing on people aged 50 years and older and their spouses, extensive data is gathered from more than 60.000 people across Europe (and Israel) [SHARE, 2013]. An overview of the methodology used to gather the SHARE data is given by Börsch-Supan and Jürges [2005]. The main aim of SHARE is to provide in depth data on ageing individuals and populations [Börsch-Supan et al., 2013]. The data is subdivided into numerous modules covering key aspects of people's lives, such as demographics, different measures of health, and social support variables. SHARE consists of panel data spanning four regular waves in which similar information is obtained from the participants, making this dataset suitable for measuring health effects of caregiving over time (there is also a fifth wave, called SHARELIFE, focusing on each participant's history; this wave is not used in this study). These waves are measured roughly in 2004, 2006, 2010 and 2013 (although there is some small variation over the countries), using computer-assisted personal interviews and a self-completion paper. This study is based on version 2.6.1 of the 2004 and 2006 waves, version 1.1.1 of the 2010 wave and version 1.0.0

of the 2013 wave.

4.2 Variable measurement

The ultimate goal of this thesis will be to examine the causal impact of providing informal care on the mental and physical health of the caregiver. A first step necessary for doing so, is operationalizing the concepts of mental/physical health and caregiver. This step is not trivial, as both concepts can be defined in various ways. One general notion about the variable measurement is that the treatment (i.e. caregiver status) is determined using information from only the 2006 wave, while the mental/physical health outcomes are determined for the years 2006, 2010 and 2013 (immediate, intermediate-term and long-term effects, respectively). All covariates on which the matching is based are measured using information from the 2004 wave only. The reader is referred to section 3.5 for the argumentation for this time structure.

4.2.1 Caregiving status

The currently available literature on health effects of informal caregiving differs in the operationalization of the concept of a caregiver. Whether or not someone provides informal care is usually determined by simply asking the respondent for this information. Usually, a minimum threshold with respect to the amount of time spend in this provision of care is used. For example, the measure used by Coe and Van Houtven [2009] is based on a threshold of 100 hours in the last 2 years. These authors also argue that reporting accuracy concerning the specific number of hours of care might be low and should therefore preferably not be used.

In this thesis, one is classified as a caregiver in two distinct situations. In the first situation, the respondent indicated (in the 2006 wave) that he/she gave help to his/her partner *within the household* daily or almost daily during at least three months (to avoid the capturing of caregiving during short-term sickness of the partner) within the period since the last interview (from the 2004 wave), where help is defined as anything related to personal care, such as washing, getting out of bed or dressing. In the second situation, the respondent indicated to have given help to his/her partner *living in another household*, the only difference with the last situation being that the concept of help is now extended to include household help (e.g. home repairs, gardening, transportation, shopping) and help with paperwork (e.g. filling out forms, settling financial or legal matters). Thus, a minimum frequency of providing help (almost daily) is used to classify persons as caregivers. When a respondent indicated that he/she did not provide help to the partner, that person is classified as a non-caregiver. Those who help their partner outside the household with a relatively small frequency (weekly or less) are removed from the analysis because they cannot be reliably categorized as either a caregiver or a non-caregiver.

4.2.2 Health outcomes

As mentioned in the section 2.1, health is fundamentally multidimensional. This results in some freedom with regard to the definition of generic health measures, which is reflected in the multitude of these measures used in the literature. Multiple mental health scales are included in the SHARE dataset, in particular the EURO-D and CES-D scales (depression) and the CASP-12 scale (quality of life). For physical health, multiple measures are used as well. These include the presence of particular diseases, anthropometric-based indicators (weight and height) and limitations with activities of daily living (ADL limitations).

In this thesis, mental health is measured using an inverted EURO-D scale. The EURO-D scale is a measure formed by the addition of 12 different items [Prince et al., 1999] and is also used in other studies measuring mental health (e.g. Di Novi and Brenna [2013]). All of these items are in some way related to depression, measuring for example pessimism, guilt, fatigue or tearfulness, and they are dichotomous (equal to one for the state related to depression, and zero otherwise). The addition of these items leads to an EURO-D scale measuring depression with outcomes ranging from 0 (not depressed at all) to 12 (severely depressed). This EURO-D scale is inverted to yield the measure of mental health used in this study, such that high values are associated with mentally healthy persons.

ADL limitations are chosen to represent physical health, as other measures seem too limited in their scope to measure such as broad concept as physical health (e.g. the presence of particular diseases) or their relation with physical health is too indirect (as is the case with the anthropometric-based indicators). ADL limitations are measured using the Global Activity Limitation Index (GALI), a dichotomized variable that indicates if respondents were limited for at least the past six months because of a health problem with 'activities people usually do'. The physical health indicator used in this thesis is equal to 0 when limited and 1 when not limited.

An often-used measure for general health is the health as experienced by the respondents: the self-perceived health (SPH). As an additional check, SPH is used as a third measure of health. For SPH, a dichotomized variable is used equal to 1 when the health is perceived as good or better, and equal to 0 when health is perceived as worse than good.

4.2.3 Covariates

In section 3.3.1, a description is given of the consideration of which variables to use for the matching step. These variables can be grouped, in accordance with Schmitz and Westphal [2013], as variables that influence (i) the need to provide care, (ii) the willingness to provide care, and (iii) the ability to provide care. Of course, this classification is artificial and there is extensive overlap between the groups; nevertheless, this distinction helps to identify the factors most likely to influence the variables related to caregiving activity.

The need to provide care to the partner is related to characteristics of the partner, as well as to the availability of alternative sources of care (both formal and informal). First of all, the need to provide care to the partner depends upon the presence of a partner; those without a partner obviously have no need (or possibility) to provide care in the definition used in this study. Since the absence of a partner perfectly predicts non-caregiving, using this variable as a covariate for the matching would violate the SITA assumption. Instead, all respondents without a partner, defined as a spouse or another kind of registered partner, are excluded from the analysis.

Other variables influencing the need to provide care are the health and age of the partner, the living country/region (each with its own formal care arrangements and cultural background regarding informal care provision [Reher, 1998]), the presence of children living at home, the number of siblings (as children and siblings can function as alternative care providers) and the type of living area. In accordance with Di Novi and Brenna [2013], three different regions of living are distinguished: northern Europe, central Europe and southern Europe. Northern Europe includes Denmark, Sweden and the Netherlands, central Europe includes Austria, Belgium, France, Germany and Switzerland, while Greece, Italy and Spain are classified as southern European countries. As in Di Novi and Brenna [2013], this particular classification is based (next to geographical location and culture) on the percentage of GDP spent on formal long-term care, with the northern countries spending the most while the southern countries spend the least [Francesca et al., 2011].

The number of children living at home is directly available in the SHARE dataset. Partner's health is measured in the same way as the health outcomes (see section 4.2.2). Partner's age is measured by subtracting their year of birth from the interview year. The total amount of living siblings of both partners together is chosen as a covariate for the matching, as, in principle, both kinds of siblings can help with providing care.

Whether one lives in a rural or urban environment potentially impacts the probability of providing care as well (for example by a difference in availability of socially connected neighbours who can function as alternative providers of care). Therefore, a living area is classified as urban if the living area is indicated to be in a big city, in the suburbs or outskirts of a big city, or in a large town. The living area is classified to be rural (i.e. not urban) if it is indicated as a small town, a village or a rural area.

The factors influencing the willingness and ability to provide care to the partner are mostly related to characteristics of the potential caregiver him- or herself. Factors included in this study are the respondent's own age, gender (which is used for sample stratification, thus not directly as a covariate for the matching), former caregiving activity regarding the partner (caregiving status in 2004; exact matching is performed for this variable), caregiving activity regarding the parents, health (mental, physical and self-perceived health), employment status, education level, income (both household and individual income), proxies for personality and the cohabitation status regarding their partner (as some partners do not live together).

Age is modelled as a continuous variable by subtracting the year of birth from the interview year. Age squared is also incorporated as a covariate in the matching model, as this non-linear effect is usually taken into account in other studies on caregiving as well (e.g. Schmitz and Westphal [2013]). Gender is directly available in the SHARE dataset. Health as a matching covariate (measured in the 2004 wave) is determined in the same way as the health outcomes (measured in later waves; see above).

Exact matching is performed on earlier caregiving status (see section 3.3.1 for details). For the measurement of this variable, the information on caregiving from the 2004 wave is used. Otherwise, it is measured in the same way as the treatment indicator variable representing caregiving activity in the 2006 wave (see above).

Caregiving activity with the parents or parents-in-law as care recipients is also regarded as a matching covariate, as these other caregiving commitments possibly impact the decision to provide care for the partner as well. The definition of caregiving to parents/parents-in-law is the same as the definition for caregiving to partners (see above), except for the difference in care recipient.

A few indicator variables are constructed representing employment status, categorizing respondents as retired, employed (including those who are self-employed or working for a family business), homemaker or unemployed (including those who are permanently sick or disabled).

The education variable in the SHARE dataset is classified by the 1997

International Standard Classification of Education (ISCED-97) in seven levels: 0 (pre-primary schooling), 1 (primary education), 2 (lower secondary education), 3 (upper secondary education), 4 (post high school education), 5 (university education) and 6 (postgraduate education). In this study, the level of education is regrouped in three different classes, as in Di Novi and Brenna [2013]; these are low (levels 0-2), medium (levels 3-4) and high (levels 5-6) levels of education.

Two different income measures are used for the matching. The first of them, household income, is a measure of socio-economic status. It is obtained by taking the nominal annual household income, and dividing it by the purchasing power-adjusted exchange rate for the year before the interview (as the household income as provided in the SHARE dataset was imputed for the prior year) to express the household income in ppp-adjusted euros. Consequently, the household income is normalized by dividing it by the household size and the natural logarithm of the result is taken as the final variable representing household income.

The second income measure is the individual income of the respondent relative to the household income. This variable is included since a high dependence of the household on the potential caregiver's income presumably leaves a smaller opportunity for this respondent to take on caregiving activities as well. This variable is computed by taking the ppp-adjusted annual personal income of the respondent (a summation of the income from regular employment and the income from self-employment; ppp-adjusted in the same way as the household income) and dividing it by the ppp-adjusted total annual household income. In a negligible amount of cases (3 in total), this fraction has a value outside the (0,1) range. For these individuals, this variable is indicated to be missing.

Personality is a very broad concept that cannot be easily described using a single variable. However, it can be hypothesized that people with certain personality traits (e.g. altruistic persons or persons feeling highly responsible) are more likely to provide informal care than those without these characteristics. Even though personality itself cannot be easily measured, proxies for the mentioned personality traits should be included as matching covariates in order to maximise the randomization regarding these traits. As no useful results of personality tests are available in the SHARE dataset, more indirect proxies for personality should be used. Some information that is available, is the motivation of respondents for performing certain social activities. These social activities include voluntary or charity work, caring for a sick or disabled adult, providing help to family, friends or neighbours, attending an educational or training course, going to a sport, social or other kind of club and taking part in a religious, political or community-related organization.

For each performed activity, respondents are asked about their motivation for that particular activity. Two binary variables are constructed that indicate if the motivations 'to contribute something useful' and 'because I am needed' are used for at least one the mentioned activities; they are proxies for altruism and responsibility, respectively. A similar way to obtain proxies for personality traits is described in Oudijk et al. [2011].

A binary variable is constructed indicating whether or not a respondent is married while simultaneously living in a different household than his/her spouse. This variable is hypothesized to influence caregiving, as in this case caregiving requires the additional barrier of physically going towards the partner.

4.3 Construction and summary statistics of final dataset

The first two waves of the used SHARE dataset contain a total of 46,788 individuals (i.e. observations). However, as mentioned in various sections of this thesis, not all of these individuals are actually taken into account in this study for various reasons. In particular, the following list indicates all individuals that are removed from the analysis (the argumentation for removing them is provided in the indicated sections):

- All respondents who are not present in both the first (2004) and the second (2006) wave (see section 3.5).
- All respondents without a living partner (see section 3.2).
- All respondents with missing values of caregiving status in 2006 (the 'treatment' indication); this includes all respondents who give care to the particular care recipient (in this case the partner) outside the household, but indicate to do so less than weekly (see section 4.2.1).
- All respondents who are classified as a caregiver later than 2006 but not in 2006 itself (see section 3.5).
- All respondents who have a missing value for one or more of the covariates (see section 3.3.1) or health outcomes (see section 3.3.4).

After removal of these individuals, the final dataset that is used for the matching contains 10,211 individuals, with roughly equal amounts of men (5132) and women (5079).

Descriptive statistics per gender are shown in table 4.1. As can be observed, a rather large decrease in panel size occurs between the second (2006) and third (2010) wave, which forms the main motivation for performing separate matches when analyzing the effect of caregiving on health on different time scales (immediate, intermediate-term and long-term effects). Of the total sample, around 6.6% of females and 5.1% of males are classified as a caregiver according to the definition in section 4.2.1.

	Females					Males				
Variable	Obs.	Mean	Std. Dev.	Min	Max	Obs.	Mean	Std. Dev.	Min	Max
Caregiver partner (2004)	5079	0.041	0.198	0	1	5132	0.034	0.183	0	1
Caregiver partner (2006)	5079	0.066	0.248	0	1	5132	0.051	0.219	0	1
Mental health (2004)	5079	9.587	2.178	0	12	5132	10.391	1.787	1	12
Mental health (2006)	5079	9.696	2.186	0	12	5132	10.35	1.901	1	12
Mental health (2010)	2907	9.384	2.216	1	12	2845	10.21	1.834	0	12
Mental health (2013)	2682	9.349	2.276	1	12	2509	10.22	1.898	1	12
Physical health (2004)	5079	0.612	0.487	0	1	5132	0.654	0.476	0	1
Physical health (2006)	5079	0.621	0.485	0	1	5132	0.626	0.484	0	1
Physical health (2010)	2962	0.511	0.500	0	1	2910	0.545	0.498	0	1
Physical health (2013)	2679	0.518	0.500	0	1	2506	0.571	0.495	0	1
Self-perceived health (2004)	5079	0.753	0.431	0	1	5132	0.774	0.419	0	1
Self-perceived health (2006)	5079	0.693	0.461	0	1	5132	0.712	0.453	0	1
Self-perceived health (2010)	2962	0.668	0.471	0	1	2908	0.682	0.466	0	1
Self-perceived health (2013)	2682	0.752	0.432	0	1	2509	0.716	0.451	0	1
Age	5079	60.36	9.181	26	91	5132	63.68	8.931	39	96
Age squared	5079	3728	1141	676	8281	5132	4135	1173	1521	9216
Living as a single	5079	0.026	0.158	0	1	5132	0.023	0.151	0	1
Child at home	5079	0.136	0.343	0	1	5132	0.192	0.394	0	1
Total number of siblings	5079	4.853	3.230	0	22	5132	4.810	3.246	0	22
Education (low)	5079	0.519	0.500	0	1	5132	0.444	0.497	0	1
Education (medium)	5079	0.302	0.459	0	1	5132	0.326	0.469	0	1
Education (high)	5079	0.179	0.383	0	1	5132	0.230	0.421	0	1
Income fraction	5079	0.132	0.215	0	1	5132	0.268	0.344	0	1
Log household income	5079	9.597	0.855	3.320	13.02	5132	9.607	0.850	3.320	13.02
Retired	5079	0.319	0.466	0	1	5132	0.586	0.493	0	1
Employed	5079	0.304	0.460	0	1	5132	0.358	0.480	0	1
Homemaker	5079	0.320	0.467	0	1	5132	0.004	0.059	0	1
Unemployed	5079	0.057	0.231	0	1	5132	0.053	0.223	0	1
Contribute	5079	0.203	0.402	0	1	5132	0.229	0.420	0	1
Needed	5079	0.294	0.455	0	1	5132	0.276	0.447	0	1
Caregiver parent	5079	0.110	0.313	0	1	5132	0.068	0.251	0	1
North	5079	0.277	0.448	0	1	5132	0.277	0.448	0	1
Central	5079	0.441	0.497	0	1	5132	0.445	0.497	0	1
South	5079	0.282	0.450	0	1	5132	0.278	0.448	0	1
Urban	5079	0.481	0.500	0	1	5132	0.486	0.500	0	1
Partner's age	5079	63.62	9.007	39	104	5132	60.28	9.321	26	91
Partner's mental health	5079	10.21	2.273	0	12	5132	9.486	2.461	0	12
Partner's physical health	5079	0.654	0.476	0	1	5132	0.624	0.485	0	1
Partner's self-perceived health	5079	0.770	0.421	0	1	5132	0.758	0.428	0	1

Table 4.1: Descriptive statistics by gender. All variables are measured using data from only the 2004 wave, unless otherwise indicated.

Chapter 5

Results

In this chapter, the results obtained by applying the 4-step procedure by Stuart [2010] are shown. First, the results of the propensity score estimation are shown, in particular the results of the associated probit regression (in which caregiving activity in 2006, the 'treatment' indication, is regressed on the covariate vector). Then, the quality of the matching (i.e. the similarity of the matched samples with respect to each other) is discussed. Finally, the estimated effects of caregiving to the partner on various aspects of health are presented for the different time scales.

5.1 Propensity score estimation

Propensity scores are estimated by a probit regression, where caregiving activity in 2006 is regressed on the complete covariate vector. As discussed in section 3.4, this regression is performed separately for each gender. The regression is also performed separately for individuals who were caregivers and non-caregivers in 2004[1], and separately for the three different samples corresponding to the three different time scales (immediate effect, intermediate-term effect and long-term effect) to minimize the impact of panel attrition. Thus, a total of twelve probit regressions are performed, with caregiving to the partner in 2006 as the dependent variable; due to length and clarity considerations, the only presented results are those of the probit regression for women who were not providing care in the 2004 wave, using the complete

[1] Although theoretically not necessary, the command used in Stata for the matching (called *psmatch2* [Leuven and Sianesi, 2014]) couples the probit regression, and thus the estimation of the propensity score, to the matching process itself. To enforce exact matching on previous caregiving activity, the probit regression and the matching are performed twice (for previous caregivers and non-caregivers), after which the two groups are merged.

sample that is later used for calculating the immediate effect of caregiving on health. These results are shown in table 5.1. Results for the other probit regressions are similar (they can be found in the log file associated with this thesis).

The results shown in table 5.1 are interesting in itself, because they show which covariates are associated with caregiving to the partner. As indicated in the table, bad health and high age of a woman's partner are strongly positively associated with caregiving activity of the particular woman (with the partner as a care recipient), as expected. The other statistically significant associations are not very surprising either; being retired or unemployed is associated with a higher probability of caregiving (relative to the employed woman, the reference category), while the presence of children living at home is associated with a lower probability of caregiving. Finally, ceteris paribus, woman are less likely to be providing care to their partner in the northern European countries (relative to the central region, the reference category).

Table 5.2 shows the sample size of the different samples used for the matching (grouping the two samples that differ on caregiving status in 2004 together). Again, panel attrition can be observed as longer-term effects are studied. Some of the 'treated' individuals are removed from their respective sample because their estimated propensity score is located off the common support (usually, their estimated propensity score is too high to be reliably matched to 'untreated' individuals, who have lower propensity score estimations). When immediate effects of caregiving on health are studied, the amount of individuals off the common support is limited compared to the total amount of treated individuals. However, when analyzing long-term effects, the relative amount of removed treated individuals becomes substantial, especially for men (this is the motivation for some other authors to focus on women only, e.g. Schmitz and Westphal [2013]). This has some implications regarding the reliability of the obtained long-term results, as discussed more thoroughly in section 6.1.

5.2 Matching quality

Although the probit regression and its associated matching is performed twelve times, matching quality has to be assessed only six[2] times, as the corresponding matched samples with different caregiving activities in 2004 are grouped together to form the final matched sample (the initial distinc-

[2]Three different time scales (immediate effect, intermediate-term effect and long-term effect) and two genders give a total of six combinations.

Caregiver partner (2006)	Coef.	Std. Err.	z	p-value
Mental health	-0.038	0.033	-1.160	0.247
Physical health	-0.279	0.162	-1.720	0.086
Self-perceived health	-0.025	0.176	-0.140	0.887
Age	-0.034	0.085	-0.400	0.691
Age squared	0.000	0.001	0.370	0.712
Living as a single	-1.003	0.610	-1.650	0.100
Child at home	-0.748	0.294	-2.540*	0.011*
Total number of siblings	0.020	0.022	0.910	0.365
Education (medium)	0.046	0.175	0.260	0.793
Education (high)	-0.215	0.249	-0.860	0.387
Income fraction	0.944	0.554	1.710	0.088
Log household income	-0.080	0.098	-0.820	0.414
Retired	0.812	0.336	2.420*	0.016*
Homemaker	0.476	0.338	1.410	0.159
Unemployed	1.137	0.365	3.110*	0.002*
Contribute	0.137	0.197	0.690	0.488
Needed	-0.272	0.195	-1.400	0.162
Caregiver parent	0.242	0.257	0.940	0.345
North	-0.393	0.191	-2.050*	0.040*
South	-0.232	0.185	-1.250	0.210
Urban	-0.211	0.147	-1.430	0.152
Partner's age	0.051	0.016	3.220*	0.001*
Partner's mental health	-0.087	0.027	-3.190*	0.001*
Partner's physical health	-0.631	0.162	-3.910*	0.000*
Partner's self-perceived health	-0.726	0.163	-4.450*	0.000*
Constant	-2.593	2.911	-0.890	0.373

Table 5.1: Probit regression results, where caregiving to the partner in 2006 is regressed on the covariate vector for the subgroup of woman who were not providing care to their partner in 2004 (a total of 4872 observations). The coefficients shown here are used to estimate the propensity score for each individual in this reduced sample. * indicates statistical significance at the 5% significance level. All covariates are measured using information from the 2004 wave only.

tion between these two groups during the matching is only there to ensure exact matching on this covariate). General measures for matching quality (i.e. Rubin's B, Rubin's R and the individual covariates that are considered to be 'of concern' or 'bad' with respect to their matching, according to the criteria mentioned in section 3.3.3) are provided for all six situations in table 5.3. Matching diagnostics for each covariate are, for brevity, shown only for women, using the complete sample that is later used for calculating the immediate effect of caregiving on health. They are presented in table 5.4.

As explained in section 3.3.3, Rubin's B is the absolute standardized difference of the means of the estimated propensity score between the treated and matched control group, and Rubin's R is the ratio of variance of the estimated propensity score of the treated group to that of the matched control group. According to Rubin [2001], B should be smaller than 0.25 and R should be between 0.5 and 2 in order to consider the samples as sufficiently balanced. Clearly, these criteria are satisfied, as can be observed in table 5.3. None of the covariates can be classified as 'bad' with respect to their matching based on the criterion by Rubin [2001]. Some covariates are classified as 'of concern', but none of these were identified as a key covariate in section 3.3.3. Thus, these matches are not rejected.

In table 5.4, it can be observed that the difference in means between the treated and control group (in this example, for the immediate effect sample containing only women) of essentially all covariates decreases substantially when matching. This is quantitatively captured in the decrease in standardized percentage bias, as defined by Rosenbaum and Rubin [1985] and shown in table 5.4. Moreover, according to the criterion regarding the ratio of the variance of regression residuals (as defined in section 3.3.3) between the treated and control group, 17 covariates are classified as 'of concern' or 'bad' before the matching (including the key covariates mental health, self-perceived health, caregiving activity with respect to the partner, and partner's mental health and self-perceived health), while only two of these are still of concern after the matching ('living as a single' and high education level). As all the key covariates are properly matched, this particular matching is accepted. Tables similar to table 5.4, but using the other five samples (other time scales and for males) can be found in the log file associated with this thesis.

All six assessments of matching quality show good general matching diagnostics (i.e. Rubin's B <0.25; $0.5 <$ Rubin's R < 2) and satisfactory matching diagnostics regarding individual covariates (all key covariates properly matched). Thus, in conclusion, all matches are accepted and the effect of caregiving to the partner on caregiver's health can be determined.

		Total observations	Treated observations off common support	Treated observations remaining
Females	Immediate effect sample	5079	7	328
	Intermediate-term effect sample	2903	15	166
	Long-term effect sample	2610	13	146
Males	Immediate effect sample	5122	6	253
	Intermediate-term effect sample	2831	19	123
	Long-term effect sample	2494	20	98

Table 5.2: The size of the different samples used for the matching are shown in the 'Total observations' column. Some of the 'treated' individuals (those who are classified in 2006 to provide care to their partner) are removed from the samples because their propensity score is not located on the common support. The amounts of removed and remaining treated individuals are indicated in the rightmost columns.

		Rubin's B	Rubin's R	Covariates classified as 'of concern'	Covariates classified as 'bad'
Females	Immediate effect sample	0.126	0.93	Living as a single; Education (high)	-
	Intermediate-term effect sample	0.204	0.81	Log household income	-
	Long-term effect sample	0.176	0.81	-	-
Males	Immediate effect sample	0.143	0.86	-	-
	Intermediate-term effect sample	0.199	0.76	Living as a single	-
	Long-term effect sample	0.224	0.93	Age; Age squared; Total siblings; Log household income	-

Table 5.3: General measures of matching quality. These measures were introduced in section 3.3.3 and are discussed again in the main text in this section.

Variable	Unmatched/ Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	$V_e(T)/V_e(C)$
Mental health	U	8.764	9.645	-38.3		1.30*
	M	8.787	8.781	0.3	99.3	1.01
Physical health	U	0.466	0.622	-31.8		1.14
	M	0.457	0.473	-3.2	90.1	0.99
Self-perceived health	U	0.594	0.765	-37.2		1.47*
	M	0.588	0.595	-1.5	96.0	1.03
Age	U	65.53	60.00	59.4		1.41*
	M	65.53	65.29	2.6	95.6	1
Age squared	U	4386	3682	59.5		1.58*
	M	4386	4353	2.8	95.3	1
Living as a single	U	0.015	0.026	-8.0		0.58*
	M	0.015	0.021	-4.1	48.8	0.67*
Child at home	U	0.075	0.140	-21.3		0.61*
	M	0.073	0.066	2.2	89.6	1.13
Total number of siblings	U	4.785	4.857	-2.2		1.12
	M	4.829	4.883	-1.6	25.7	1.06
Education (medium)	U	0.266	0.304	-8.6		0.92
	M	0.259	0.253	1.4	84.2	1
Education (high)	U	0.093	0.185	-26.9		0.58*
	M	0.095	0.117	-6.4	76.1	0.78*
Income fraction	U	0.079	0.136	-27.7		0.86
	M	0.079	0.079	0.3	98.9	1.1
Log household income	U	9.382	9.613	-28.8		0.75*
	M	9.382	9.384	-0.2	99.3	0.82
Retired	U	0.504	0.306	41.3		1.29*
	M	0.494	0.485	1.8	95.7	0.97
Homemaker	U	0.284	0.323	-8.6		0.92
	M	0.290	0.306	-3.6	57.4	0.94
Unemployed	U	0.087	0.054	12.6		1.57*
	M	0.088	0.077	4.5	64.5	1.16

Contribute	U	0.152	0.206	-14.1		0.78*
	M	0.155	0.165	-2.5	82.5	0.93
Needed	U	0.233	0.298	-14.8		0.87
	M	0.229	0.241	-2.7	81.5	0.92
Caregiver partner	U	0.287	0.023	78.0		4.03**
	M	0.274	0.274	0.0	100	1
Caregiver parent	U	0.072	0.112	-14.1		0.65*
	M	0.073	0.075	-0.6	95.6	0.98
North	U	0.188	0.284	-22.6		0.77*
	M	0.192	0.190	0.6	97.3	1.01
South	U	0.334	0.278	12.1		1.08
	M	0.329	0.344	-3.3	72.9	0.99
Urban	U	0.427	0.485	-11.6		1
	M	0.424	0.435	-2.3	80.0	1
Partner's age	U	69.53	63.21	68.0		1.60*
	M	69.41	69.25	1.7	97.4	1.09
Partner's mental health	U	8.836	10.30	-56.1		1.79*
	M	8.887	8.877	0.4	99.3	0.97
Partner's physical health	U	0.296	0.679	-83.1		1.01
	M	0.302	0.313	-2.5	97.0	0.97
Partner's self-perceived health	U	0.397	0.796	-89.0		1.45*
	M	0.405	0.409	-0.7	99.2	0.99

Table 5.4: Matching quality, individual covariates. This table contains diagnostics regarding the matching of the immediate effect sample containing only women. Within this subgroup, it shows the mean of each covariate for the treated and control group, both before and after the matching. It also shows the standardized percentage bias, which is the difference of the sample means in the treated and control (before or after matching) sub-samples as a percentage of the square root of the average of the sample variances in the treated and control groups (see Rosenbaum and Rubin [1985] for the formula). The standardized percentage bias is reduced significantly for essentially all covariates due to the matching. The rightmost column shows the ratio of the variance of regression residuals (as defined in section 3.3.3) between the treated and control group before and after the matching. For essentially all covariates, this ratio moves towards one due to the matching. * indicates covariates 'of concern', while ** indicates 'bad' covariates, according to the definition of Rubin [2001]. In this example, 17 covariates are classified as belonging to one of these categories before the matching, but only two of them remain of concern after the matching ('Living as a single' and 'Education (high)').

5.3 Effect of caregiving on health

The effect of caregiving to the partner on caregiver's health is determined by regressing the various aspects of caregiver's health (mental, physical and self-perceived health) in different times (for estimating the immediate, intermediate-term and long-term effects) on the treatment indication (caregiving activity to the partner in 2006), controlling for the complete covariate vector. OLS regressions are performed for each gender, each health outcome and each time scale, a total of eighteen regressions. The complete regression results are provided in the log file associated with this thesis; the relevant outcomes regarding the effect of caregiving to the partner on caregiver's health are shown in table 5.5. As can be observed, on the 5% significance level, only two of the eighteen regressions shows a statistically significant effect of caregiving on health: the immediate effects of providing care by women to the partner on mental health are negative and statistically significant (t=-3.52, p=0.000 and t=-2.20, p=0.028, respectively); for this subgroup, caregiving is associated with an immediate decrease of 0.542 of mental health (on a scale of 0 to 12) and an immediately decreased probability to self-classify health as good or better[3]. On the 10% significance level, the immediate effect of caregiving to the partner on mental health for men is also negative and significant (t=-1.92, p=0.054); for this subgroup, caregiving is associated with an immediate decrease of 0.291 of mental health (on a scale of 0 to 12). All other effects of caregiving on health (e.g. effects on physical health, intermediate- and long-term effects) are not statistically significant.

[3]According to the OLS model, this probably decreases with about 7 percentage points, but since the OLS regression model is not ideal for estimating these effects on binary variables, this value has to be interpreted as an approximation.

			Coef	Robust Std. Err.	t	p-value
Females	Mental health	Immediate effect	-0.542	0.154	-3.52**	0.000**
		Intermediate-term effect	-0.096	0.193	-0.50	0.619
		Long-term effect	-0.203	0.196	-1.03	0.301
	Physical health	Immediate effect	-0.039	0.033	-1.19	0.234
		Intermediate-term effect	0.032	0.040	0.81	0.419
		Long-term effect	0.053	0.039	1.37	0.170
	Self-perceived health	Immediate effect	-0.070	0.032	-2.20**	0.028**
		Intermediate-term effect	-0.047	0.041	-1.14	0.254
		Long-term effect	0.030	0.027	1.11	0.269
Males	Mental health	Immediate effect	-0.291	0.151	-1.92*	0.054*
		Intermediate-term effect	0.243	0.180	1.35	0.178
		Long-term effect	-0.132	0.228	-0.58	0.563
	Physical health	Immediate effect	-0.013	0.031	-0.41	0.680
		Intermediate-term effect	0.017	0.040	0.43	0.667
		Long-term effect	0.075	0.046	1.64	0.101
	Self-perceived health	Immediate effect	-0.036	0.030	-1.18	0.237
		Intermediate-term effect	0.058	0.042	1.37	0.172
		Long-term effect	-0.019	0.038	-0.50	0.618

Table 5.5: Regression coeffients and other statistics of 18 regressions, representing the effect of caregiving to the partner on all three aspects of health for both genders and all time scales. The coefficients shown are the regression coefficients obtained when regressing the particular aspect of health at a particular year (2006 for the immediate effect, 2010 for the intermediate-term effect, and 2013 for the long-term effect) on caregiver status (with the partner as care recipient) in 2006, while controlling for the complete covariate vector and using the obtained matched samples. ** indicates significance at the 5% level, while * indicates significance at the 10% level.

Chapter 6

Discussion and Conclusion

In this chapter, the research question stated in chapter 1 is answered using the results presented in chapter 5. The economic relevance of this answer is discussed and the findings are compared to the those obtained from the literature on caregiving presented in section 2.2. Finally, some indications for future research are provided.

6.1 Conclusion regarding the research question

In chapter 1, the research question studied in this thesis was stated:

> *"What is the causal impact of providing informal care to one's partner on the mental and physical health of the caregiver?"*

The results presented in chapter 5, and in particular those shown in table 5.5, are now used to answer this question. For women, the results suggest that providing care to their partner does indeed have a negative impact on the health of the care provider, in particular on mental health and self-perceived health (which can be considered to be a measure of health that contains aspects of both mental and physical health). These results are significant at the 5% significance level. Physical health (as measured by limitations in daily activities) does not seem to be affected by care providing. For men, the health impact of providing care to their partner is less clear; on the 10% significance level, a negative impact of care providing on mental health seems to exist for men. For other kinds of health, no statistically significant effects exist.

For both women and men, the mentioned (immediate) negative health effects apparently dissipate over time. Both four and seven years after the treatment indication, the health (mental, physical and self-perceived) of those who were caregivers initially was not significantly different from the health of the matched control group (after controlling for covariates).

In principle, this forms an answer to the research question. However, as in all studies, the reliability of the answer is highly dependent on the research methodology, which is therefore critically discussed in this section. Most aspects of the research methodology, such as the assumptions that are implicitly made by using the matching method of this study, are discussed elaborately in chapter 3 (including the argumentation that these assumptions are satisfied) and are not repeated here. The treated and matched control groups seem to be sufficiently similar (as shown in section 5.2), so it seems unlikely that the results of this study are significantly biased due to improper matching.

However, there is one aspect of the methodology that does have the potential to impact the obtained results; this is the removal of treated observations that have an estimated propensity score outside the region of common support. This removal is indicated in table 5.2. For estimating the immediate effect of caregiving on health, the removal of these individuals is likely to be of only minor importance, as their number is small compared to the total number of treated individuals. However, for estimating longer-term effects, the number of treated individuals with estimated propensity scores off the common support rises in an absolute as well as relative sense. The shrinkage of the region of common support itself (due to panel attrition) is the cause of this finding. As a consequence, those individuals with the highest propensity scores are removed from the analysis when estimating intermediate- and long-term effects of caregiving on health, especially in the case of men. As indicated in table 5.1, the propensity score is highly impacted by characteristics of the partner; high estimated propensity scores are associated with unhealthy partners and thus a high burden of care. By predominantly dropping these individuals, the average care burden in the dataset decreases. If care burden is associated with negative health effects (as it is hypothesized to be), it is likely that the negative intermediate- and long-term effects of providing care to the partner on the caregiver's health are underestimated in this study, especially for men.

It is therefore concluded that providing care to the partner has a significant and negative immediate effect on the mental and self-perceived health of the female care provider, but there is no significant effect on physical health. For the male care provider, such a negative immediate effect is present only for mental health (on the 10% significance level). The magnitudes of these effects are discussed in section 6.2. Longer-term effects (four or seven years) of caregiving on health cannot be completely excluded due to common support issues, but no evidence for their presence is obtained in this study.

6.2 Economic relevance

So far, the statistical significance of health effects of caregiving have been the main concern of this chapter. In order to make a statement concerning their economic relevance, the magnitudes of these health effects have to be analysed. Caring for the partner leads to a decrease in mental health (defined as the inverse of the EURO-D scale for depression, which ranges from 0 to 12) of 0.542 (females) or 0.291 (males). Considering that the difference between a person with optimal mental health and a person that is likely to be clinically identified as depressed is only around 3 or 4 [Prince et al., 1999] on this scale, providing care to the partner can contribute substantially to mental morbidity. For women, a negative immediate effect of caregiving on self-perceived health is observed; providing care to the partner leads to approximately a 7 percentage point decreased probability for a woman to self-classify her health as good or better, which seems appreciable. Although informal care might be considered by some to be a 'cheap' form of care and is therefore increasingly relied on in some countries (for example in the Netherlands [de Klerk et al., 2014]), it seems reasonable to take these 'health costs' into account when comparing the pros and cons of formal and informal care.

6.3 Comparison of results to other literature

In section 2.2, empirical findings of similar studies were presented. Comparing the results of this study to those obtained by other authors is interesting, but is also hindered somewhat by the different care recipients and health outcomes; other studies, such as Coe and Van Houtven [2009] and Do et al. [2013], mainly focused on parents as the care recipients.

Coe and Van Houtven [2009] stratify by marital status and observed increases in depressive symptoms and decreases in self-rated health due to caregiving (with the parent as a care recipient) in both married men and women, while they do not find effects of caregiving on physical health for these groups. This roughly corresponds to the results shown in this thesis, in which the effect of caregiving on mental health is substantial but no effect for physical health is observed. In contrast, Do et al. [2013] focus exclusively on the effect of caregiving to the parent on physical health and do find significant effects. Their results, however, are not necessarily contradictory to the results of this study, as the care recipient differs. Instead, this thesis adds to the mentioned literature by studying the health effects of caregiving to the partner, a care recipient that has (to the best of my knowledge) not been considered in isolation yet in the recent caregiving studies that control for self-selection.

6.4 Indications for future research

As informal care provision is becoming more important to meet the growing demand for care in the near future, it becomes increasingly important to understand all the relevant costs of this informal care provision. One aspect that is quite important for the economic relevance of the health effects of caregiving is their persistence over time. Although no evidence for the existence of long-term effects of caregiving on health is found in this thesis, they cannot be completely excluded either based on this analysis due to a limited region of common support in the matching procedure. Thus, the main suggestion for future research is to study the persistence of health effects over time using a more complete dataset. Identification of these longer-term effects will bring us one step closer to understanding all the costs associated with informal care provision, and thus to better-justified policy regarding this exciting subject.

Bibliography

Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

Barbara M Barer and Colleen L Johnson. A critique of the caregiving literature. *The Gerontologist*, 30(1):26–29, 1990.

Ana Bobinac, N Job A van Exel, Frans FH Rutten, and Werner BF Brouwer. Caring for and caring about: disentangling the caregiver effect and the family effect. *Journal of health economics*, 29(4):549–556, 2010.

Axel Börsch-Supan and Hendrik Jürges. The survey of health, aging, and retirement in Europe. *Methodology, Mannheim Research Institute for the Economics of Aging (MEA)*, 2005.

Axel Börsch-Supan, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie Korbmacher, Frederic Malter, Barbara Schaan, Stephanie Stuck, and Sabrina Zuber. Data resource profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, page dyt088, 2013.

Norma B Coe and Courtney Harold Van Houtven. Caring for mom and neglecting yourself? The health effects of caring for an elderly parent. *Health economics*, 18(9):991–1010, 2009.

Mirjam de Klerk, Alice de Boer, Sjoerd Kooiker, Inger Plaisier, and Peggy Schyns. Hulp geboden - een verkenning van de mogelijkheden en grenzen van (meer) informele hulp. *Sociaal en Cultureel Planbureau*, 2014.

Rajeev H Dehejia and Sadek Wahba. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, 84(1):151–161, 2002.

Cinzia Di Novi and Elenka Brenna. Is caring for elderly parents detrimental for women's mental health? The influence of the European North-South gradient. *The Influence of the European North-South Gradient (October 18, 2013). University Ca'Foscari of Venice, Dept. of Economics Research Paper Series*, (23), 2013.

Young Kyung Do, Edward C Norton, Sally Stearns, and Courtney H Van Houtven. Informal care and caregiver's health. Technical report, National Bureau of Economic Research, 2013.

Colombo Francesca, Llena-Nozal Ana, Mercier Jérôme, and Tjadens Frits. *OECD Health Policy Studies Help Wanted? Providing and Paying for Long-Term Care*, volume 2011. OECD Publishing, 2011.

Michael Lechner. Long-run labour market and health effects of individual sports activities. *Journal of Health Economics*, 28(4):839–854, 2009.

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

Edwin Leuven and Barbara Sianesi. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Statistical Software Components*, 2014.

Debbie Oudijk, Isolde Woittiez, and Alice de Boer. More family responsibility, more informal care? The effect of motivation on the giving of informal care by people aged over 50 in the Netherlands compared to other european countries. *Health policy*, 101(3):228–235, 2011.

Martin Pinquart and Silvia Sörensen. Correlates of physical health of informal caregivers: a meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 62(2):P126–P137, 2007.

Martin J Prince, Friedel Reischies, AT Beekman, Rebecca Fuhrer, Catholijn Jonker, Sirkka-Liisa Kivela, BA Lawlor, Anthony Lobo, Halgrimur Magnusson, Manfred Fichter, et al. Development of the EURO-D scale–a European Union initiative to compare symptoms of depression in 14 European centres. *The British Journal of Psychiatry*, 174(4):330–338, 1999.

David Sven Reher. Family ties in western Europe: persistent contrasts. *Population and development review*, pages 203–234, 1998.

Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Paul R Rosenbaum and Donald B Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38, 1985.

Donald B Rubin. The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, pages 185–203, 1973.

Donald B Rubin. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3-4):169–188, 2001.

Donald B Rubin and Elizabeth A Stuart. Affinely invariant matching methods with discriminant mixtures of proportional ellipsoidally symmetric distributions. *The Annals of Statistics*, pages 1814–1826, 2006.

Donald B Rubin and Neal Thomas. Matching using estimated propensity scores: relating theory to practice. *Biometrics*, pages 249–264, 1996.

Hendrik Schmitz and Magdalena A Stroka. Health and the double burden of full-time work and informal care provisionEvidence from administrative data. *Labour Economics*, 24:305–322, 2013.

Hendrik Schmitz and Matthias Westphal. Short-and medium-term effects of informal care provision on health. 2013.

Richard Schulz and Scott R Beach. Caregiving as a risk factor for mortality: the caregiver health effects study. *Jama*, 282(23):2215–2219, 1999.

SHARE. Survey of Health, Ageing and Retirement in Europe, brochure. *www.share-project.org*, 2013.

Helene Starks and Melissa M Garrido. Observational & quasi-experimental research methods. 2014.

Elizabeth A Stuart. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1):1, 2010.

Bernard Van den Berg and Ada Ferrer-i Carbonell. Monetary valuation of informal care: the well-being valuation method. *Health economics*, 16(11): 1227–1244, 2007.

Courtney Harold Van Houtven, Michele R Wilson, and Elizabeth C Clipp. Informal care intensity and caregiver drug utilization. *Review of Economics of the Household*, 3(4):415–433, 2005.