MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany · Tel +49 (0) 3 81 20 81 - 0 · Fax +49 (0) 3 81 20 81 - 202 · www.demogr.mpg.de

MPIDR Working Paper WP 2021-005 I March 2021
https://doi.org/10.4054/MPIDR-WP-2021-005

Care or self-care? The impact of informal care provision on health behaviour

Peter Eibich I eibich@demogr.mpg.de

Care or self-care? The impact of informal care provision on health behaviour

Peter Eibich[1]

Abstract

This study examines the impact of informal care provision on carer's health behaviour in England using data on men and women aged 40-69 from the English Longitudinal Study of Ageing (ELSA) covering the period 2002-2017. I evaluate the validity of several instrumental variables for care provision suggested in the literature. I find little evidence of negative effects of informal care provision on health behaviour. The preferred instrumental variable specifications suggest that informal caregivers are more likely to exercise, and women providing care are less likely to smoke.

Keywords: informal care; health behaviour; England; ELSA; instrumental variables

JEL codes: I12, J14, C26

Acknowledgements: I would like to thank Hendrik Schmitz and Matthias Westphal for their detailed comments and suggestions on a previous version of the paper. I also gratefully acknowledge comments and suggestions by seminar participants at the University of Oxford, Hitotsubashi Institute for Advanced Studies, Université Paris Dauphine, and the Max Planck Institute for Demographic Research as well as conference participants at HESG Winter 2021.

[1] Max Planck Institute for Demographic Research, Konrad-Zuse-Str. 1, 18057 Rostock, Germany. E-mail: eibich@demogr.mpg.de.

1 Introduction

As populations in many countries are ageing, an increasing share of older adults depend on social (or long-term) care to cope with activities of their daily life, including shopping, preparing meals or dressing and washing. A large (and growing) share of this care is provided informally, i.e., by family members and friends rather than paid professionals. For example, the UK Office for National Statistics (ONS) estimated the value of informal care (i.e., provided by unpaid adult carers) as £56.9 billion per year (ONS, 2017). This is more than twice as high as public spending on social care in England, Scotland and Wales at £21.3 billion. According to the same report, the amount of time spent by individuals aged 50 and above on informal care has increased by 15% for men and 21% for women between 2000 and 2015 (ONS, 2017).

Policy makers in several countries, including the UK, have promoted and encouraged informal care provision to lower the financial burden on social care services imposed by population ageing. The academic literature largely confirms that informal care provision can substitute for formal care (e.g., Bonsang, 2009; Van Houtven and Norton, 2004), yet there is also mounting evidence of the costs borne by informal caregivers. Informal care is often provided at the expense of the carer's career (Bolin et al., 2008). Caregivers might reduce their working hours or exit the labour market (temporarily or permanently) to cope with the double-burden of employment and care-giving (Van Houtven et al., 2013). Moreover, in the long-run caregivers also face substantial wage penalties (Schmitz and Westphal, 2017). Beyond wages and employment, there is also substantial evidence of a negative health impact of informal caregiving, in particular for mental health (Chen et al., 2019; Coe and Van Houtven, 2009; de Zwart et al., 2017; Do et al., 2015; Schmitz and Stroka, 2013; Schmitz and Westphal, 2015; Van Houtven et al., 2005; Wakabayashi and Kureishi, 2018).

Yet, the mechanisms behind these health effects remain unclear. In this study, I examine the effect of informal care provision on health behaviour in England as a potential mechanism. In addition, I also contribute to the literature by systematically evaluating the validity of several instrumental variables for informal care provision that have been proposed in previous studies. Thus, the results reported in this study might inform identification strategies of future studies on the effects of informal care provision.

Theoretical economic models, such as the Health Capital Model (Grossman, 1972), suggest time trade-offs as a potential mechanism. Caregivers might reduce their engagement in time-intensive health behaviours (such as exercise or preventive care use) to cope with the demand

for informal care, similar to the trade-off between employment and caregiving. Following this line of thought, a negative impact of caregiving on health behaviour might (partly) explain the negative health effects. Alternatively, the Health Belief Model in health psychology (Rosenstock, 1966) emphasizes the role of perceived risk and severity of diseases and "cues to action" for health behaviour adoption. Informal caregivers are more frequently confronted with the consequences of (chronic) diseases in old age, which might prompt them to adopt a healthier lifestyle. Similar learning effects have, e.g., been observed in response to health shocks within the family (Fadlon and Nielsen, 2019).

The potential impact of informal care provision on health behaviour matters for health and social care policy. Negative effects of informal care provision on health behaviour might imply that caregiving has long-lasting negative consequences for health. Public health authorities and policy makers could consider interventions to allow individuals to meet the demand for care while maintaining a healthy lifestyle, similar to interventions such as carer's allowances and caregiving leave that aim to reconcile caregiving and employment.

Evaluating the consequences of informal care provision is complicated by selection into caregiving. Caregiving can be conceptualised as a process that only occurs if there is a demand for informal care, and an individual who is willing to meet this demand and provide care. Both the demand and supply of informal care might be affected by selection. The demand for informal care often arises due to the presence of a family member (typically a spouse or parent) with health problems or functional limitations. The incidence of such limitations might be correlated with unobserved characteristics of the care recipient (e.g., education or socioeconomic status), which might be correlated with the caregiver's own characteristics. There are also likely characteristics of the caregiver that influence their willingness or capability to provide care as well as their health behaviour. The most salient example is health: Individuals are more likely to provide care if they are in good health themselves, and health is both influenced by health behaviour and shapes health behaviour, e.g., by enabling individuals to engage in exercise.

Studies on the consequences of informal care provision have drawn on a variety of approaches to address these selection issues, such as matching (Brenna and Di Novi, 2016; de Zwart et al., 2017; Schmitz and Westphal, 2017, 2015), panel data methods (Kaschowitz and Brandt, 2017; Schmitz and Stroka, 2013) and instrumental variable estimation (Coe and Van Houtven, 2009; Do et al., 2015; Heger, 2017; Van Houtven and Norton, 2004; Wakabayashi and Kureishi,

2018). Instrumental variable estimation is a powerful approach that can potentially address all possible sources of bias, but it requires a valid instrument. The literature has proposed several such instruments, including physical limitations of close family members (Do et al., 2015), death of a parent (Coe and Van Houtven, 2009), sibling characteristics (Coe and Van Houtven, 2009; Van Houtven and Norton, 2004) or the presence of a single parent (Heger, 2017). Yet, the validity of these instruments remains unclear. For example, it seems plausible that instruments based on the health status of family members might have a direct effect on informal caregivers. Indeed, recent studies have distinguished between the negative health effects of informal care provision and the separate impact of care recipient's health shocks on their caregiver's health (Bobinac et al., 2010; Bom et al., 2018; Heger, 2017).

This study uses eight waves of data from the English Longitudinal Study of Ageing (ELSA), covering the period 2002-2017. I estimate the effect of informal care provision on caregiver's health behaviour for men and women aged 40-69 in England. I systematically examine the validity of several instruments for informal caregiving based on family characteristics, drawing on a novel test of instrument validity proposed by Mourifié and Wan (2017) as well as heteroskedasticity-based instruments (Lewbel, 2012). The first-stage results as well as the instrument validity test suggest that the validity of the instruments remains questionable. The test fails to reject the null hypothesis of instrument validity only when the sample is restricted to individuals for whom a care spell is observed in the panel. This suggests that the instruments are affected by selection into caregiving, but they can identify valid variation in the timing of caregiving.

Results from the preferred IV specification provide little evidence for a negative effect of caregiving on health behaviour. Instead, informal care provision rather improves health behaviour, such as the frequency of moderate exercise for men. These conclusions prove robust to several changes of the model specification as well as an alternative identification strategy using the post-double selection lasso with fixed effects. Interestingly, I also do not find evidence for negative effects of informal care provision on health. This finding is in line with a recent study on care provision in the U.K. (Bom and Stöckel, 2021). Although in contrast to Bom and Stöckel (2021), I also find no evidence of negative effects of high intensity care provision, these differences might simply reflect differences in the underlying data and should therefore not be overinterpreted.

Thus, informal care provision does not seem to negatively affect carer's health behaviour. On the contrary, there is some evidence that caregivers adopt healthier behaviour, which (in line with the Health Belief Model) might be explained by learning effects through caregiving, or because care provision acts as a trigger for health behaviour change.

The remainder of the paper is organized as follows: Section 2 describes the methodological approaches used in the analysis. Section 3 provides an overview of the data. The main results are presented in section 4. Section 5 discusses the findings and concludes.

2 Methods

2.1 IV estimation

Informal care provision is not randomly assigned, and instead it depends on the caregiver's capabilities and willingness to provide care (i.e., the supply of care) as well as on the care recipient's needs (i.e., the demand for care). Both demand and supply of care are likely to be correlated with observable and unobservable characteristics of the caregiver, which in turn might be related to differences in health behaviour. For example, poor health might limit an individual's capabilities to provide care for someone else, and it might also limit their ability to exercise. Such characteristics can introduce omitted variable bias. In addition, there might be simultaneity, if the beginning or end of a care spell coincides with health shocks of the care recipient (Bobinac et al., 2010). In this paper, I use instrumental variable (IV) estimation to address the endogeneity of caregiving.

IV models can resolve all potential sources of bias by exploiting variation in an instrumental variable that is correlated with informal care provision but unrelated to health behaviour. The model can be represented as follows:

$$Care_{it} = X_{it}\beta + Z_{it}\theta + c_i + \delta_t + \nu_{it} \quad (1)$$

$$y_{it} = Care_i\tau + X_{it}\beta + W_{jt}\gamma + c_i + \delta_t + \epsilon_{it} \quad (2)$$

The first stage in eq. (1) predicts informal care provision using observed characteristics of the individual[2] i (X_{it})as well as an additional, instrumental variable Z_{it}. Intuitively, the second

[2] This might include characteristics of the potential care recipient, but would require that the potential care recipient can be identified, e.g., because the study focuses on parental or spousal caregiving, and the potential care recipient is also observed in the data.

stage model in eq. (2) then estimates the causal effect of care provision by regressing health behaviour on predicted care provision.

The instrumental variable Z needs to satisfy three conditions to ensure that the IV model identifies a *local average treatment effect* (Farbmacher et al., 2020):

1. *Relevance:* The instrument should have an effect on informal care provision independently of all other terms included in eq. (1).
2. *Validity*: The instrument should be assigned as good as random (2a), and it should not affect the outcome through any other mechanism than through its effect on the treatment (2b).[3]
3. *Monotonicity*: The instrument should affect the treatment propensity of all individuals in the same direction.[4]

The relevance of an instrument can be readily assessed in empirical studies using the estimates from the first stage of the model, and by comparing the value of the F-statistic on the strength of the excluded instruments to commonly used benchmark values (Staiger and Stock, 1997). However, evaluating the validity and monotonicity conditions is not trivial, and suitable tests of these conditions have only recently been proposed in the econometric literature (Farbmacher et al., 2020; Huber and Mellace, 2014; Kitagawa, 2015; Mourifié and Wan, 2017).

Previous studies on the effects of informal care provision have proposed several instruments, such as health status of parents or parents-in-law (Coe and Van Houtven, 2009; Do et al., 2015) or family characteristics (Coe and Van Houtven, 2009; Heger, 2017; Van Houtven and Norton, 2004). However, especially the validity of these instruments remains unclear. For example, if health shocks experienced by a close relative indeed exert an impact on a person's health regardless of whether they provide informal care or not, then instruments based on the potential care recipient's health might not be valid as they violate the exclusion restriction. Similarly, family characteristics such as the number of siblings might be correlated with other individual traits than informal care provision.

In this study, I consider four candidate instruments derived from the existing literature: whether the respondent has living siblings (Coe and Van Houtven, 2009; Van Houtven and Norton, 2004), whether the individual's mother is alive, whether the father is alive (Coe and Van

[3] Condition 2a is sometimes referred to as "unconfoundedness", and condition 2b is frequently referred to as the "exclusion restriction".
[4] This assumption is colloquially referred to as the "no defier" assumption".

Houtven, 2009; Heger, 2017), and whether only one parent is still alive (Heger, 2017).[5] The intuition behind these candidate instruments is that having a living mother or father should increase the probability to provide informal care. Similarly, having only one living parent should increase the caregiving probability, because in families where both elderly parents are still alive one of the parents might act as the primary caregiver for their spouse. In contrast, the number of living siblings is expected to decrease the caregiving probability, because the siblings might assume the role of caregivers.

These instruments might be influenced by variation in the survival of family members (which might in turn be related to health and health behaviour of the individual), thus raising concerns about the validity of the instruments. In particular, it is plausible that the death of a parent might coincide with the end of a caregiving episode (or indeed the beginning of a care spell if the surviving parent has previously received care from their deceased spouse). It is necessary to disentangle changes in caregiving from such events, because the loss of a parent might exert a direct impact on the potential caregiver's health and health behaviour. In previous studies, this has been referred to as the "family effect" or the "effect of caring about someone" (Bobinac et al., 2010; Heger, 2017). To address this effect, I control for the recent death of the respondent's father or mother in all models. Controlling for the recent death of a parent implies that the variation in caregiving identified by the instrument should not be driven by short-term mortality shocks, rather these instruments rely on variation in long-term survival. Consequently, these control variables should mitigate concerns about the "family effect", but they do not address all concerns about the validity of these instruments.

2.2 Testing IV assumptions

I assess the validity and monotonicity of these instruments using the test proposed by Mourifié and Wan (2017). The test builds on work by Balke and Pearl (1997) and Heckman and Vytlacil (2005), who derived testable implications of the joint conditions of instrument validity and monotonicity. These implications can be expressed in the form of two inequalities (Mourifié and Wan, 2017):

$$P(Y, D = 1|Z = 0) \leq P(Y, D = 1|Z = 1) \quad (3)$$

$$P(Y, D = 0|Z = 1) \leq P(Y, D = 0|Z = 1) \quad (4)$$

[5] Instruments based on the potential care recipient's health would require me to restrict the sample to individuals living in the same household as the potential care recipient, as otherwise their health status is not observed. 73% of caregiving spells in the data are extraresidential (see section 3 for details), and I therefore do not consider such instruments here.

Equations (3) and (4) essentially state that the joint distributions of the outcome (*Y*) and treatment (*D*) (or the "treated" and "untreated" outcomes) for those affected by the instrument ($Z = 1$) and those unaffected by the instrument ($Z = 0$) are nested within each other (Kitagawa, 2015). Intuitively, these equations imply that the probability of observing a treated outcome should be higher for those affected by the instrument than for those unaffected by the instrument for any interval in the support of *Y*, and vice versa the probability of observing an untreated outcome should be higher for those unaffected by the instrument. These inequalities should hold if validity and monotonicity hold, because the instrument (without loss of generality) increases the probability of treatment but is otherwise independent of the outcome. Tests based on these inequalities are not sufficient, i.e., it is possible that these inequalities hold even though the validity or monotonicity condition are violated. Therefore, the test should be considered as a falsification exercise to detect certain violations of the validity and monotonicity conditions.

Mourifié and Wan (2017) show that these inequalities in eq. (3) and (4) can be characterised as conditional moment inequalities as follows:

$$\theta(y,1) \equiv E[c_1 D(1-Z) - c_0 DZ | Y = y] \leq 0 \qquad (5)$$

$$\theta(y,0) \equiv E[c_0(1-D)Z - c_1(1-D)(1-Z) | Y = y] \leq 0 \qquad (6)$$

with $c_k = \Pr(Z = k)\,, k = 0{,}1$.

This characterisation has the advantage that the proposed test can be implemented using the intersection bounds framework by Chernozhukov et al. (Chernozhukov et al., 2015, 2013), for which software packages readily exist. Formally, the proposed test can be stated as follows:

$$H_0: \theta_0 \equiv \sup_{v \in \mathcal{V}} \theta(v) \leq 0 \qquad \text{and} \qquad H_1: \theta_0 > 0;$$

i.e., the test considers whether the supremum of the conditional moment inequalities described in eq. (5) and (6) is non-positive for all intervals in the support of the outcome variable.

2.3 Lewbel instruments

I also consider an alternative approach to test the validity of these candidate instruments, which exploits heteroskedasticity-based instruments (referred to as "Lewbel instruments" in the following). Lewbel (2012) shows that in certain models identification can be achieved even when valid excluded instruments are not available by exploiting information that is contained within the heteroskedastic standard errors of the model.

Under the assumptions that the standard errors of the first stage, ν_{it}, are heteroskedastic and that $Cov(X, \nu\epsilon) = 0$ and $Cov(X, \nu^2) \neq 0$, it is possible to construct instruments of the form $\hat{Z} = (X - \bar{X})\hat{\nu}$ (Lewbel, 2018, 2012), which can then be used in a linear two-stage least squares model (Baum and Lewbel, 2019). Lewbel (2012) argues that the required assumptions on heteroskedasticity are likely to be relevant for many common economic models (at least for continuous endogenous variables).

Identification based upon technical restrictions rather than exclusion restrictions may be met with (reasonable) scepticism. Yet, Lewbel instruments offer an important advantage. It is possible to include both excluded instruments and Lewbel instruments in the same model, which then allows calculating test statistics for overidentification (e.g., the Sargan-Hansen test) as an alternative to assess the joint validity of the excluded and the constructed instruments.

Even if the necessary assumptions for both the excluded and the constructed instruments are violated at the same time, it seems unlikely that the bias should operate in the same direction. Therefore, I use the Lewbel instruments for two purposes – *(i)* to test the validity of the excluded instruments using the Sargan-Hansen test of overidentification, and *(ii)* to examine the robustness of the conclusions towards a different set of identifying assumptions.

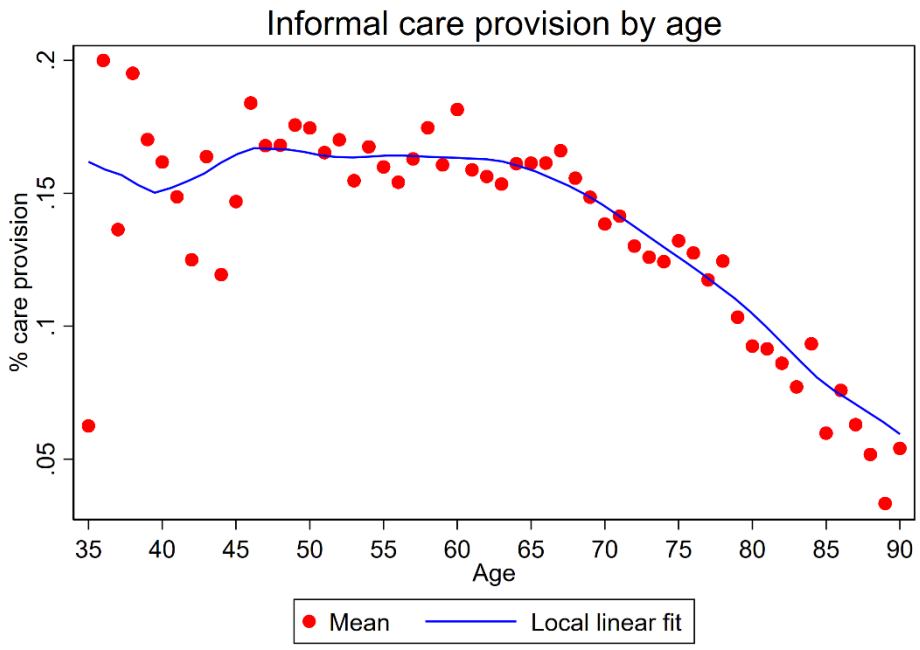
3 Data

3.1 Study description

This study draws on data from the English Longitudinal Study of Ageing (ELSA) (Clemens et al., 2019). ELSA is designed to be representative of individuals aged 50 and above living in private households in England. Cohabiting partners of respondents are included in the survey regardless of age. Data is collected every two years since 2002. I use the 30th release of the study, which includes data from eight waves covering the period 2002-2017. This study does not involve primary data collection, and ethical approval was therefore not deemed necessary.

The working sample includes all individuals aged 40-69. ELSA includes very few individuals below age 40, and Figure 1 shows that from age 70 onwards, informal care provision becomes less and less common (and likely increasingly selective). Therefore, I exclude individuals outside this age range from the sample. I also exclude individuals that were living in an institution at the time of follow-up.

Fig. 1 Informal care provision by age



Source: ELSAv30, own calculations. The markers show the share of individuals providing informal care at each age (in years). The line shows a local linear fit.

3.2 Outcomes

3.2.1 Health behaviour

The outcomes measure participation in primary and secondary prevention. In particular, I examine smoking behaviour, exercise, and screening for breast, bowel and prostate cancer.

Smoking behaviour is included in every wave of ELSA. Respondents are asked whether they have ever smoked in the past or whether they are currently smoking. I construct a binary measure that takes on the value of 1 if respondents report to smoke currently, and 0 otherwise.

For exercise participation, respondents are asked how often they participate in sports or activities that are mildly energetic, moderately energetic, and vigorously energetic. Possible answers are “more than once a week”, “once a week”, “one to three times per month” and “hardly ever or never”. I construct two binary indicators that measure whether individuals report to participate at least once a week in *(i)* moderate exercise, and *(ii)* vigorous exercise. These measures are likewise available in every wave.

Screening participation was only included from wave 5 (2010/11) onwards. Respondents are asked whether they ever had a mammography (screening for breast cancer), a prostate-specific

antigen (PSA) test (screening for prostate cancer) or whether they had completed a test kit for bowel cancer. If the respondent report having participated in screening in the past, they are asked how long ago their last screening was. Unfortunately, the exact wording of these questions as well as the filters of the survey changed over the years. In wave 5, the questions asked more generally about any type of screening, whereas in wave 7 these questions specifically referred to the NHS screening programs and were only asked to respondents within the relevant age range of the screening program. Therefore, I construct outcome measures based on the NHS screening guidelines. For each type of screening, I construct two outcomes: First, a binary outcome that measures whether individuals ever participated in screening. Second, I construct an indicator whether individuals have recently screened for the respective type of cancer based on the screening interval of the NHS programmes. For breast cancer, this interval is 3 years for women aged 50 to 70. Bowel cancer screening is recommended to individuals aged 60-74 every two years. There is no prostate cancer screening programme, and instead men need to request a PSA test from their physician. I therefore construct an indicator whether men have completed a PSA test in the last 12 months. For all measures, the sample is restricted to individuals eligible in the NHS screening programs. For breast cancer screening, I estimate models using a sample of women aged 50-69. For bowel cancer screening, I estimate models for men and women aged 60-69 (due to the upper age limit for the working sample), and for prostate cancer screening I estimate models for men aged 40-69.

3.2.2 Health outcomes

I also consider five measures of health to examine whether findings on the health effects of informal care provision can be replicated in the ELSA data (Bom and Stöckel, 2021). I use a measure of poor self-reported health, which takes on the value of 1 if respondents reported their general health status as “fair” or “poor”, and 0 if they report their health status as “excellent”, “very good” or “good”. This measure of self-reported health was adopted from the Health and Retirement Study, and it was included in ELSA waves 1, 2 and 4 through 8, but not in wave 3 where a different measure adopted from the Health Survey for England was included.

In addition, I consider two measures derived from the eight item CES-D scale, which assesses the presence of depressive symptoms. I use these eight items to construct two measures – *(i)* a CES-D score, which measures the total number of symptoms reported by the respondent (ranging from 0 to 8), and *(ii)* a binary indicator for individuals who reported 3 or more symptoms and who would thus be considered to suffer from depression (White et al., 2016).

I use two scores to measure respondent's functional limitations – *(i)* limitations in activities of daily living (ADLs), *(ii)* and limitations in instrumental activities of daily living (IADLs). These scores are derived from a battery of questions, which ask individuals whether they had difficulties with certain activities due to "physical, mental, emotional or memory problems". Difficulties that were expected to last less than three months should be excluded. The following activities were included in the ADL score: *(i)* dressing, *(ii)* walking across a room, *(iii)* bathing or showering, *(iv)* eating, *(v)* getting in and out of bed, and *(vi)* using the toilet. The IADL score referred to the following activities: *(i)* using a map to figure out how to get around a strange place, *(ii)* recognising when in physical danger, *(iii)* preparing a hot meal, *(iv)* shopping for groceries, *(v)* making telephone calls, *(vi)* communication, *(vii)* taking medication, *(viii)* work around house and garden, and *(ix)* managing money. The ADL and IADL score were derived by summing over all available items. It should be noted that two items in the IADL score (communication and physical danger) were only included from wave 4 onwards, and therefore the IADL scores in waves 1-3 have a lower maximum value than from wave 4 onwards.[6]

3.3 Informal care

For the main analysis, I compare individuals providing any amount of informal care to individuals not providing informal care (i.e., the external margin). This binary measure of informal care provision is based on a question which asked respondents: *"Did you look after anyone in the past week (including your partner or other people in your household)?"*

If respondents state that they did provide care in the past week, they are asked a number of follow-up questions, including the number of hours they provided care, their relationship to the care recipient and whether the care recipient lives with the caregiver. Based on these questions, I also examine the intensity of caregiving by using the self-reported number of hours of care per week as a continuous variable, as well as a set of binary variables measuring whether individuals provided *(i)* any amount of informal care, but less than 10 hours per week ("low intensity care"), *(ii)* between 10 and 39 hours of care per week ("medium intensity"), or *(iii)* more than 40 hours of informal care per week ("high intensity").

[6] I do not exclude these items, because it seems plausible that respondents from wave 4 onwards reported limitations in these categories that were reported under another category in waves 1-3. For example, respondents reporting difficulties in communication in wave 4 might have reported difficulties in making telephone calls in wave 3, but did not report this category in wave 4. In this scenario, excluding communication from the list would artificially reduce the IADL score.

3.4 Covariates

All models control for age in years, a binary indicator for married individuals (and those in a civil partnership), and educational attainment. Educational attainment is coded in seven levels: *(i)* degree or equivalent, *(ii)* higher education below degree level, *(iii)* A-levels or equivalent, *(iv)* O-levels or equivalent, *(v)* compulsory school education, *(vi)* foreign or other educational attainment, and *(vii)* no qualification. I also control for the recent death of a father or mother. These indicators show whether the respondent reported having a living parent in the previous wave but not in the current wave.

3.5 Instruments

I evaluate four potential instrumental variables based on previous studies (Coe and Van Houtven, 2009; Heger, 2017; Van Houtven and Norton, 2004). A binary indicator for whether respondents have at least one living sibling is based on the self-reported number of living siblings. Unfortunately, the data does not allow refining the instrument based on the sibling's gender, age or distance to the respondent's parents. Such information is only available for siblings that live in the same household as the respondent.

Respondents are also asked at every wave whether their father and their mother are still alive. I also construct a binary indicator for whether only one parent is still alive. This indicator takes on the value of 0 if a respondent reported that their father and mother are either both alive or both deceased, and it takes the value of 1 if they reported that either the mother is alive and the father deceased, or the father is alive and the mother deceased.

3.6 Descriptive statistics

Table 1 below shows descriptive statistics for our working sample. Note that about 15% of the sample provide any amount of informal care. The unconditional mean intensity is relatively low at 5.5 hours per week, however, among those that provide any amount of care the average duration is considerably higher at 38 hours per week. This high conditional average is driven by a considerable share of caregivers that report providing care "all the time", which is coded as 168 hours of care per week. The conditional median of care intensity is lower at about 10 hours per week.

Table 1: Summary Statistics

Variable	Mean	SD	Difference by caregiving		N
A. Outcomes					
Non-smoker	0.830	0.376	-0.002		50,240
Moderate exercise	0.812	0.391	0.036	***	50,877
Vigorous exercise	0.349	0.477	0.010	*	50,874
Recent bowel cancer screening	0.407	0.491	0.032	***	18,590
Ever bowel cancer screening	0.509	0.500	0.031	***	18,590
Recent PSA test	0.149	0.356	0.026	**	8,147
Ever PSA test	0.349	0.477	0.039	**	8,147
Recent mammography	0.679	0.467	0.009		11,140
Ever mammography	0.939	0.240	0.018	***	11,140
B. Caregiving					
Informal caregiving	0.146	0.354	1.000	***	49,939
Caregiving hours per week	5.546	25.512	38.111	***	49,885
C. Demographic characteristics					
Age	59.664	5.886	-0.010		51,063
Women	0.553	0.497	0.142	***	51,063
Married	0.723	0.448	0.038	***	51,063
Working	0.500	0.500	-0.091	***	50,857
D. Education					
University degree	0.188	0.391	-0.028	***	50,235
Higher education below degree	0.140	0.347	0.007		50,235
A-levels	0.094	0.291	0.002		50,235
O-Levels	0.206	0.405	0.014	***	50,235
NVQ1/CSE	0.037	0.189	-0.006	***	50,235
Foreign qualification	0.101	0.301	0.014	***	50,235
No qualification	0.234	0.423	-0.004		50,235
E. Income and Wealth					
Decile equiv. income	6.056	2.894	-0.230	***	42,462
Decile net financial wealth	5.729	3.007	-0.069	*	42,462
F. Health					
Poor health (0/1)	0.224	0.417	-0.025	***	42,957
CES-D score	0.916	1.672	0.034		51,063
Depressed (0/1)	0.128	0.334	0.011	**	51,063
ADL score	0.283	0.876	-0.054	***	50,943
IADL score	0.273	0.858	-0.063	***	50,943
Recent death of father	0.029	0.167	0.006	**	35,199
Recent death of mother	0.051	0.219	-0.010	***	35,483
G. Instrumental Variables					
At least one living sibling	0.854	0.353	0.000		50,900
Father alive	0.142	0.349	0.041	***	50,314
Mother alive	0.308	0.462	0.138	***	50,705
Both parents alive	0.091	0.287	0.016	***	50,157
Only one parent alive	0.268	0.443	0.148	***	50,157

Source: ELSA Waves 1-8, own calculations. Column 4 shows the estimated difference between caregivers and non-caregivers, and Column 5 shows the results from a t-test for mean differences. Significance: * p<0.1, ** p<0.05, *** p<0.01.

About 73% of the caregivers in our sample provide care for someone living outside their own household. 33% of caregivers report looking after a parent, 20% provide care for a spouse, 22% care for a grandchild and 14% report caring for a friend.

Looking at differences between caregivers and non-caregivers, I note that caregivers are more likely to be female and married. Caregivers are less likely to be working, they are less likely to have a university degree, and their wealth and income are significantly lower than those of non-caregivers. This might suggest that individuals select into caregiving based on their (lack of) financial capability to purchase formal care. Evidence on the selection on health is mixed – caregivers report fewer limitations in ADLs and IADLs, and they are less likely to report their health as fair or poor. On the other hand, they also report higher CES-D scores and they are more likely to report at least three depressive symptoms. Looking at health behaviour, caregivers seem to be more likely to exercise frequently, and they report higher rates of participation in cancer screening (with the exception of recent mammography screening).

4 Results

4.1 First-stage estimates and instrument validity

Before estimating the effect of informal care provision on health behaviour, I examine the validity of the four candidate instrumental variables. First, I conduct the instrument validity test by Mourifié and Wan (2017) separately for men and women for all outcome variables. Although the outcomes are conceptually closely related, it is nevertheless possible that, e.g., the exclusion restriction holds for some but not all outcomes. Therefore, employing the test for all outcomes should be more informative than only considering one selected outcome. I include dummy variables for the recent death of a father or mother in all models to account for potential violations of the exclusion restriction that might occur if the death of a father or mother influences the caregiving probability and has a direct impact on the outcome (i.e., the "family effect").

The results of the test are shown in Table 2 below. Each cell shows the confidence level at which the test rejects the null hypothesis of joint instrument validity and monotonicity for the specified outcome and candidate instrument. To facilitate the interpretation of the results, cells in which the null hypothesis is not rejected at the ten percent level are highlighted in bold (**p>0.1**), and cells in which the null hypothesis is rejected at the ten percent level or five percent level (but not at the one percent level) are highlighted in italics (*p<0.1* or *p<0.05*). In the full sample (shown in panel A), the test rejects the null hypothesis for all combinations of candidate

instrument and outcome considered here. This implies that the four candidate instruments cannot be used to identify exogenous variation in informal care provision. The test by Mourifié and Wan (2017) is a joint test of instrument validity and monotonicity, and it is not possible to determine from the test which of the two conditions is violated individually. A direct effect of these instrumental variables on health behaviour (i.e., a violation of the exclusion restriction) may seem unlikely, especially after controlling for the recent death of a father or mother. In contrast, it appears plausible that these instruments are not unconfounded. For example, parental socioeconomic status may influence both survival of the parents (i.e., the potential care recipients) and health behaviour of their children (i.e., the potential caregivers).

To address such concerns, I consider an alternative sample, which only includes observations from individuals for whom a care spell is observed in the data. In this sample, all individuals provide care at some point in their lives, and therefore the candidate instruments do not identify variation in treatment status (i.e., variation between individuals that provide care and those who do not), but rather variation in the timing of treatment. If violations of instrument validity are indeed caused by selection on (largely time-invariant) characteristics such as parental socioeconomic status, then we might expect that considering only variation in the timing of treatment will reduce such concerns. The results for the instrument validity test in this alternative sample are shown in panel B of Table 2. The test fails to reject the null hypothesis for "mother alive" and "only one living parent" for almost all outcomes. For these two instruments, the null hypothesis is only rejected for ADLs and IADLs for men (at the one percent level) and women (at the ten percent level) as well as for poor health for men (at the five percent level). In contrast, for "father alive" and "at least one living sibling" the null hypothesis is rejected at the one percent level for all outcomes.

Table 2: Instrument Validity

	Men				*Women*			
Outcome	**Father alive**	**Mother alive**	**Only one parent alive**	**At least one living sibling**	**Father alive**	**Mother alive**	**Only one parent alive**	**At least one living sibling**
A. Full sample								
Non-smoker	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01
Frequent moderate exercise	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01
Frequent vigorous exercise	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01
Recent bowel cancer screening	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01
Ever screened for bowel cancer	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01
Recent prostate cancer screening	p<0.01	p<0.01	p<0.01	p<0.01				
Ever screened for prostate cancer	p<0.01	p<0.01	p<0.01	p<0.01				
Recent mammography					p<0.01	p<0.01	p<0.01	p<0.01
Ever mammography					p<0.01	p<0.01	p<0.01	p<0.01
Poor health	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01
CES-D score	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01
Depressed	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01	p<0.01
B. Only individuals with a care spell								
Non-smoker	p<0.01	**p>0.1**	**p>0.1**	p<0.01	p<0.01	**p>0.1**	**p>0.1**	p<0.01
Frequent moderate exercise	p<0.01	**p>0.1**	**p>0.1**	p<0.01	p<0.01	**p>0.1**	**p>0.1**	p<0.01
Frequent vigorous exercise	p<0.01	**p>0.1**	**p>0.1**	p<0.01	p<0.01	**p>0.1**	**p>0.1**	p<0.01
Recent bowel cancer screening	p<0.01	**p>0.1**	**p>0.1**	p<0.01	p<0.01	**p>0.1**	**p>0.1**	p<0.01
Ever screened for bowel cancer	p<0.01	**p>0.1**	**p>0.1**	p<0.01	p<0.01	**p>0.1**	**p>0.1**	p<0.01
Recent prostate cancer screening	p<0.01	**p>0.1**	**p>0.1**	p<0.01				
Ever screened for prostate cancer	p<0.01	**p>0.1**	**p>0.1**	p<0.01				
Recent mammography					p<0.01	**p>0.1**	**p>0.1**	p<0.01
Ever mammography					p<0.01	**p>0.1**	**p>0.1**	p<0.01
ADLs	p<0.01	p<0.01	**p>0.1**	p<0.01	p<0.01	**p>0.1**	*p<0.1*	p<0.01
IADLs	p<0.01	p<0.01	**p>0.1**	p<0.01	p<0.01	**p>0.1**	*p<0.1*	p<0.01
Poor health	p<0.01	*p<0.05*	**p>0.1**	p<0.01	p<0.01	**p>0.1**	**p>0.1**	p<0.01
CES-D score	p<0.01	**p>0.1**	**p>0.1**	p<0.01	p<0.01	**p>0.1**	**p>0.1**	p<0.01
Depressed	p<0.01	**p>0.1**	**p>0.1**	p<0.01	p<0.01	**p>0.1**	**p>0.1**	p<0.01

Source: ELSA waves 1-8, own calculations based on Mourifié and Wan (2017). Each cell shows the p-value for the null hypothesis of joint instrument validity and monotonicity for the specified sample, outcome and instrument. The recent death of a father and mother are included as covariates in all models.

I cannot draw definitive conclusions why the test rejects the null hypothesis for "father alive" but fails to reject the null hypothesis for "mother alive" and "only one living parent". It seems possible that father's survival is more strongly affected by selection than mother's survival, since inequality in life expectancy in Europe is stronger for men than for women (Smits and Monden, 2009). Moreover, the impact of having a living parent on the probability of informal care provision might depend on the survival of the other parent. If both parents are still alive, then any demand for informal care might be met by the spouse rather than children. This might imply potential violations of the monotonicity conditions for both "father alive" and "mother alive". However, having a living father often implies that both parents are still alive due to the higher life expectancy of women, while the same is not true for mothers. In the data, 63% of observations with a living father come from respondents with both living parents, whereas only 30% of observations with a living mother also have a living father. Thus, any violations of the monotonicity conditions are likely to be more severe for having a living father than for having a living mother.

Next, I consider the relevance of the candidate instruments by inspecting the first stage results. Table 3 below reports first-stage estimates from a linear 2SLS model.[7] All models control for age (quadratic trend), education, marital status, the recent death of the father and mother as well as year and month of the interview. I report the estimated coefficient and significance level of the candidate instrument as well as the Kleibergen-Paap (KP) Wald-F statistic for the strength of the excluded instrument.

[7] The model reported here uses "non-smoking" as the outcome variable to maximise sample size. The first-stage estimates do not depend on the values of the outcome, but may depend on, e.g., changes in sample size.

Table 3: Instrumental Variables - First-stage estimates

	Men				*Women*			
Instrument	**First-stage estimate**		**KP Wald F**	**N**	**First-stage estimate**		**KP Wald F**	**N**
			A. Full sample					
Father alive	0.048	***	15.500	14,919	0.068	***	34.074	19,039
Mother alive	0.049	***	30.168	14,919	0.136	***	210.517	19,039
Only one parent alive	0.063	***	52.886	14,919	0.137	***	237.995	19,039
Number of living siblings	-0.020	**	4.210	14,884	0.004		0.155	19,016
			B. Only individuals with a care spell					
Father alive	0.057	***	7.269	5,759	0.074	***	21.895	10,400
Mother alive	0.085	***	25.706	5,759	0.149	***	148.913	10,400
Only one parent alive	0.122	***	55.666	5,759	0.146	***	156.447	10,400
Number of living siblings	0.002		0.011	5,744	0.003		0.045	10,383

Source: ELSA waves 1-8, own calculations. Columns 1 and 4 show the estimated effect of the instrument on the caregiving probability from a 2SLS model for men and women. All models use current smoking as the outcome. "KP Wald F" shows the Kleibergen-Paap F-statistic for weak instruments. *** p<0.01, ** p<0.05, * p<0.1. Panel A shows results for the full sample, while panel B shows results for a sample of individuals for whom a caregiving spell is observed in the data.

When looking at the full sample in panel A, I note that a living father, a living mother or having only one living parent are all predictive of informal care provision. For men, all three instruments increase the probability to provide care by 5-6 percentage points, while for women the increase is between 7 (*father alive*) and 14 (*mother alive, only one parent alive*) percentage points. As expected, the Kleibergen-Paap Wald F-statistic also confirms that the instruments are stronger for women than for men, likely because more women provide care than men. However, even for men the F-statistics are above 10, which is often used as a rule of thumb to detect weak instruments. Interestingly, having at least one living sibling has little influence on informal care provision. For men, having a living sibling has a significant negative effect on informal care provision, although the Wald F-statistic indicates that the instrument is very weak. For women, the estimated coefficient on the instrument is insignificant and very close to zero.

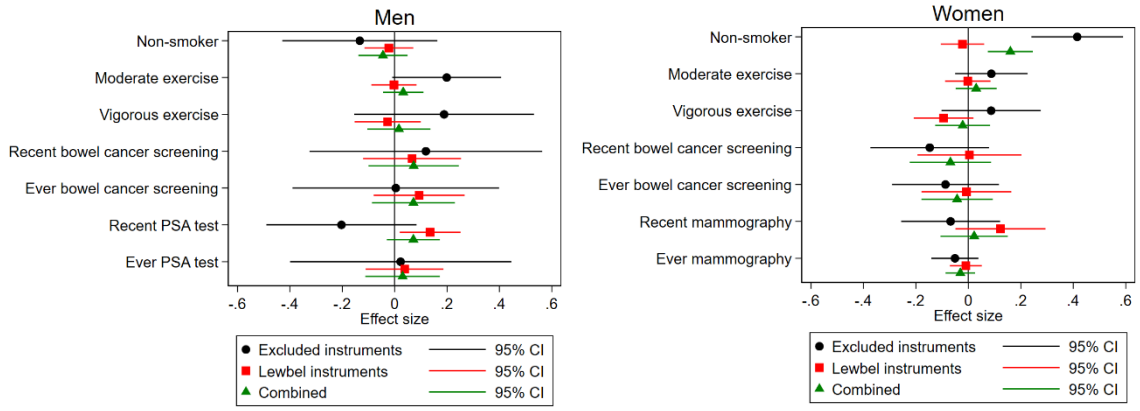
In the second sample, which includes only individuals for whom a care spell was observed in the panel, instruments based on the respondent's parents are still significant predictors of informal care provision. The point estimates are larger for all three instruments, although the Wald F-statistic is reduced for most candidate instruments. This is due to the considerable reduction in sample size. However, the Wald F-statistic on the strength of the excluded

instruments is above the commonly used threshold value for almost all instruments, with the exception of "father alive" for men. Having only one living parent is the strongest instrument for men and women. For both men and women having a living sibling does not seem to predict informal care provision in this sample.

In summary, the candidate instruments considered here do not seem to be valid instruments to identify variation in treatment status, because the instrument validity test suggests that in the full sample instrument validity and monotonicity do not jointly hold. However, when I condition on the presence of a care spell in the panel, it seems that having a living mother and having only one living parent can be used to identify exogenous variation in the timing of informal care provision. Therefore, I will use both of these instruments jointly and only consider the sample restricted to individuals with a care spell within the panel for the following IV analyses.

4.2 Informal care provision and health behaviour

Fig. 2 Effects of informal care provision on health behaviour



Source: ELSA, own calculations. Notes: The figures show regression results from a 2SLS model. Markers show point estimates and the lines show 95% confidence intervals. The excluded instruments are *only one parent* and *mother alive*. All models control for a quadratic trend in age, education, marital status, survey wave as well as the recent death of a mother or father. The sample includes all individuals aged 40-69 who were observed providing informal care in the panel. For bowel cancer screening outcomes the sample is restricted to ages 60-69, and for mammography screening to ages 50-69.

Fig. 2 shows the estimated effects of informal care provision on health behaviour using 2SLS IV models. The models are estimated on the sample restricted to individuals that provided care at any point in the panel. The figure shows estimates from three different models: *(i)* IV models using "mother alive" and "only one living parent" as excluded instruments, *(ii)* models using

Lewbel instruments constructed from heteroskedastic residuals, and *(iii)* models using both excluded and Lewbel instruments.

First, I note that across all specifications there is little evidence of a significant negative effect of informal care provision on health behaviour. For men, the model using only excluded instruments suggests a reduction in the probability of recent PSA testing by about 20 percentage points, which is significant at the 10 percent level. However, both the model using only Lewbel instruments and the model using excluded and Lewbel instruments imply an increase in the likelihood of recent PSA testing, and the Sargan-Hansen test of overidentifying restrictions does not reject the null hypothesis of joint validity of the excluded and constructed instruments (see Table A.1 in the appendix). In contrast, there is some evidence for positive effects of informal care provision on health behaviour. For men, the models using excluded instruments suggest an increase in the probability of frequent moderate exercise, and for women the likelihood to be a non-smoker increases. Both of these effects are significant at the 5 percent level. The Sargan-Hansen test does not reject the null for joint validity of the excluded and constructed instruments for almost all models. For women, the test rejects the null for "non-smoking" at the 10 percent level. Here, the excluded instruments suggest a large increase in the probability to quit smoking, whereas the Lewbel instruments suggest an insignificant effect that is very close to zero. Therefore, the large positive effect implies by the excluded instruments should be interpreted with caution. For women and frequent moderate exercise, the Sargan-Hansen test rejects the null of joint validity at the 5 percent level, although both excluded and constructed instruments imply an insignificant and relatively small effect.

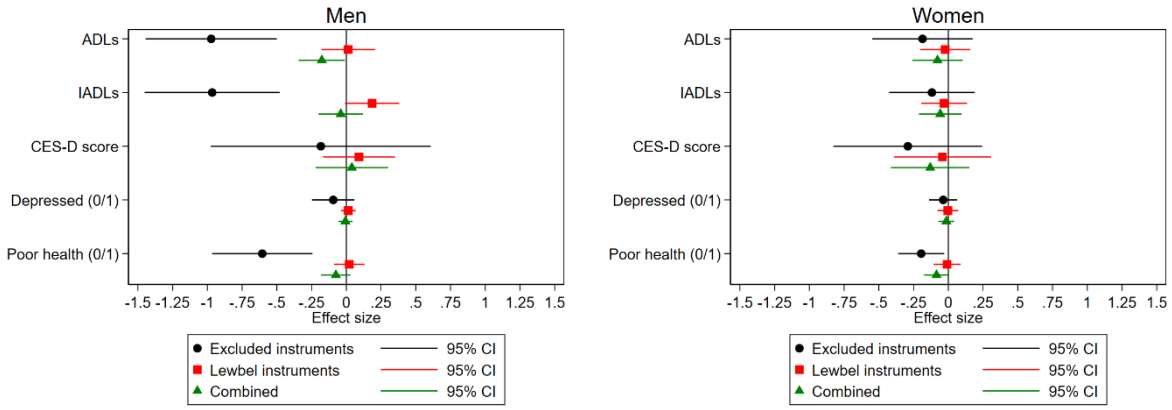
In summary, I conclude that these estimates suggest that there is very little evidence for a negative effect of informal care provision on health behaviour, and instead it seems that informal care provision might have positive effects on certain health behaviour outcomes. In the next step, I examine health outcomes to examine whether I can replicate findings of negative effects of care provision on health reported in previous studies.

4.3 Informal care provision and health

The estimated effects of informal care provision on health are shown in Fig. 3. It should be noted that the outcomes shown in Fig. 3 are measured on different scales – the ADL score ranges from 0 to 6, the IADL score ranges from 0 to 9, and the CES-D score ranges from 0 to 8. In contrast, "poor health" and "depressed" are binary indicators. Interestingly, the results in Fig. 3 suggest that informal care has a positive impact on caregiver's health in England. For

men, I observe a significant reduction in functional limitations (both ADLs and IADLs) in the model using only excluded instruments, and I observe a reduction in the likelihood to report poor health for both men and women in the models using only excluded instruments. Estimates from models using Lewbel instruments or using both excluded and constructed instruments are mostly insignificant and close to zero, with the exception of IADLs for men. Here, the model using only Lewbel instruments indicates an increase in functional limitations in instrumental activities of daily living.

Fig. 3 Effects of informal care provision on health



Source: ELSA, own calculations. Notes: The figures show regression results from 2SLS models. Markers show point estimates and the lines show 95% confidence intervals. The excluded instruments are *only one parent* and *mother alive*. All models control for a quadratic trend in age, education, marital status, survey wave as well as the recent death of a mother or father. The sample includes all individuals aged 40-69 who were observed providing informal care in the panel.

Although such positive effects of informal care provision on health might seem surprising, they are in line with earlier findings by Bom and Stöckel (2021), who report a positive effect of informal care provision on physical health in the U.K. In their study, negative effects of informal care provision on mental health only emerge for high-intensity care provision in the U.K. It seems plausible that the estimated effects in Fig. 3 might conceal similar heterogeneity. However, the instrumental variable models considered above are not well-suited to examine the effects of informal care intensity. While it is possible to consider a linear trend in care intensity (see Fig. A.1 in the appendix), there is no reason to expect that the effect of informal care intensity on health behaviour or health should be linear. In contrast, considering different categories of care intensity requires one instrument per category. Crucially, all excluded instruments considered in this paper identify variation in the probability of care provision (i.e., at the external margin), but they are unlikely to identify variation in the intensity of care provision (i.e., at the intensive margin).

4.4 Robustness

I conduct three additional robustness checks. First, I re-estimate all IV models using "mother alive" and "only one parent alive" separately as instruments for informal care provision. The results (Fig. A.2 in the appendix) confirm that the estimated effects from both instruments are very similar to each other. Second, I re-estimate the IV models in Fig. 2 with individual fixed effects using the full sample. Both in the fixed effects IV model and in my preferred IV specification individuals without a care episode do not contribute to the identification of the effects of informal care provision. However, the two approaches are not equivalent – e.g., individuals without a care episode do contribute to the identification of the coefficients on covariates in the model. The results in Fig. A.3 in the online appendix confirm that there are no negative effects of informal care provision on health behaviour.

Finally, I consider an alternative identification strategy based on selection on observables using the post-double selection lasso estimator (Ahrens et al., 2018; Belloni et al., 2014) to select relevant covariates that are predictive of health behaviour and informal care provision from a candidate set of up to 3,346 terms (Schmitz and Westphal, 2017). A detailed description is provided in online appendix B. In short, the results confirm that there is little evidence of negative effects of informal care provision on health behaviour or health in England. They also reveal that the positive effects of informal care provision on health behaviour are driven by low and medium intensity care provision.

5 Discussion

This paper examines the effect of informal care provision on engagement in health behaviour in England. I contribute to the existing literature along two dimensions. First, I systematically evaluate several candidate instrumental variables that have been proposed in previous empirical studies using a novel test of instrument validity and monotonicity. Second, I examine engagement in health behaviour as a potential mechanism for the health effects of informal care provision reported in previous studies.

The results suggest that the validity of instrumental variables based on a potential caregiver's family background remains questionable. While living parents increase the likelihood of informal care provision, the test of joint instrument validity and monotonicity rejects the null hypothesis for all instruments considered for the full sample. Moreover, having a living sibling is in itself not predictive of informal care provision. If anything, instrumental variables based on survival of the parents might be valid to identify variation in the timing of informal care

provision. When I restrict the sample to individuals that provide care at any point during the observation period, the test fails to reject the null hypothesis of joint validity and monotonicity for having a living mother and having only one living parent, but it rejects the null hypothesis for having a living father and for having at least one living sibling. Although I cannot draw definitive conclusions, it seems plausible that mother's survival is less selective than father's survival, and therefore mother's survival may be a valid instrument whereas father's survival is not.

Looking at the effects of informal care provision on health behaviour, I find no systematic evidence of a negative effect of informal care provision. Rather, I observe certain improvements in health behaviour, in particular in the frequency of moderate exercise. I arrive at the same conclusions when considering the post-double selection lasso as an alternative identification strategy. Both sets of models involve different identifying assumptions (exclusion restrictions vs. selection on observables), and there is no reason to suspect that the remaining bias in either approach should operate in the same direction. Therefore, I argue that taken together these two approaches confirm the robustness of these conclusions.

Interestingly, I also do not find evidence for negative effects of informal care provision on health. In line with a recent study on the U.K. (Bom and Stöckel, 2021), I find that informal care provision has small positive effects on caregiver's health. These effects disappear at higher intensities of care provision, but in contrast to Bom and Stöckel (2021) I do not find evidence of negative health effects even for high intensity caregiving. These results do not necessarily contradict each other, since both studies differ in their methodological approaches as well as their study populations. Bom and Stöckel (2021)'s sample includes individuals aged 45 to 65 over the period 2009-2014 for the entire U.K., whereas in this study I consider individuals aged 40 to 69 during the period 2002-2017 in England.

There are several important limitations of this study. First, previous studies have considered alternative instruments that I could not assess in this study. For example, information on the health status of the potential care recipient is not available in ELSA, unless I restrict the sample to parents that are cohabitating with their children. Since a large majority of informal care in ELSA is provided outside the caregiver's household, this would considerably reduce the sample size as well as the external validity of the findings. Perhaps more importantly, previous studies have drawn on additional information on the respondent's siblings, e.g., gender, birth order, or distance to the parents. It seems plausible that such additional information could be

used to strengthen the sibling instrument, but unfortunately this information is not available in ELSA. Second, although I do not find evidence of negative significant effects of care provision on health behaviour or health, the effects are often not precisely identified. It is possible that the lack of precision is due to unobserved heterogeneity in the effects of care provision. Examining such heterogeneity in IV models is difficult, because instrumental variables identify specific local variation in treatment assignment. Third, I do not find evidence of negative effects of care provision on health, although earlier studies have reported negative effects for countries such as, e.g., Germany (Schmitz and Westphal, 2015) or continental Europe (Heger, 2017). Hence, it is possible that my conclusions on health behaviour might not generalise to these countries, and it would be interesting to replicate the current study with data from countries for which a negative effect of informal care provision in health was previously reported.

In summary, I find that informal care provision does not have negative effects on health behaviour for caregivers in England, but rather seems to improve certain behaviours such as moderate exercise. These improvements seem driven by low intensity caregiving. Such positive effects are in line with learning effects and cues to action as proposed in the Health Belief Model.

References

Ahrens, A., Hansen, C.B., Schaffer, M.E., 2018. PDSLASSO: Stata module for post-selection and post-regularization OLS or IV estimation and inference. Boston College Department of Economics.

Balke, A., Pearl, J., 1997. Bounds on Treatment Effects from Studies with Imperfect Compliance. Journal of the American Statistical Association 92, 1171–1176. https://doi.org/10.1080/01621459.1997.10474074

Baum, C.F., Lewbel, A., 2019. Advice on using heteroskedasticity-based identification. The Stata Journal 19, 757–767. https://doi.org/10.1177/1536867X19893614

Belloni, A., Chernozhukov, V., Hansen, C., 2014. Inference on Treatment Effects after Selection among High-Dimensional Controls†. The Review of Economic Studies 81, 608–650. https://doi.org/10.1093/restud/rdt044

Bobinac, A., van Exel, N.J.A., Rutten, F.F.H., Brouwer, W.B.F., 2010. Caring for and caring about: Disentangling the caregiver effect and the family effect. Journal of Health Economics 29, 549–556. https://doi.org/10.1016/j.jhealeco.2010.05.003

Bolin, K., Lindgren, B., Lundborg, P., 2008. Your next of kin or your own career? Caring and working among the 50+ of Europe. J Health Econ 27, 718–738. https://doi.org/10.1016/j.jhealeco.2007.10.004

Bom, J., Bakx, P., Schut, F., van Doorslaer, E., 2018. The Impact of Informal Caregiving for Older Adults on the Health of Various Types of Caregivers: A Systematic Review. The Gerontologist 59, e629–e642. https://doi.org/10.1093/geront/gny137

Bom, J., Stöckel, J., 2021. Is the grass greener on the other side? The health impact of providing informal care in the UK and the Netherlands. Social Science & Medicine 269, 113562. https://doi.org/10.1016/j.socscimed.2020.113562

Bonsang, E., 2009. Does informal care from children to their elderly parents substitute for formal care in Europe? Journal of Health Economics 28, 143–154. https://doi.org/10.1016/j.jhealeco.2008.09.002

Brenna, E., Di Novi, C., 2016. Is caring for older parents detrimental to women's mental health? The role of the European North–South gradient. Review of Economics of the Household 14, 745–778. https://doi.org/10.1007/s11150-015-9296-7

Chen, L., Fan, H., Chu, L., 2019. The hidden cost of informal care: An empirical study on female caregivers' subjective well-being. Social Science & Medicine 224, 85–93. https://doi.org/10.1016/j.socscimed.2019.01.051

Chernozhukov, V., Kim, W., Lee, S., Rosen, A., 2015. Implementing intersection bounds in Stata. Stata Journal 15, 21–44.

Chernozhukov, V., Lee, S., Rosen, A.M., 2013. INTERSECTION BOUNDS: ESTIMATION AND INFERENCE. Econometrica 81, 667–737.

Clemens, S., Phelps, A., Oldfield, Z., Blake, M., Oskala, A., Marmot, M., Rogers, N., Banks, J., Steptoe, A., Nazroo, J., 2019. English Longitudinal Study of Ageing: Waves 0-8, 1998-2017, [data collection], UK Data Service, 30th Edition. https://doi.org/10.5255/UKDA-SN-5050-17

Coe, N.B., Van Houtven, C.H., 2009. Caring for mom and neglecting yourself? The health effects of caring for an elderly parent. Health Economics 18, 991–1010. https://doi.org/10.1002/hec.1512

de Zwart, P.L., Bakx, P., van Doorslaer, E.K.A., 2017. Will you still need me, will you still feed me when I'm 64? The health impact of caregiving to one's spouse. Health Econ 26 Suppl 2, 127–138. https://doi.org/10.1002/hec.3542

Do, Y.K., Norton, E.C., Stearns, S.C., Van Houtven, C.H., 2015. Informal Care and Caregiver's Health. Health Economics 24, 224–237. https://doi.org/10.1002/hec.3012

Fadlon, I., Nielsen, T.H., 2019. Family Health Behaviors. American Economic Review 109, 3162–3191.
Farbmacher, H., Guber, R., Klaassen, S., 2020. Instrument Validity Tests With Causal Forests. Journal of Business & Economic Statistics 1–10. https://doi.org/10.1080/07350015.2020.1847122
Grossman, M., 1972. On the Concept of Health Capital and the Demand for Health. Journal of Political Economy 80, 223–255.
Heckman, J.J., Vytlacil, E., 2005. Structural Equations, Treatment Effects, and Econometric Policy Evaluation. Econometrica 73, 669–738.
Heger, D., 2017. The Mental Health of Children Providing Care to their Elderly Parent. Health Economics 26, 1617–1629. https://doi.org/10.1002/hec.3457
Huber, M., Mellace, G., 2014. Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints. The Review of Economics and Statistics 97, 398–411. https://doi.org/10.1162/REST_a_00450
Kaschowitz, J., Brandt, M., 2017. Health effects of informal caregiving across Europe: A longitudinal approach. Social Science & Medicine 173, 72–80. https://doi.org/10.1016/j.socscimed.2016.11.036
Kitagawa, T., 2015. A Test for Instrument Validity. Econometrica 83, 2043–2063. https://doi.org/10.3982/ECTA11974
Lewbel, A., 2018. Identification and estimation using heteroscedasticity without instruments: The binary endogenous regressor case. Economics Letters 165, 10–12. https://doi.org/10.1016/j.econlet.2018.01.003
Lewbel, A., 2012. Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. Journal of Business & Economic Statistics 30, 67–80. https://doi.org/10.1080/07350015.2012.643126
Mourifié, I., Wan, Y., 2017. Testing Local Average Treatment Effect Assumptions. The Review of Economics and Statistics 99, 305–313. https://doi.org/10.1162/REST_a_00622
ONS, 2017. Unpaid carers provide social care worth £57 billion.
Rosenstock, I.M., 1966. Why People Use Health Services. Milbank Q 44, 10.1111/j.1468–0009.2005.00425.x. https://doi.org/10.1111/j.1468-0009.2005.00425.x
Schmitz, H., Stroka, M.A., 2013. Health and the double burden of full-time work and informal care provision — Evidence from administrative data. Labour Economics 24, 305–322. https://doi.org/10.1016/j.labeco.2013.09.006
Schmitz, H., Westphal, M., 2017. Informal care and long-term labor market outcomes. Journal of Health Economics 56, 1–18. https://doi.org/10.1016/j.jhealeco.2017.09.002
Schmitz, H., Westphal, M., 2015. Short- and medium-term effects of informal care provision on female caregivers' health. Journal of Health Economics 42, 174–185. https://doi.org/10.1016/j.jhealeco.2015.03.002
Smits, J., Monden, C., 2009. Length of life inequality around the globe. Social Science & Medicine 68, 1114–1123. https://doi.org/10.1016/j.socscimed.2008.12.034
Staiger, D., Stock, J.H., 1997. Instrumental Variables Regression with Weak Instruments. Econometrica 65, 557–586. https://doi.org/10.2307/2171753
Van Houtven, C.H., Coe, N.B., Skira, M.M., 2013. The effect of informal care on work and wages. J Health Econ 32, 240–252. https://doi.org/10.1016/j.jhealeco.2012.10.006
Van Houtven, C.H., Norton, E.C., 2004. Informal care and health care use of older adults. Journal of Health Economics 23, 1159–1180. https://doi.org/10.1016/j.jhealeco.2004.04.008

Van Houtven, C.H., Wilson, M.R., Clipp, E.C., 2005. Informal Care Intensity and Caregiver Drug Utilization. Review of Economics of the Household 3, 415–433. https://doi.org/10.1007/s11150-005-4942-0

Wakabayashi, M., Kureishi, W., 2018. Differences in the effects of informal family caregiving on health and life satisfaction between wives and husbands as caregivers. Review of Development Economics 22, 1063–1080. https://doi.org/10.1111/rode.12390

White, J., Zaninotto, P., Walters, K., Kivimäki, M., Demakakos, P., Biddulph, J., Kumari, M., De Oliveira, C., Gallacher, J., Batty, G.D., 2016. Duration of depressive symptoms and mortality risk: the English Longitudinal Study of Ageing (ELSA). Br J Psychiatry 208, 337–342. https://doi.org/10.1192/bjp.bp.114.155333

Online Appendix

A. Additional tables and figures

Table A.1: IV estimates

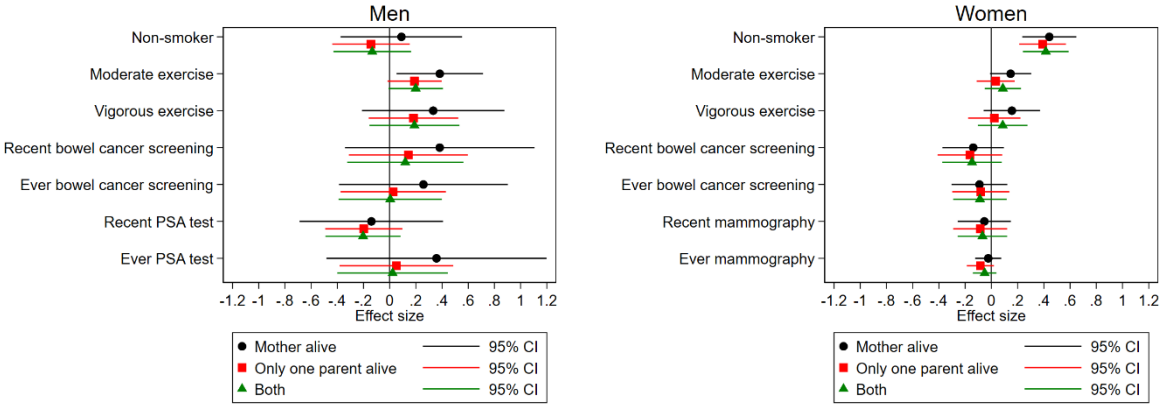
Outcome	Excluded instruments		Lewbel instruments		Combined model		Sargan-Hansen test
A. Men							
Non-smoking	-0.133		-0.022		-0.045		
Moderate exercise	0.199	**	-0.003		0.033		
Vigorous exercise	0.189		-0.027		0.016		
Recent bowel cancer screening	0.119		0.066		0.073		
Ever bowel cancer screening	0.004		0.094		0.072		
Recent PSA test	-0.203	*	0.135	**	0.071	*	
Ever PSA test	0.023		0.038		0.030		
B. Women							
Non-smoking	0.415	***	-0.022		0.160	***	*
Moderate exercise	0.087		-0.002		0.030		**
Vigorous exercise	0.087		-0.094	*	-0.022		
Recent bowel cancer screening	-0.147		0.004		-0.068		
Ever bowel cancer screening	-0.087		-0.007		-0.043		
Recent mammography	-0.068		0.122	*	0.022		
Ever mammography	-0.051		-0.009		-0.031		

Source: ELSA waves 1-8, own calculations. All models are estimated on individuals aged 40-69 with a care spell observed within the panel. All models include control variables for quadratic age, education, the recent death of a father or mother and survey year and month. The exluded instruments are "mother alive" and "only one living parent". The last column reports the significance level of the Sargan-Hansen test for overidentifying restrictions for the model including both excluded instruments and Lewbel instruments. *** p<0.01, ** p<0.05, * p<0.1.

Fig. A.1 Intensity of caregiving – IV estimates

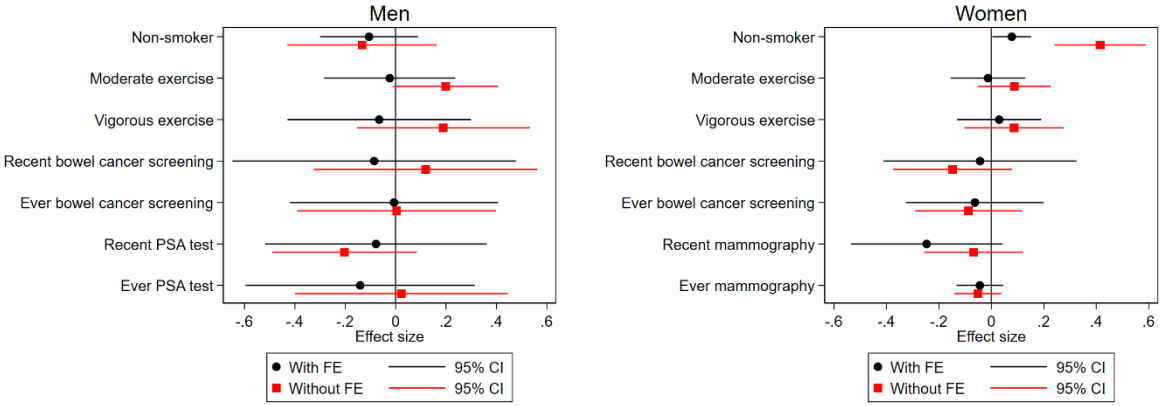
Source: ELSA, own calculations. Notes: The figures show regression results from 2SLS models with hours of informal care provision as the endogenous variable. Markers show point estimates and the lines show 95% confidence intervals. The excluded instruments are *only one parent* and *mother alive*. All models control for a quadratic trend in age, education, marital status, survey wave as well as the recent death of a mother or father. The sample includes all individuals aged 40-69 who were observed providing informal care in the panel. For bowel cancer screening outcomes the sample is restricted to ages 60-69, and for mammography screening to ages 50-69.

Fig. A.2 Robustness – Choice of instruments



Source: ELSA, own calculations. Notes: The figures show regression results from a 2SLS model. Markers show point estimates and the lines show 95% confidence intervals. The excluded instruments are "*only one parent alive"* and "*mother alive"*. Model 3 includes both instruments jointly. All models control for a quadratic trend in age, education, marital status, survey wave as well as the recent death of a mother or father. The sample includes all individuals aged 40-69 who were observed providing informal care in the panel. For bowel cancer screening outcomes the sample is restricted to ages 60-69, and for mammography screening to ages 50-69.

Fig. A.3 Robustness – Individual Fixed Effects



Source: ELSA, own calculations. Notes: The figures show regression results from a 2SLS model. Markers show point estimates and the lines show 95% confidence intervals. The excluded instruments are "*only one parent alive"* and "*mother alive"*. *"With FE"* indicates that the model includes individual fixed effects and is estimated on the full sample. *"Without FE"* does not include individual fixed effects and is only estimated on a sample of individuals that provide care at any point in the panel. All models control for a quadratic trend in age, education, marital status, survey wave as well as the recent death of a mother or father. For bowel cancer screening outcomes the sample is restricted to ages 60-69, and for mammography screening to ages 50-69.

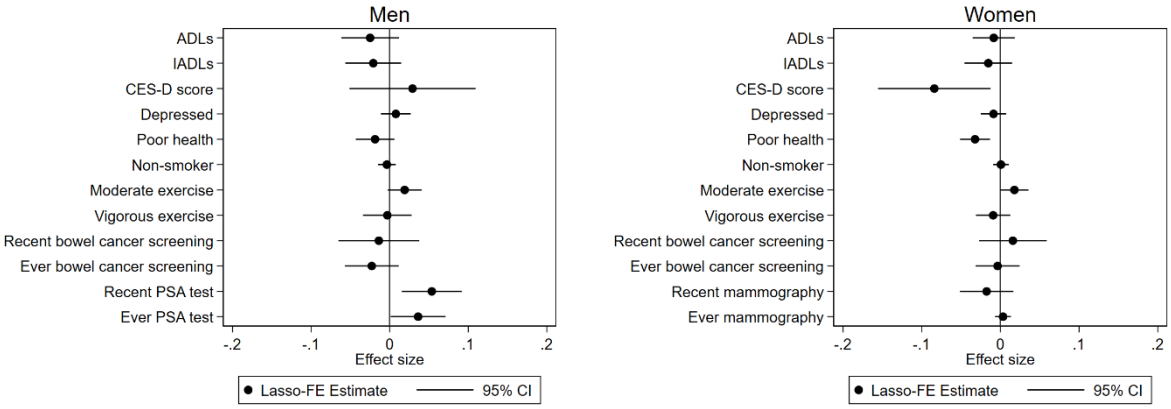
B. Post-double selection lasso

I consider an alternative identification strategy based on the conditional independence assumption, which closely follows the approach by Schmitz and Westphal (2017). Under the assumption that all relevant covariates that affect both informal care provision and health behaviour are observed and included in the model, it is possible to estimate the causal effect of informal care provision on health behaviour without the use of instrumental variables. This assumption might not be credible in parsimonious models. Therefore, I combine fixed effects estimation with the post-double selection lasso proposed by Belloni et al. (2014). The lasso is a machine-learning technique that, intuitively, imposes a penalty on non-zero regression coefficients for a high-dimensional set of control variables, thus shrinking most of these coefficients to zero and selecting a sparse set of influential (i.e., non-zero) regression coefficients. The "post-double selection" approach developed by Belloni et al. (2014) and implemented by Ahrens et al. (2018) in STATA involves two lasso steps to select covariates predicting informal care provision (i.e., the treatment) as well as the outcome of interest. The union of the covariates selected by these two lasso steps are then included as control variables in a fixed effects regression model.

I consider a quadratic age trend, month and year of the survey, education, marital status, income decile, wealth decile, family structure (using the instrumental variables described above and an indicator for the presence of a dependent child under 13), health behaviour and existing functional limitations as measured by the ADL and IADL score. All variables (with the exception of age and interview date) are included as lags to reduce potential problems around reverse causality. Moreover, I also consider all two-way interactions between these covariates, which results in a set of up to 3,346 candidate terms.

The results shown in Fig. B.1 confirm the conclusions of the IV models. There is little evidence of negative effects of informal care provision on health or health behaviour, rather there are positive and significant effects on certain health behaviour outcomes, such as a decrease in the likelihood to report a poor health status and an increase in the probability of frequent moderate exercise for men and women. The post-double selection lasso approach also allows me to consider nonlinear heterogeneity in the intensity of caregiving. Fig. B.2 suggests that these positive effects on health and health behaviour are primarily driven by low and medium intensity care provision. However, even for high intensity care provision I find no systematic evidence of negative effects on health or health behaviour.

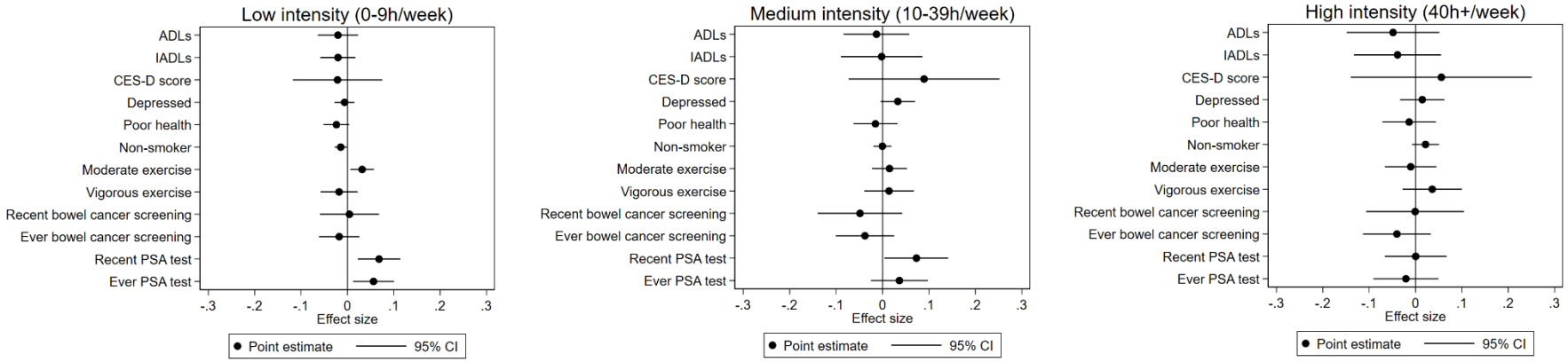
Fig. B.1 Lasso estimates of the impact of informal care on health and health behaviour



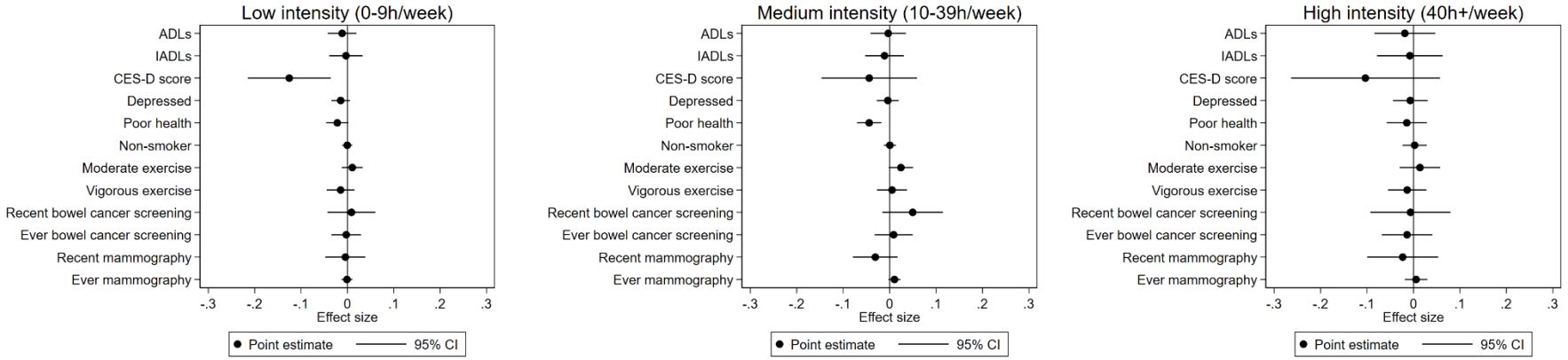
Source: ELSA, own calculations. Notes: The figures show regression results from fixed effects regression models with covariates selected through post-double selection lasso. Markers show point estimates and the lines show 95% confidence intervals. The sample includes all individuals aged 40-69. For bowel cancer screening outcomes the sample is restricted to ages 60-69, and for mammography screening to ages 50-69.

Fig. B.2 Intensity of care giving – Lasso-FE estimates

A. Men



B. Women



Source: ELSA, own calculations. Notes: The figures show regression results from fixed effects regression models with covariates selected through post-double selection lasso. Markers show point estimates and the lines show 95% confidence intervals. Effects for low intensity, medium intensity, and high intensity caregiving for each outcome are estimated in the same model. The sample includes all individuals aged 40-69. For bowel cancer screening outcomes the sample is restricted to ages 60-69, and for mammography screening to ages 50-69.

References

Ahrens, A., Hansen, C.B., Schaffer, M.E., 2018. PDSLASSO: Stata module for post-selection and post-regularization OLS or IV estimation and inference. Boston College Department of Economics.

Belloni, A., Chernozhukov, V., Hansen, C., 2014. Inference on Treatment Effects after Selection among High-Dimensional Controls†. The Review of Economic Studies 81, 608–650. https://doi.org/10.1093/restud/rdt044

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