HEDG Working Paper 09/11

How Does Retirement Affect Health?

Stefanie Behncke

June 2009

ISSN 1751-1976

http://www.york.ac.uk/res/herc/research/hedg/wp.htm

How Does Retirement Affect Health?

Stefanie Behncke *

June 23, 2009

Abstract: This paper investigates the effects of retirement on various health outcomes. Data stem from the first three waves of the English Longitudinal Study of Ageing (ELSA). With this informative data, non-parametric matching methods can be applied to identify causal effects. It is found that retirement significantly increases the risk of being diagnosed with a chronic condition. In particular, it raises the risk of developing a cardiovascular disease and being diagnosed with cancer. Estimates also indicate that retirement has quite diverse effects for different individuals.

Keywords: retirement, health, matching methods, ELSA
JEL classification: I10, J14, J26

*The author is also affiliated with IZA, Bonn. She was visiting the Centre for Health Economics (CHE) in York while a substantial part of this paper was written. She acknowledges financial support from the Swiss National Science Foundation (SNSF). She is grateful to Andrew Jones, Michael Lechner, Nigel Rice and Laura Vallejo-Torres for helpful comments. She also thanks Susan Nunn for providing further information on the data set ELSA which was supplied by the ESCR data archive. This paper has been presented at the Health Economist's Study Group (HESG) meeting in Manchester. The usual disclaimer applies. Address for correspondence: Stefanie Behncke, University of St. Gallen, SEW, Varnbüelstrasse 14, 9000 St. Gallen, Switzerland; stefanie.behncke@unisg.ch

1 Introduction

In the UK, as in many other industrialised countries, the labour force participation of the elderly has been dramatically decreasing in the last decades (Bloendal and Scarpetta, 1999). The median retirement age has decreased from 65 to 61 years, while average life expectancy has risen by three years and age of finishing full time education has risen by two years (Banks and Smith, 2006). The declining ratio of economically active to inactive individuals raises concerns about the sustainability of pensions and other welfare systems. As a consequence, British policymakers have implemented a bundle of policies that aim to raise the economic activity of the elderly: eligibility rules for disability benefits have been tightened, special measures for unemployed people over 50 have been introduced, mandatory early retirement and age discrimination have been abolished. Furthermore, the British government has raised the official state pension age: it will gradually increase from 60 to 65 by 2020 for women; and from 65 to 68 for both men and women between 2024 and 2046.

Delaying the retirement age, though, might have consequences for the health of the elderly. Retirement is accompanied by many basic life changes: time availability, income streams, social networks, social status and so on. Lifestyle changes are likely to affect the retiree's state of health. This aspect of retirement is often neglected when implementing policies which aim to raise the effective retirement age. However, if retirement affects health outcomes, the welfare effects of policy changes may differ in crucial ways. Firstly, the well-being of the growing number of elderly people is determined by policies regulating retirement to which many elderly people are exposed. Secondly, health care spending depends on how retirement affects health. Thirdly, policies regulating retirement could widen or close health inequalities between the wealthy and the poor, as they probably have a larger impact on retirement decisions made by the elderly with sparse financial resources.

Little is known about the health effects of retirement. Establishing causal links is aggravated by selection into retirement. For instance, it has been well established that poor health causes economic inactivity (see Disney, Emmerson, and Wakefield (2006) and Rice,

Roberts, and Jones (2007) for the UK; the seminal study by Bound, Schoenbaum, Stinebrickner, and Waidmann (1999) or Conti, Berndt, and Frank (2006) for the US; Riphahn (1999) and Lechner and Vazquez-Alvarez (2004) for Germany; Lindeboom and Kerkhofs (2002) for the Netherlands; and Au, Crossley, and Schellhorn (2005) for Canada). Poor health, though, is also likely to affect later health outcomes and hence confounds the analysis. Besides health itself, there are also other factors that simultaneously affect the decision to retire and later health outcomes. When employing different identification strategies to this non-random selection into retirement, previous studies find conflicting evidence (see Charles (2002); Dave, Rashad, and Spasojevic (2006) and Coe and Lindeboom (2008) for the US; and Morris, Cook, and Shaper (1994) and Bound and Waidmann (2007) for the UK).

This paper aims to estimate health effects of retirement in England. It contributes to the existing literature by using a very comprehensive data set: the English Longitudinal Study of Ageing (ELSA). This data allows us to control for potential confounders, e.g. initial health, income, pensions, and anticipations. Furthermore, it exploits information on the dates of retirement and the diagnoses of chronic disease. Thus, reverse causality can be ruled out, i.e. we can observe whether individuals retire before or after a new diagnosis was made. Moreover, this paper provides estimates for a subpopulation of general interest, namely the retirees. This is in contrast to instrumental variable studies that provide estimates for the subpopulation who responded to the instrument. Furthermore, this paper applies non-parametric matching methods, i.e. it does not need to rely on any functional form assumptions and also allows for individual effect heterogeneity. In addition, the applied method provides estimates on the selection into the retirement state. Understanding the selection process per se is important from a policy perspective: if policies aim to raise the economic activity of elderly people, having insights into its determinants is crucial. Thus, this paper also contributes to the debate on whether financial or health variables are more important for determining the retirement decision (see Lindeboom (2006) for a discussion).

We find that retirement increases the risk of being diagnosed with a chronic condition, e.g. heart attack, stroke, or cancer. However, we can only observe the date of the diagnosis and not the onset of the disease. Under the assumption that onset and diagnosis coincide, this result suggests that retirement *on average* harms individuals. Furthermore, there is evidence suggesting that there is a considerable effect heterogeneity: retirement does not affect every individual in the same way. These results are in contrast to the study by Bound and Waidmann (2007), who find no evidence for the negative health effects of retirement in the UK and some positive effects for men. Results are in line with evidence by Morris, Cook, and Shaper (1994), who find that mortality due to cardiovascular diseases and cancer is significantly increased for male retirees in the UK.

In Section 2, retirement in the UK is described, the potential mechanism through which retirement affects health is discussed and the literature on the health effects of retirement is reviewed. Section 3 describes the data, the sample selection mechanism, and the definition of retirement and the health outcomes. It provides descriptive statistics and discusses health-related attrition. Section 4 is dedicated to the methodology: the identification strategy and its plausibility are discussed and the matching estimator is described. Section 5 presents estimation results for the determinants and health effects of retirement. The final section offers a conclusion.

2 Background

2.1 Retirement in the United Kingdom

Individuals decide to retire based on their personal situation within an institutional context. The pension system is likely to affect retirement decisions by setting incentives. The UK pension system has two tiers.[1] The first tier is a Basic State Pension (BSP). It is a very low (less than £ 100 per week) flat-rate benefit. Individuals above the state pension age (currently 65 for men and 60 for women) are entitled if they have paid National Insurance contributions for at least 25% of their working life. The second tier is compulsory for

[1] Blundell, Meghir, and Smith (2002), and Emmerson and Johnson (2001) and Banks and Smith (2006) provide an extensive overview on the UK pension system.

all employees earning more than a certain amount. They have to contribute either to the State Second Pension (default option) or to a private pension scheme. In the State Second Pension, the entitlement depends on employment and earnings history, and is basically in proportion to average earnings. Alternatively, individuals can opt out and have either an employer provided or an individually arranged pension. Since many employees chose this option, private pensions matter more for retirement decision in the UK compared to other countries. Private pensions can take either the form of a defined benefit (DB) or a defined contribution (DC) scheme. In DB schemes, pensions are based on a formula involving age, years of salary and services. They have been found to encourage early retirement (Blundell, Meghir, and Smith, 2002). They also set incentives (i) to stay in work until the early retirement age (which is usually below the state pension age) and (ii) to work full time until that age instead of taking gradual retirement. In DC schemes, money is invested in a fund and the size of the fund determines income at retirement. They usually do not set incentives to retire at a certain age.

2.2 Mechanisms through which retirement affects health

In economic theory, health is considered to be both an investment good, allowing for higher productivity and fewer sick days, and a consumption good providing direct utility (Grossman, 1972). With retirement, incentives to invest in health in order to raise productivity cease to exist. Still, individuals gain from investing in their health as it raises their utility. The effects of retirement depend on how the marginal cost of investing in health capital and marginal benefits from enjoying better health change, which in turn depends on how the marginal value of time changes (Dave, Rashad, and Spasojevic, 2006). For instance, if the value of time declines after retirement, the time cost for investing in health such as visiting a physician or cooking healthy food is reduced. Simultaneously, the marginal benefit of having better health declines when time is less valued after retirement. Thus, the theoretical effect of retirement is ambiguous and likely to be heterogeneous for individuals with different preferences.

There are also other channels through which retirement might affect health besides chang-

ing incentives to invest in health. Retirement might have psycho-social consequences. It is often viewed as a stressful life event (Minkler, 1981) and has been ranked in tenth place out of 43 life events in an attempt to measure stress as a predictor of illness (Holmes and Rahe, 1967). Life events are defined as discrete experiences that disrupt an individual's usual activities, causing a substantial change and readjustment. Depending on individual coping resources, life events can cause stress, even if they are perceived as positive. Stress, however, is known to influence health by affecting health habits such as smoking, drinking, sleeping, eating and exercising (Brannon and Feist, 1997), by deteriorating the immune system, endocrine and cardio-vascular reactivity (Ader, Felten, and Cohen, 2001) and by causing negative affect such as anxiety or depression which itself negatively affects health outcomes (Scheier and Bridges, 1995). Health consequences of stress are inter alia depression, cardio-vascular diseases, infection (Cohen, Janicki-Deverts, and Miller, 2007), and possibly cancer (Kiecolt-Glaser and Glaser, 1999). On the other hand, retirement might not cause stress at all, but might be instead a relief from a stressful working life. Then, retirement would consequently be expected to have positive health effects.

Irrespective of stress, there is evidence suggesting that health behaviour changes with economic activity; these changes, though, can harm or promote health. For instance, retirees were found to be more likely to quit smoking (Lang, Rice, Wallace, Guralnik, and Melzer, 2007); to engage in regular exercise; and to have lower stress levels (Midanik, Soghikian, Ransom, and Tekawa, 1995). But it has also been found that they have reduced social interactions (Sugisawa, Sugisawa, Nakatani, and Shibata, 1997), increased problematic alcohol consumption (Perreira and Sloan, 2001) and gain weight (Forman-Hoffman, Richardson, Yankey, Hillis, Wallace, and Wolinsky, 2008).

In general, we would not expect that retirement categorically harms or benefits individuals. Most likely, individual characteristics and circumstances of retirement influence health effects of retirement. For instance, there is evidence that employees' failure to control retirement according to their wishes adversely affects health (Solinge, 2007) and that marital context plays an important role on retirement well-being (Szinovacz and Davey, 2004).

2.3 Literature

While the mere association between retirement and health is found to be negative, empirical studies have provided conflicting evidence on the causal effects of retirement. Besides the presumption that there is a considerable effect heterogeneity, another potential explanation for this discrepancy is that the applied strategies identify different effects. For instance, when addressing the problem that health and retirement are jointly determined, one strategy is to find an instrumental variable that only affects the endogenous retirement decision, but does not affect the health outcome. This approach identifies the *local average treatment effect* (LATE), i.e. the effect of retirement for the subpopulation whom the instrument has induced to retire. However, the LATE can considerably deviate from the *average treatment effect* (ATE), i.e. the effects of retirement for the whole population, and the *average treatment effect on the treated* (ATET), i.e. the effects of retirement for the retirees.[2] It can even have the opposite sign. Indeed, instrumental variables studies tend to find positive effects of retirement, while studies that control for confounding variables tend to find negative effects.

One example for an instrumental variable strategy is Charles (2002). He proposes age specific retirement incentives in the US, a change in the US pension system, and coverage by a retirement rule as instruments. He finds that men feel 'lonely' or 'depressed' less often after retirement. Effects are considerable and range between 6 and 28 percentage points. However, if individuals anticipate that there are financial incentives to retire at a certain age, or that there will be a change in the pension system[3], they adjust their behaviour before retirement. This may result in estimates biased upwards, i.e. more beneficial health effects of retirement are found than are actually the case. There is also the concern whether the instruments are valid, i.e. whether they have no direct effect on the health outcome and whether they are uncorrelated with any other variable that affects health. For instance, the first assumption would be violated if reaching a certain age (turning '60') had direct psychological effects; the second assumption would be violated if

[2]For a discussion on the instrumental variables approach see Deaton (2009); Heckman and Urzua (2009) and the response by Imbens (2009).

[3]The social security amendment was signed in 1983 and was enacted in the early 90's.

individuals with a certain latent health state are more likely to chose jobs with mandatory retirement.

A similar approach has been adopted by Bound and Waidmann (2007) for England, by Neuman (2007) for the US, and by Coe and Zamarro (2008) for eleven European countries. These studies use pension eligibility as a source of exogenous variation and find overall positive health effects of retirement. As above, there is the concern that anticipation of the pension age might bias estimates upwards. While the latter studies use a standard regression discontinuity approach, Bound and Waidmann (2007) compare linear trends of health indicators before and after state pension age. However, many elderly people cease to work before they have reached the state pension age[4]. To the extent that they retire earlier due to ill-health, this would bias retirement effects upwards. There is also the concern that their estimates are based on rather different cohorts.[5] Furthermore, fitting a linear trend for health variables seems to be a strong assumption.

Coe and Lindeboom (2008) circumvent the problem of anticipation by using the offer of retirement windows as instrument. Using US data, they find no significant retirement effects. A potential concern, though, is that firms offer retirement incentives to workers who vary in their health or in other characteristics that determine their health, which would lead to biased estimates.

Other studies base their identification on controlling for the selection into the retirement state. One example for this approach is Dave, Rashad, and Spasojevic (2006). Using US data, they find that retirement increases depression by 6 to 9 percent, illness conditions by 5 to 6 percent and difficulties associated with mobility and daily activities by 5 to 16 percent. They estimate fixed effects models, which can control for unobserved time-invariant (but not for time-varying) heterogeneity across individuals. In order to avoid reverse causality, they stratify their sample across individuals who had no major illnesses or health problems in the waves prior to the retirement wave. They argue that for these individuals retirement is more likely to be exogenous to health. However, they cannot

[4]In ELSA wave 2 (which is the sample in Bound and Waidmann (2007)), 18% of males and 8% of females are retired before they reached state pension age.

[5]Their sample includes 53 to 89-year-olds.

control for health shocks that occurred between the last interview before retirement and time of retirement. To the extent that these unobserved health shocks drove individuals out of the labour market, the negative effects of retirement might be overestimated.
The same caveat applies to the findings by Bamia, Trichopoulou, and Trichopoulos (2007) who can only condition on initial health, but cannot rule out reverse causation. Using Greek data, they find that in comparison to individuals still employed, retirees had a 51% increase in overall mortality. Findings were similar for cardiovascular mortality and cancer mortality, although the latter was not significant.
In contrast, Morris, Cook, and Shaper (1994) can exclude that loss of employment was due to ill health. Furthermore, they can control for age, region, social class, health behaviour and health at baseline. Using British data, they find that men who retire due to other reasons than ill health and who were apparently healthy at baseline had an 86% increase in mortality compared to men who remained continuously employed. In particular, they had a significant increase in both cardiovascular mortality and cancer mortality. Still some sceptics might argue that the findings rather reflect subtle unreported health problems that have contributed to the decision to retire, as noted by Trichopoulos (1996).

In this paper, a wider range of confounding factors is taken into account. In particular, the data allow us to control for anticipations with respect to health and work as well as for financial variables. Furthermore, initial health is not only measured by subjective means, but also by objective means and at more points in time. Another advantage of this paper is that we can exploit information on the date of diagnosis and of retirement to avoid reverse causality. Furthermore, our estimates do not rely on any functional form assumptions and allow for arbitrary effect heterogeneity. Finally, it provides estimates on the average treatment effects for *all* retirees, which might be more interesting than estimates for the subpopulation for which the instrument induced retirement.

3 Data

3.1 Sample selection, definition of retirement and health outcomes

Data stems from the English Longitudinal Study of Ageing (ELSA). ELSA is a panel survey of people born before 1952 who live in England. It collects substantial information on health measures (including nurse visits), health behaviour, employment, earnings, wealth, pensions, expectations about future health and paid work, and other household and individual demographics. The Appendix provides further information on ELSA.

We consider a sample of individuals who are employed at the first ELSA wave. For some of them, we observe transition to retirement at the second wave. For these individuals, we aim to estimate the health effects of retirement by comparing their subsequent health states with similar individuals who remain employed. At wave 1, we observe 2906 employed individuals, where an individuals is considered to be employed if he responds that (i) being "employed" describes his current situation best and (ii) "paid work" was one of his activities in the last month. Out of them, 1803 individuals remain employed at wave 2 and 268 become retired, where an individual is considered to be retired if she responds that (i) "retired" describes her current situation best and (ii) "not in paid work" was her activity in the last month. Out of the individuals who retired at wave 2, 192 are still retired at wave 3. For these individuals, we will analyse the health effects in most of our specifications. Out of the individuals who have been employed at the two prior waves, 1247 individuals are still employed at the third wave. Based on these individuals, we will estimate counterfactual health outcomes that we would observe in the absence of retirement. Thus, our main sample consists of the 1439 individuals who are either still employed or still retired at wave 3. (Table A.1 in the Appendix provides the detailed sample selection mechanism.) As we lose many individuals due to non-response, we will discuss attrition as a potential source of bias in Section 3.3.

A range of health measures is provided in ELSA: diagnosed cardiovascular or other chronic conditions, symptomatic angina or depression, biological measures of physical health such as body mass index, blood pressure, risk factors or lung function. However, most of these

measures are only observed at the time of the interview. In order to rule out reverse causality, we would like to know whether ill health drove individuals out of work or whether they experienced a health shock after retirement. Unfortunately, due to a coding error almost all of the retired individuals (262 out of 268) were not asked for the reasons of their (early) retirement. For newly diagnosed conditions, however, it is possible to exploit the timing of events in order to avoid reverse causality. The respondents are asked in which month a new diagnosis was made. Retirees are also asked when their job ended. Under the assumption that recall bias is minimal (which seems to be reasonable for life events such as retirement or the diagnosis of a chronic condition), it is possible to determine whether retired individuals experienced a health shock before or after their retirement. The diagnosis of a new condition is considered to be an outcome variable only if it was made *after* retirement. If the diagnosis was made before retirement, it is considered to be a potentially confounding variable, since it might affect the retirement decision and also increase vulnerability to other health conditions. In order to avoid bias, we want to control for this confounding health shock prior to retirement. For individuals in the control group, though, the retirement date does not exist. We follow the approach suggested by Lechner (1999) and simulate retirement dates for the control group. We regress the month of retirement of retired individuals on a set of time invariant personal and regional characteristics. We use the estimated coefficients plus a draw in the residual distribution to predict a corresponding value for non retired individuals. Thus, by finding a control observation that is comparable at the assigned retirement date, we can control for confounding due to health shocks prior to retirement.

Table 1 shows means of new health conditions that have been diagnosed after retirement. This is based on the actual retirement date for the retired individuals and the simulated retirement date for the employed individuals. The first two columns show the mean health outcomes for our main sample. For every retired individual, we can observe new diagnoses *at least* up to two years after retirement, since retirement took place before the wave 2 interview and we have also information from the wave 3 interview two years later. Around 11% of the employed individuals have been diagnosed with any chronic

Table 1: Mean health outcomes after retirement

	wave 1 to wave 3		wave 1 to wave 2	
	employed	retired	employed	retired
observations	1247	192	1803	268
any newly diagnosed disease	0.111	0.208	0.052	0.097
new angina, heart attack, or stroke	0.012	0.042	0.007	0.037
new arthritis or cancer	0.064	0.156	0.028	0.056
newly diagnosed angina	0.003	0.010	0.003	0.019
new heart attack	0.006	0.021	0.004	0.019
new stroke	0.004	0.021	0.001	0.015
newly diagnosed diabetes	0.022	0.026	0.008	0.007
newly diagnosed arthritis	0.059	0.120	0.025	0.037
newly diagnosed cancer	0.008	0.042	0.004	0.019
newly diagnosed psychiatric condition	0.019	0.026	0.011	0.011

condition (either diabetes, angina, heart attack, stroke, arthritis, cancer or psychiatric problem). The corresponding percentage for retired individuals is almost twice as high. Severe cardiovascular diseases (angina, heart attack or stroke) have been diagnosed for around 1% of the employed and 4% of the retired individuals. When considering the chronic diseases separately, we notice that in particular arthritis followed by cancer have been often diagnosed. All considered conditions have been relatively more often diagnosed for retired than employed individuals. We also have information when a heart failure, dementia and Alzheimer's have been diagnosed, but these conditions were not present for any individual in our sample.

The last two columns show mean health outcomes when we only take wave 1 and wave 2 into account, but not wave 3. The means are different for two reasons: (i) health outcomes are only observed *at most* for two years after retirement and (ii) the sample also contains individuals who do not respond at wave 3 or switch their state. While the former necessarily results in smaller means, the latter could potentially result in larger means if drop-outs have worse health than wave 3 respondents. Indeed, their health is slightly worse, as will be discussed in Section 3.3. Overall, mean health outcomes are much lower when wave 3 is not taken into account. Still, the same pattern emerges: retired individuals have more diagnosed conditions than the employed.

3.2 Descriptives

The differences in newly diagnosed conditions in Table 1, however, do not account for the fact that individuals who retire are quite distinct from individuals who remain employed. Table A.2 in the Appendix provides means of socio-economic and health variables for these two groups. The third column shows whether means are significantly different according to a standard t-test. As age is one important difference, we show in the fourth column whether age adjusted means in both groups are different. We derive age adjusted variables by balancing the respective variables on age, age squared and four age dummies via propensity score matching using an approach similar to the one explained in Section 4.3. Then we apply a t-test. This shows us the extent to which differences between retired and employed are driven by their age differences. As all variables are measured before retirement took place, they are informative on factors playing a role for selection into retirement.

From the first row, we see that 50% of the employed and 54% of the retired individuals are females, but that this difference is not statistically significant. Retired individuals are on average 60 years old and hence 5 years older than employed individuals. Their household size is smaller. Their children are less likely to (still) live with them and they more often have grandchildren. However, the latter is mainly driven by age differences, as age adjusted means are not significantly different. Retired individuals less often have an academic degree and more often no qualifications. These educational differences stem from age differences, as younger cohorts are more likely to have received a higher education. No major differences are apparent with respect to social classification.

Retired individuals also had different types of jobs when they were still working. They worked longer in their current jobs, were less likely to have permanent jobs and worked for fewer hours per week. They worked less often in firms with 20 to 99 employees and more often in firms with more than 1000 employees. There are no significant differences in the type of work related physical activity.

Retired and employed individuals are not much different with respect to regional characteristics such as government office region, population density and an index of deprivation

with regard to income, employment, health, housing and crime.

There are significant differences with regard to pensions and other financial variables. Retired individuals are more likely to have reached the official state pension age at the second ELSA interview. But 66% of the retired individuals takes retirement before they are eligible to draw state pensions. This is in line with the finding that state pensions play a minor role compared to private pensions for retirement incentives (Blundell, Meghir, and Smith, 2002) in the UK. Retired individuals are more likely to have been offered early retirement incentives before the first ELSA wave (which they had decided not to take). But this difference vanishes once one adjust for age differences, which reflects that such incentives have mainly been offered to early cohorts. Retirees are more likely to have defined benefit schemes and less likely to have defined contribution schemes or any private pension schemes. Interestingly, differences in defined benefit and contribution schemes remain significant once one adjusts for age. This suggests that they do not only reflect cohort effects, but might also indicate preferences for schemes that set incentives to retire at a certain age (DB) or to avoid this (DC). Retirees have a higher total pension wealth[6] compared to employed individuals. In particular, their private pension wealth is higher. On average, state pensions contribute to only 30% of their total pension wealth, which again reflects their less pronounced role compared to private pensions. Differences in pension wealth remain significant even after adjustment for age differences. Thus, retired individuals have higher pension funds at their disposal compared to employed individuals of the same age. Their employment income is lower, but not significantly different once one adjusts for age.

Retired individuals differ also with respect to health variables before retirement. Their memory is not as good and they have more difficulties with the activities of daily living. Most of these differences stem from the age difference. Retirees describe their health more often as fair as compared to very good or excellent. This difference remains significant when one controls for age. They more often have a diagnosis of angina, arthritis, osteoporosis and cancer, but these differences mostly vanish when one adjust for age. There

[6]Total pension wealth is calculated as the sum of private and state pension wealth assuming retirement in 2002 (see Banks, Emmerson, and Tetlow (2005) for the exact calculation).

are no systematic differences in symptomatic conditions such as heart attacks and angina (assessed by the Rose questionnaire), depression (assessed by the Center of Epidemiologic Studies Depression Scale) and mental illness (assessed by the twelve item General Health Questionnaire, see Goldberg and Williams (1988)). Retired individuals already had more diagnosed conditions at time of the HSE interview.

With regard to health habits, we observe that retirees exercise less often. There are almost no significant differences in smoking and drinking behaviour or in having private health insurance.

When asked to estimate chances to stay alive, employed and retired individuals are not much different. Retired individuals' subjective probability that they stay in work was significantly lower than the one of employed individuals. This difference remains significant even if one adjusts for age. Interestingly, their subjective probability that health limits their ability to work is also lower, although other health variables suggest that, on average, their health is worse. One potential explanation for this is that individuals retiring at wave 2 plan at wave 1 to retire in the next years. Hence, they weigh the probability of experiencing a health shock that forces them to leave work lower than individuals who plan to remain in work. However, retired individuals are less likely to respond to questions about their expectations and it is not clear why they do not respond.

In sum, these differences indicate that retired and employed individuals differ in important characteristics even before retirement. Age is one important difference between them. Beyond their age differences, they also differ in living arrangements, job characteristics, pension schemes and pension wealth, health and retirement expectations even after adjustment for age differences. Since these variables are likely to affect health outcomes, the worse health states of retired individuals in Table 1 cannot be interpreted by any means as causal. Instead, when estimating the causal effects of retirement, we need to control for these confounding variables to avoid selection bias. Attrition as another potential source of confounding is discussed in detail in the next paragraph.

3.3 Health-related attrition

18% of individuals who are employed at wave 1 are lost due to attrition at wave 2 and another 10% due to attrition at wave 3. High attrition rates are a concern if non-response is related to health and retirement. Then estimates on the health effects of retirement would be biased. Health-related attrition might occur due to death or serious illness which prevents individuals from responding or due to moving to institutional care which aggravates tracing individuals. It is another source of selection bias, since observed individuals are different to individuals in the population.

The reason for non-response at wave 2 is reported in ELSA.[7] Table A.3 in the Appendix reports non-response reasons for the sample under consideration: refusals and not traceable are main reasons for drop-outs. Ill health and death seem to play only a minor role. But, an unknown fraction of individuals with other drop-out reasons than health could still have dropped out due to health without mentioning it.

Ideally, the response probability should be independent of health given a set of observed variables. Then the data are said to be *missing at random* (MAR). However, since the missing data are not observed, we cannot verify whether they are missing at random (Nicoletti, 2005). It is only possible to test for selection on observables, but not on unobservables. However, if we observe selection on observable health variables, this might give us further insights on the direction of a potential bias due to unobservables.

In order to analyse this issue, the likelihood of response at wave 2 and wave 3 is regressed on variables from the HSE interview and the first ELSA wave. These estimates represent response patterns for the sample under consideration, i.e. they are conditional on responding and being employed at wave 1. Results from a probit are reported in Table A.4 in the Appendix. They indicate that individuals with children in their households, with higher education and from the East Midlands are more likely to respond at wave 2 and wave 3. Individuals who were born outside the UK, who have higher debts and who do physical work are less likely to respond. Likelihood of response is also reflected in compliance at the HSE or first ELSA interview as measured by missing HSE interviews,

[7] At the time this paper was written, reason for non-response at wave 3 was not available.

failure of interviewing all members of the household, refusal of nurse visit or later interview dates. Most importantly, health variables at the first ELSA wave explain response rates at wave 2, but - with the exception of bad mental health - not at wave 3. This suggests that newly experienced health shocks have a rather immediate effect on response rates. Furthermore, we note that the effects of wave 1 health variables on wave 2 response rates are ambiguous: individuals with diagnosed diabetes and who regard a work limiting health shock as more probable are more likely to drop out; individuals with bad mental health, diagnosed angina or psychiatric conditions are less likely to drop out. Other health variables are not significant and hence were not included in the final specification. While other studies have found that individuals in poor initial health are more likely to drop out (Jones, Koolman, and Rice, 2006), this pattern does not appear to be very pronounced in our sample. A possible explanation for the positive effects of health conditions on response rates might be that individuals with certain conditions have an increased intrinsic motivation to participate in ELSA, since they would like to benefit from the nurse checks or since they appreciate the value of a health survey for society.

With respect to health-related attrition, we cannot infer much from the observed selection pattern: health shocks that are caused by retirement are not likely to be systematic related to variables that explain attrition. For example, suppose that retirement would cause heart attacks and suppose further that individuals experiencing heart attacks are more likely to drop out due to deaths. This non-response would only be ignorable if it was captured by controlling for lagged variables. However, two major risk factors of heart attack (angina and diabetes) affect attrition rates in opposite directions. Still we can infer from Table A.4 that there is scope for health-related attrition. Its direction, however, remains undetermined. The type of the health shock and its severity are most likely to be crucial for the direction of bias due to non-response.

In principal, one could narrow the degree of attrition by using a balanced sample of two instead of three waves. However, an extensive analysis suggests that this does not contribute to solving the problem (see Section 7.2.3 in the Appendix).

4 Methodology

4.1 Conditional independence assumption as identification strategy

We base our analysis on the prototypical model of the evaluation literature with a binary treatment variable (Neyman, 1923; Fisher, 1935; Rubin, 1974, 1979). Let us consider an individual i who works. In one state of the world the individual retires in the next period t and in the other state the individual still works. Let D_i denote the retirement state of this individual, i.e. $D_i = 1$ if the individual retires and $D_i = 0$ if not. We are interested in the effect of retirement on a health outcome $Y_{i,t+\tau}$ in period τ after retirement has taken place. Health outcomes are, for instance, developing a certain condition such as cancer, a cardiovascular disease or depression. We would like to compare for the same individual the health status if retirement had taken place with the potential health status if it had not taken place. Let $\{Y^0_{i,t+\tau}, Y^1_{i,t+\tau}\}$ denote the potential health outcomes for this particular individual. Only one of these potential outcomes can be observed, since the individual either retires or does not retire. With this notation, we can define the average treatment effect for the treated (ATET) as the difference between potential outcomes for those individuals who retired:

$$E[Y^1_{t+\tau} - Y^0_{t+\tau}|D = 1].$$

For being able to estimate the ATET, we need to identify $E[Y^0_{t+\tau}|D = 1]$. Generally, we would expect this to be different from the observed value for those who did not retire:

$$E[Y^0_{t+\tau}|D = 1] \neq E[Y^0_{t+\tau}|D = 0],$$

since individuals with $D_i = 1$ and individuals with $D_i = 0$ might differ in various other characteristics which are directly related to health outcomes. For instance, individuals who retire are on average older than individuals who do not, but age has a direct impact on the likelihood of developing a new condition. Our identification strategy is based on controlling for all variables X that jointly affect D as well as health outcomes $Y^0_{t+\tau}$, such

that conditional on X

$$E[Y^0_{t+\tau}|X = x, D = 1] = E[Y^0_{t+\tau}|X = x, D = 0] \quad \forall x \in \chi,$$

where $\chi \subseteq \sup(X|D = 1)$. This assumption is referred to as conditional independence assumption (CIA) in the following. It is also called unconfoundedness in the statistical literature (Rubin, 1974). We assume that CIA holds for every value of x that lies in the support of X in the $D = 1$ population, i.e. $\chi = \sup(X|D = 1) \subseteq \sup(X|D = 0)$.

4.2 Plausibility of the conditional independence assumption given our data

The most crucial aspect of the identification strategy relies on being able to observe all confounding variables X that jointly determine selection into retirement D and health outcomes $Y^0_{t+\tau}$. Thus, the question arises, whether this is a plausible assumption with our data.

From an economic viewpoint there are different sets of factors that might jointly affect the retirement decision and later health outcomes. Age is in many pension schemes a key determinant for the entitlement of pensions. Age and gender together determine state pension age. With rising age, the vulnerability to certain health shocks increases. Gender also affects vulnerability to certain diseases. Having children and grandchildren could affect a retirement decision, but might also play a role for the onset of certain conditions. Education is considered to have a causal effect on health (Grossman and Kaestner, 1997), since more educated individuals are more efficient producers of health (Grossman, 2000) or since education increases information about true effects of health (Kenkel, 2000). Education also determines occupation choices and might affect retirement decision. Current and expected income, wealth and pension benefits clearly determine the retirement decision. Simultaneously, they determine the extent to which individuals can and want to afford a healthy life style and have access to most effective treatment. Geographical factors such as air pollution, traffic noise, proximity to recreation areas, and supply of health services

(number and quality of GPs) have a direct effect on health. Simultaneously, they affect retirement decisions, since retirement depends on local labour market conditions such as unemployment rates, average wages, and job offers and also on regional leisure opportunities. Lifestyle itself is likely to be a confounder. Activity level within the job, physical activity in leisure time, diet, smoking and drinking behaviour are important input factors for health production. But they are also related to job characteristics and propensity to retirement. Another set of confounding variables are health variables themselves. Health has been found to be an important determinant for retiring. Future health outcomes depend on current and past health, since negative health shocks make individuals more vulnerable to further health shocks (comorbidities). Therefore, we clearly want to control for health measures prior to retirement. Anticipation is another issue that needs to be addressed in this context. If individuals expect deteriorations in their health in the future and reduce their life expectations, they might be more likely to retire earlier. Then, shocks to their health are not to be considered as caused by their retirement. Therefore, it seems to be crucial to control for individuals expectations when estimating selection into retirement.

4.3 Non-parametric matching estimation

In the empirical analysis, we use a matching estimator as implemented in Lechner, Miquel, and Wunsch (2006). The advantage of matching estimators is that they are non-parametric and that they allow for arbitrary individual effect heterogeneity (see Heckman, Lalonde, and Smith (1999) for matching with binary treatment and Imbens (2004) for a review on recent advances in this field). To ease notation we drop time subscripts $t + \tau$ from the notation. Then, the ATET can be written as

$$E[Y^1 - Y^0|D = 1] = E[Y|D = 1] - E[Y^0|D = 1].$$

The first term can be estimated by the sample mean in the treated population. The second term is more difficult to estimate as this is the average counterfactual outcome,

i.e. the health of retired individuals that would have been realised if they had not retired. Using iterated expectations, we can re-write the previous expression as

$$= E[Y|D=1] - \int E[Y^0|X=x, D=1] \cdot dF_{X|D=1},$$

where $dF_{X|D=1}$ represents the distribution function of X in the $D=1$ population. The conditional independence assumption (CIA) implies that for individuals with the same value of X, the potential outcomes are on average identical in the $D=1$ and $D=0$ population. It follows that

$$\begin{aligned} &= E[Y|D=1] - \int E[Y^0|X=x, D=0] \cdot dF_{X|D=1} \\ &= E[Y|D=1] - \int E[Y|X=x, D=0] \cdot dF_{X|D=1}. \end{aligned}$$

We thus have identified the ATET by transforming the unobserved potential outcome into an expression that depends on the distribution of observed random variables.
Let $m_0(x) = E[Y|X=x, D=0]$ denote the conditional expectation function of x in the $D=0$ population. Then, the ATET can be expressed as:

$$E[Y^1 - Y^0|D=1] = E[Y|D=1] - \int m_0(x) \cdot dF_{X|D=1}.$$

Hence, we can estimate the ATET as the average of $Y - m_0(x)$ in the $D=1$ population:

$$\widehat{E[Y^1 - Y^0|D=1]} = \frac{1}{N_1} \sum_{i:D_i=1} (y_i - \widehat{m}_0(x_i)).$$

where $\widehat{m}_0(x_i)$ is a non-parametric estimate of $m_0(x_i)$. As we search for each individual in $D=1$ population for the nearest neighbours in the $D=0$ population, this is usually referred to as "matching" estimator, since it matches observations from one subsample to another. Rosenbaum and Rubin (1983) have shown that instead of matching on the high-dimensional vector X, consistent estimates are also obtained by matching on the one-dimensional propensity score $p(x) = Pr(D=1|X=x)$. The small sample properties of

matching estimators have been well explored and appeared to be quite robust in different practical applications (Larsson, 2003; Gerfin, Lechner, and Steiger, 2005). Moreover, it was subjected to Monte Carlo studies investigating small sample problems and sensitivity issues (Lechner, 2002; Frölich, 2004).

In this paper we use an extension of conventional matching estimation as introduced by Lechner, Miquel, and Wunsch (2006). It extends the first-nearest neighbour propensity score matching estimator in several directions. First, matching does not only proceed with respect to the propensity score $p(x)$, but also incorporates additionally some other covariates $\widetilde{X}$ that are suspected to be potentially highly correlated with the outcome variable Y^0 as well as with D. These so called *balancing scores* can help to ensure that a misspecification of the functional form of the propensity score has only a minor impact on the estimation of ATET. In this application we use age and gender as balancing scores.

As a second extension, we use *radius matching* instead of first-nearest neighbour matching. While first nearest neighbour matching uses only the first nearest control observation for each treated, we use all neighbours within a pre-specified radius. When there are other comparison observations that are similar to the matched comparison observation, there are straightforward efficiency gains by considering these additional very close neighbours without paying a high price in terms of additional bias. To increase efficiency, we form an "averaged matched comparison" observation. Thus, the estimator $\widehat{m}_0(x)$ is essentially a kernel estimator

$$\widehat{m_0}(x) = \frac{\sum_{j:D_j=0} y_j \cdot K_h(\parallel x_j - x \parallel_M)}{\sum_{j:D_j=0} K_h(\parallel x_j - x \parallel_M)},$$

where K_h refers to a kernel function and $\parallel x_j - x \parallel_m$ to the Mahalonobis distance between x_j and x. Since we match on the propensity score $p(x)$ and the balancing scores $\widetilde{x}$ we define this distance as follows. Let $(p_j, \widetilde{x}_j')'$ be the column vector of the estimated propensity score $\widehat{p}(x_j)$ and $\widetilde{x}_j$ of individual j. Let Ω be the empirical covariance matrix of $(p_j, \widetilde{x}_j')'$

in the $D = 0$ subsample. We define the Mahalanobis distance with respect to $(p_j, \widetilde{x}_j')'$ as:

$$\| x_j - x \|_M = [(p_j - p(x), (\widetilde{x}_j - \widetilde{x})')] \cdot \Omega^{-1} \cdot [(p_j - p(x), (\widetilde{x}_j - \widetilde{x})')]'.$$

Through the covariance matrix Ω, Mahalanobis distance is a weighted distance metric which takes the different covariances of its components into account. It gives less weight to variables with a higher variance. To ensure that the propensity score is not dominated by the additional variables, we modify the Mahalanobis distance by multiplying the weight implicitly assigned to the propensity score by five, as suggested in Lechner, Miquel, and Wunsch (2006).

The kernel function now gives different weights to observations that are close to x in terms of Mahalanobis distance. All $D = 0$ observations within a radius h around x receive a positive weight in $\widehat{m}_0(x)$, whereas all other observations are ignored. Of course, there are many ways to determine the radius h in practice. We suggest being more cautious with respect to additional bias than with respect to additional variance because the variance of the estimator is visible after the estimation, whereas the bias generally is not. To be conservative, we only consider observations that have a distance to their treated observation of no more than 90% (denoted by R in the following) of the worst match that we had obtained by one-to-one matching after enforcing common support ($R = 0$ is the case of one-to-one matching). To be even more conservative, we weigh the observations proportionally to their distance from the treated (corresponding to a triangular kernel).

As a third extension, the matching quality is increased by implementing a weighted regression based bias removal procedure on top of matching. Appropriately weighted regressions that use the sampling weights from matching have the so-called *double robustness property*. This property implies that the estimator remains consistent if the matching step is based on a correctly specified selection model or the regression model is correctly specified (Rubin, 1979; Joffe, Have, Feldman, and Kimmel, 2004). Moreover, this procedure should increase precision and may reduce small sample as well as asymptotic bias of matching estimators and thus increase robustness of the estimator in this dimension as well. Note that Abadie and Imbens (2006) have shown that the usual 1-to-K matching estimators,

where K is a fixed number, may exhibit an asymptotic bias, because matches are not exact. Our weighted radius matching estimator does not necessarily imply a fixed K and is thus probably less subject to this problem. Nevertheless, we follow their proposal and implement a weighted regression based bias removal procedure on top of the matching. The regression is done in the comparison sample only. Outcomes are predicted for the attributes observed in treated and control samples. Specifically, the outcome variable is regressed on the propensity score and the additional variables with weights coming from the matching step (Imbens, 2004). The difference between the mean of the predicted outcomes using the observed X of the treated and the weighted X of the comparison observations gives an estimate of the bias. Without the theoretical justification given by Abadie and Imbens (2006), a somewhat similar procedure has been used by Rubin (1979) and Lechner (2000).

The different steps of the estimator are described in Table A.6 in the Appendix. In the first step, a probit model is used to estimate the propensity score. Step 2 ensures that we estimate only effects in the region of common support. In the third step, the counterfactual outcome is estimated. Standard errors are calculated conditional on the weights for the comparison observation. As retirement decision of couples is likely to be not independent of each other, standard errors are clustered at the household level. Table A.7 describes the calculation of the variance.

5 Results

5.1 Determinants of retirement

In this section we present our estimation results for the determinants of retirement, before turning to the estimated effects of retirement on health in the next section. They are interesting for two reasons. First, they are a key factor on which our treatment effects are estimated. Hence, we will discuss the criteria we used when specifying our propensity score model. Second, they give policymakers insights on factors that affect retirement decisions of employed individuals over 50.

We have identified confounding variables by our theoretical considerations in Section 4.2 and found them to be present in our data as illustrated by the descriptives statistics in Section 3.2. When specifying the propensity score, we seek to find a specification which is comprehensive enough to control for them. We do so by taking into account all (and even more) variables from Table A.2. But we also seek to find a parsimonious specification given our sample size. Score tests for omitted variables help us to eliminate variables without additional explanatory power given a set of regressors. For instance, we do not include the number of employees working for the same employer in our preferred specification as it does not explain the retirement decision once one controls for other job characteristics. Some of the health variables, however, remain in our final model even when they are farther away from conventional significance levels, as we consider them to be important confounders. We also aim to find the best functional forms for our variables, e.g. age. We check for heteroskedasticity and for normality. Finally, we seek to find a specification which provides a good match quality, i.e. which balances characteristics in the subsamples of employed and retired individuals (see Section 7.4 in the Appendix).

Table 2 shows Probit estimates for the likelihood of retiring at wave 2. As couples' decision to retire is most likely not independent of each other, we report clustered standard errors. Older individuals are more likely to retire with a slightly decreasing likelihood after having reached a certain age. Individuals with children in their household are less likely to retire. Educational variables do not significantly explain the retirement decision when other socio-economic variables are included. The longer individuals have worked in their current jobs, the higher is their probability of retiring. Individuals who are less attached to the labour market, as expressed by having a non-permanent job or working for fewer hours per week, are more likely to retire. Individuals living in London or in relatively more deprived areas are also more likely to retire. Pension and financial variables play a dominant role in explaining the retirement decision. Having a higher total pension wealth and having a DB scheme increases the likelihood of retiring, while having a defined DC scheme reduces it. Initial health variables explain retirement decisions only to a minor extent. One exemption is having experienced a severe cardiovascular related disease at

Table 2: Probit estimates for retirement at wave 2

	coefficient	t-stat	
constant	-41.41	5.50	***
female	0.20	1.28	
age in year 2002	1.17	4.67	***
squared age in year 2002	-0.01	3.87	***
child lives in household	-0.25	1.91	*
country of birth outside the UK	-0.31	1.07	
degree	-0.27	1.42	
number of years worked in current job	0.01	2.74	***
current job is permanent	-0.25	1.17	
1-30 hours work per week	0.44	2.90	***
London	0.33	1.75	*
fourth quintile index of multiple deprivation	0.42	2.87	***
fifth quintile (most deprived) index of multiple deprivation	-0.15	0.78	
state pension age reached at wave 2 interview	0.42	1.26	
has been offered early retirement incentives and did not take it	-0.24	1.07	
has a defined benefit pension scheme	0.33	2.10	**
has a defined contribution pension scheme	-0.26	1.77	*
total pension wealth/10000	0.01	3.25	***
fair or poor self-assessed health	0.21	1.18	
has limiting long-standing illness	0.23	1.32	
has diabetes and/or hypertension	-0.17	1.27	
has angina, heart attack or stroke	0.67	2.73	***
has diagnosed arthritis or osteoporosis	0.14	1.04	
has diagnosed cancer	0.16	0.67	
health shock in any condition before retirement	-0.08	0.35	
number of conditions of long-standing illnesses at HSE	0.13	1.67	*
hypertensive blood pressure at HSE	-0.24	1.77	*
current smoker	-0.09	0.57	
drinks over limit	0.16	1.21	
never engages either in moderate or vigorous activity	-0.08	0.33	
likelihood of working	-1.33	8.09	***
missing likelihood of working	-1.32	3.70	***
likelihood that health limits ability to work	0.15	0.73	
missing likelihood that health limits work ability	-0.42	1.16	

Note: ***,**,* denote significance at the 1, 5, 10% level. Standard errors are clustered at household level. Estimates are based on 1439 observations. Out of them 192 individuals retire at wave 2. Log-likelihood is -330, Efron (1978) R^2 is 0.38.

the first ELSA interview, which raises retirement probability. The likelihood of retiring also increases with the number of health conditions which were already present at the time of the HSE interview. It is smaller for individuals with hypertensive blood pressure as measured by the nurse at the HSE interview. The latter coefficients, however, are

only significant at the 10% level. Health-related behaviour does not significantly explain retirement. Anticipation about work, however, is an important predictor. Individuals who expect to have a higher chance of remaining in work are less likely to retire. Having missing expectations about future work reduces the retirement likelihood. Expectations about experiencing a work limiting health shock are not significant. With an Efron's R^2 of 0.38, the overall explanatory power of the regressors is relatively high. This suggests that retirement is highly selective and that we need to carefully control for the selection mechanism. Section 7.4 in the Appendix provides a further discussion on this issue.

With respect to policy implications, Probit estimates suggest that individuals' decisions to retire might be affected to some extent by institutional changes. In our specification, the state pension age does not significantly affect retirement, but its sign points in the right direction. Policymakers who aim to increase retirement age might also do so by promoting pension schemes with appropriate incentives. To some degree, retirement can be postponed by improving health of employees, in particular with respect to cardiovascular diseases. We note, though, that we cannot fully assess potential gains of investing in health, since the sample under consideration has been employed and hence had at least an initial health state that enabled to be in work.

5.2 Health effects of retirement

The estimated ATET are presented in Table 3. The first column shows the estimated effects for the whole sample, while the other columns show results for subsamples. Results are obtained by applying the matching estimator as described in Section 4.3. We match on the propensity score as estimated in Table 2 and on gender, age, and squared age.[8] Retirement significantly increases the risk of being diagnosed any new chronic disease (either angina, heart attack, diabetes, stroke, arthritis, cancer or psychiatric problem) by 9 percentage points. It raises the likelihood of being diagnosed with a severe cardiovascular related disease by 3 percentage points. It also raises the risk for other somatic, but non-

[8]We slightly modulate the propensity score and the balancing scores in the subsamples.

cardiovascular related diseases (cancer or arthritis) by 8 percentage points. When looking at the specific conditions separately, we find significant effects for having a stroke or a cancer diagnosis. The other effects are not significant, but the signs never become negative.

Table 3: Estimated average treatment effects on the treated

	sample	gender		social class		age	
		female	male	low	high	50-59	60-69
any chronic condition	0.09**	0.09	0.07	0.02	0.11*	0.03	0.16**
	(2.06)	(1.49)	(0.77)	(0.23)	(1.87)	(0.54)	(2.48)
angina, heart attack, or stroke	0.03*	0.02	0.04	0.02	0.03	0.01	0.04
	(1.66)	(1.10)	(1.05)	(0.90)	(1.15)	(0.37)	(1.45)
arthritis, or cancer	0.08**	0.08	0.02	-0.02	0.11**	0.03	0.13**
	(1.96)	(1.49)	(0.33)	(-0.43)	(2.16)	(0.55)	(2.51)
angina	0.01	0.00	0.00	0.01	0.01	0.01	.
	(0.56)	(-0.27)	(0.19)	(0.48)	(0.78)	(0.78)	
heart attack	0.01	0.01	0.02	.	0.02	0.01	0.01
	(0.80)	(1.00)	(1.27)		(1.23)	(0.40)	(0.38)
stroke	0.02*	0.01	0.04	0.04	0.01	.	0.04*
	(1.87)	(1.00)	(1.50)	(1.62)	(0.82)		(1.83)
diabetes	0.02	0.02	0.01	0.03	0.00	0.02	0.04*
	(0.85)	(0.83)	(0.38)	(1.28)	(-0.17)	(0.76)	(1.82)
arthritis	0.05	0.03	0.01	-0.06	0.09*	0.01	0.10*
	(1.36)	(0.70)	(0.19)	(-1.31)	(1.84)	(0.17)	(1.81)
cancer	0.03*	0.04*	0.02	0.04	0.03	0.03	0.04
	(1.81)	(1.66)	(0.73)	(1.21)	(1.49)	(1.13)	(1.58)
psychiatric	0.01	-0.01	0.03	-0.01	0.03	0.01	-0.02
	(0.34)	(-0.44)	(0.91)	(-0.28)	(0.96)	(0.31)	(-0.62)
number of treated	192	103	89	90	102	94	93
common support	0.96	0.94	0.92	0.94	0.96	0.94	0.91

Note: T-stats in parentheses. ***,**,* denote significance at the 1, 5, 10% level. Standard errors are clustered at household level. Last row shows percentage remaining after common support.

Some of these conditions, e.g. heart attack or stroke, rapidly develop, others such as cancer grow over longer periods. In any case, retirement might be only one among other risk factors. However, we can only estimate the effects of retirement on *diagnosis*, but not on *onset* of the disease. It seems to be likely that diagnosis and onset coincide for conditions with a rapid onset. However, for the other conditions it is not clear, whether the risk of the actual condition or whether the likelihood of diagnosis has increased due to

retirement. It could just be the case that individuals have more time to see their physician after retirement. However, policy implications would be different then, as early diagnosis might help to more effectively treat the disease, so that retirement would actually be beneficial for health in the long run. Unfortunately, we are not able to distinguish between both possible explanations. But we keep in mind that visiting a physician is likely to be induced by some complaints and not only by time availability and that other studies have found a positive association between retirement and cancer mortality that cannot be explained by time availability (Morris, Cook, and Shaper, 1994; Bamia, Trichopoulou, and Trichopoulos, 2007).
We performed a profound robustness analysis, which is described in Section 7.5 in the Appendix. We found the results for our main population to be quite stable.

In order to analyse effect heterogeneity, we split the sample into further subsamples. However, as the sample size becomes smaller, finding statistical significant effects is less likely. Furthermore, effects become less robust. Despite this, we decided to report them as they illustrate that retirement is likely to have diverse effects for different individuals. We compare the levels of the ATETs, as they might be informative for detecting tendencies among the different subgroups. However, we are cautious with our interpretation as confidence intervals are wide. Compared to the overall sample, we find that females have a higher risk of being diagnosed with cancer and a lower risk of developing cardiovascular diseases. For male retirees, estimated effects are higher for heart attack, stroke and psychiatric problems, but smaller for arthritis and cancer. We report ATETs for different socio-economic classes, since the presence of health inequalities has been well established in the literature. We observe that individuals from a low social class are more likely to experience a stroke or to have a cancer diagnosis, but less likely to have arthritis. This might indicate that retirement has mixed health effects for them. In contrast, retirement clearly worsens the health for individuals with a higher socio-economic background. We report treatment effects for different ages of retirement, since it is controversial whether earlier or later retirement is associated with mortality (Tsai, Wendt, Donnelly, de Jong,

and Ahmed, 2005). We find that retirement raises the likelihood for being diagnosed with a condition for individuals who were between 50 to 59 years old in 2002. However, adverse health effects are much higher for individuals who were older when they retired.

All these estimates do not take a potential bias due to health-related attrition into account. Table A.4 shows that drop-outs have different characteristics than our sample. In particular, they differ with respect to their socio-economic status. Thus, it might be legitimate to assume that retirement affects drop-outs in a similar way as individuals in lower socio-economic classes. Under this assumption, effects are likely to be underestimated for all conditions with the exception of arthritis.

5.3 Additional results

In this section, we discuss further supporting evidence. ELSA additionally provides information on risk factors that are related to some of the chronic conditions for which effects of retirement have been estimated. These risk factors are measured for all individuals who participated in the nurse interview at wave 2. They complement the health outcomes considered above, as they reflect the health states of individuals, irrespective of whether a diagnosis has been made. However, we cannot observe whether these risk factors were already high before retirement. Thus, reverse causation cannot be fully ruled out. Still we can control for the other confounding variables. When adjusting the risk factors for the same confounding variables as above, we find that retirees have significantly higher LDL cholesterol and fibrinogen levels (see Table A.8), which are risk factors for cardiovascular diseases. However, the interpretation of these results is not straightforward, as will be discussed in Section 7.6 in the Appendix. Still, they can be judged as further support for our findings: retirees appear to have worse health than employed individuals after adjusting for confounding variables.

We also check the robustness of our results with respect to our sample choice. The first column in Table A.9 in the Appendix shows estimates when we only use information from the first two waves, but discard any information from wave 3. This specification reduces the scope of attrition. But it only allows us to observe health outcomes in the short-run.

Furthermore, retirement and employment are less strictly defined. We find that the adverse effects of retirement are usually smaller. One exception are the effects for severe cardiovascular diseases, which are 1 percentage point higher.

Finally, we estimate the short-run health effects of retirement by looking at new health conditions that have been diagnosed within the first year after retirement. Table A.9 illustrates that the adverse health effects of retirement are smaller, but still considerable. They also indicate that - while most of the adverse health effects accumulate within the first year after retirement - retirement continues to have a detrimental effect on health even in the medium-term.

6 Conclusion

In this paper, we examine the effects of retirement on different health outcomes. Data stem from the first three waves of ELSA and allow us to control for potential confounders. Estimates are obtained by non-parametric matching estimators. Reverse causality is ruled out by exploiting the timing of retirement and of a new diagnosis. Health-related attrition might still be present in the estimates, but is most likely to bias estimates towards zero.

First, we provide estimates on the determinants of retirement. We find that employees propensity to retire is largely affected by their age, job characteristics and anticipations to stay in work. Our results also indicate that policymakers can affect the retirement decision by promoting certain pension schemes or employees' health.

Then, we provide estimates on the health effects of retirement for the retirees. We find that retirement increases the risk of being diagnosed with a chronic condition. In particular, it raises the risk of severe cardiovascular diseases such as heart attacks, strokes or angina. We also find that it increases the risk of being diagnosed with cancer, but we cannot disentangle whether the likelihood of diagnosis or the actual incidence has increased. We do not find any significant effects for diagnosis of psychiatric problems, but are aware that psychiatric conditions often remain undiagnosed. These results are fairly robust with regard to different specifications. They are also reflected by risk factors which have been measured by nurses.

Under the assumption that onset and diagnosis of a disease rather coincide, our findings suggest that, *on average*, retirement has detrimental health effects. Furthermore, there is evidence for a substantial effect heterogeneity: retirement does not harm *every* retiree, and - if it does - it provokes *different* conditions. Thus, we do not expect that there is a single *optimal* retirement age that minimises health risks for everyone.

Our findings are in contrast to studies that employ instruments to address the endogeneity of retirement. One explanation for this might be that the subpopulation for which the instrument induced retirement experiences different health effects than retirees on average. Our results raise further questions as to whether the adverse health effects of retirement can be avoided. For this purpose it is crucial to understand through which mechanisms adverse health effects are transmitted. Once larger sample sizes are available, analysing effect heterogeneity in more detail might give us more insights.

7 Appendix

7.1 Further details on data and sample

Data stems from the English Longitudinal Study of Ageing (ELSA). ELSA has been designed to be representative for people born before 29th February 1952, living in private households in England. It was drawn from households that had previously responded to the Health Survey for England (HSE) in 1998, 1999, or 2001. So in addition to the ELSA interviews, information on the ELSA members is available from the HSE. ELSA contains very detailed information on employment, earnings, wealth, pensions, expectations about future health and paid work, and other household and individual demographics. It consists of personal face-to-face computer assisted interviews, self-completion questionnaires and nurse visits. Wave 1 fieldwork was conducted in 2002-2003, wave 2 in 2004-2005 and wave 3 in 2006-2007.

At wave 1, 11392 individuals were interviewed. Only individuals who were not interviewed by a proxy are retained, since proxy interviews are an indication of a serious mental or physical impairment which prevented the respondent from a face-to-face interview. As a second selection criterion, only the 2906 individuals who were employed at the first wave are kept. An individual is considered to be employed if he responds that (i) being "employed" describes his current situation best and (ii) "paid work" was one of his activities in the last month.[9] We deliberately discard self-employed individuals, since they are likely to have more leeway in deciding on retirement and thus selection into retirement might be different.

At wave 2, around 18% of the employed individuals are lost due to attrition.[10] We observe 1803 individuals who remain employed. 268 individuals are newly retired. We consider an individual as retired if she responds that (i) "retired" describes her current situation best and (ii) "not in paid work" was her activity in the last month.[11]

At the third wave, we lose 15% of the individuals who were either retired or employed at wave 2 due to non-response. Out of the individuals who have been employed at the two prior waves, 1247 individuals are still employed at the third wave. Based on these individuals, we will estimate counterfactual health outcomes that we would observe in the absence of retirement. Out of the individuals who retired at wave 2, 192 are still retired at wave 3. For these individuals, we will analyse the health effects in most of our specifications. Thus, our main sample consists of the 1439 individuals who are either still employed or still retired at wave 3. This sample allows us to analyse health outcomes not only in the short, but also in the medium term. In addition, we will analyse effects of

[9] While 99% of employed individuals are also in paid work, only 86% of individuals in paid work also describe their current situation as being employed. By imposing both criteria, we discard individuals in paid work who describe their current situation as retired (6%), self-employed (4%), looking after home or family (2%), being semi-retired (1%), being unemployed or being disabled. As a consequence, selected individuals should be more likely to be working at the first wave.

[10] See Section 3.3

[11] All retired individuals were not in paid work, while individuals who were not in paid work have also described themselves as unemployed (7%), disabled (9%), looking after home or family (10%), being semi-retired (1%)), besides being retired (71%). Thus, by imposing both criteria, we are more likely to estimate the effect of economic inactivity due to regular retirement. We focus on regular in contrast to partial retirement, because potential health effects are more likely to occur if individuals experience a more drastic change in their lifestyle.

retirement when only considering the first two (instead of the first three waves).

Table A.1: Sample selection

	Number of deleted	individuals remaining
wave 1 interview (individuals older than 50 in 2002)		11392
not interviewed by a proxy	-158	11234
employed at wave 1	-8328	2906
no attrition at wave 2	-537	2369
employed at wave 2		1803
retired at wave 2		268
sample at wave 2		2071
no attrition at wave 3	-302	1769
employed at wave 1 to wave 3		1247
retired at wave 2 and wave 3		192
sample at wave 3		1439

Table A.2: Means of selected variables for employed and retired individuals

	Means		Test for Mean Differences	
	Employed	Retired	not adjusted	age adjusted
observations	1247	192		
Socio-demographics				
female	0.50	0.54		
age in year 2002	55.07	60.20	***	.
between 50-54 years old in 2002	0.51	0.05	***	.
between 55-59 years old in 2002	0.38	0.44		.
between 60-64 years old in 2002	0.09	0.40	***	.
older than 64 in 2002	0.02	0.11	***	.
married or cohabiting	0.82	0.79		
number of individuals living in household	2.54	2.11	***	**
child lives in household	0.44	0.22	***	**
no children	0.10	0.11		
has grandchildren	0.36	0.56	***	
country of birth outside the UK	0.05	0.05		

Continued on next page

Continued ...

	Means		Test for Mean Differences	
	Employed	Retired	not adjusted	age adjusted
Education				
degree	0.19	0.11	***	
higher education below degree	0.16	0.15		
A levels	0.10	0.08		
O levels	0.22	0.24		
other grade equivalent	0.03	0.06		
no qualification	0.21	0.29	**	
Socio-economic classification				
professional	0.05	0.05		
managerial and technical	0.34	0.29		
skilled manual	0.24	0.25		
skilled non-manual	0.18	0.20		
semi-skilled	0.14	0.13		*
unskilled	0.04	0.06		
Job characteristics				
number of years worked in current job	11.25	14.67	***	***
current job is permanent	0.95	0.91	*	
1-30 hours work per week at current job	0.25	0.43	***	**
Number of employees working for the same employer in the UK				
1-4 employees	0.04	0.05		
5-19 employees	0.10	0.10		
20-99 employees	0.15	0.10	**	***
100-499 employees	0.12	0.10		
500-999 employees	0.07	0.05		
1000 and more employees	0.51	0.58	*	**
Physical activity at work				
sedentary occupation	0.42	0.40		
standing occupation	0.32	0.34		
physical work	0.22	0.24		
heavy manual work	0.04	0.03		
Regional characteristics				
Government office region				
North East	0.06	0.06		
North West	0.11	0.13		
Yorkshire and the Humber	0.11	0.11		
East Midlands	0.12	0.10		
West Midlands	0.10	0.09		
East of England	0.12	0.11		
London	0.09	0.15	*	
South East	0.17	0.16		
South West	0.11	0.08		

Continued on next page

Continued ...

	Means		Test for Mean Differences	
	Employed	Retired	not adjusted	age adjusted
Population density of postcode sector				
first quintile (least dense)	0.14	0.10		
second quintile	0.23	0.22		
third quintile	0.23	0.26		
fourth quintile	0.23	0.22		*
fifth quintile	0.16	0.19		
Index of multiple deprivation				
first quintile (least deprived)	0.21	0.18		
second quintile	0.17	0.19		
third quintile	0.21	0.15	**	
fourth quintile	0.21	0.23		
fifth quintile	0.19	0.25	*	
Pension characteristics and financial variables				
state pension age reached at wave 2 interview date	0.06	0.34	***	
has been offered early retirement incentives	0.04	0.09	***	
has defined benefit pension scheme	0.39	0.46	*	***
has defined contribution pension scheme	0.35	0.19	***	***
has any private pension scheme	0.74	0.57	**	
has public sector pension scheme	0.25	0.22		
total pension wealth in 2002	185541	221342	**	***
private pension wealth in 2002	127768	158393	**	**
state pension wealth in 2002	57773	62949	*	*
net (non-pension) total wealth	208845	207568		
employment income per week	414	337	***	
Health characteristics				
Cognitive function				
no full score on date question	0.13	0.18		
number of words that could be recalled	6.23	5.92	**	
Number of difficulties in activities of daily living				
due to physical problems	0.70	1.02	***	*
due to mental and physical problems	0.13	0.20		
Self-assessed health				
poor	0.01	0.01		
fair	0.09	0.18	***	**
good	0.31	0.31		
very good	0.39	0.34		
excellent	0.19	0.16		

Continued on next page

Continued ...

	Means		Test for Mean Differences	
	Employed	Retired	not adjusted	age adjusted
Diagnosed conditions at wave 1				
high blood pressure	0.27	0.29		
angina	0.02	0.05	**	
heart attack	0.01	0.03		
congestive heart failure	0.00	0.00		
heart murmur	0.03	0.05		
abnormal heart rhythm	0.03	0.05		
diabetes	0.03	0.06		
stroke	0.01	0.02		
chronic lung disease	0.02	0.03		
asthma	0.09	0.10		
arthritis	0.19	0.27	**	
osteoporosis	0.01	0.05	**	*
cancer	0.03	0.06	*	
psychiatric problem	0.07	0.05		
Symptomatic conditions at wave 1				
heart attack (Rose questionnaire)	0.06	0.09		
angina (Rose questionnaire)	0.00	0.00		
depression (CES depression scale)	0.09	0.08		
bad general health (GHQ12)	0.10	0.08		
Conditions at HSE interview				
number of conditions	0.53	0.73	***	**
hypertensive blood pressure	0.26	0.24		
body-mass index	25.53	24.84		
takes medication	0.36	0.36		
takes vitamins, minerals	0.31	0.28		**
Health habits				
Amount of vigorous physical activity (cycling, jogging, etc.)				
more than once a week	0.27	0.19	**	
once a week	0.14	0.15		
once to three times a month	0.12	0.08	*	
hardly ever or never	0.48	0.58	***	
Smoking and drinking				
current smoker	0.20	0.18		
ex-smoker	0.40	0.39		
no alcohol units per week	0.05	0.05		
under 1 alcohol units per week	0.10	0.11		
1-7 alcohol units per week	0.32	0.25	**	
7-10 alcohol units per week	0.11	0.10		
10-21 alcohol units per week	0.21	0.23		
21-38 alcohol units per week	0.11	0.16		
more than 38 alcohol units per week	0.09	0.06		
has private health insurance	0.23	0.22		

Continued on next page

Continued ...

	Means		Test for Mean Differences	
	Employed	Retired	not adjusted	age adjusted
Expectations about health and work				
likelihood of being alive	0.65	0.66		
missing likelihood of being alive	0.01	0.01		
likelihood of working	0.68	0.28	***	***
missing likelihood of working	0.06	0.26	***	
likelihood that health limits ability to work	0.38	0.31	***	
missing likelihood that health limits work ability	0.04	0.10	***	
Feed forward problems at wave 1 and missing HSE interview				
was not asked to confirm cvd condition at wave 2	0.17	0.19		
wrong information was fed forward	0.00	0.01		
HSE interview missing	0.01	0.03	*	

Note: ***,**,* denote that means are significantly different at the 1, 5, 10% level. Standard errors are clustered at household level. Third column uses standard t-test for mean differences. Fourth column tests for mean differences in age adjusted variables, where variables are balanced on age, age squared and four age dummies.

7.2 Further details on attrition

7.2.1 Reason for non-response at wave 2

Table A.3: Reason for non-response at wave 2, sample drop-outs

	Frequency	Percent
Ill at home during survey period	12	2.2
Away/in hospital during survey period	6	1.1
Physically or mentally incompetent	1	0.2
Died	6	1.1
Not issued	44	8.2
No contact	21	3.9
Refusal before interview	266	49.5
Refusal during interview	8	1.5
Broken appointment	23	4.3
Office refusal	56	10.4
Language difficulties	1	0.2
Other unproductive	15	2.8
Untraced	60	11.2
Moved – out of Britain	18	3.4
Total	537	100.0

7.2.2 Probability of non-response at wave 2 and wave 3

Table A.4: Probit estimates for response rate at wave 2 and wave 3, sample: individuals who are employed at wave 1

	in wave 2	in wave 3
Observations	2906	2906
Percentage with response	0.82	0.72
Log-likelihood	-1297	-1607
Pseudo R2	0.07	0.06
female	0.04	0.09*
	(0.74)	(1.88)
age in year 2002	0.01	0.01
	(0.64)	(0.74)
married	-0.08	-0.01
	(1.28)	(0.20)
child lives in household	0.16**	0.18***
	(2.34)	(2.79)
country of birth outside the UK	-0.36***	-0.27**
	(3.27)	(2.47)
degree	0.18*	0.35***
	(1.83)	(3.88)
higher education below degree	0.34***	0.46***
	(3.76)	(5.37)
A levels	0.42***	0.40***
	(3.55)	(3.93)
O levels	0.03	0.23***
	(0.41)	(3.14)
total income per week/1000	0.13	0.08
	(1.29)	(0.92)
imputation quantity for total income	-0.01	-0.00
	(1.28)	(0.18)
debt/10000	-0.16***	-0.06
	(3.18)	(1.34)
physical work	-0.17**	-0.14**
	(2.51)	(2.19)
heavy manual work	-0.05	-0.13
	(0.33)	(1.04)
East Midlands	0.26**	0.19*
	(2.31)	(1.91)
South East	-0.14	-0.09
	(1.61)	(1.12)
most deprived area (fifth quintile)	-0.14	-0.33***
	(1.50)	(3.72)

Continued on next page

Continued ...

	in wave 2	in wave 3
has any long-standing illness	-0.05	-0.10
	(0.73)	(1.61)
fair or poor self-assessed health	0.14	-0.09
	(1.15)	(0.79)
good self-assessed health	0.08	-0.13
	(0.86)	(1.57)
very good self-assessed health	0.04	-0.10
	(0.46)	(1.33)
bad mental health (GHQ12)	0.25**	0.19**
	(2.36)	(2.06)
diagnosed angina	0.35*	0.14
	(1.87)	(0.91)
diagnosed diabetes	-0.26*	-0.12
	(1.84)	(0.91)
diagnosed arthritis	0.16**	0.08
	(2.04)	(1.14)
diagnosed psychiatric problem	0.28**	0.15
	(2.00)	(1.32)
chances that health limits ability to work	-0.25**	0.02
	(2.36)	(0.24)
missing chances that health limits work ability	-0.29**	-0.15
	(2.21)	(1.21)
missing whether current smoker	0.05	-0.00
	(0.88)	(0.08)
never engaging either in moderate or vigorous activity	-0.16	-0.13
	(1.46)	(1.23)
1998 HSE source year	-0.20***	-0.27***
	(2.61)	(3.76)
1999 HSE year	0.40**	0.07
	(2.25)	(0.42)
HSE interview missing	-0.91***	-1.00***
	(5.22)	(5.77)
not all household members interviewed at HSE	-0.28***	-0.27***
	(2.96)	(3.08)
interview date wave 1	-0.06***	-0.05***
	(3.82)	(3.07)
has not fully completed sc booklet at HSE	0.09	0.20
	(0.37)	(0.86)
did not agree to nurse visit at individual HSE	-0.28	-0.29
	(1.36)	(1.45)
nurse interview not completed at HSE	-0.25*	-0.14
	(1.69)	(1.05)
Constant	1.23**	0.77*
	(2.45)	(1.65)

Note: T-stats in parentheses. ***,**,* denote significance at the 1, 5, 10% level. Standard errors are clustered at household level.

7.2.3 Differences between two and three waves balanced sample

One problem we face in our analysis is the substantial attrition rate, as has been illustrated by Table A.1. Attrition does not only reduce the sample size and hence aggravates statistical inference, but is also a potential threat for the consistency of our estimates: if attrition is not at random, but related to health and retirement, estimates might be biased.

In principal, one could narrow the degree of attrition by using a balanced sample of two instead of three waves. This might result in a sample with different characteristics because of potentially systematic non-response patterns. Characteristics might also differ because groups would be less strictly defined when using only two waves. For instance, individuals who retired at wave 2, but switched back to work at wave 3 would be considered to be retired, while they would not be when using three waves.

To assess this issue, we estimate a probit where the dependent variable is whether an individual who was employed (retired) at wave 2 is still employed (retired) at wave 3. Thus, the dependent variable takes the value zero if the individual is lost due to non-response at wave 3 or due to switching to another state. As explanatory variables we can now use health variables from wave 2, since this sample is conditional on having responded at wave 2. Table A.5 below indicates that observed health variables at wave 2 do not contribute much to explain whether individuals still belong to the same group at wave 3. For employed individuals, the only significant health variable is the diagnosis of an arthritis, which is only significant at the 10% level. For retired individuals, a worse health state (measured by not good self-assessed health, diagnosed angina and a high C-reactive protein) reduces the likelihood of responding to be retired at wave 3. However, we note that most of the coefficients are small and insignificant. In particular, our health outcomes of interest as diagnosed at wave 2 do not add much explanatory power.

As argued before, this does not rule out that there is health-related attrition at wave 3, but only suggests that attrition is not strongly related to observed health variables at wave 2. Still, we would expect that attrition with regard to our outcomes of interest should be correlated with the same variables as diagnosed one wave before. Thus, it seems to be highly unlikely that health-related attrition at wave 3 results in systematically different estimation results as compared to using only two waves. On the other hand, using three waves allows observations of health outcomes for a longer time horizon. For these reasons, we opt to use three waves in our main specification. However, we keep in mind that health-related attrition might be present in our sample. We would expect, though, that most of the health-related attrition has accrued before wave 2, while at wave 3 there seems to be less scope for it.

Table A.5: Probit estimates for probability of belonging to the same group at wave 3 as in wave 2

sample	employed at wave 2	retired at wave 2
Observations	1803	268
Percentage with (same) response	0.69	0.72
Log-likelihood	-1053	-142
Pseudo R2	0.05	0.11
Constant	3.32***	1.00
	(6.05)	(0.67)
female	-0.04	0.07
	(0.68)	(0.38)
age in year 2002	-0.05***	-0.01
	(5.58)	(0.63)
limiting long-standing illness at wave 2	-0.06	0.08
	(0.68)	(0.33)
good self-assessed health at wave 2	0.15	0.82***
	(1.36)	(2.93)
very good self-assessed health at wave 2	0.12	0.34
	(1.07)	(1.20)
excellent self-assessed health at wave 2	0.12	0.60*
	(0.91)	(1.77)
missing nurse interview at wave 2	-0.62***	-0.17
	(5.17)	(0.49)
hypertensive blood pressure at wave 2	-0.08	0.25
	(1.07)	(1.11)
high total cholesterol (more than 5 mmol/l) at wave 2	0.10	0.17
	(1.34)	(0.81)
high waist hip ratio at wave 2	-0.05	0.09
	(0.74)	(0.48)
high c-reactive protein (more than 3 mg/l) at wave 2	-0.11	-0.43**
	(1.36)	(2.03)
Health outcomes after assigned/actual retirement at wave 2		
newly diagnosed angina	-0.30	-1.50*
	(0.47)	(1.80)
heart attack	-0.15	0.07
	(0.26)	(0.07)
stroke		0.76
		(1.44)
newly diagnosed arthritis	-0.35*	1.02
	(1.87)	(1.59)
newly diagnosed cancer	-0.14	
	(0.32)	
newly psychiatric problem	-0.05	
	(0.15)	

Note: T-stats in parentheses. ***,**,* denote significance at the 1, 5, 10% level. Standard errors are clustered at household level. We also control for marriage, child in household, three educational dummies, employment income, debts, government office region. Stroke is not included as a regressor in the first column, as none of the employed drop-outs had this condition. Cancer and psychiatric problems are not contained as regressors in the second column, since none of the retired drop-outs had these conditions.

7.3 Further details on the estimator

Table A.6: Implementation of the estimator

step 1	We estimate a probit model to obtain the choice probabilities $\widehat{p_i} = Pr(D = 1\|X = x_i)$ for every observation.
step 2	We restrict sample to common support. All $D = 1$ observations with $\widehat{p_i}$ larger than the largest propensity score among the $D = 0$ observations are deleted.
step 3	We estimate the counterfactual expectation of the outcome variable $E[Y^0\|D = 1]$ for the common support.
step 3a	We find for every observation i of the $D = 1$ sample the nearest $D = 0$ observation in terms of the Mahalanobis distance. Mahalanobis distance is defined with respect to the propensity score and a subset of covariates $(\widehat{p_i}, \widetilde{x}_i')'$. This distance is denoted as δ_i for observation i from the $D = 1$ sample. The maximum distance among all $D = 1$ observations is denoted as δ. In other words, δ is the smallest possible value such that for every $D = 1$ observation there is at least one $D = 0$ observation within a δ neighbourhood with respect to Mahalanobis distance. This value δ would be a natural choice for the bandwidth of the kernel. However, a smaller bandwidth often leads to better finite sample properties (Lechner, Miquel, and Wunsch, 2006). In our preferred specification, we choose $h = 0.9\delta$ as the bandwidth value for the kernel. This of course implies that there are some (at least one) $D = 1$ observations for which the neighbourhood as defined by h does not contain any $D = 0$ observations. In this case, the bandwidth is extended locally such that exactly one $D = 0$ observation is included.
step 3b	With the choice of h, we calculate the ATET according to $$E[Y^1 - \widehat{Y^0}\|D = 1] = \frac{1}{N_1} \sum_{i:D_i=1} (y_i - \widehat{m}_0(x_i)),$$ where $\widehat{m}_0$ is a kernel estimator as described in Section 4.3.
step 3c	As a final step, we reduce bias through a weighted regression following a recommendation of Imbens (2004). This procedure takes into account a potential bias if treated and matched comparison are not exactly identical in values of X. Then some regression adjustment for differences in X can help to reduce bias. Let i be an observation from the $D = 1$ sample and $j(i)$ be the nearest observation from the $D = 0$ sample. The estimated counterfactual outcome for i is thus $y_{j(i)}$ and the mismatch in X is $x_{j(i)} - x_i$. If we assume that $E[Y\|X, D = 0] = X'\beta_0$ is linear, the adjusted counterfactual for i is $y_{j(i)} - (x_{j(i)} - x_i)'\beta_0$. The ATET is therefore $$\frac{1}{N_1} \sum_{i:D_i=1} (y_i - y_{j(i)} - (x_{j(i)} - x_i)'\beta_0.$$ With radius matching, we need to replace $y_{j(i)}$ with the expression for $\widehat{m}_0(x_i)$ and similarly for $x_{j(i)}$.

Table A.7: Estimation of the variance for clustered data

Using estimators of $\widehat{m}_0(x_i)$, it is readily seen that the ATET estimator may be expressed as the weighted average of y_i in the form

$\sum_{i:D_i=1} \frac{y_i}{N_i} - \sum_{i:D_i=0} y_i w_i$, where $\sum_{i:D_i=0} w_i = 1$ and $\sum_{i:D_i=1} N_i = 1$.

Following Lechner (2001), since the two terms are based on independent subsamples and treating the weights as given, the variance of the ATET is the sum of the variance of these two terms. In the following, we focus on the calculation of the variance for the second term and note that the calculation for the first is analogous by replacing w_i with $1/N_i$. We introduce the cluster structure by writing

$\sum_{i:D_i=0} y_i w_i = \sum_{j=1}^{J} \sum_{i=1}^{N} I(C_i = j)(1 - D_i) y_i w_i$ where the sums are over all N individuals and over the J households. The variable $C_i \in 1, ...J$ indicates the household to which individual i belongs. The indicator function I takes value if the household of individual i is household j. The number of individuals for household j in the $D = 0$ and weighted by w_i is thus given by

$N^j = \sum_{i=1}^{N} I(C_i = j)(1 - D_i) y_i w_i$. We now compute the variance by allowing that the outcomes across individuals in the same household are dependent, but assume that observations across households are independent:

$$Var[\sum_{i:D_i=0} y_i w_i] = \sum_{j=1}^{J} Var[\sum_{i=1}^{N} I(C_i = j)(1 - D_i) y_i w_i]$$

$$= \sum_{j=1}^{J} (N^j)^2 Var[\frac{1}{N^j} \sum_{i=1}^{N} I(C_i = j)(1 - D_i) y_i w_i] = \sum_{j=1}^{J} (N^j)^2 Var(A_j)$$

where $A_j = \frac{1}{N^j} \sum_{i=1}^{N} I(C_i = j)(1 - D_i) y_i w_i$. Hence, the variance is obtained by summing over the households the variance of expression A_j. Since the A_j are independent across households we obtain

$$Var[\sum_{i:D_i=0} y_i w_i] = Var(A) \sum_{j=1}^{J} (N^j)^2$$

and can estimate $Var(A)$ as

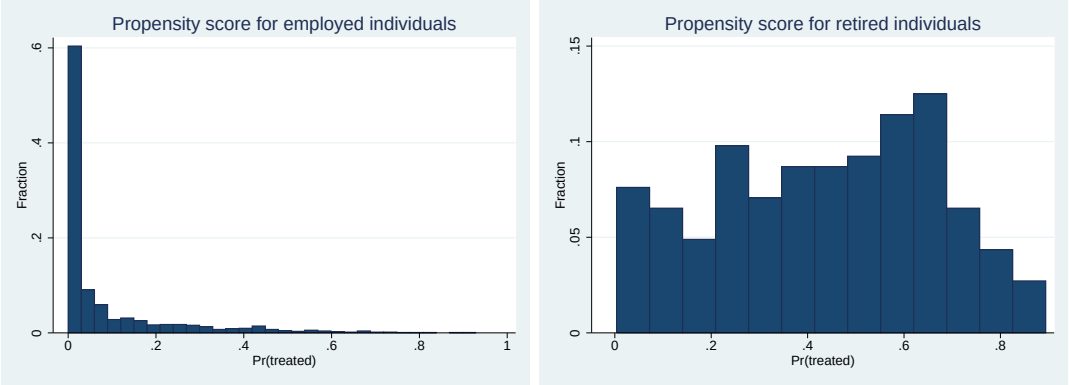
$$Var(A) = \frac{1}{J} \sum_{j=1}^{J} [A_j - \frac{1}{J} \sum_{j=1}^{J} A_j]^2$$

which we now plug in the previous formula.

7.4 Further details on selectivity of retirement and match quality

The high selectivity of retirement is also illustrated by Figure A.1 below. It shows the distribution of propensity scores for employed and retired individuals. Most of the individuals who remain employed throughout the three ELSA waves have a very low retirement probability with a mean of 8%. For retired individuals the support of the propensity score ranges from 0 to 0.96 with quite considerable probability mass in the middle. Differences in distributions reflect the differences between retired and employed individuals. To apply propensity score matching, however, the shape of the distribution can be different, as long as there are comparison individuals with propensity scores similar to retired individuals. In our case, even very high values of the propensity score are observed in the control group (its maximum is at 0.91). We delete retired individuals with a propensity score larger than this maximum from the estimation. This common support restriction turns out to be not very restrictive: we delete eight observations (around 4% of the retired individuals).

Figure A.1: Distribution of propensity scores



Furthermore, we do not use all employed individuals, but only those who are *similar* to the retired individuals (as explained in detail in Section 4.3). Thus, most of the employed individuals with a very low propensity to retire are not included in our estimation, while those with a higher propensity score are used more often. In our case, estimation is based on a quite diverse comparison sample: 597 different individuals are used in the estimation and most of them (569) are used around once as comparison.

In order to assess the match quality, we check whether the marginal distributions of the covariates X are identical in the treated and the matched weighted comparison subsample. As suggested by Rosenbaum and Rubin (1983), matching on the propensity score leads to a balancing of the covariates X in the $D = 1$ and $D = 0$ population. Hence, after matching treated and control, the joint distribution of X and thereby the marginal distributions of X should be identical in both matched subsamples. As argued by Imai, King, and Stuart (2008), traditional hypothesis tests for equality of means such as t-, F-, Chi2- or Kolmogorov-Smirnov tests are problematic for evaluating balance. Since our estimation process also includes some bias adjustment, we prefer to base our balancing test on this estimation procedure instead of using equality of means tests. More precisely, we treat one of the covariates X as if it were the outcome variable and estimate the treatment effect of D on this covariate. The treatment effect on this covariate must be zero by definition of the estimator. We thus test whether the estimated treatment effect of D on this covariate

is indeed zero. Since this test has no power for the balancing scores $\widetilde{X}$, we also follow Imai, King, and Stuart (2008) and look at the quantile-quantile plots. Tests suggests that marginal distributions between treated and matched comparison are similar.

7.5 Robustness analysis

We performed a profound robustness analysis to check the sensitivity of our results. First, we reduce the radius to $R = 10$. Estimated treatment effects hardly change, while variance increases. The only exception is severe cardiovascular disease: the ATET increases to 0.04 and is significant at the 1% level. This is mainly due to an increased risk of having a heart attack. As the bias should be reduced by a smaller radius due to matching more similar individuals, we consider our estimates for cardiovascular diseases in Table 3 as rather conservative.
Second, we use different weights for the balancing scores as compared to the propensity score. We find that coefficients are not much affected.
Third, we chose different balancing scores. When adding expectations about work, coefficients and significance increase. When adding expectations about health shocks or having experienced a severe cardiovascular disease in the past, treatment effects remain comparable and even rise for cardiovascular diseases. Adding other health variables does not considerably change treatment effects either.
Fourth, we vary the propensity score. We find that our estimates are not much affected by (i) eliminating insignificant health variables or (ii) adding further health variables such as limitations in physical activities, memory problems or life expectations.
Fifth, we use age dummies instead of linear and quadratic age terms. This results in a less binding common support restriction: we lose only one retired individual. Estimated effects slightly decrease compared to our main specification. In our main specification, however, weighted means of age indicators in the control group are more similar to those of retired.

7.6 Further details on the additional results

Table A.8 provides mean differences in risk factors that are related to some of the chronic conditions for which ATETs have been estimated. They are adjusted for the same confounding variables which are used to identify the ATETs, as they have been estimated in the same way as the ATET. However, they should not be interpreted as causal effects, since reverse causality cannot be ruled out. We only observe risk factors as measured during the nurse interview, but do not know whether they were already raised before retirement. Still we report them as they show differences in health between retired and employed individuals which are present irrespective of a diagnosis. Compared to unadjusted means, adjusted mean differences are more informative, since they control for the confounding variables used to estimate the propensity score. In other words, when finding significant differences between treated and control groups, it can be ruled out, that these differences are driven by age, gender and so on. Unfortunately, 7% of the individuals in our sample have not participated at the nurse interview at wave 2; so our results are based on a smaller sample.
We do not find any significant effects for all, but two risk factors: retired individuals are 7% more likely to have a high fibrinogen concentration and 6% more likely to have a high LDL cholesterol in their blood sample. Fibrinogen is a soluble protein essential to the

blood clotting mechanism. High fibrinogen is related to increased risk of cardiovascular disease (Smith, Patterson, Yarnell, Rumley, Ben-Shlomo, and Lowe, 2005). LDL cholesterol is a risk factor for cardiovascular disease and intervention studies have shown that a reduction of LDL cholesterol with statins leads to a reduction in these diseases (Baigent, Keech, Kearney, Blackwell, Buck, Pollicino, Kirby, Sourjina, Peto, Collins, and Simes, 2005). We also note that retired individuals are more likely to be overweight or obese. This is not significant when estimated separately, but becomes significant when aggregated. While not significant, other biomarkers such as C-reactive protein, triglycerides and haemoglobin also suggest that retirees have a higher risk for cardiovascular diseases. However, it is not straightforward to interpret even significant differences. First, reverse causality cannot be ruled out. Second, diagnosis of a disease and risk factor need not be correlated. Suppose an individual was diagnosed with a diabetes, then he would also be likely to receive some treatment so that his level of glucose is controlled. Correspondingly, high fibrinogen might be predictive of cardiovascular diseases, but might fall after having experienced a cardiovascular shock.
Although the interpretation of differences in biomarkers is not straightforward, they support our estimated treatment effects: retirees appear to have a worse health than employed individuals after adjusting for confounding variables.

Table A.8: Adjusted mean differences in risk factors

	adjusted mean difference	t-stat	number treated	common support
hypertensive blood pressure	0.01	0.22	161	0.96
high total cholesterol (more than 5 mmol/l)	0.03	1.07	141	0.94
high LDL cholesterol (more than 3 mmol/l)	0.06*	1.69	141	0.94
high triglycerides (more than 1.6 mmol/l)	0.07	0.90	141	0.94
high C-reactive protein (more than 3 mg/l)	0.08	1.21	141	0.94
high fibrinogen (more than 4.2 g/l)	0.07**	2.12	140	0.94
high glucose (more than 7 mmol/l)	0.01	0.22	101	0.94
low haemoglobin (less than 12g/ dl)	0.03	1.11	138	0.94
overweight (bmi between 25 and 30)	0.07	1.09	170	0.96
obese (bmi higher than 30)	0.05	0.84	170	0.96

Note: ***,**,* denote significance at the 1, 5, 10% level. Standard errors are clustered at household level. Last column shows percentage remaining after common support.

We also check the robustness of our results with respect to our sample choice. The first column in Table A.9 below shows estimates when we only use information from the first two waves, but discard any information from wave 3. The second column shows our estimation results from above when we use all three waves. We would expect estimates in the first column to be different for three reasons: (i) retirement and employment are less strictly defined; (ii) there is less attrition; (iii) the health outcomes are observed for a shorter time horizon. However, we would be concerned if our estimates did fundamentally change.
We find that ATETs are usually smaller when using only two waves. However, for developing a severe cardiovascular disease, in particular a heart attack, effects are 1-% point

Table A.9: ATETs for balanced two wave, three wave sample and in the short-run

sample time horizon	wave 1 and wave 2 cumulate	wave 1 to wave 3 cumulate	wave 1 to wave 3 short-run
any chronic condition	0.06*	0.09**	0.06*
	(1.89)	(2.06)	(1.94)
angina, heart attack, or stroke	0.04**	0.03*	0.02
	(2.40)	(1.66)	(1.05)
arthritis, or cancer	0.04	0.08**	0.04
	(1.28)	(1.96)	(1.45)
angina	0.01	0.01	0.01
	(1.30)	(0.56)	(0.55)
heart attack	0.02*	0.01	0.01
	(1.76)	(0.80)	(0.60)
stroke	0.02*	0.02*	0.01
	(1.82)	(1.87)	(1.13)
diabetes	0.01	0.02	0
	(0.44)	(0.85)	(0.45)
arthritis	0.02	0.05	0.02
	(0.76)	(1.36)	(0.68)
cancer	0.02*	0.03*	0.02*
	(1.79)	(1.81)	(1.71)
psychiatric	0.02	0.01	0.01
	(1.19)	(0.34)	(0.56)

Note: T-stats in parentheses. ***,**,* denote significance at the 1, 5, 10% level. Standard errors are clustered at household level.

higher. Overall, estimates confirm that retirement raises the risk of developing certain conditions.

Finally, we estimate the short-run health effects of retirement. We define a short-run health condition to be present if it has been diagnosed up to one year after retirement. The third column in Table A.9 shows the respective ATETs. These are smaller and less often significant than the cumulative health effects in the second column. Calculating the difference between the cumulative and the short-run effects indicates that (i) much of the adverse health effects accumulate within the first year after retirement, but also that (ii) retirement continues to have adverse effects on health even one year after retirement. In particular, the higher risk of developing arthritis or cancer would be underestimated if we were only to observe health outcomes in the short-run.

References

Abadie, A., and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74(1), 235–267.

Ader, R., D. Felten, and N. Cohen (2001): *Psychoneuroimmunology*. Academic Press.

Au, D., T. Crossley, and M. Schellhorn (2005): "The Effects of Health Shocks and long-term Health on the Work Activity of Older Canadians," *Health Economics*, 14, 999–1018.

Baigent, C., A. Keech, P. Kearney, L. Blackwell, G. Buck, C. Pollicino, A. Kirby, T. Sourjina, R. Peto, R. Collins, and R. Simes (2005): "Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins," *Lancet*, 366, 1267–1278.

Bamia, C., A. Trichopoulou, and D. Trichopoulos (2007): "Age at Retirement and Mortality in a General Population Sample. The Greek EPIC Study," *American Journal of Epidemiology*, 167, 561–569.

Banks, J., C. Emmerson, and G. Tetlow (2005): "Estimating pension wealth of ELSA respondents," IFS Working Papers W05/09, Institute for Fiscal Studies.

Banks, J., and S. Smith (2006): "Retirement in the UK," *Oxford Review of Economic Policy*, 22, 40–56.

Bloendal, S., and S. Scarpetta (1999): "The Retirement Decision in OECD Countries," *OECD Economics Department Working Papers*, (No. 202).

Blundell, R., C. Meghir, and S. Smith (2002): "Pension incentives and the patterns of early retirement," *The Economic Journal*, 112, 153–170.

Bound, J., M. Schoenbaum, T. R. Stinebrickner, and T. Waidmann (1999): "The dynamic effects of health on the labor force transitions of older workers," *Labour Economics*, 6(2), 179–202.

Bound, J., and T. Waidmann (2007): "Estimating the Health Effects of Retirements," Working Papers wp168, University of Michigan, Michigan Retirement Research Center.

Brannon, L., and J. Feist (1997): *Health psychology. An introduction to behavior and health.* Pacific Grove, CA: Brooks/Cole.

Charles, K. K. (2002): "Is Retirement Depressing?: Labor Force Inactivity and Psychological Well-Being in Later Life," NBER Working Papers 9033, National Bureau of Economic Research, Inc.

Coe, N. B., and M. Lindeboom (2008): "Does Retirement Kill You? Evidence from Early Retirement Windows," IZA Discussion Papers 3817, Institute for the Study of Labor (IZA).

Coe, N. B., and G. Zamarro (2008): "Retirement Effects on Health in Europe," Working Papers 588, RAND Corporation Publications Department.

Cohen, S., D. Janicki-Deverts, and G. Miller (2007): "Psychological Stress and Disease," *Journal of the American Medical Association*, 298(14), 1685–1687.

Conti, R. M., E. R. Berndt, and R. G. Frank (2006): "Early Retirement and Public Disability Insurance Applications: Exploring the Impact of Depression," NBER Working Papers 12237, National Bureau of Economic Research, Inc.

Dave, D., I. Rashad, and J. Spasojevic (2006): "The Effects of Retirement on Physical and Mental Health Outcomes," NBER Working Papers 12123, National Bureau of Economic Research, Inc.

Deaton, A. S. (2009): "Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development," NBER Working Papers 14690, National Bureau of Economic Research, Inc.

Disney, R., C. Emmerson, and M. Wakefield (2006): "Ill-health and retirement in Britain – a panel data analysis," *Journal of Health Economics*, 25, 621–49.

Efron, B. (1978): "Regression and ANOVA with Zero-One Data: Measures of Residual Variation," *Journal of the American Statistical Association*, 73, 113–212.

Emmerson, C., and P. Johnson (2001): "Pension provision in the United Kingdom," in *Pension systems and retirement incomes across OECD countries*. R. Disney and P. Johnson.

Fisher, R. (1935): *Design of Experiments*. Oliver and Boyd:Edinburgh.

Forman-Hoffman, V., K. Richardson, J. Yankey, S. Hillis, R. Wallace, and F. Wolinsky (2008): "Retirement and weight changes among men and women in the health and retirement study," *Journals of Gerontology Series B: Psychological Sciences and Social Sciences Online*, 63, 146–153.

Frölich, M. (2004): "Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators," *The Review of Economics and Statistics*, 86(1), 77–90.

Gerfin, M., M. Lechner, and H. Steiger (2005): "Does Subsidised Temporary Employment Get the Unemployed Back to Work? An Econometric Analysis of Two Different Schemes," *Labour Economics*, 12, 807–835.

Goldberg, D., and P. Williams (1988): "Users Guide to the General Health Questionnaire," *NFER Nelson*.

Grossman, M. (1972): "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 2(223-255).

——— (2000): *Handbook of Health Economics*chap. 7. Elsvier, Amsterdam.

Grossman, M., and R. Kaestner (1997): *Effects of education on health* pp. 69–123. University of Michigan Press, Ann Arbor.

Heckman, J. J., R. J. Lalonde, and J. A. Smith (1999): "The economics and econometrics of active labor market programs," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3 of *Handbook of Labor Economics*, chap. 31, pp. 1865–2097. Elsevier.

Heckman, J. J., and S. Urzua (2009): "Comparing IV With Structural Models: What Simple IV Can and Cannot Identify," NBER Working Papers 14706, National Bureau of Economic Research, Inc.

Holmes, T., and R. Rahe (1967): "The Social Readjustment Rating Scale," *Journal of Psychosomatic Research*, 11, 213–218.

Imai, K., G. King, and E. A. Stuart (2008): "Misunderstandings between experimentalists and observationalists about causal inference," *Journal Of The Royal Statistical Society Series A*, 171(2), 481–502.

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review," *The Review of Economics and Statistics*, 86(1), 4–29.

——— (2009): "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," NBER Working Papers 14896, National Bureau of Economic Research, Inc.

Joffe, M., T. Have, H. Feldman, and S. Kimmel (2004): "Model Selection, Confounder Control, and Marginal Structural Models," *The American Statistician*, 58, 272–279.

Jones, A., X. Koolman, and N. Rice (2006): "Health-related non-response in the British Household Panel Survey and European Community Panel: using inverse-probability-weighed estimators in non-linear models," *Journal of the Royal Statistical Society, Series A*, 169, 543–569.

Kenkel, D. (2000): *Handbook of Health Economics* chap. 31. Elsevier, Amsterdam.

Kiecolt-Glaser, J., and Glaser (1999): "Psychoneuroimmunology and cancer: fact or fiction? - effects of an early structured psychiatric intervention, coping and affective state on recurrence and survival 6 years later," *European Journal of Cancer*, 35(11), 1603–1607.

Lang, I., N. Rice, R. Wallace, J. Guralnik, and D. Melzer (2007): "Smoking cessation and transition into retirement: analyses from the English Longitudinal Study of Ageing," *Age and Aging*, 36, 638–643.

Larsson, L. (2003): "Evaluation of Swedish Youth Labor Market Programs," *Journal of Human Resources*, 38, 891–927.

Lechner, M. (1999): "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification," *Journal of Business and Economic Statistics*, 17, 74–90.

——— (2000): "An Evaluation of Public-Sector-Sponsored Training Continuous Vocational Training Programs in East Germany," *Journal of Human Resources*, 35(2), 347–375.

——— (2001): "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption," in *Econometric Evaluations of Labour Market Policies*, ed. by M. Lechner, and F. Pfeiffer, pp. 85–124. Heidelberg: Physica Verlag.

——— (2002): "Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods," *Journal Of The Royal Statistical Society Series A*, 165(1), 59–82.

Lechner, M., R. Miquel, and C. Wunsch (2006): "Long-run Effects of Public Sector Sponsored Training in West Germany," University of St. Gallen Department of Economics working paper series 2004 2004-19, Department of Economics, University of St. Gallen.

Lechner, M., and R. Vazquez-Alvarez (2004): "The Effect of Disability on Labour Market Outcomes in Germany: Evidence from Matching," CEPR Discussion Papers 4223, C.E.P.R. Discussion Papers.

Lindeboom, M. (2006): *Health and Work of Older Workers* Elgar Companion to Health Economics.

Lindeboom, M., and M. Kerkhofs (2002): "Health and Work of the Elderly," Tinbergen Institute Discussion Papers 02-025/3, Tinbergen Institute.

Midanik, L., K. Soghikian, L. Ransom, and I. Tekawa (1995): "The effect of retirement on mental health and health behaviors: the Kaiser Permanente Retirement Study," *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 50(1), S59–S61.

Minkler, M. (1981): "Research on the Health Effects of Retirement: An Uncertain Legacy," *Journal of Health and Social Behaviour*, 22(2), 117–130.

Morris, J., D. Cook, and A. Shaper (1994): "Loss of employment and mortality," *British Medical Journal*, 308, 1135–1139.

Neuman, K. (2007): "Quit Your Job and Get Healthier? The Effect of Retirement on Health," *Journal of Labor Research*, 29, 177–201.

Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles," *Statistical Science*, Reprint(5), 463–480.

Nicoletti, C. (2005): "Nonresponse in dynamic panel data models," *Journal of Econometrics*, 132, 461–489.

Perreira, K., and F. Sloan (2001): "Life events and alcohol consumption among mature adults: A longitudinal study.," *Journal of Studies on Alcohol*, 62, 501–508.

Rice, N., J. Roberts, and A. M. Jones (2007): "Sick of work or too sick to work? Evidence on health shocks and early retirement from the BHPS," Working Papers 2007002, The University of Sheffield, Department of Economics.

Riphahn, R. T. (1999): "Income and employment effects of health shocks A test case for the German welfare state," *Journal of Population Economics*, 12(3), 363–389.

Rosenbaum, P., and D. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

Rubin, D. (1974): "Estimating Causal Effects of Treatments in randomized and nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

——— (1979): "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies," *Journal of the American Statistical Association*, 74, 318–328.

Scheier, M., and M. Bridges (1995): "Person variables and health: Personality predispositions and acute psychological states as shared determinants for disease," *Psychosomatic Medicine*, 57, 255–268.

Smith, A., C. Patterson, J. Yarnell, A. Rumley, Y. Ben-Shlomo, and G. Lowe (2005): "Which hemostatic markers add to the predictive value of conventional risk factors for coronary heart disease and ischemic stroke? The Caerphilly Study," *Circulation*, pp. 3080–3087.

Solinge, H. (2007): "Health Change in Retirement A Longitudinal Study among Older Workers in the Netherlands," *Research on Aging*, 29, 225–256.

Sugisawa, A., H. Sugisawa, Y. Nakatani, and H. Shibata (1997): "Effect of retirement on mental health and social well-being among elderly Japanese," *Nippon Koshu Eisei Zasshi*, 44, 123–130.

Szinovacz, M., and A. Davey (2004): "Honeymoons and joint lunches: effects of retirement and spouse's employment on depressive symptoms," *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 59(5), 233–245.

Trichopoulos, D. (1996): "Any questions," *British Medical Journal*, 312, 632–632.

Tsai, S., J. Wendt, R. Donnelly, G. de Jong, and F. Ahmed (2005): "Age at retirement and long term survival of an industrial population: prospective cohort study," *British Medical Journal*, 331, 995–999.