Gutenberg School of Management and Economics

& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

No “Honeymoon Phase”

Whose health benefits from retirement and when

Birgit Leimer

October, 2017

Discussion paper number 1718

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
wiwi.uni-mainz.de

Contact Details:

Birgit Leimer
Chair of Statistics and Econometrics
Johannes Gutenberg University Mainz
Jakob-Welder-Weg 4
55128 Mainz
Germany

bleimer@uni-mainz.de

No "Honeymoon Phase"
Whose health benefits from retirement and when

Birgit Leimer *†

October, 2017

Abstract

I use a fixed effects instrumental variable approach to determine the effect retirement has on health. The exogenous variation in the probability to retire at the normal and early retirement age thresholds is exploited to instrument for the otherwise endogenous retirement decision. Six health aspects are considered: self-assessed health, depression, limitations in (instrumental) activities of daily living, mobility limitations, grip strength and number of words recalled. Using data for 10 countries from the Survey of Health, Retirement and Ageing in Europe (SHARE), I find that retiring both at the normal and early retirement eligibility ages significantly improves all health aspects, including the objective measure grip strength. Results do not generally support the theory that previous research was biased towards zero due to behavioral changes during the anticipation phase prior to retirement. Results also do not show the presence of a honeymoon phase directly following the start of retirement, in which individuals are believed to experience a euphoric state leading health improvements. It appears that individuals, especially blue collar workers, go through an adjustment period after retirement in which they experience more health problems, before stabilizing and improving. Overall, retirement has a health preserving effect for both genders and all occupations in the long term. Neither blue collar workers nor workers with physically or psychologically demanding jobs benefit more from retirement than others.

Keywords: retirement, health, honeymoon, retirement phases, SHARE, fixed effects, instrumental variables
JEL classification: I10, J14, J26

*bleimer@uni-mainz.de
†Faculty of Law and Economics, Johannes Gutenberg University Mainz

1 Introduction

Even though life expectancy has doubled over the last century (World Bank 2016*a*), retirement age thresholds have decreased since Otto von Bismarck introduced a retirement age of 70 in 1889 (von Herbay 2014)[1]. This has lead to a continuously increasing number of retirees (pension benefit recipients) alongside a decreasing number of workers (contributors). As a result, it has become increasingly difficult to fund retirement systems. To oppose this trend, reforms to eliminate early retirement options or to increase the normal retirement eligibility ages have been introduced by many European countries over the past two decades (Hofäcker 2015). Although successful in raising labor force participation among the elderly, these reforms have not been sufficient to establish financial sustainability of pension systems.

Before further changes to the pension systems are implemented, the impact of potential reforms on the retirees' health should be analyzed. As increasing medical expenditures are also putting a financial strain on social security systems, changes to the pension system need to consider the impact on the social security system as a whole. Retirement can be thought to relieve individuals of work-related stress and strain, thereby improving a person's well being. Particularly individuals with physically and mentally straining jobs are expected to benefit from retirement. If retirement improves health, delaying the onset of retirement will delay the health improvement. This may increase health care expenditure prior to retirement and may cause individuals to follow other pathways to exit the labor force, as their health does not allow them to work until reaching the retirement age thresholds. However, others argue that retirement is a break in life structure, leading to a loss of identity and purpose, negatively affecting health. A delayed onset of retirement would then delay the worsening of health, leading to lower or at least delayed medical expenditure. Following this argument, postponing retirement might be beneficial for retirees. It remains unclear, whether health is preserved, unchanged or harmed by retirement.

Providing causal evidence on the impact of retirement on health is not straightforward. Poor health and health shocks influence a person's decision to retire (Dwyer & Mitchell 1999). Additionally, an individual's observable and unobservable characteristics may drive the retirement decision and influence the health status. Both pathways will confound the identification of the effect retirement has on health. Several studies have attempted to account for these endogeneity concerns by using stratified samples or instrumental variables. While these approaches allow for the identification of the causal relationship between retirement and health, no definite conclusions can be drawn as opposing results have been presented.

The inconclusive results can potentially be explained by violations to the homogeneity assumption, which previous research has implicitly made. Retirement does not affect all individuals in the same manner, as the transition to retirement implies different lifestyle and behavioral changes. The heterogeneity of the retirement effect needs to be considered not only across different groups, but also across different retirement phases. Based on the work by Atchley (1980), several economic studies have discussed the presence of a honeymoon effect, during which retirees are thought to experience an idealistic state immediately after retiring. This is expected

[1] Retirement age was lowered to 65 around two decades later, where is has remained since (von Herbay 2014).

to have a health improving effect, especially on perceived health measures such as self-assessed health and depression (see for example Bonsang et al. (2012) or Heller-Sahlgren (2017)). Others, including Behncke (2012) and Coe & Lindeboom (2008), have discussed the potential bias from what Atchley (1980) called the anticipation phase. They argued that individuals may plan their retirement transition before retiring, which could in turn lead to health changes prior to retirement, biasing results. Allowing for heterogeneity in the retirement effect could clear up the inconclusive evidence presented to date.

In this paper, I separate the effect of retirement on health by gender and occupational characteristics. Six different subjective and objective health measures are used as outcome variables, covering both mental and physical health. The health variables include self-assessed health, depression, limitations in (instrumental) activities of daily living, mobility limitations as well as maximum grip strength and a word recall test. The retirement effect is split into an anticipation phase, honeymoon phase and long-term retirement. I exploit the financial incentives to retire at the normal and early retirement age (NRA and ERA), which exogenously increase the likelihood to retire, to instrument for the retirement decision. Individual and wave fixed effects are controlled for to ensure unbiased estimates. The analysis is completed using data from Waves 1, 2, 4, 5 and 6 of the Survey of Health, Ageing and Retirement in Europe (SHARE). The sample is restricted to 50-80 year olds who have been employed, self-employed or retired in all waves, living in the original countries from Wave 1 (except Greece). These countries include Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden and Switzerland.

Results of my analysis show that retiring at the NRA or ERA lead to improvements in health, both in terms of subjective and objective health measures. Other than previous literature, my results show that retirees are significantly stronger in terms of maximum grip strength, which is the only truly objective and not self-reported health measure. The health preserving effect remains in the long term when separating retirement into the anticipation, honeymoon and long-term phases. However, results do not support the theory that health improves prior to retirement, nor do they support the honeymoon effect. Instead, there is evidence of the opposite occurring - retirees, especially blue collar workers, first experience significantly worse health upon retiring. I further find that both genders experience health improvements. Women experience greater improvements in terms of self-assessed health and maximum grip strength, while men are less likely to suffer from depression and limitations in (instrumental) activities of daily living. Contrary to previous work and theoretical considerations, results do not suggest that blue collar workers or workers who consider their job either physically or mentally straining experience greater health improvements. Having children living close by or having grandchildren increases the positive effect retirement has on health. Overall, all individuals appear to benefit from retirement in the long term.

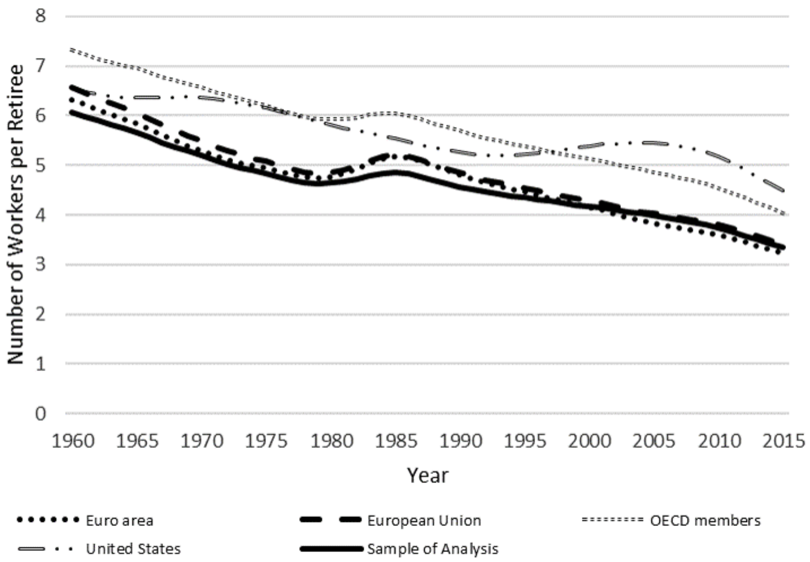
The paper proceeds as follows. Section 2 discusses the concept of retirement, presents the theoretical impact of retirement on health and gives an overview of the current literature. Section 3 introduces the dataset and gives definitions of key variables. The econometric model is described in Section 4. Results are presented in Section 5 and discussed in Section 6.

2 Background

2.1 Retirement Eligibility Rules

The most common retirement eligibility criteria is the normal retirement age (NRA). Reaching the NRA allows individuals to retire and to receive full pension benefits. Most countries also offer an option to retire a younger age, the early retirement age (ERA). The cost of retiring early is reflected in a reduction of the pension benefits for every month they retire before the NRA. Early retirement options became more popular in the 1970s, leading to declining labor force participation of individuals 60+ (OECD 2016). Due to a fertility rates declining simultaneously, there has been a shift in the population distribution away from more workers per retiree to fewer workers. This is captured by the old age dependency ratio, which relates the number of retirees to active workers (see Figure 1). In 1960, there were 6.7 workers per pensioner in Europe. By 2015, the ratio has dropped to 3.3 workers per pensioner and is expected to fall even further (World Bank 2016*b*). As a result, it is becoming increasingly difficult to finance pay-as-you-go (PAYG) public pension systems, where the contributions of current workers pay for the benefits of current pensioners.

Figure 1: Old Age Dependency Ratio - Number of Workers Per Retiree



Source: World Bank Development Indicators

Over the past two decades, many countries have realized that the costs attached to earlier retirement options and the longer retirement durations were causing financial instability in their social security systems. Reforms implemented since this realization have been successful in raising the labor force participation of the elderly (OECD 2016), however they have not been sufficient to establish financial sustainability. Governments continue to discuss reforming their pension systems, including eliminating early retirement options, increasing age thresholds further or linking the retirement age to life expectancy. This raises the question if there is a limit to how long individuals can work. Recent research suggests that there is significant additional health capacity to work at older ages (Coile et al. 2016). However, Coile et al. clearly state that their findings are not intended to suggest what retirement age thresholds should be. The health status of a 65 year old today may be better than the health status of a 65 year old several decades ago, but it is unclear how longer working lives and a delayed entry into retirement will impact their health.

2.2 Theoretical Impact of Retirement on Health

A priori it is not clear how retirement will affect the health of an individual. The theoretical framework proposed by Grossman (1972) views health to be both an investment and a consumption good. Investing in health, either through health-promoting activities or by seeking medical care, decreases the number of sick days, thereby increasing productivity and consequently earnings. After retirement, earnings are no longer dependent on productivity, so the incentive to invest in health to increase earnings disappears. An individual now values consuming health, as better health improves the quality of life. Retirement may also change the marginal value of time, making it cheaper to spend time on health promoting activities. This could potentially lead to health improvements. For some individuals, the marginal cost of time may still be too high to spend on health promoting activities. Depending on the size of the different effects, retirement will therefore improve or worsen health.

Role theory supports this heterogeneity. As Kim & Moen (2002) point out, the role enhancement perspective explains that transitioning into retirement could mean a feeling of identity loss for those individuals whose work was a central part of their identity. On the other hand, retirement can also be seen as a major life-course role exit, which reduces role strain and overload. Being relieved from the stress of their job, individuals may experience less depressive symptoms and feel healthier overall. The direction of the effect, therefore, depends on the circumstances of retirement. A person who had little control in his job, may finally be able to fulfill himself in retirement, while individuals who felt they were in control at their job, will lose their meaning in life with retirement.

Besides changing an individual's role, retirement can alter both the type and amount of social interactions. Social interactions have been shown to improve health (Petrou & Kupek 2008), suggesting that increases in social interaction will lead to better health. There is evidence that especially women benefit from the additional free time after retiring to spend with friends and family (Thomas 2011). If women are able to uphold and even expand their social interactions after retirement, they may experience health improvements. Men, on the other hand, may have more difficulties upholding social interactions, leading to feelings of loneliness.

Retirement also influences other lifestyle aspects, including activity level, stress, smoking behavior, alcohol consumption and dietary habits (Zantinge et al. 2014). Behavior can either change to become healthier or unhealthier, depending on an individual's preferences and work history. For example, individuals with physically demanding jobs may experience a drastic drop in physical activity, leading to severe weight gain and health issues after retirement. Individuals who had great responsibility and pressure in their occupations, may finally experience relief in retirement, leading to better overall health.

Heterogeneity in the retirement effect does not only exist in terms of personal and occupational characteristics, but may also depend on the distance to or from the point of retirement. Psychological literature suggests that the effect of retirement can be separated into the following phases: (1) Preretirement, (2) Honeymoon, (3) Disenchantment, (4) Reorientation and (5) Stability (Atchley 1980). The preretirement phase is split into a remote phase, in which retirement is seen as something occurring far in the future and a near phase, in which individuals begin planning for retirement. During this anticipation and planning phase, individuals may start altering their lifestyle in order to make it healthier, by spending more time exercising and caring less about the stress at work. The honeymoon phase is believed to be a euphoric period, in which the retiree enjoys the new-found freedom, time and space. Especially self-assessed health measures and mental health could improve drastically through this feeling of euphoria. The third stage, disenchantment, encompasses a period of feelings of letdown, possibly depression, when an individual realizes retirement is not only an extended vacation, but has its downsides as well. This is followed by a period of reorientation, where new alternatives are developed, leading right into a stability phase.

2.3 Literature Review

The ambiguity of theoretical considerations is reflected in the empirical literature. Early correlational work identified a negative association between retirement and health (Dave et al. 2008). This relationship cannot be considered causal, as the results can, in part, be explained by poor health or unexpected health shocks increasing the likelihood of an individual to retire (Dwyer & Mitchell 1999). To fully account for the endogeneity caused by reverse causality, it has become widely accepted to use an instrumental variables (IV) approach. Coe & Lindeboom (2008) and Coe et al. (2012) used self-reported offers of early retirement in the Health and Retirement Survey (HRS) to instrument for the retirement decision. They found no statistically significant worsening of health, with a slight health improvement for certain sub-groups, including blue-collar workers. Hallberg et al. (2015) exploited an early retirement offer made to military officers in Sweden and found that retirement decreased mortality as well as the number of days of inpatient care. The unexpected nature of these retirement offers prevents individuals from preparing for retirement, thereby excluding potential bias due to behavioral adjustments prior to retirement. However, individuals who are offered early retirement packages are not representative of all workers. Early retirement packages are more common in large companies in the manufacturing, utilities and banking industries and among white collar workers.

Insler (2014) instrumented for the retirement decision using self-reported probabilities to work past the ages 62 and 65. He argued that the instruments fulfill the exogeneity requirement, as individuals answer these question before retiring and are therefore unaware of unanticipated retirement-inducing health shocks. However, it is easily argued that these instruments fail to fulfill the exogeneity requirement. It is very likely that individuals consider their current health and their expected future health, based on family history and health behavior, in their evaluation of this question.

By far the most common instruments are the normal and early retirement eligibility ages. Reaching these eligibility ages increases the probability of an individual to retire, without having a separate effect on health (see section 5 for a detailed explanation of the IV-strategy). Early studies used cross-sectional data and found no effect on health (Bound & Waidmann 2007), worse health (Rohwedder & Willis 2010) and better health (Coe & Zamarro 2011). To better account for endogeneity issues and to obtain more precise results, further studies used panel data in the hope to identify the true direction of the effect (Neuman (2008), Behncke (2012), Mazzonna & Peracchi (2012), Bonsang et al. (2012), Gorry et al. (2015), Heller-Sahlgren (2017), and Mazzonna & Peracchi (2017)).

One limitation in many of these studies is the failure to control for unobserved individual heterogeneity. Unobserved individual characteristics, such as time preferences, influence both the decision to retire and the health of a person. Results from estimations without individual level fixed effects are therefore biased and have to be considered with caution. Those who do control for individual level fixed effects alongside their IV strategy still found opposing effects. Using HRS data, Bonsang et al. (2012) found that retirement had a diminishing effect on cognitive abilities. Further negative effects on mental health were found by Heller-Sahlgren (2017), while Mazzonna & Peracchi (2017) suggested that while retirement worsens cognitive ability and self-assessed health for a large part of the population, it also improved the health of those who were previously in physically demanding occupations. Furthermore, Gorry et al. (2015) found that several different health aspects improved with retirement.

Another potential problem in previous research is a possible violation to the exogeneity assumption when using the NRA as an instrument in US datasets (Gorry et al. (2015), Bonsang et al. (2012) and Rohwedder & Willis (2010)). The NRA coincides with the eligibility age to receive health insurance coverage through Medicare, which has been shown to have a separate, health-improving effect (Card et al. 2008). As a result, turning 65[2] affects health through other paths than just its effect on the likelihood to retire, leading to biased results. To circumvent this issue, Neuman (2008) excluded the NRA in his set of instruments. It is to date unclear if results excluding the NRA will be externally valid, as it is not known whether the health effects of retiring at the NRA or at the ERA are the same. It is possible that individuals with poorer health self-select into occupations in which early retirement is common, thereby leading to different effects of retirement at the NRA and ERA.

Not only could the health effect of retirement depend on retiring at the NRA or ERA, but also on the aspect of health which is considered. Several studies focus only on cognitive abilities and mostly found that cognitive abilities declined with retirement (Mazzonna & Peracchi (2012), Rohwedder & Willis (2010), Bonsang et al. (2012)) or that retirement had no effect on cognitive abilities (Coe et al. 2012). The results when considering depression as the health outcomes have been mixed, although the overall tendency is no significant effect (Neuman (2008), Coe & Zamarro (2011), Gorry et al. (2015)). A overwhelmingly positive effect on health has been measured when self-assessed health (SAH) was used. Most studies looking at SAH found that health was perceived to be better after retirement (Neuman (2008), Coe & Zamarro (2011)). Results using other health outcomes such as the number of chronic conditions, disease diagnosis, or limitations in (instrumental) activities in daily living, have lead to mixed results without an overall trend becoming apparent.

[2]The NRA is currently incrementally increasing to reach 67 for the cohorts born in 1960 or later. Most work that has been done in the US has used cohorts for whom 65 was still the NRA.

Studies differ not only in respect to health outcomes, but also their sample restrictions. Most studies either restricted their sample to men or split their analysis by gender. Some evidence has been found that the effects of retirement on health differ by gender, however little evidence exists for the impact on women. Very few studies extended their heterogeneity analysis further. Results by Mazzonna & Peracchi (2012) suggest that individuals with more years of education seem to experience greater health improvements when they retire, while no difference was detected by Heller-Sahlgren (2017). Hallberg et al. (2015), on the other hand, found that individuals without a college education experience a greater health improvement. Coe et al. (2012) explored the heterogeneity among occupational groups and found that blue collar workers experienced a significant health improvement, while there was no significant effect for white collar workers. The study by Mazzonna & Peracchi (2017) explored further occupational differences by considering the physical and psycho-social burden of the last job an individual held before retirement. They found that while retirement overall affects health negatively, those in particularly physically burdensome jobs experience a health improvement.

Several studies have attempt to consider the dynamic effect retirement has on health. A few studies, such as Mazzonna & Peracchi (2017), included the retirement duration to capture long-term effects. However, Mazzonna & Peracchi specified a linear age trend, so it remains unclear whether their long-term detrimental health effect is truly due to retirement worsening health, or if the negative effect captures the negative effect aging has on health. Others have split the analysis into a short and long-term effect by analyzing the effect retiring had after one wave and after two waves (Coe & Lindeboom (2008), Insler (2014), Gorry et al. (2015)). None of these studies explicitly looked at the presence of a honeymoon effect, although the relevance of the honeymoon phase is mentioned several times in the literature.

One main argument against using the NRA and/or ERA thresholds to instrument for the decision to retire is that these age thresholds are well known and can therefore be anticipated and planned for. According to this argument, health effects will already take place before retirement, biasing results (for example Coe & Lindeboom (2008) or Behncke (2012)). The only attempt to identify the presence of an anticipation effect was by Coe & Lindeboom (2008), who compare results using unexpected early retirement offers to instrument for the retirement decision to using the ERA and NRA as instruments. They found some evidence that the anticipation effect may bias results towards zero if the ERA and NRA are used. Although important, no study has managed to instrument for the anticipation and honeymoon phase to explore the dynamic effect retirement has on health.

3 Contribution

This paper contributes to the literature by exploring the heterogeneity of the effect, in terms of occupation and personal characteristics. Furthermore, the analysis differentiates between the effect on health of retiring at the NRA and at the ERA. Most importantly, the effect of retirement is split into three phases: anticipation, honeymoon and long-term effect. To identify which aspect of health is affected by retirement, six objective and subjective health measures are used separately, covering mental and physical health. An instrumental variable approach with individual and wave fixed effects ensures that all endogeneity concerns are addressed. Using the SHARE dataset avoids possible confounding by Medicare effects. Lastly, more SHARE waves are included in the analysis compared to previous work, allowing for the analysis of more individuals over a longer period of time.

4 Data

This paper uses data from SHARE Waves 1, 2, 4, 5 and 6[3] (Börsch-Supan 2017), see Börsch-Supan et al. (2013) for methodological details[4]. SHARE is a multidisciplinary, cross-national, individual-level dataset on health, well-being, socio-economic status as well as social and family networks of the population aged 50+ in several European countries. The third wave, SHARE-LIFE, cannot be used since it is a retrospective survey asking individuals about their life history.

4.1 Sample Selection and Retirement Definition

Some sample restrictions are necessary for this study. Only those countries which were surveyed in all five waves were included, as these original countries are more similar to each other than those added in later waves[5]. The sample includes Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden and Switzerland. Only individuals between the ages of 50 and 80 are considered. Individuals over 80 are excluded since health trends are very different among the very old and selective mortality becomes a greater issue at higher ages.

Retirement is defined using a question about self-declared job situation, in which respondents are asked which of the following best describes their employment situation: retired, employed or self-employed, unemployed, permanently sick or disabled, homemaker or other. All individuals declaring themselves to be retired are considered to be retired, while those declaring themselves to be employed or self-employed are considered to be working. To measure the effect of transitioning into retirement from employment, individuals who ever report any other job status are dropped from the analysis. In the sensitivity analysis, other definitions for retirement and other sample restrictions are used to ensure the robustness of the results.

There are a total of 106 904 observations for 48 616 individuals in the main analysis. The main analysis includes 90 045 observations for the 31 757 individuals who are observed in at least two waves. Around 30% of these individuals are working in all waves, while 54% are retired in all waves. The other individuals retire between interviews [6]. The percentage of retired individuals differs between countries, as is shown in Table 1. The differences in retirees is in part explained

[3]DOIs: 10.6103/SHARE.w1.600, 10.6103/SHARE.w2.600, 10.6103/SHARE.w4.600, 10.6103/SHARE.w5.600, 10.6103/SHARE.w6.600

[4]The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARE-LIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: N°211909, SHARE-LEAP: N°227822, SHARE M4: N°261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169,Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

[5]The Netherlands conducted an experiment using an online survey or telephone interviewing instead of face-to-face interviews for WAVE 6 and therefore data for the Netherlands are excluded in Wave 6.

[6]Only eight individuals change their working status from being retired to working. The rest of the individuals change from working to being retired.

Table 1: Descriptive Statistics by Country

	Avg. Age (in yrs)	Mean Ret. Age	Retired (in %)	Female (in %)	Education (in yrs)	Number Children	Sample size
Austria	65.6	58.2	76.7%	50.9%	9.4	2.0	10 676
Belgium	63.8	60.0	60.0%	43.4%	12.8	2.1	13 765
Denmark	63.8	62.7	49.6%	50.5%	13.7	2.3	10 416
France	64.8	59.4	66.5%	50.4%	11.6	2.2	13 234
Germany	64.9	61.2	59.3%	46.5%	13.1	2.0	11 841
Italy	65.4	58.6	69.3%	38.4%	9.4	1.8	10 623
Netherlands	63.7	61.8	52.9%	37.7%	12.1	2.2	6 838
Spain	66.1	62.2	62.6%	32.4%	9.3	2.2	8 871
Sweden	66.5	63.1	61.9%	52.8%	11.7	2.4	12 498
Switzerland	64.7	62.6	51.9%	45.1%	9.4	2.0	8 142
Total	65.0	60.7	61.7%	45.5%	11.4	2.1	106 904

by the different retirement eligibility ages as well as the different attitudes toward retirement in the different countries.

4.2 Retirement Eligibility Ages of the Sample

The SHARE dataset is supplemented with the relevant NRAs and ERAs. These eligibility age thresholds are gender, cohort, year and country specific. Table 2 gives an overview (by gender) of the most common eligibility ages in the interview years, incorporating the reforms currently being implemented in several countries. There is relatively little variation in the NRA. Men retire at age 65(+) in all countries except France. The variation among women is slightly higher, ranging from 60 in France, to 65(+) in Germany, Sweden, the Netherlands, Spain and Denmark. The ERA shows greater variation, ranging from 56 to 64 for men and 56 to 63 for women[7]. Over the time span of the interviews, retirement ages have increased and some early retirement schemes have already been abolished.

4.3 Health Measures

SHARE provides a variety of health variables, covering different aspects of health. To gain a comprehensive understanding, six health measures are used separately as the outcome variable. They include both subjective and objective measures as well as physical and mental health aspects.

The first health measure, capturing general health, is the self-assessed health status (SAH). It is based on a question asking individuals to rate their health on a scale from 1 (excellent) to 5 (poor). Following convention, an indicator variable is generated which is equal to 1 if a person reports being in very good to excellent health and 0 otherwise. The disadvantage of using SAH is its susceptibility to justification bias, where individuals report poorer health to justify being retired (McGarry 2004). This would downward-bias the results. If a health preserving effect is measured nonetheless, it means the true health preserving effect is larger. Despite its drawbacks,

[7]Even though Denmark does not have an official early retirement age, a voluntary early retirement scheme is available to the majority of the population (OECD 2015).

Table 2: Applicable Retirement Age Thresholds in Europe by Gender

Panel A: Normal Retirement

Male / Female	2004	2006	2007	2011	2012	2013	2015
Austria	65/60	65/60	65/60	65/60	65/60	65/60	65/60
Belgium	65/63	65/64	65/64	65/65	65/65	65/65	65/65
Denmark	65/65	65/65	65/65	65/65	65/65	65/65	65/65
France	60/60	60/60	60/60	60/60	60.3/60.3	60.8/60.8	61.6/61.6
Germany	65/65	65/65	65/65	65/65	65.1/65.1	65.2/65.2	65.3/65.3
Italy	65/60	65/60	65/60	65/60	66/62	66/62	66.3/63.3
Netherlands	65/65	65/65	65/65	65/65	65/65	65/65	65/65
Spain	65/65	65/65	65/65	65/65	65/65	65/65	65/65
Sweden	65/65	65/65	65/65	65/65	65/65	65/65	65/65
Switzerland	65/64	65/64	65/64	65/64	65/64	65/64	65/64

Panel B: Early Retirement

Male / Female	2004	2006	2007	2011	2012	2013	2015
Austria	61/56	62/57	62/57	62/58	63/59	63/59	64/60
Belgium	60/60	60/60	60/60	60/60	60/60	60.5/60.5	61.5/61.5
Denmark	60.5/60.5	60.5/60.5	60.5/60.5	60.5/60.5	60.5/60.5	60.5/60.5	60.5/60.5
France	56/56	56/56	56/56	56/56	56/56	56/56	56/56
Germany	63/60	63/60	63/60	63/60	63/63	63/63	63/63
Italy	57/57	57/57	58/58	60/60	60/60	61/61	61/61
Netherlands	60/60	60/60	60/60	-/-	-/-	-/-	-/-
Spain	61/61	61/61	61/61	61/61	61/61	61/61	61/61
Sweden	61/61	61/61	61/61	61/61	61/61	61/61	61/61
Switzerland	63/62	63/62	63/62	63/62	63/62	63/62	63/62

Source: SHARE job episode panel supplemented by retirement ages provided by the Mutual Information System on Social Protection (http://www.missoc.org/), the US Official Social Security Website (https://www.ssa.gov/) and the websites of the governments of the respective countries.

SAH has been shown to be an independent predictor of mortality, particularly among the elderly and therefore is an important measure to consider (see for example Idler & Benyamini (1997)).

Mental health is captured in two variables - depression and cognitive ability. According to the EURO-D scale, a person is categorized as depressed if at least four out of the twelve symptoms[8] are experienced (Prince et al. 1999). The indicator variable is equal to 1 if a person is not categorized as depressed, i.e. has less than four symptoms. Cognitive ability is captured by the total word recall test, in which respondents are read a list of 10 words and asked to repeat them immediately afterwards and with a small delay. These two word recalls are summed up, giving a maximum score of 20. SHARE also provides other variables measuring cognitive ability, such as numeracy. However, total word recall is used as it measures episodic memory, which is particularly affected by aging (Bonsang et al. 2012).

Physical health is analyzed using three different measures: limitations in (instrumental) activities of daily living, limitations in mobility and maximum grip strength. The first two variables are based on self-reported limitations in activities of daily living (ADL), instrumental activities of daily living (IADL) and mobility[9]. One indicator variable is generated that is equal to 1, if a

[8]Variables forming the EURO-D scale: depression, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment, tearfulness.

[9]ADL include dressing, including putting on shoes and socks; walking across a room; bathing or showering; eating, such as cutting up your food; getting in and out of bed; using the toilet, including getting up or down.

person does not report any limitations in ADL nor IADL. Another indicator variable is equal to 1, if no mobility limitations are reported. Grip strength (0-100 kg) is measured by the interviewer using a dynamometer (Smedley, S Dynamometer, TTM, Tokyo, 100 kg). Respondents are instructed to hold their elbow at a 90° with the upper arms pressed to their body, in a standing (or sitting) position. Respondents are then asked to squeeze as hard as possible - twice with each hand. The maximum grip force is recorded - which can reach values up to 100 kg. It has been shown that it is a good, independent predictor of mortality (Ambrasat et al. (2011), Hank et al. (n.d.) and Leong et al. (2015) among others). The test is constructed so that even the weakest subjects can participate, ensuring a high participation rate. Grip strength is the only truly objective health measure.

Two other widely used health measures, number of chronic diseases (or indicator variables for the presence of certain diseases) and a health index, are not used in this study. As discussed in subsection 2.2, retirement changes the marginal value of time. As time is less restricted, the cost of going to see a doctor decreases. Conditions may be diagnosed that were present before entering retirement, but had gone undiagnosed. This leads to a diagnosis bias, as the diagnosis indicates worse health after retirement, even though the health of the individual was just as poor before. I do not include a health index in which several health variables are used to predict a person's general health, as this will not allow for a heterogeneity analysis. Using an index may hide important differences, as certain groups may experiences changes in one health aspect, while another group experiences the change in a different health aspect.

5 Econometric Model

This section presents the theoretical foundation for the empirical analysis that will follow. First, the baseline ordinary least squares (OLS) model is presented, including its limitations. It will then be followed by the corrected model.

5.1 Baseline Model - Ordinary Least Squares Model

To identify the effect of retirement on health, it would be ideal to compare the health of an individual i as he/she retires in one state of the world (Y_i^1) with the health of that same individual if he/she were to continue working in another state of the world (Y_i^0). As an individual cannot retire and continue working, the average treatment effect (ATE) is identified instead, in which the health status of the retirees (the "treated") is compared to those still working (the "untreated"), by estimating the following equation using OLS:

$$Y_i = \beta_0 + \beta_1 R_i + u_i \tag{1}$$

IADL include using a map to figure out how to get around in a strange place; preparing a hot meal; shopping for groceries; making telephone calls; taking medications; doing work around the house or garden; managing money, such as paying bills and keeping track of expenses. Mobility includes walking 100 meters; sitting for about two hours; getting up from a chair after sitting for long periods; climbing several flights of stairs without resting; climbing one flight of stairs without resting; stooping, kneeling, or crouching; reaching or extending your arms above shoulder level; pulling or pushing large objects like a living room chair; lifting or carrying weights over 10 pounds/ 5 kilos, like a heavy bag of groceries; picking up a small coin from a table.

where i designates the individual, Y_i a health measure and R_i is an indicator variable equal to 1 if an individual is retired and 0 otherwise.

The estimated coefficient of retirement in Equation 1, $\widehat{\beta}_1$, will be biased, unless the average health status of those who are not retired is equal to the health status of those who are, had they not retired. It is unlikely that this holds, as retired and working individuals differ in various characteristics, including age, which have direct effects on health outcomes. One solution is to condition on all confounding variables, X, that jointly affect Y and R by running the following regression:

$$Y_i = \beta_0 + \beta_1 R_i + \theta X_i + u_i \tag{2}$$

where X_i is a set of exogenous controls. In the following analyses I will control for age, age squared, gender, years of education, number of children, survey wave and country of residence. The rest is as described above. Using this regression will lead to biased results, as some variables which influence both the decision to retire and a person's health, are unobserved. Furthermore, an endogeneity problem arises from reverse causality. It has been shown that the retirement decision is in part driven by poor health (Dwyer & Mitchell 1999). The baseline model will therefore result in downward biased effects of retirement on health.

5.2 Corrected Model - Fixed Effects Instrumental Variable Model

5.2.1 Binary Retirement Decision

In order to identify the causal relationship between retirement and health, an individual level fixed effects (FE) approach is used to control for unobserved time-invariant heterogeneity:

$$Y_{it} = \beta_0 + \beta_1 R_{it} + \alpha X_{it} + \mu_i + \tau_t + u_{it} \tag{3}$$

where i is the individual and t is the survey period. Therefore μ_i are individual fixed effects and τ_t wave fixed effects. The other variables are as described above. Using an FE-approach ensures that time-invariant confounding factors, such as genes, health history and environmental factors, are controlled for. The common approach in the literature has been to use age and age squared as control variables. I follow this approach, but test different age trend specifications to ensure the robustness of my results.

This model may still suffer from endogeneity, as individual fixed effects will not remove reverse causality. I exploit the fact that the regressor of interest, retirement (R_{it}), is partially determined by a known, discontinuous function of age, which is not directly related to an individual's health (Y_{it}). Policies determining the normal and early age thresholds at which an individual becomes eligible for old-age pension (see subsection 4.2) change the probability of retiring discontinuously as a function of gender and age. A set of instruments will be used in which there is one indicator variable per gender for being above the relevant age thresholds (either NRA or ERA).

A two stage least square (2SLS) estimation procedure is used. The following first stage regression is estimated:

$$R_{it} = \delta_0 + \delta_1 Z_{igt} + \varphi X_{it} + \mu_i + \tau_t + \epsilon_{igt} \tag{4}$$

where g is the gender of person i, Z_{igt} is the vector of instruments, $I(age_{igt} \geq NRA_{igt})$ and $I(age_{igt} \geq ERA_{igt})$, and the rest is as discussed above. The fitted values, $\widehat{R}_{it}$, are then used to estimate:

$$Y_{it} = \beta_0 + \beta_1 \widehat{R}_{it} + \rho X_{it} + \zeta_i + \theta_t + u_{igt}. \tag{5}$$

5.2.2 Retirement Phases

In a second step, the effect of retirement is separated into anticipation, honeymoon and long-term retirement phases. SHARE provides the exact retirement date, which can be used to calculate the exact time until and since retirement and determines which phase an individual is in. To instrument for these phases, gender-specific indicator variables are used, which are equal to 1 if an individual is within two years before (anticipation) or after (honeymoon) the ERA or NRA. In regressions controlling for the honeymoon effect, the retirement instrument is equal to 1 if an individual is older than the NRA or ERA plus the honeymoon phase. The main analysis will use a length of two years for both the anticipation and honeymoon phases. The robustness of the results will be checked using different phase durations. Controlling for the retirement phases leads to the following first stage regressions:

$$R_{it} = \delta_0 + \delta_1 Z_{igt} + \varphi X_{it} + \mu_i + \tau_t + \epsilon_{igt} \tag{6}$$

$$A_{it} = \alpha_0 + \alpha_1 Z_{igt} + \sigma X_{it} + \xi_i + \pi_t + \nu_{igt} \tag{7}$$

$$H_{it} = \eta_0 + \eta_1 Z_{igt} + \upsilon X_{it} + \chi_i + \psi_t + o_{igt} \tag{8}$$

where A_{it} is an indicator variable equal to 1 if a person will retire within the next two years, while H_{it} is an indicator variable indicating whether a person has been retired for two or fewer years. The vector of instruments, Z_{igt}, now includes indicator variables for being zero to two years below the NRA/ ERA[10], for being zero to two years from the NRA/ ERA[11] and for being over the NRA/ ERA plus two years[12]. The second stage estimation therefore becomes:

$$Y_{it} = \beta_0 + \beta_1 \widehat{R}_{igt} + \beta_2 \widehat{H}_{it} + \beta_3 \widehat{A}_{it} + \rho X_{it} + \zeta_i + \theta_t + u_{igt}. \tag{9}$$

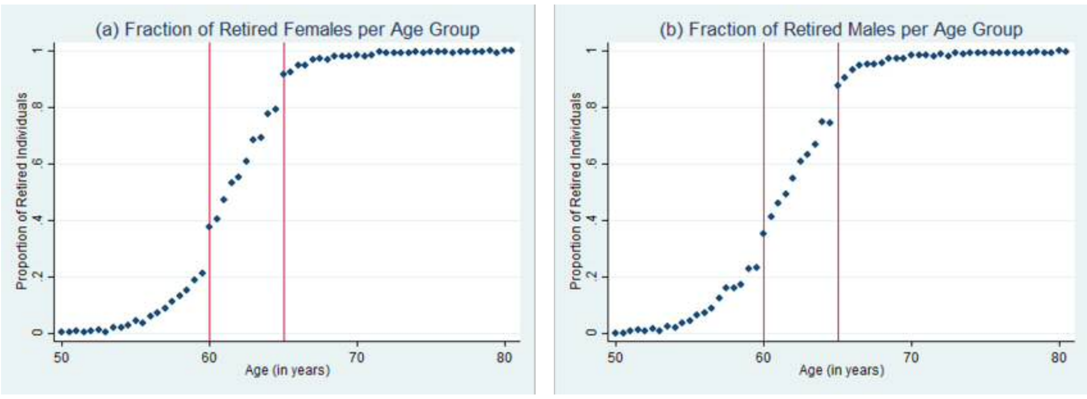
5.2.3 Instrument Validity

Instrument validity depends on three assumptions: relevance, exogeneity and monotonicity. An instrument is considered relevant if it causes a shift in the regressor of interest. Figure 2 gives a visualization of the relevance assumption. It shows the fraction of individuals per age group (divided into 6 month bins) who are retired. The fraction of retirees clearly increases around the average NRA and ERA eligibility thresholds (most common retirement ages illustrated by the vertical lines). There is a jump in the fraction of retired individuals of around 15-20 percentage

[10] $I(NRA_{igt} - 2 \leq age_{igt} < NRA_{igt})$ and $I(ERA_{igt} - 2 \leq age_{igt} < ERA_{igt})$
[11] $I(NRA_{igt} \leq age_{igt} < NRA_{igt} + 2)$ and $I(ERA_{igt} \leq age_{igt} < ERA_{igt} + 2)$
[12] $I(age_{igt} \geq NRA_{igt} + 2)$ and $I(age_{igt} \geq ERA_{igt} + 2)$

points from age 59 to 60 and of around 15 percentage points from age 64 to 65 for both genders. The relevance assumption will also be confirmed by the relevant F-statistics.

Figure 2: Proportion of Retirees Per Age in 6 Months Bins



Note: Due to the difference in ERA and NRA, the vertical lines indicate only the most common age thresholds and therefore these jumps underestimate the true variation caused by reaching the retirement age thresholds.

Instrument exogeneity requires that the NRA and ERA do not impact health through other channels than through their effect on the decision to retire. While health does deteriorate with age, it is unlikely that turning a particular age has a direct effect on health, especially physical health. It could be argued that turning a milestone age, such as 60, negatively affects mental health and therefore biases results when using depression as an outcome variable. Assuming that turning 60 increases the probability to be depressed and that retirement decreases the likelihood to be depressed, this bias would result in the lower bound of the true effect. Cross-country variation in retirement ages allows for an abstraction from this potential effect, as not all NRA and ERA can be considered milestone ages. However, several threats to the exogeneity assumption remain in the literature.

Some studies argue that the NRA and ERA are known ahead of time, thereby causing an individual to alter behavior prior to retirement, which could affect health. Knowing there is only a limited time left in their job, individuals may be less stressed by their job as they know any problem will soon not be theirs to solve. Being less stressed could lead to better health outcomes. It is also possible, that a person will take up new hobbies to ensure a smooth transition into retirement, which would also improve health prior to retirement. A positive effect of retirement would therefore be a lower bound of the true effect. This potential source of bias is more problematic if retirement worsens health. If that is the case, it is necessary to check if there is actually an adverse effect of retirement or if the adverse effect captures mean reversion - that reaching retirement brings individuals back to the health level they experienced before the planning phase. Instrumenting for the anticipation phase will check whether such a bias is present.

Another threat to exogeneity is presented by de Grip et al. (2012). They show that large, discontinuous changes in retirement ages can have a separate effect on health. While most changes to eligibility ages have been phased in slowly with many years of advance notice, some of the more abrupt changes, such as the increased early retirement age in Italy or the complete discontinuation of early retirement in the Netherlands, may negatively affect health. As a robustness check, the affected cohorts of these countries in the years of the jump and

The last requirement an instrument has to fulfill is monotonicity. Monotonicity is fulfilled if all people who are affected by the instrument are affected in the same direction. Either reaching the age threshold has no effect on an individual or it has to be positive for all individuals (or negative for all individuals). It cannot be that some persons are more likely to retire while others are less likely to retire. It is assumed that this holds and the first stage regression can be used to check that there is no indication that this may not be the case.

While using an instrument allows for causal inference to be drawn, it only allows for conclusions about the effect of retirement on those individuals who retire due to reaching the eligibility age, not on those who retire for other reasons. In other words, this strategy allows for the identification of the local average treatment effect (LATE). However, this effect is most important when considering changes to existing pension policies, as those retiring due to reaching the official age thresholds, the compliers, will be most directly affected by policy changes. Those who retire due to other reasons will likely do so even if the retirement age is increased.

5.3 Heterogeneity

The approach discussed above does not take into consideration that retirement may have diverse effects on health for different individuals. It assumes that $Y_i^1 - Y_i^0$ is the same $\forall i$[13]. It is possible that the effect of retiring at the NRA and the ERA differ, as individuals who accept a cut in pension benefits in return for retiring early may differ from those who choose to retire without a cut in benefits. As can be seen in Table 3, 71.2% of those retiring upon reaching the ERA are white collar workers, compared to 65.5% of those retiring at the NRA. This in turn means a larger fraction of blue collar workers postpone retirement until they reach full pension benefits[14]. Table 3 also shows that those retiring at the ERA are less likely to report working in a physically demanding job than those retiring later, however more individuals who felt time pressure to do their work retire when reaching the ERA. In summary, there are slight differences among those who retire when reaching the ERA and the NRA, which could lead to different retirement effects around each threshold.

As indicated by subsection 2.2, the impact of retirement on health is expected to differ by personal as well as job characteristics. The analysis will be split by gender, as well as other personal characteristics such as having grandchildren. The effect of retirement will further be broken down by job characteristics. First, the analysis will be completed for white and blue collar workers. As can be seen in Table 3, 69.5% of the sample are considered white collar workers, while only 30.3% are classified as blue collar workers. Furthermore, Table 4 shows that blue collar workers, on average, experience greater health issues among retired and non-retired individuals. If blue collar workers benefit more than white collar workers, an average

[13] Y_i^1 is the health of person i if he/she is retired, while Y_i^0 is the health of person i if he/she is working.

[14] Individuals are categorized as blue or white collar worker using the ISCO-88 categorization given in the first interview. The categorization therefore uses the current job for those still employed or the last job for those who are already retired during their first interview. Armed forces are excluded.

Table 3: Personal and Occupational Descriptive Statistics by Retirement Age

	Entire Population	Retirement Age: ERA	NRA	>ERA <NRA	Other
Female	45.5%	47.9%	47.0%	44.0%	40.1%
Married	74.8%	73.7%	72.6%	75.4%	74.3%
White Collar	69.5%	71.2%	65.5%	65.4%	64.2%
Blue Collar	30.3%	28.6%	34.3%	34.4%	35.2%
Physically Demanding	42.9%	40.1%	44.0%	43.4%	42.2%
Time Pressure in Job	49.7%	51.6%	48.6%	46.0%	44.1%
N	106 904	7 579	13 969	18 858	30 584

Note: Married includes individuals living with a partner or spouse. Due to missing data, some individuals could not be classified as white or blue collar workers and therefore the fraction of white collar workers plus the fraction of blue collar workers is not 100%.

treatment effect may be insignificant due to the larger sample size of white collar workers. Splitting the sample to identify the separate effects is therefore sensible. The effect is further differentiated using questions about the characteristics of the job to see if those who agree that their job is physically straining, or who agree they feel time pressure for a large workload or feel like they have no freedom in their work benefit more from retirement. These questions asks the respondent to strongly agree, agree, disagree, and strongly disagree with the respective characteristic. The answers are dichotomized into strongly agree / agree and disagree / strongly disagree. Individuals were classified using the answers they gave in their first interview.

Table 4: Average Health Measures - Retired vs Working Individuals

	All Workers		White Collar		Blue Collar	
	Working	Retired	Working	Retired	Working	Retired
Age	56.8	70.0	57.0	69.7	56.8	70.7
Very good to excellent SAH	47.2%	27.8%	50.7%	32.8%	36.0%	19.0%
No Depression	83.0%	79.2%	83.3%	80.8%	83.1%	76.0%
No (I)ADL Limitations	94.6%	83.1%	95.0%	85.4%	93.3%	79.0%
No Mobility Limitations	75.1%	53.0%	76.1%	55.9%	72.1%	47.4%
Maximum Grip Strength	39.9	34.7	38.8	34.4	43.5	35.3
Total Words Recalled	10.8	8.9	11.3	9.7	9.4	7.4

Individuals who are unemployed are excluded in the main analysis, allowing the identification of the effect of transitioning from working to not working due to retirement. However, as many other studies include these individuals such as Mazzonna & Peracchi (2017), it is important to consider the effect of this sample selection. Of those who report being unemployed, 49.0% were working in a blue collar job when they were employed. Also, 55.3% reported having worked in a physically demanding job. Including individuals who transition into retirement via unemployment is likely to impact the results, as it is a group with a higher ratio of blue collar workers with physically demanding jobs. This group is more vulnerable and likely to have difficulties reaching the official retirement ages if they are increased. This has to be kept in mind during policy reforms as it will be costly to social security systems when a growing number of individuals exit the work force through unemployment or in the worst case, end up on disability insurance before entering retirement.

6 Results

6.1 Retirement Effect on Health Outcomes

A negative association between retirement and health is suggested by Table 4. It shows that on average, workers are in better health. Retirees are less likely to perceive their health as very good to excellent, and more likely to be depressed and to suffer from limitations. They are also weaker and recall fewer words. This negative correlation is confirmed by the POLS regression results shown in column (1) of Table 5. Retirement leads to a significantly worse health status in all six health outcomes. However, retirement appears to improve health once individual level fixed effects are controlled for (see column (2) of Table 5). Even though the self-selection into retirement is not yet completely accounted for, there is either a sign reversal or a reduction in magnitude of the negative effect of retirement on all health measures.

Table 5: Binary Effect of Retirement on Health of Entire Population

	POLS	FE	Instrumental Variable Approach		
	(1)	(2)	(3)	(4)	(5)
Very good to excellent SAH					
retired	-0.051***	0.024***	0.078***	0.100**	0.082***
	(0.006)	(0.007)	(0.021)	(0.033)	(0.018)
KP F-stat			[1265]	[591]	[1022]
N	104 733	104 733	89 674	89 674	89 674
No clinical depression					
retired	-0.008*	0.020***	0.022	0.054**	0.029*
	(0.005)	(0.006)	(0.018)	(0.027)	(0.016)
KP F-stat			[1249]	[597]	[1018]
N	103 016	103 016	87 820	87 820	87 820
No Limitations in (I)ADL					
retired	-0.020***	0.014***	0.050***	0.030	0.045***
	(0.003)	(0.005)	(0.014)	(0.019)	(0.012)
KP F-stat			[1265]	[591]	[1022]
N	104 742	104 742	89 689	89 689	89 689
No Mobility Limitations					
retired	-0.052***	0.010	0.054***	0.075**	0.059***
	(0.006)	(0.007)	(0.021)	(0.031)	(0.018)
KP F-stat			[1264]	[592]	[1021]
N	104 744	104 744	89 692	89 692	89 692
Maximum Grip Strength					
retired	-0.369***	-0.186**	0.813***	0.054	0.558**
	(0.096)	(0.093)	(0.245)	(0.413)	(0.224)
KP F-stat			[1225]	[580]	[999]
N	99 308	99 308	83 907	83 907	83 907
Total words recalled (cognitive ability)					
retired	-0.152***	0.071	0.286**	0.110	0.236**
	(0.038)	(0.047)	(0.131)	(0.210)	(0.115)
KP F-stat			[1258]	[591]	[1020]
N	103 122	103 122	87 888	87 888	87 888
NRA	No	No	Yes	No	Yes
ERA	No	No	No	Yes	Yes
FE	No	Yes	Yes	Yes	Yes

Note: Positive coefficients imply a health improvement and the first four health measures are binary. Robust standard errors, clustered at the individual level, are given in parentheses. Kleibergen-Paap (KP) rk Wald F-statistics are reported in brackets. All regressions control for age, age squared, female, number of children, years of education, interview wave and country of residence. POLS stands for a pooled ordinary least square regression. FE is an individual-level fixed effects regression. Column (3) uses NRA, (4) ERA and (5) both to instrument the retirement decision.
Significance levels: ***p<0.01, **p<0.05, *p<0.1.

To remove the bias due to reverse causality, columns (3), (4) and (5) of Table 5 show the results when using NRA (column (3)), ERA (column (4)) and both (column (5)) to instrument for the retirement decision. The sample size decreases, as all individuals with only one observation are dropped from the analysis. The exact sample size differs between health outcome variables due to some missing data. Instrument validity is checked using the Kleibergen-Paap rk Wald F statistic as well as the test for over- and underidentification. The Kleibergen-Paap rk Wald F statistic is the robust analog of the Cragg-Donald statistic, which can be used to test instrument validity when using one or more endogenous regressors. The critical values developed by Stock & Yogo (n.d.) are only applicable when homogeneity is assumed. As this is unlikely to hold, the suggestion by Baum et al. (2007) is followed and the well-known rule-of-thumb, that instruments are weak if the F-statistic is smaller than 10, is applied. The Kleibergen-Paap rk Wald F-statistic is always larger than 10, indicating strong instruments. The Hansen J statistic can be used to test for overidentification. The joint null hypothesis that the instruments are valid cannot be rejected. The Kleibergen-Paap rk LM statistic is used to test for underidentification. The null hypothesis that the model is under-identified is rejected. The first stage regression results demonstrate the relevance of the instruments. Reaching the normal retirement age increases the probability to retire by around 35% for women and 26% for men, while reaching the early retirement age threshold increases the probability by 19% for women and 21% for men (see Table A1 in Appendix A).

Columns (3) and (4) in Table 5 suggest that the effects of retiring at the NRA and ERA are in the same direction and of comparable magnitude . Small differences remain, which should be considered if only one of the age thresholds is used for the analysis. For the remainder of the analyses, only the results using both instruments jointly are presented (results for NRA and ERA separately are available upon request). Retiring leads to an 8.2% increase in reporting very good to excellent health, decreases depression by 2.9% and the probability to experience limitations in (I)ADL or mobility by 4.5% and 5.9% respectively. Retirees are also significantly stronger, increasing their grip strength by 0.558 kg on average, while they recall 0.236 more words. While several studies have identified significant health preserving effects in a variety of health measures, this is the first to identify a positive effect on the objective health measure maximum grip strength. Result of this analysis also suggest a positive effect on cognitive abilities, contrary to most of the current literature.

Having established a health improving binary effect of retirement on health, Table 6 separates the retirement effect into different phases. Column (2) additionally includes the anticipation phase. Results do not generally support the idea that individuals experience health benefits shortly before retiring. The only health measure which provides evidence in favor of this theory is the depression measure. In the two years prior to retirement, individuals are 8.8% less likely to be categorized depressed. Retirement then leads to a 6.6% decrease in the likelihood to be considered depressed. Controlling for the anticipation phase also leads to a greater increase in grip strength, although the effect of the anticipation phase is not statistically significant. The anticipation coefficient is negative for self-assessed health, limitations in (I)ADL and cognitive ability, suggesting health actually worsens prior to retirement.

Table 6: The Effect of Different Retirement Phases on Health

	Binary (1)	Anticipation (2)	Honeymoon (3)	Anticipation & Honeymoon (4)
Very good to excellent SAH				
retired	0.083***	0.076***	0.075***	0.066***
	(0.018)	(0.024)	(0.018)	(0.025)
anticipation (2yrs)		-0.006		-0.020
		(0.048)		(0.050)
honeymoon (2 yrs)			0.021	0.021
			(0.019)	(0.020)
KP F-stat	[1022]	[109]	[268]	[62]
N	89 674	89 674	89 674	89 674
No clinical depression				
retired	0.029*	0.066***	0.038**	0.078***
	(0.016)	(0.020)	(0.015)	(0.021)
anticipation (2yrs)		0.088**		0.104**
		(0.041)		(0.043)
honeymoon (2 yrs)			-0.018	-0.031*
			(0.017)	(0.018)
KP F-stat	[1018]	[107]	[264]	[60]
N	87 820	87 820	87 820	87 820
No Limitations in (I)ADL				
retired	0.045***	0.036**	0.045***	0.040**
	(0.012)	(0.015)	(0.012)	(0.016)
anticipation (2yrs)		-0.022		-0.017
		(0.031)		(0.032)
honeymoon (2 yrs)			-0.023*	-0.020
			(0.014)	(0.014)
KP F-stat	[1022]	[109]	[268]	[62]
N	89 689	89 689	89 689	89 689
No Mobility Limitations				
retired	0.059***	0.067**	0.062***	0.078***
	(0.018)	(0.024)	(0.018)	(0.025)
anticipation (2yrs)		0.021		0.033
		(0.048)		(0.049)
honeymoon (2 yrs)			-0.040*	-0.043**
			(0.020)	(0.021)
KP F-stat	[1021]	[109]	[268]	[62]
N	89 692	89 692	89 692	89 692
Maximum Grip Strength				
retired	0.558**	0.767***	0.662***	0.935***
	(0.224)	(0.297)	(0.222)	(0.307)
anticipation (2yrs)		0.586		0.770
		(0.579)		(0.591)
honeymoon (2 yrs)			-0.356	-0.410*
			(0.234)	(0.240)
KP F-stat	[999]	[105]	[259]	[59]
N	83 907	83 907	83 907	83 907
Total words recalled (cognitive ability)				
retired	0.236**	0.206	0.223*	0.153
	(0.115)	(0.157)	(0.116)	(0.165)
anticipation (2yrs)		-0.104		-0.158
		(0.310)		(0.318)
honeymoon (2 yrs)			0.207	0.223*
			(0.128)	(0.132)
KP F-stat	[1020]	[109]	[265]	[61]
N	87 888	87 888	87 888	87 888

Note: ERA and NRA used jointly to instrument for the retirement decision. Positive coefficients imply a health improvement and the first four health measures are binary. Robust standard errors, clustered at the individual level, are given in parentheses. Kleibergen-Paap (KP) rk Wald F-statistics are reported in brackets. All regressions control for age, age squared, female, number of children, years of education, interview wave and country of residence.
Significance levels: ***p<0.01, **p<0.05, *p<0.1.

The model in column (3) splits the retirement effect into the honeymoon effect and the long-term retirement effect. The effect of retirement remains largely unchanged, both in magnitude and significance levels. There is no indication that retirees enter a blissful stage directly after retirement. Results instead point towards the opposite effect occurring. Within the first two years of retirement, the likelihood to suffer from limitations in (I)ADL increases by 2.3% and to suffer from mobility limitations increases by 4.0%.

Column (4) in Table 6 shows the results of the preferred specification, including both the honeymoon and anticipation phase. Compared to the standard model used in most literature (column (1)), retirement leads to a smaller improvement in terms of self-assessed health, limitations in (I)ADL and total words recalled. On the other hand, the improvements in terms of maximum grip strength, mobility limitations and depression become stronger in magnitude and size. This supports the idea that the effect of retirement is not linear and not homogeneous across different health measures.

6.2 Robustness Checks

Results are robust to various other specifications (see Table 7). The model in column (2) uses an alternative retirement definition, where individuals are only considered to be retired, if they claim to be retired and who do not report having done paid work within the last four weeks[15]. The magnitude of the retirement effect increases in all six health measures. Furthermore, a significant health improving effect in the anticipation phase is now measured both in terms of depression and maximum grip strength. Health also improves more during the honeymoon phase. This alternate retirement definition leads to the identification of a purer effect of the transition from working to not working due to retirement, as those individuals who continue working will not experience a drastic change in their life. Those who continue working are not expected to experience much of a health change, as their life will not adjust as much, if at all.

Results are also robust to redefining the instruments. In column (3), the instrument does not differ by gender, while the model in column (4) includes country and gender specific instruments (i.e. one indicator for females in Austria reaching NRA and one for reaching ERA, one for males in Austria reaching NRA, etc.). Both models yield similar results. There is even stronger evidence against health improvements during the honeymoon phase when using country-gender specific instruments. During the honeymoon phase, retirees are 2.8% more likely to be classified as depressed, 2.8% more likely to suffer from limitations in (I)ADL, 5.2% more likely to suffer from mobility limitations and are 0.484 kg weaker in their grip strength.

Results are also robust to using different age specifications, as can be seen in columns (5) through (8). Allowing for a country specific age trend or a linear, cubic or quartic age trend leads to similar results, both in terms of magnitude and significance. Shortening the anticipation and honeymoon to one year also gives similar results (see column (9) in Table 7). Shortening the

[15] Around 11.8% of those individuals who claim to be retired also reported having done paid work in the last four weeks. These individuals are considered to be working in the robustness check. Dropping this individuals completely from the sample also leads to similar results.

Table 7: Robustness Checks for Main Specification

			Other Instruments		Different Age Specifications				
	Main	Alt. Ret. Definition	One IV	Country* Gender	Country* Age	Linear	Cubic	Quartic	1 Year Phases
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Very good to excellent SAH									
retired	0.066**	0.115***	0.058**	0.049**	0.081***	0.065***	0.083***	0.061	0.064***
	(0.025)	(0.040)	(0.026)	(0.022)	(0.026)	(0.021)	(0.029)	(0.041)	(0.022)
anticipation	-0.020	-0.017	-0.034	-0.033	-0.012	-0.022	-0.021	-0.043	-0.063
	(0.050)	(0.048)	(0.051)	(0.040)	(0.050)	(0.045)	(0.050)	(0.056)	(0.084)
honeymoon	0.021	0.044**	0.021	0.014	0.013	0.021	0.009	0.008	0.018
	(0.020)	(0.019)	(0.020)	(0.017)	(0.020)	(0.020)	(0.024)	(0.024)	(0.029)
KP F-stat	[62]	[53]	[121]	[9]	[59]	[68]	[61]	[58]	[27]
N	89 674	89 640	89 674	89 674	89 674	89 674	89 674	89 674	89 674
No clinical depression									
retired	0.078***	0.126***	0.073***	0.066***	0.069***	0.143***	0.067**	0.077**	0.065***
	(0.021)	(0.035)	(0.021)	(0.018)	(0.022)	(0.018)	(0.026)	(0.035)	(0.018)
anticipation	0.104**	0.099**	0.092**	0.069**	0.096**	0.202***	0.104**	0.113**	0.171**
	(0.043)	(0.042)	(0.034)	(0.034)	(0.043)	(0.039)	(0.043)	(0.048)	(0.068)
honeymoon	-0.031*	-0.005	-0.028*	-0.028*	-0.030	-0.023	-0.024	-0.024	-0.027
	(0.018)	(0.017)	(0.018)	(0.015)	(0.018)	(0.018)	(0.021)	(0.021)	(0.027)
KP F-stat	[60]	[51]	[117]	[9]	[57]	[66]	[60]	[56]	[26]
N	87 820	87 793	87 820	87 820	87 820	87 820	87 820	87 820	87 820
No Limitations in (I)ADL									
retired	0.040**	0.058**	0.042***	0.042***	0.036**	0.143***	0.001	-0.006	0.050***
	(0.016)	(0.026)	(0.016)	(0.014)	(0.017)	(0.015)	(0.021)	(0.027)	(0.014)
anticipation	-0.017	-0.026	-0.015	-0.000	-0.019	0.135***	-0.005	-0.022	-0.004
	(0.032)	(0.031)	(0.032)	(0.025)	(0.032)	(0.030)	(0.032)	(0.036)	(0.054)
honeymoon	-0.020	-0.006	-0.021	-0.028**	-0.018	-0.008	0.009	0.009	-0.034*
	(0.014)	(0.013)	(0.014)	(0.012)	(0.014)	(0.014)	(0.017)	(0.017)	(0.020)
KP F-stat	[62]	[53]	[121]	[9]	[59]	[68]	[61]	[58]	[27]
N	89 689	89 656	89 689	89 689	89 689	89 689	89 689	89 689	89 689
No Mobility Limitations									
retired	0.078***	0.115***	0.080***	0.061***	0.081***	0.126***	0.064**	0.075*	0.069***
	(0.025)	(0.040)	(0.025)	(0.022)	(0.026)	(0.021)	(0.030)	(0.040)	(0.022)
anticipation	0.033	0.017	0.033	0.039	0.043	0.105**	0.035	0.047	0.047
	(0.049)	(0.048)	(0.050)	(0.039)	(0.050)	(0.045)	(0.049)	(0.055)	(0.082)
honeymoon	-0.043**	-0.017	-0.044**	-0.052***	-0.040*	-0.038*	-0.033	-0.033	-0.035
	(0.021)	(0.020)	(0.021)	(0.018)	(0.021)	(0.021)	(0.025)	(0.025)	(0.031)
KP F-stat	[62]	[53]	[120]	[9]	[59]	[68]	[61]	[58]	[26]
N	89 692	89 658	89 692	89 692	89 692	89 692	89 692	89 692	89 689
Maximum Grip Strength									
retired	0.935***	2.116***	0.386	0.647**	0.840***	2.417***	0.896**	1.531***	0.715**
	(0.307)	(0.485)	(0.320)	(0.268)	(0.316)	(0.263)	(0.369)	(0.498)	(0.277)
anticipation	0.770	1.224**	-0.102	0.732	0.744	3.011***	0.739	1.381**	0.601
	(0.591)	(0.574)	(0.618)	(0.476)	(0.597)	(0.547)	(0.591)	(0.660)	(1.051)
honeymoon	-0.410*	-0.073	-0.253	-0.484**	-0.343	-0.209	-0.424	-0.432	-0.513
	(0.240)	(0.233)	(0.250)	(0.208)	(0.242)	(0.241)	(0.286)	(0.286)	(0.349)
KP F-stat	[59]	[51]	[115]	[9]	[56]	[65]	[59]	[55]	[26]
N	83 907	83 878	83 907	83 907	83 907	83 907	83 907	83 907	83 907
Total words recalled (cognitive ability)									
retired	0.153	0.317	0.126	-0.013	0.369**	1.294***	-0.031	0.042	0.200
	(0.165)	(0.264)	(0.166)	(0.144)	(0.169)	(0.142)	(0.199)	(0.264)	(0.143)
anticipation	-0.158	-0.100	-0.226	-0.142	-0.032	1.536***	-0.156	-0.086	-0.171
	(0.318)	(0.309)	(0.319)	(0.259)	(0.321)	(0.291)	(0.318)	(0.355)	(0.533)
honeymoon	0.223*	0.277**	0.229*	0.060	0.120	0.371***	0.351**	0.348**	-0.043
	(0.132)	(0.125)	(0.133)	(0.111)	(0.133)	(0.133)	(0.160)	(0.160)	(0.191)
KP F-stat	[61]	[52]	[119]	[9]	[58]	[67]	[61]	[57]	[26]
N	83 907	83 878	83 907	83 907	83 907	83 907	83 907	83 907	83 907

Note: ERA and NRA used jointly to instrument for the retirement decision. Positive coefficients imply a health improvement and the first four health measures are binary. Robust standard errors, clustered at the individual level, are given in parentheses. Kleibergen-Paap (KP) rk Wald F-statistics are reported in brackets. All regressions control for age, age squared, female, number of children, years of education, interview wave and country of residence. Significance levels: ***p<0.01, **p<0.05, *p<0.1.

phases to one year lowers the F-statistic, as fewer observations are within the honeymoon and anticipation phases.

In another robustness check, countries in the years of drastic retirement age adjustments are excluded (the Netherlands in the years 2011-2015, Italy during 2011-2015, and Germany during 2012-2015). Results are robust and become stronger. Results are also robust when keeping all individuals who report being retired or employed/ self-employed instead of dropping those who ever report being unemployed, disabled, homemakers or do not declare their employment status. Including individuals who report being unemployed increases the health preserving effect retirement has on health. Conducting a placebo test using 70 as the normal retirement age does not yield any significant effect of retirement on health (see columns (2) through (5) of Table A2 in Appendix B).

To check if attrition due to the death of individuals is biasing results, the SHARE end of life questionnaire is used to identify individuals who have passed away. In the sample, 1,452 individuals pass away for whom the end of life survey was recorded (around 4.6% of the individuals included in the main analysis). In the main analysis, individuals who have died are not considered in the estimation once they have died. As a robustness check, these individuals are included by assigning them a value of 0 for the four binary health measures (self-assessed health, not depressed, no limitations in (I)ADL, no mobility limitations) and the bottom tenth percentile value of maximum grip strength and number of words recalled. The control variables take on the value of the last wave they participated in the survey alive. The employment status reported in the last interview is used in a first analysis. In a second analysis, those who were working before passing away and have then surpassed the ERA eligibility age are considered retired. In a third analysis, retirement status is assigned to those surpassing the NRA eligibility age. In all three analyses, the health preserving effect of retirement becomes even stronger across all health measures. Attrition due to death is biasing results downwards, as those with the worst health drop out of the sample. Overall, the results of my main analysis appear to be very robust to different specifications (see columns (6) through (8) of Table A2 in Appendix B).

6.3 Heterogeneity Analysis

After having established the robustness of the main specification, the effect of retirement is analyzed within different populations (see Table 8). The effect of retirement is generally similar between males and females. However, there are some notable differences, especially in terms of depression and grip strength. While the probability for women to be categorized as depressed decreases by 6.6%, it decreases 8.3% for men. The difference in the anticipation and honeymoon phases is more noteworthy. Men are significantly less likely to be categorized as depressed during the anticipation phase, yet fall into a slump immediately after retirement, when the probability to suffer from depression increases 5.9%. Maximum grip strength, however, appears to be significantly impacted by retirement only among women. While women gain 1.365 kg in strength during the anticipation phase and 4.426 kg in long-term retirement, they lose some strength in the honeymoon phase. For men there is no significant effect, but the signs of the coefficients go into the opposite direction.

Table 8: The Heterogeneity of the Effect of Retirement on Health

	Gender		Occupation		Physical Job		Time Pressure in Job		Freedom in Job	
	Female	Male	White Collar	Blue Collar	Yes	No	Yes	No	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Very good to excellent SAH										
retired	0.081**	0.038	0.075***	0.065	0.071	0.081**	0.106**	0.059	0.212***	0.054*
	(0.032)	(0.041)	(0.029)	(0.054)	(0.044)	(0.040)	(0.046)	(0.040)	(0.074)	(0.032)
antici.	-0.014	-0.042	-0.031	0.051	-0.070	0.004	-0.029	-0.008	0.214*	-0.065
	(0.063)	(0.080)	(0.058)	(0.099)	(0.074)	(0.068)	(0.074)	(0.070)	(0.122)	(0.056)
hnym.	0.031	0.011	0.028	0.005	-0.002	-0.003	-0.061	0.054	-0.135*	0.038
	(0.026)	(0.031)	(0.024)	(0.035)	(0.046)	(0.044)	(0.048)	(0.043)	(0.071)	(0.035)
KP F-stat	[66]	[56]	[45]	[15]	[19]	[26]	[19]	[25]	[10]	[35]
N	40 765	48 909	61 129	25 894	19 154	25 974	22 349	22 772	10 436	34 685
No clinical depression										
retired	0.066**	0.083***	0.079***	0.076	0.034	0.111***	0.090**	0.060**	0.034	0.086***
	(0.031)	(0.028)	(0.024)	(0.048)	(0.038)	(0.032)	(0.040)	(0.030)	(0.061)	(0.026)
antici.	0.079	0.111**	0.082*	0.143	0.040	0.142***	0.080	0.104**	0.042	0.111**
	(0.062)	(0.056)	(0.048)	(0.092)	(0.064)	(0.054)	(0.065)	(0.052)	(0.100)	(0.045)
hnym.	-0.009	-0.059**	-0.003	-0.118***	-0.009	-0.027	-0.024	-0.010	-0.016	-0.012
	(0.026)	(0.024)	(0.021)	(0.037)	(0.041)	(0.034)	(0.042)	(0.033)	(0.063)	(0.028)
KP F-stat	[63]	[55]	[44]	[15]	[19]	[26]	[18]	[25]	[10]	[35]
N	40 149	47 671	60 153	25 094	18 835	25 690	22 036	22 482	10 237	34 275
No Limitations in (I)ADL										
retired	0.031	0.052**	0.036**	0.032	0.025	0.023	0.031	0.019	0.006	0.030
	(0.022)	(0.023)	(0.017)	(0.040)	(0.030)	(0.023)	(0.031)	(0.024)	(0.047)	(0.020)
antici.	-0.018	-0.015	-0.017	-0.049	-0.010	0.014	0.021	0.011	-0.012	0.012
	(0.044)	(0.045)	(0.035)	(0.073)	(0.047)	(0.038)	(0.047)	(0.039)	(0.076)	(0.032)
hnym.	-0.011	-0.031	-0.009	-0.057*	0.027	-0.010	0.013	0.002	0.027	0.002
	(0.020)	(0.020)	(0.016)	(0.030)	(0.030)	(0.026)	(0.031)	(0.025)	(0.045)	(0.022)
KP F-stat	[66]	[56]	[45]	[15]	[19]	[26]	[19]	[25]	[10]	[35]
N	40 769	48 920	61 142	25 896	19 148	25 983	22 351	22 773	10 428	34 686
No Mobility Limitations										
retired	0.082**	0.072**	0.063**	0.114**	0.096**	0.028	0.089**	0.030	0.123*	0.034
	(0.033)	(0.037)	(0.028)	(0.056)	(0.042)	(0.038)	(0.045)	(0.038)	(0.070)	(0.031)
antici.	0.063	-0.000	-0.005	0.081	0.136*	-0.013	0.088	0.016	0.054	0.043
	(0.066)	(0.074)	(0.056)	(0.106)	(0.073)	(0.064)	(0.073)	(0.066)	(0.119)	(0.053)
hnym.	-0.032	-0.058*	-0.046*	-0.012	-0.028	-0.040	-0.035	-0.052	-0.166**	0.002
	(0.029)	(0.030)	(0.025)	(0.041)	(0.046)	(0.043)	(0.049)	(0.042)	(0.073)	(0.035)
KP F-stat	[66]	[56]	[45]	[15]	[19]	[26]	[19]	[25]	[10]	[35]
N	40 772	48 920	61 143	25 898	19 151	25 982	22 352	22 774	10 429	34 687
Maximum Grip Strength										
retired	1.426***	-0.267	0.446	2.262***	0.919*	0.575	0.546	0.921**	1.790**	0.531
	(0.330)	(0.558)	(0.331)	(0.758)	(0.515)	(0.456)	(0.531)	(0.457)	(0.851)	(0.370)
antici.	1.365**	-1.171	-0.018	2.727**	1.258	0.305	0.841	0.748	2.283	0.442
	(0.651)	(1.049)	(0.675)	(1.319)	(0.873)	(0.768)	(0.861)	(0.784)	(1.473)	(0.617)
hnym.	-0.719***	0.119	-0.231	-1.073*	-0.215	-0.755	-0.014	-0.838*	-0.687	-0.500
	(0.270)	(0.417)	(0.277)	(0.509)	(0.539)	(0.485)	(0.561)	(0.475)	(0.800)	(0.400)
KP F-stat	[61]	[56]	[43]	[14]	[18]	[26]	[18]	[24]	[9]	[34]
N	37 912	45 995	57 831	23 619	18 135	25 054	21 306	21 876	9 823	33 359
Total words recalled (cognitive ability)										
retired	0.179	0.097	0.046	0.331	0.470	0.383	0.305	0.470*	-0.013	0.526**
	(0.216)	(0.254)	(0.185)	(0.366)	(0.298)	(0.247)	(0.309)	(0.252)	(0.477)	(0.208)
antici.	-0.067	-0.332	-0.388	0.210	0.203	0.064	-0.620	0.759*	0.284	0.273
	(0.425)	(0.477)	(0.368)	(0.649)	(0.487)	(0.413)	(0.489)	(0.423)	(0.779)	(0.343)
hnym.	0.143	0.322	0.212	0.319	0.124	-0.132	0.098	-0.074	0.017	0.004
	(0.179)	(0.196)	(0.160)	(0.247)	(0.308)	(0.270)	(0.326)	(0.264)	(0.470)	(0.224)
KP F-stat	[65]	[55]	[44]	[15]	[19]	[26]	[19]	[25]	[10]	[35]
N	40 191	47 697	60 158	25 139	18 813	25 694	22 020	22 480	10 227	34 267

Note: ERA and NRA used jointly to instrument for the retirement decision. Positive coefficients imply a health improvement and the first four health measures are binary. Robust standard errors, clustered at the individual level, are given in parentheses. Kleibergen-Paap (KP) rk Wald F-statistics are reported in brackets. All regressions control for age, age squared, female, number of children, years of education, interview wave and country of residence. Significance levels: ***p<0.01, **p<0.05, *p<0.1.

Splitting up the sample into blue and white collar workers reduces the sample size, as not all individuals answered the question regarding job type. Due to a smaller sample size, the F-statistic for blue collar workers is lower. Other than expected, results suggest that white and blue collar experience similar health effects upon retirement. Blue collar workers tend to have more health problems during the honeymoon phase, but experience similar health improvements in the long term. However, standard errors are larger among blue collar individuals, leading to fewer statistically significant results. The magnitude of the coefficients is similar to that of white collar workers. The biggest difference between the two groups is measured in terms of maximum grip strength. Blue collar workers gain in strength immediately prior to retirement, but lose some strength during their honeymoon phase, which they regain afterwards. As a larger share of blue collar workers enter retirement from unemployment, including those individuals who were unemployed before retirement strengthens the effect for blue collar workers. The results now show a significant health improvement for blue collar workers in the long term, although they also experience significantly more health issues within the honeymoon phase. These results suggests that blue collar workers go through a rough adjustment phase after retirement, before adjusting back to better health.

Estimating the effect for further groups in terms of job characteristics leads to less precise results due to smaller sample sizes (see Table 8). Only a subset of individuals answered the relevant questions, so the following results should be interpreted as a first indication of the heterogeneity of the effect retirement has on health. Results do not generally support the theory that individuals who considered their job physically demanding benefit more from retirement than those who did not believe their job to be physically demanding (see columns (5) and (6)). They do however experience a 9.6% decrease in mobility limitations, with a large honeymoon effect of 13.6% reduced likelihood to experience mobility limitations. This is evidence that once the physical strain of their job ends, these individuals experience fewer physical ailments. They also gain 0.919 kg in maximum grip strength. Those who did not consider their jobs physically demanding experience greater improvements both in self-assessed health and being categorized as depressed. These results do not concur with the results by Mazzonna & Peracchi (2017), who found that those with physically straining jobs experience greater health improvements upon entering retirement. While Mazzonna & Peracchi (2017) do not rely on the self-reported job characteristics and were able to use a more precise measurement, they state that using the self-reported measures yielded the same results. The results also do not support the idea that for those whose job was mentally straining, either due to feeling time pressured or feeling like there was no freedom to decide how to do their work, retirement brings a greater relief and therefore greater health improvements (see columns (7) through (10)). Although there are slight differences depending on which health aspect is considered, no general trend can be identified.

Looking at certain sub-groups based on personal characteristics, results show that individuals with grandchildren experience greater health improvements in terms of self-assessed health, depression and grip strength. At the same time, individuals with grandchildren are more likely to suffer from limitations in (I)ADL. Retirees whose children live less than 25 km away also experience greater health improvements in terms of self-assessed health, limitations in (I)ADL and grip strength (see Table A3 in Appendix C).

7 Discussion and Conclusion

Using SHARE data from waves 1, 2, 4, 5, and 6, I use a fixed effects instrumental variable approach to determine the effect retirement has on health. The exogenous variation in the probability to retire at the normal and early retirement age thresholds, NRA and ERA respectively, is exploited to instrument for the otherwise endogenous retirement decision. The baseline OLS model suggest a negative association between retirement and health. A large part of this negative association is driven by unobserved individual heterogeneity, as demonstrated by the fixed effects regression. Even without using an instrumental variable approach to account for reverse causality, retirement preserves health once individual fixed effects are included in the estimation. Instrumenting for the retirement decision leads to an even stronger health preserving effect. The results of this paper are in line with those studies finding overwhelmingly positive effects of retirement on health. Unlike previous literature, a significant positive effect is identified in the objective health outcome maximum grip strength. Furthermore, contrary to most literature, cognitive ability, measured by the word recall test, also improves with retirement.

Retiring either at the NRA or ERA significantly improves health. However, there are slight difference in the magnitude and significance level depending on which health outcome it considered. While results for self-assessed health and mobility limitations are similar for both retirement age thresholds, the likelihood to be categorized as depressed is only significant at the ERA, while improvements in (I)ADL limitations, maximum grip strength and total words recalled are only measured for retirement at the NRA. Therefore, using only one of the eligibility ages may lead to results that do not reflect the different nuances of the retirement effect on health.

The specification of the main model is robust. Using an alternate retirement definition which excludes those individuals who report being retired and having done paid work in the past four weeks strengthens the results. Changing to a gender-neutral or gender-country specific instrument also leads to similar results. Adjusting the age specifications, using a country-age specific age trend, or a linear, cubic or quartic age trend, confirms the results of the main specification which includes a quadratic age trend. The results are therefore not driven by incorrectly specifying the effect age has on health. Using several different sample restrictions, such as including those who reported being unemployed or excluding those countries who experience sharp changes in their retirement age thresholds, confirms and strengthens the findings of the main analysis.

On average, the health preserving effect of retirement is experienced in the long term. I do not find that health improves during the anticipation phase. There is only weak support for the argument that using NRA and ERA as instruments will lead to biased results. When depression, mobility limitations and maximum grip strength are used as outcome variables, there is some evidence that the effect is biased downwards. However, the effect of retirement on self-assessed health, limitations in (I)ADL and total words recalled is stronger when the anticipation phase is not controlled for. In these outcome measures, the changes prior to retirement lead to an overestimation of the effect. Furthermore, there is no evidence of the honeymoon effect. The opposite occurs - retirees are more often depressed and suffer from more mobility limitations during the honeymoon phase. Controlling for the anticipation and honeymoon phases lowers the health improvement of retirement in self-assessed health, limitations in (I)ADL and number of words recalled, while it increases the effect measured in depression, mobility limitations and grip strength. The effect of retirement on health is not linear, but depends on the time until or since the start of retirement.

The results further show that on average, all individuals - both men and women - experience health improvements upon retirement. Retirement has a health preserving effect for white and blue collar workers. It is not the case that those who worked in physically and psychologically straining jobs benefit more from retirement than others. Certain personal characteristics including having grandchildren or children living close by seem to increase this health preserving effect.

One main drawback of my approach is that the identified effect is not an average treatment effect, but a local average treatment effect (LATE). The average effect of all persons who retire due to reaching the official retirement thresholds is measured. No conclusions can be drawn about the impact retirement has on individuals who retire due to different reasons at another age. Nevertheless, I believe the LATE is important for policy makers when deciding about further increases to the retirement ages. It is these compliers that are most likely to extend their working life, as those who retire due to other reasons will likely continue doing so in the future. This analysis does not give insight into whether fewer people will work until the retirement age, if it is increased even further. Further research is necessary to understand whether raising retirement ages will increase the uptake on unemployment or disability benefits right before retirement.

These results are relevant for policy makers discussing pushing back retirement ages and eliminating early retirement options. Since retirement leads to an improvement in health, pushing back this health boost could lead to greater health issues in the years prior to retirement and therefore a greater strain on the health care system. Furthermore, without this health boost, more individuals may be driven to seek alternative exit routes, such as unemployment or disability leave, that will put a further strain on the social security systems. If a policy maker wants to target inequality among the elderly, it may also be important to consider that all workers seem to have the same health boost once they retire. However, an improvement for a blue collar worker does not imply the same health status as an improvement for a white collar worker, as they start from a different level of health. Retirement does not act as an equalizer between different groups. If that is one goal of policy makers, differentiated retirement rules for populations should be considered.

References

Ambrasat, J., Schupp, J. & Wagner, G. G. (2011), 'Comparing the predictive power of subjective and objective health indicators: changes in handgrip strength and overall satisfaction with life as predictors of mortality', *SOEP paper* .

Atchley, R. C. (1980), *The social forces in later life: An introduction to social gerontology*, Wadsworth Publishing Company.

Baum, C. F., Schaffer, M. E. & Stillman, S. (2007), 'Enhanced routines for instrumental/generalized method of moments estimation and testing', *The Stata Journal* **7**(4), 465–506.

Behncke, S. (2012), 'Does retirement trigger ill health?', *Health Economics* **21**(3), 282–300.

Bonsang, E., Adam, S. & Perelman, S. (2012), 'Does retirement affect cognitive functioning?', *Journal of Health Economics* **31**(3), 490–501.

Börsch-Supan, A. (2017), *Survey of Health, Ageing and Retirement in Europe (SHARE) Waves 1, 2, 4, 5, and 6. Release version: 6.0.0.*, SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w1.600, 10.6103/SHARE.w2.600, 10.6103/SHARE.w4.600, 10.6103/SHARE.w5.600, 10.6103/SHARE.w6.600.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S. & Zuber, S. (2013), 'Data resource profile: the Survey of Health, Ageing and Retirement in Europe (SHARE)', *International Journal of Epidemiology* **42**(4), 992–1001.

Bound, J. & Waidmann, T. (2007), 'Estimating the health effects of retirement', *Michigan Retirement Research Center Research Paper No. UM WP* **168**.

Card, D., Dobkin, C. & Maestas, N. (2008), 'The impact of nearly universal insurance coverage on health care utilization: Evidence from medicare', *American Economic Review* **98**(5), 2242–2258.

Coe, N. B. & Lindeboom, M. (2008), 'Does retirement kill you? Evidence from Early Retirement Windows', *IZA Discussion Paper 3817* .

Coe, N. B., von Gaudecker, H.-M., Lindeboom, M. & Maurer, J. (2012), 'The effect of retirement on cognitive functioning', *Health Economics* **21**(8), 913–927.

Coe, N. B. & Zamarro, G. (2011), 'Retirement effects on health in Europe', *Journal of Health Economics* **30**(1), 77–86.

Coile, C., Milligan, K. S. & Wise, D. A. (2016), 'Health Capacity to Work at Older Ages: Evidence from the US', *National Bureau of Economic Research* .

Dave, D., Rashad, I. & Spasojevic, J. (2008), 'The effects of retirement on physical and mental health outcomes', *Southern Economic Journal* **75**(2), 497–524.

de Grip, A., Lindeboom, M. & Montizaan, R. (2012), 'Shattered dreams: the effects of changing the pension system late in the game', *The Economic Journal* **122**(559), 1–25.

Dwyer, D. S. & Mitchell, O. S. (1999), 'Health problems as determinants of retirement: Are self-rated measures endogenous?', *Journal of Health Economics* **18**(2), 173–193.

Gorry, A., Gorry, D. & Slavov, S. (2015), 'Does retirement improve health and life satisfaction?', *National Bureau of Economic Research* .

Grossman, M. (1972), 'On the concept of health capital and the demand for health', *Journal of Political Economy* **80**(2), 223–255.

Hallberg, D., Johansson, P. & Josephson, M. (2015), 'Is an early retirement offer good for your health? Quasi-experimental evidence from the army', *Journal of Health Economics* **44**, 274–285.

Hank, K., Jürges, H., Schupp, J. & Wagner, G. G. (n.d.), 'Isometrische Greifkraft und sozialgerontologische Forschung', *Zeitschrift für Gerontologie und Geriatrie* .

Heller-Sahlgren, G. (2017), 'Retirement blues', *Journal of Health Economics* **54**, 66–78.

Hofäcker, D. (2015), 'In line or at odds with active ageing policies? Exploring patterns of retirement preferences in Europe', *Ageing & Society* **35**(7), 1529–1556.

Idler, E. L. & Benyamini, Y. (1997), 'Self-rated health and mortality: a review of twenty-seven community studies', *Journal of Health and Social Behavior* **38**(1), 21–37.

Insler, M. (2014), 'The health consequences of retirement', *Journal of Human Resources* **49**(1), 195–233.

Kim, J. E. & Moen, P. (2002), 'Retirement transitions, gender, and psychological well-being: A life-course, ecological model', *Journal of Gerontology* **57B**(3), 212–222.

Leong, D. P., Teo, K. K., Rangarajan, S., Lopez-Jaramillo, P., Avezum, A., Orlandini, A., Seron, P., Ahmed, S. H., Rosengren, A., Kelishadi, R. et al. (2015), 'Prognostic value of grip strength: findings from the Prospective Urban Rural Epidemiology (PURE) study', *The Lancet* **386**(9990), 266–273.

Malter, F. & Börsch-Supan, A. (2015), *SHARE Wave 5: Innovations & Methodology*, MEA, Max Planck Institute for Social Loaw and Social Policy.

Mazzonna, F. & Peracchi, F. (2012), 'Ageing, cognitive abilities and retirement', *European Economic Review* **56**(4), 691–710.

Mazzonna, F. & Peracchi, F. (2017), 'Unhealthy retirement?', *The Journal of Human Resources* **52**(1), 128–151.

McGarry, K. (2004), 'Health and retirement: do changes in health affect retirement expectations?', *The Journal of Human Resources* **39**(3), 624–648.

Neuman, K. (2008), 'Quit your job and get healthier? The effect of retirement on health', *Journal of Labor Research* **29**(2), 177–201.

OECD (2015), *Ageing and Employment Policies: Denmark 2015 - Working Better with Age*, OECD Publishing.

OECD (2016), 'Labour force participation rate'.
URL: *https://data.oecd.org/emp/labour-force-participation-rate.htm*

Petrou, S. & Kupek, E. (2008), 'Social capital and its relationship with measures of health status: evidence from the Health Survey for England 2003', *Health Economics* **17**(1), 127–143.

Prince, M. J., Reischies, F., Beekman, A., Fuhrer, R., Jonker, C., Kivela, S.-L., Lawlor, B., Lobo, A., Magnusson, H., Fichter, M. et al. (1999), 'Development of the EURO-D scale–a European Union initiative to compare symptoms of depression in 14 European centres', *The British Journal of Psychiatry* **174**(4), 330–338.

Rohwedder, S. & Willis, R. J. (2010), 'Mental retirement', *The Journal of Economic Perspectives* **24**(1), 119.

Stock, J. & Yogo, M. (n.d.), Testing for weak instruments in linear IV] regression, booktitle = Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, year = 2005, publisher = Cambridge University Press, location = Cambridge, pages = 80-108,.

Thomas, P. A. (2011), 'Gender, social engagement, and limitations in late life', *Social Science & Medicine* **73**(9), 1428–1435.

von Herbay, A. (2014), 'Otto Von Bismarck is not the origin of old age at 65', *The Gerontologist* **54**(1), 5–5.

World Bank (2016*a*), 'Life expectancy'.
URL: *http://data.worldbank.org/*

World Bank (2016*b*), 'World development indicators - age dependency ratio'.
URL: *https://data.worldbank.org/indicator/SP.POP.DPND.OL/*

Zantinge, E. M., van den Berg, M., Smit, H. A. & Picavet, H. S. J. (2014), 'Retirement and a healthy lifestyle: opportunity or pitfall? A narrative review of the literature', *The European Journal of Public Health* **24**(3), 433–439.

Appendix A - First Stage Results

Table A1: First Stage Regression of Binary Retirement Effect

	(1)	(2)	(3)
Very good to excellent SAH			
overNRA_F	0.357***		0.347***
	(0.009)		(0.009)
overNRA_M	0.276***		0.260***
	(0.008)		(0.008)
overERA_F		0.228***	0.186***
		(0.009)	(0.009)
overERA_M		0.219***	0.206***
		(0.008)	(0.008)
N	89 674	89 674	89 674

Note: Model (1) uses only NRA to instrument the retirement decision, Model (2) uses only ERA and model (3) uses both. Robust standard errors, clustered at the individual level, are given in parentheses. All regressions control for age, age squared, female, number of children, years of education, interview wave and country of residence. Significance levels: ***p<0.01, **p<0.05, *p<0.1.

Appendix B - Further Robustness Checks

Table A2: Robustness Checks

						Including Deceased		
	Main	Countries w/o big changes	Ever unemplyd etc. incl.	Unemployed Included	Placebo Test (NRA of 70)	Ret. Stat Last Wave	Ret. Stat. if Age≥ERA	Ret. Stat. if Age≥NRA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Very good to excellent SAH								
retired	0.066***	0.073***	0.078***	0.081***	0.264	0.070***	0.068***	0.070***
	(0.025)	(0.028)	(0.024)	(0.023)	(0.380)	(0.025)	(0.025)	(0.025)
anticipation	-0.020	-0.026	-0.003	0.010	0.693	-0.007	-0.009	-0.006
	(0.050)	(0.051)	(0.048)	(0.043)	(1.461)	(0.051)	(0.050)	(0.051)
honeymoon	0.021	0.008	0.023	0.024	-0.116	0.015	0.015	0.015
	(0.020)	(0.021)	(0.017)	(0.017)	(0.290)	(0.020)	(0.020)	(0.020)
KP F-stat	[62]	[55]	[68]	[81]	[4]	[56]	[56]	[56]
N	89 674	73 975	99 818	104 350	89 674	91 783	91 783	91 783
No clinical depression								
retired	0.078***	0.084***	0.099***	0.109***	-0.145	0.093***	0.090***	0.092***
	(0.021)	(0.023)	(0.021)	(0.020)	(0.347)	(0.022)	(0.021)	(0.022)
anticipation	0.104**	0.114***	0.128***	0.133***	-0.526	0.108**	0.105**	0.109**
	(0.043)	(0.043)	(0.043)	(0.039)	(1.343)	(0.044)	(0.044)	(0.045)
honeymoon	-0.031*	-0.027	-0.028*	-0.024	0.079	-0.035*	-0.034*	-0.035*
	(0.018)	(0.018)	(0.017)	(0.016)	(0.266)	(0.019)	(0.019)	(0.019)
KP F-stat	[60]	[53]	[66]	[79]	[4]	[55]	[55]	[54]
N	87 820	72 369	97 745	102 170	87 820	89 878	89 878	89 878
No Limitations in (I)ADL								
retired	0.040**	0.043**	0.050***	0.058***	0.209	0.056***	0.055***	0.056***
	(0.016)	(0.018)	(0.016)	(0.016)	(0.326)	(0.017)	(0.016)	(0.017)
anticipation	-0.017	-0.011	-0.010	0.001	0.725	-0.004	-0.005	-0.003
	(0.032)	(0.033)	(0.033)	(0.030)	(1.259)	(0.034)	(0.034)	(0.034)
honeymoon	-0.020	-0.021	-0.013	-0.016	-0.240	-0.021	-0.021	-0.021
	(0.014)	(0.015)	(0.013)	(0.013)	(0.249)	(0.015)	(0.015)	(0.015)
KP F-stat	[62]	[55]	[68]	[81]	[4]	[56]	[56]	[56]
N	89 689	73 988	99 831	104 361	89 689	91 797	91 797	91 797
No Mobility Limitations								
retired	0.078***	0.070***	0.078***	0.082***	-0.062	0.084***	0.082***	0.083***
	(0.025)	(0.027)	(0.024)	(0.023)	(0.414)	(0.025)	(0.024)	(0.025)
anticipation	0.033	0.030	0.025	0.026	-0.455	0.042	0.040	0.043
	(0.049)	(0.050)	(0.049)	(0.044)	(1.599)	(0.051)	(0.050)	(0.051)
honeymoon	-0.043**	-0.043**	-0.035*	-0.033*	0.002	-0.043**	-0.043**	-0.043**
	(0.021)	(0.021)	(0.019)	(0.018)	(0.316)	(0.022)	(0.022)	(0.022)
KP F-stat	[62]	[55]	[68]	[81]	[4]	[56]	[56]	[56]
N	89 692	73 992	99 837	104 368	89 692	91 800	91 800	91 800
Maximum Grip Strength								
retired	0.935***	1.250***	1.167***	1.326***	-4.058	1.074***	0.989***	1.056***
	(0.307)	(0.336)	(0.312)	(0.298)	(4.881)	(0.330)	(0.324)	(0.327)
anticipation	0.770	1.013*	1.194*	1.490***	-15.009	1.053	0.939	1.051
	(0.591)	(0.604)	(0.612)	(0.553)	(18.801)	(0.645)	(0.643)	(0.647)
honeymoon	-0.410*	-0.533**	-0.258	-0.316	2.455	-0.249	-0.229	-0.247
	(0.240)	(0.247)	(0.225)	(0.223)	(3.733)	(0.263)	(0.264)	(0.264)
KP F-stat	[59]	[52]	[65]	[78]	[4]	[54]	[54]	[53]
N	83 907	69 315	93 180	97 453	83 907	85 833	85 833	85 833
Total words recalled (cognitive ability)								
retired	0.153	0.193	0.062	0.115	-4.800	0.166	0.160	0.165
	(0.165)	(0.180)	(0.159)	(0.151)	(2.968)	(0.168)	(0.164)	(0.166)
anticipation	-0.158	0.188	-0.272	-0.103	-17.701	-0.160	-0.166	-0.156
	(0.318)	(0.324)	(0.315)	(0.285)	(11.415)	(0.329)	(0.327)	(0.330)
honeymoon	0.223*	0.191	0.233**	0.235**	3.975*	0.275**	0.275**	0.274**
	(0.132)	(0.136)	(0.118)	(0.116)	(2.295)	(0.137)	(0.137)	(0.137)
KP F-stat	[61]	[54]	[67]	[80]	[4]	[56]	[56]	[55]
N	87 888	72 483	97 835	102 292	87 888	89 948	89 948	89 948

Note: ERA and NRA used jointly to instrument for the retirement decision. Anticipation and honeymoon phase 2 years long. Positive coefficients imply a health improvement and the first four health measures are binary. Robust standard errors, clustered at the individual level, are given in parentheses. Kleibergen-Paap (KP) rk Wald F-statistics are reported in brackets. All regressions control for age, age squared, female, number of children, years of education, interview wave and country of residence.
Significance levels: ***p<0.01, **p<0.05, *p<0.1.

Explanation of the different model specifications in Table A2:

- Model (2): This model excludes data from the interview years 2011-2015 in the Netherlands and Italy as well as the data from 2011-2015 in Germany.
- Model (3): This model includes observations of individuals who were ever unemployed, disabled, homemakers or did not declare their job status (besides being employed, self-employed or retired). These were dropped in the main analysis. The observations with employment/ retirement status are included, not the observation in which the respondent reported another job status.
- Model (4): In addition to including observations of those who were ever unemployed, homemakers, etc., this model also includes observations in which the job status is unemployed.
- Model (5): This model is a placebo test, in which the NRA was set to 70 for all persons.
- Models(6)-(9): The main analysis only includes observations of respondents who are alive. In this robustness check, those individuals who passed away and for whom the end-of-life survey was filled out are included in the analysis. They are assigned the worst health status, which is 0 for the binary health outcomes and the lowest 10th percentile value for maximum grip strength and total words recalled. In model (6), the retirement status of the last wave the individual was alive is used, while model (7) also considers those retired who were working in the previous wave but have surpassed the ERA. Model (8) then considers those as retired, who either were retired in the last wave they were alive or who have passed the NRA.

Appendix C - Further Heterogeneity Analysis

Table A3: Further Heterogeneity Analysis

	Main Analysis	Child Living Close	Grand-Children	Living With Partner/Spouse
	(1)	(2)	(3)	(4)
Very good to excellent SAH				
retired	0.066***	0.145*	0.071*	0.034
	(0.025)	(0.080)	(0.037)	(0.029)
anticipation	-0.020	0.199	-0.013	-0.066
	(0.050)	(0.147)	(0.063)	(0.057)
honeymoon	0.021	0.087	0.011	0.030
	(0.020)	(0.053)	(0.024)	(0.023)
KP F-stat	[62]	[10]	[36]	[49]
N	89 674	13 635	53 795	65 636
No clinical depression				
retired	0.078***	0.013	0.100***	0.061**
	(0.021)	(0.065)	(0.033)	(0.024)
anticipation	0.104**	-0.034	0.120**	0.040
	(0.043)	(0.122)	(0.056)	(0.047)
honeymoon	-0.031*	-0.041	-0.045**	-0.030
	(0.018)	(0.045)	(0.022)	(0.020)
KP F-stat	[60]	[10]	[35]	[48]
N	87 820	13 445	52 618	64 150
No Limitations in (I)ADL				
retired	0.040**	0.160***	0.006	0.041**
	(0.016)	(0.054)	(0.026)	(0.018)
anticipation	-0.017	0.125	-0.054	-0.013
	(0.032)	(0.095)	(0.044)	(0.035)
honeymoon	-0.020	-0.059	-0.028	-0.024
	(0.014)	(0.041)	(0.018)	(0.016)
KP F-stat	[62]	[10]	[36]	[49]
N	89 689	13 635	53 805	65 645
No Mobility Limitations				
retired	0.078***	-0.024	0.079**	0.065**
	(0.025)	(0.078)	(0.038)	(0.028)
anticipation	0.033	-0.030	0.026	0.036
	(0.049)	(0.141)	(0.065)	(0.056)
honeymoon	-0.043**	-0.012	-0.053**	-0.022
	(0.021)	(0.056)	(0.026)	(0.024)
KP F-stat	[62]	[10]	[36]	[49]
N	89 692	13 634	53 805	65 647
Maximum Grip Strength				
retired	0.935***	3.169***	1.600***	1.426***
	(0.307)	(1.035)	(0.450)	(0.354)
anticipation	0.770	5.260**	1.715**	1.371**
	(0.591)	(2.070)	(0.733)	(0.679)
honeymoon	-0.410*	-0.673	-0.790***	-0.535*
	(0.240)	(0.681)	(0.292)	(0.289)
KP F-stat	[59]	[9]	[34]	[47]
N	83 907	12 918	50 288	61 507
Total words recalled (cognitive ability)				
retired	0.153	0.346	0.088	0.215
	(0.165)	(0.498)	(0.240)	(0.189)
anticipation	-0.158	0.734	0.099	-0.004
	(0.318)	(0.888)	(0.407)	(0.362)
honeymoon	0.223**	0.329	0.024	0.188
	(0.132)	(0.340)	(0.163)	(0.155)
KP F-stat	[61]	[10]	[36]	[48]
N	87 88	13 547	52 719	64 151

Note: ERA and NRA used jointly to instrument for the retirement decision. Anticipation and honeymoon phase 2 years long. Positive coefficients imply a health improvement and the first four health measures are binary. Robust standard errors, clustered at the individual level, are given in parentheses. Kleibergen-Paap (KP) rk Wald F-statistics are reported in brackets. All regressions control for age, age squared, female, number of children, years of education, interview wave and country of residence.
Significance levels: ***p<0.01, **p<0.05, *p<0.1.